Fahed Yoseph

New Approach for Market Intelligence Using Artificial and Computational Intelligence

Åbo Akademi University

Fahed Yoseph

New Approach for Market Intelligence Using Artificial and Computational Intelligence

Fahed Yoseph

Information Systems
Faculty of Social Sciences, Business and Economics, and Law
Åbo Akademi University
Åbo, Finland, 2023

Supervisors

Docent, University Lecturer Markku Heikkilä
Professor Jozsef Mezei
Senior University Lecturer Anna Sell
Information Systems

Faculty of Social Sciences, Business and Economics, and Law
Åbo Akademi University
Vänrikinkatu 3, 20500, Turku, Finland

Reviewers

Professor Markku Kuula
Department of Information and Service Management
Aalto University
School of Business
Helsinki, Finland

Associate Professor Francisco Cabrerizo Lorite
Department of Computer Science and Artificial Intelligence
University of Granada
Granada, Spain

Opponent

Professor Markku Kuula
Department of Information and Service Management
Aalto University
School of Business
Helsinki, Finland

ISBN 978-952-12-4257-1 (printed)
ISBN 978-952-12-4258-8 (digital)
Painosalama, Turku, Finland 2023

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Docent Markku Heikkilä, for his invaluable guidance during the course of my study. The opportunity he gave me to explore a wide range of topics and focus on what I am always passionate about. Docent Markku's vast wisdom and in-depth scientific research knowledge were very instrumental in shaping my character as an independent researcher and helped me realize the power of critical reasoning.

I would also like to thank my Co-Superviors, Professor Jozsef Mezei and Senior University Lecturer Anna Sell for their broad knowledge and inspiring ideas. The completion of my study would not have been possible without their guidance, knowledge, and expertise.

I'm eternally grateful to Åbo Akademi University for giving me the opportunity to pursue a Ph.D. at the University. I acknowledge the Faculty of Social Sciences, Business and Economics for their astonishingly supportive academic environment and engagement in my study.

My sincere thanks also go to Professor Markku Kuula in Aalto University. He kindly accepted to be a reviewer of my Ph.D. dissertation and take the opponent's role at my Ph.D. defense. I am also very grateful to Associate Professor Francisco Cabrerizo Lorite in the University of Granada for accepting to be a reviewer of my Ph.D. dissertation. I would like to thank both of them for the time and effort they dedicated to reviewing my dissertation and for their valuable academic and scientific remarks and encouraging comments which helped me improve the manuscript.

I would also like to extend my sincere thanks to Mrs. Johanna Hedenborg, the research coordinator for always being nice and helpful, especially throughout the preparation for my public defense.

Lastly, I would like to thank my family, their belief in me has kept my spirits and motivation high during my study in Finland.

Abstract

Small and medium sized retailers are central to the private sector and a vital contributor to economic growth, but often they face enormous challenges in unleashing their full potential. Financial pitfalls, lack of adequate access to markets, and difficulties in exploiting technology have prevented them from achieving optimal productivity. Market Intelligence (MI) is the knowledge extracted from numerous internal and external data sources, aimed at providing a holistic view of the state of the market and influence marketing related decision-making processes in real-time. A related, burgeoning phenomenon and crucial topic in the field of marketing is Artificial Intelligence (AI) that entails fundamental changes to the skills-sets marketers require.

A vast amount of knowledge is stored in retailers' point-of-sales databases. The format of this data often makes the knowledge they store hard to access and identify. As a powerful AI technique, Association Rules Mining helps to identify frequently associated patterns stored in large databases to predict customers' shopping journeys. Consequently, the method has emerged as the key driver of cross-selling and upselling in the retail industry. At the core of this approach is the Market Basket Analysis that captures knowledge from heterogeneous customer shopping patterns and examines the effects of marketing initiatives. Apriori, that enumerates frequent itemsets purchased together (as market baskets), is the central algorithm in the analysis process.

Problems occur, as Apriori lacks computational speed and has weaknesses in providing intelligent decision support. With the growth of simultaneous database scans, the computation cost increases and results in dramatically decreasing performance. Moreover, there are shortages in decision support, especially in the methods of finding rarely occurring events and identifying the brand trending popularity before it peaks.

As the objective of this research is to find intelligent ways to assist small and medium sized retailers grow with MI strategy, we demonstrate the effects of AI, with algorithms in data preprocessing, market segmentation, and finding market trends. We show with a sales database of a small, local retailer how our Åbo algorithm increases mining performance and intelligence, as well as how it helps to extract valuable marketing insights to assess demand dynamics and product popularity trends. We also show how this results in commercial advantage and tangible return on investment. Additionally, an enhanced normal distribution method assists data pre-processing and helps to explore different types of potential anomalies.

Keywords Small and medium sized retailing · Market Intelligence · Association rules mining · Market Basket Analysis · Enhanced Apriori

Sammanfattning

Små och medelstora detaljhandlare är centrala aktörer i den privata sektorn och bidrar starkt till den ekonomiska tillväxten, men de möter ofta enorma utmaningar i att uppnå sin fulla potential. Finansiella svårigheter, brist på marknadstillträde och svårigheter att utnyttja teknologi har ofta hindrat dem från att nå optimal produktivitet. *Marknadsintelligens* (MI) består av kunskap som samlats in från olika interna och externa källor av data och som syftar till att erbjuda en helhetssyn av marknadsläget samt möjliggöra beslutsfattande i realtid. Ett relaterat och växande fenomen, samt ett viktigt tema inom marknadsföring är *artificiell intelligens* (AI) som ställer nya krav på marknadsförarnas färdigheter.

Enorma mängder kunskap finns sparade i databaser av transaktioner samlade från detaljhandlarnas försäljningsplatser. Ändå är formatet på dessa data ofta sådant att det inte är lätt att tillgå och utnyttja kunskapen. Som AI-verktyg erbjuder *affinitetsanalys* en effektiv teknik för att identifiera upprepade mönster som statistiska associationer i data lagrade i stora försäljningsdatabaser. De hittade mönstren kan sedan utnyttjas som regler som förutser kundernas köpbeteende. I detaljhandel har affinitetsanalys blivit en nyckelfaktor bakom kors- och uppförsäljning. Som den centrala metoden i denna process fungerar *marknadskorgsanalys* som fångar upp kunskap från de heterogena köpbeteendena i data och hjälper till att utreda hur effektiva marknadsföringsplaner är. *Apriori*, som räknar upp de vanligt förekommande produktkombinationerna som köps tillsammans (marknadskorgen), är den centrala algoritmen i analysprocessen.

Trots detta har Apriori brister som algoritm gällande låg beräkningshastighet och svag intelligens. När antalet parallella databassökningar stiger, ökar också beräkningskostnaden, vilket har negativa effekter på prestanda. Dessutom finns det brister i beslutstödet, speciellt gällande metoder att hitta sällan förekommande produktkombinationer, och i att identifiera ökande popularitet av varumärken från trenddata och utnyttja det innan det når sin höjdpunkt.

Eftersom målet för denna forskning är att hjälpa små och medelstora detaljhandlare att växa med hjälp av MI-strategier, demonstreras effekter av AI med hjälp av algoritmer i förberedelsen av data, marknadssegmentering och trendanalys. Med hjälp av försäljningsdata från en liten, lokal detaljhandlare visar vi hur Åbo-algoritmen ökar prestanda och intelligens i datautvinningsprocessen och hjälper till att avslöja värdefulla insikter för marknadsföring, framför allt gällande dynamiken i efterfrågan och trender i populariteten av produkterna. Ytterligare visas hur detta resulterar i kommersiella fördelar och konkret avkastning på investering. Dessutom hjälper den utvidgade normalfördelningsmetoden i förberedelsen av data och med att hitta olika slags anomalier.

Nyckelord: små och medelstora detaljhandlare · marknadsintelligens · affinitetsanalys · marknadskorganalys · utvidgad Apriori

Table of contents

1 Introduction 1

1.1. Background and Motivation for Research 11
1.2. Research Problem 13
1.3. Research Objectives 14
1.4. Research Questions 16
1.5. Research Articles 16
1.6. The Structure of the Thesis 23

2 Research Approach and Methods 26

2. Research Approach in Information System 26
2.1. Research Methods in Predictive Analytics..30
2.2. Approach and Methods in This Research 33
2.3. Summary 34

3 Market Intelligence 37

3. Marketing 37
3.1. Market Intelligence 39
3.1.1. Generation of Market Knowledge 42
3.1.2. Market Knowledge for Small and Medium-Sized Retailers 44
3.2. Analytics in Market Intelligence 42
3.2.1. Market Segmentation 47
3.2.2. Market Basket Analysis...52
3.2.2.1. Found Benefits of Market Basket Analysis (MBA)...................55
3.2.2.2. Market Basket Analysis for Small to Medium Size Retailers 57
3.2.3. Integrating Market Intelligence with Database 58
3.3. Market Intelligence from Business-to-Consumer (B2C) Relationship 59
3.4. Scenarios of Market Intelligence 60
3.5. Summary 62

4 Artificial Intelligence 65

4. Introducing Computational Tools For Marketing Analysis 65
4.1. Intelligent Decision Support for Marketing 66
4.2. Artificial Intelligence 69

4.2.1. Development of AI 70
4.2.2. AI as Decision Support for Marketing 71
4.2.3. Computational Intelligence And Marketing 73
4.3. Data Mining and Knowledge Discovery 77
4.3.1. Data Mining Process And Its Outcomes 79
4.3.2. Market Segmentation Project with Data Mining 83
4.3.3. Data Mining for Anomaly Detection 84

5 Association Rules Mining 87
5. Market Basket Analysis 87
5.1. Association Rules Mining Process 89
5.1.1. Rules Extraction 89
5.1.2. Rules Evaluation 90
5.1.3. Challenges in the Association Rules Mining 91
5.2. Apriori Algorithm and Its Variants 92
5.2.1. Deficiency and improvements of the Apriori Algorithm 94
5.3. Research Gap 98
5.4. Summary 99

6 Advanced Association Analysis 102
6. Advanced Association Analysis 102
6.1. Data Exploration and Transformation Practice for Market Basket Analysis 107
6.1.1. Anomaly Detection 112
6.2. Market Basket Analysis Implementation 116
6.2.1. Computational Performance and Efficiency 119
6.2.2. Computational and Artificial Intelligence 123
6.3. Performance Analysis 125
6.3.1. Algorithmic Performance 125
6.4. Performance of Decision Support 127
6.4.1. Finding Trends and Popularity in Itemset 129
6.4.2. Products and Services Recommendation Strategies 132
6.5. MBA Analysis 134

7 Conclusion And Answers to The Research Questions 138
7. Introduction 138
7.1. Answers to the Research Questions 139
7.2. Contribution 142
7.3. The Proposed Market Intelligence Process for Data Mining 144
7.4. Future Research 146

8 References 148

9 Appendices 155

List of Figures

1 Figure 1.1 Research Structure 25

2 Figure 2.1 Steps in building an empirical model 32

3 Figure 3.1 Marketing philosophy eras....................40

4 Figure 3.2 MI Intelligence cycle 45

5 Figure 3.3 Market segmentation bases....................52

6 Figure 4.1 The relationship between AI and ML 71

7 Figure 4.2 Data mining taxonomy....................79

8 Figure 4.3 Knowledge discovery in database process....................81

9 Figure 4.4 The phases of CRISP-DM 83

10 Figure 6.1 The proposed market basket analysis model process 105

11 Figure 6.2 Sample of point-of-sales data structure....................108

12 Figure 6.3 Sample from the target POS data warehouse....................110

13 Figure 6.4 ETL process from the source POS to the target pos data warehouse 111

14 Figure 6.5 Sample pos transactions fall near or far standard deviation from the mean. 116

15 Figure 6.6 Åbo algorithm as compared to the cyclical process of CRISP-DM 118

16 Figure 6.7 Database performance phases (Apriori vs. Åbo) 128

17 Figure 6.8 The top ten rules in terms of lift 130

18 Figure 6.9 Apriori trend attenuation, itemset 1 132

19 Figure 6.10 Apriori trend attenuation, itemset 2 133

20 Figure 7.1 Market intelligence implementation phases for the SMEs retailers 146

List of Tables

1 Table 1.1 Summary of articles 20

2 Table 2.1 The difference between quantitative and qualitative methods 31

3 Table 3.1 The evolution of marketing since 1800s 40

4 Table 3.2 Market intelligence literature reviews key theories and practices 43

5 Table 3.3 Market basket analysis practices 57

6 Table 4.1 AI market methods 68

7 Table 4.2 AI deployment in marketing 69

8 Table 5.1 Products, basket 1 and itemset 1 89

9 Table 6.1 Original POS dataset attributes description 110

10 Table 6.2 Operation performance metrics description 127

11 Table 6.3 Operation performance metrics 128

10 Table 6.4 Suggested recommendation strategies 135

11 Table 6.5 Excerpt from case data: Top 5 Infrequent itemsets 138

List of Equations

1 Equation 5.1 Apriori pseudocode........94

2 Equation 6.1 Enhanced normal distribution anomalies detection........115

3 Equation 6.2 Variables in the pseudocode of åbo algorithm........122

4 Equation 6.3 Variables in the pseudocode of our novel approach........123

5 Equation 6.4 Analysis steps in the Åbo algorithm........124

List of Abbreviations

Åbo Algorithm	The Proposed Enhanced Apriori Algorithm
AI	Artificial Intelligence
ARM	Association Rules Mining
CGB	Candidate-Generation Breadth-First
CI	Computational Intelligence
CLV	Customer Lifetime Value
CPG	Consumers Packaged Goods
DIC	Dynamic Itemsets Candidates
DM	Data Mining
DSS	Decision Support System
DWH	Data Warehouse
ETL	Extract, Transform, Load
FIS	Frequent Itemset
GA	Generic Algorithm
KDP	Knowledge Discovery in Database
MBA	Market Basket Analysis
MI	Market Intelligence
OLAP	Online Analytical Processing Engine
OLTP	Online Transaction Processing Systems
PGB	Pattern Growth
RFM	Recency, Frequency, and Monetary
ROI	Return on Investment
SME	Small- and Medium-Sized Retailers
SMEs	Small and Medium Retail Industry
ND	Normal Distribution

Chapter 1

1 Introduction

The US Small to Medium Business Administration has traditionally defined small to medium-sized enterprises (SMEs) as “non-subsidiary, independent organization, which employs fewer than 500 employees” (Nwankwo & Gbadamosi, 2010). According to the European Commission, SMEs are the backbone of the economy due to accounting for over 50% of all private workforce (Gal, 2010). In other words, the SMEs industry is largely considered the engine of a country's social and economic development (Turner & Ledwith, 2018). Furthermore, there is a growing division between large retailers and SMEs, which is the increase in expenditures of large retailers on analytical solutions and R&D. The SMEs has faced fierce competition and saturated markets from large retailers, as evidenced by the severe decline in shopping consumption and store closures during economic downturns of any kind (Bennett & Robson, 2004; Paolanti, 2018).

Moreover, there are many differences between small to medium and large retailers in terms of size and processes carried out in each business and in annual growth per employee position target (Bennett & Robson, 2004). Large retailers can develop areas of specialized knowledge, finance, sales, and marketing, because of the availability of substantial resources, mainly in capital and human assets. In contrast, SMEs have to learn some basic skills in every business element through experience (Bennett & Robson, 2004; Paolanti, 2018; Turner & Ledwith, 2018).

The financial crisis in the retail industry is not a new phenomenon. Many SMEs retailers have been driven out of business, primarily due to dramatic shifts in economics, demographics, media, and technology, which all seem to conflict with the traditional retail shopping paradigms and models (Omar & Fraser, 2010).

Traditional consumer shopping preferences are now on a fast decline due to new customer shopping rituals (Klaus, 2018). Also, a substantial cultural change is reshaping the customer's expectations and loyalty, allowing customers to increasingly value experience over vendors, low prices, convenience over vast choices, and personalization over standardization (Leslie, 2011). However, it is a time of unique opportunity for SMEs to reinvent successful marketing strategies (Goyat, 2011).

The SME retailers have historically been focused on niche markets, whereas larger retailers tend to offer more products and services to a wider variety of customers (Bennett & Robson, 2004). Nevertheless, today's retail landscape is quite different from how it was years or decades ago. Today, the retail industry is highly competitive, and customers are faced with a variety of products; demand tends to be higher and more complex (Yongjie Yang, Pengfei Yang, Qun Duan & Lina Han, 2016). Additionally, modern customers are better informed when making shopping decisions. Added with the increasingly shorter product life cycle, retailers have less time to make a profit on a given product (Jones & Jones, 2018).

SME retailers are keen to understand customer purchase behaviour and require measurement skills to continuously learn from data. As marketing and innovation are inseparable pairs, ensuring one complements the other is the key to a successful business (Kotler, 2019). While new forms of marketing channels are emerging, such as digital marketing, most SME retailers still adopt traditional marketing, also known as mass marketing, due to their limited marketing budgets (Marketing-schools.org, 2012). This marketing concept encompasses many forms of marketing like print advertisements, flyers, newspapers, billboards, TV, radio, etc., to market their products and services (Dahlén & Edenius, 2007; Raja, Pamina, Madhavan & Kumar, 2019). According to Kotler & Piercy (2013), traditional marketing enabled organizations to approach a wide customer base with different purchase needs into buying the same product (one offer). However, it is costly and usually ignores demographic differences in a market segment in its efforts to reach the highest number of potential customers (Tengberg, 2013).

Traditional Marketing Strategies

Marketing plays an essential role in any organization's success, small or large. Through marketing, an organization achieves success by satisfying the aspirations, wants, and needs of their targeted customer base (Walsh & Lipinski, 2009). Traditional marketing falls under the push category, also known as outbound marketing. The push strategy is more concerned with short-term sales, and the objective is to push products and services toward customers at the point of purchase using big advertisements, such as window displays at department stores that highlight how they can be used (Lambin, 2013). The pull inbounds marketing attempts to create brand loyalty and keep customers coming back (Kim & Aggarwal, 2016). However, as marketing advances, there is a paradigmatic macro shift into how customers find their information and make purchasing decisions. Kim & Aggarwal (2016) stated that traditional marketing is simply not enough anymore.

The industry is now moving from traditional marketing to modern marketing to reach out to customers and generate more revenue for their brand rather than spending marketing efforts and budgets on the traditional ways of marketing (Chiu, Kim & Won, 2018). According to Liam Damien (2019), there is an urgent need for SME retailers to revisit the relevance of traditional innovation models to see that they only allow for a blanketed approach with limited assurance to reach the right audience (Lambin, 2013). Berry & Linoff (2004) predicted that demand for more precise product profitability knowledge would become necessary, and all the developments described above seem to have proven this true.

Therefore, it is imperative for SME retailers to implement more formalized marketing strategies like market intelligence (MI) to gain a comprehensive view of the market landscape, where brands and retailers adjust to the relevant and sometimes drastic changes that otherwise render them less competitive (Stubseid & Arandjelovic, 2018). According to Zhan, Tan & Huo (2019), MI is the process of sifting through and analyzing enormous transactional data incredibly fast to leverage customers' data and improve the customer shopping experience. The process consists of recognizing complex patterns that humans are unable to detect in order to identify product trends and accurately accommodate the consumer's specific demands. This was once an insurmountable process, but it is now a reality (Marwan, 2011; Xuan Li, Xiaoyan & Zhuang 2002). Studies have shown that MI strategies powered by AI solutions such as data mining are the greatest opportunity for identifying profitable market segments and leveling customer shopping journeys between large and SME retailers (Hedin, Hirvensalo & Vaarnas, 2011).

For MI architectures, the latest trend has been one of the increasingly ubiquitous predictive analytics applications. There is an entire, broad research area for determining how to best mine the most interesting and complex patterns from big transactional databases. However, as will be shown in the review of the articles analysed here, there seems to be little justification for this trend in contemporary literature. Specifically, association rules mining (ARM) applications, in the context of customer shopping behavior, revolve around developing computational algorithms and AI capable of reviewing the work and achievement in real time. Even for the most complex AIs, however, when generating and counting the frequency of candidate sets, they vary in performance. Moreover, even the best AIs ultimately rely on association rules algorithms, each of which has its strengths and weaknesses. As expanded on below, this makes AI development a very costly and time-consuming endeavor.

Artificial Intelligence (AI)

McCarthy (2007) defined AI as the science and engineering of making intelligent machines in which its computational part has the ability to achieve goals in the world. AI offers marketers a systematic process to bridge the gap between data science and execution (Hall, 2020). In the world of digital marketing, applying AI across customer life-cycle will continue to streamline and change the way marketing campaigns are driven and eliminate the risk of human errors (Conroy, Porter, Nanda, Renner & Narula, 2015; Liam Damien, 2019; Wen & Shuai, 2018; Hall, 2020). AI models provide retailers with the ability to interact directly with customers and transform the customers' data into a blueprint of how to serve them in the future better. Today's digital marketing strategies would be impossible without a rudimentary form of AI-based solutions (Hall, 2020).

Consequently, with the existence of transactional Point of Sales (POS) data in giga and terabyte ranges, the ability to transform raw data into actionable and predictable solutions is now possible. This sharply contrasts traditional marketing, which can be aimless; modern marketing is more effective and customer oriented. In short, since modern marketing incorporates a digital platform, it is more customizable, easily adapts to change, allows customers to get endorsements for the brands, and enables businesses to have a higher chance of reaching a global audience. All of these factors help build customer loyalty and trust and worldwide brand recognition.

One of the central AI methods applied to the field of marketing is association rules mining (ARM). Large retailers and mass merchandisers primarily use ARM to create market strategies addressed to their different customer segments. The method is used in discovering knowledge from a database to find regularities in the customer's shopping behavior and help uncover relationships between seemingly unrelated data in a relational database (Borgelt, 2012). ARM applications have been applied to various domains like market segmentation (MS) and market basket analysis (MBA). MBA and MS models are the two most powerful AI techniques used in the area of marketing. Both methods encompass a broad set of analytics techniques aimed at extracting unknown knowledge about customers' lifetime shopping behavior and lifestyle (Adhikari, 2018).

MS is based on the assumption that potential customers are not identical and that the retailer should address their shopping preferences with a tailor-made marketing strategy (Dipanjan, Satish & Goutam, 2011). Instead of blending in with other brands, the strategy involves promoting awareness of differences between one retailer's brands and competitors' brands (Jobber & Ellis-Chadwick, 2012). As such, the technique relies on a process of dividing a target market into smaller, more defined categories that share similar characteristics such as demographics, interests, shopping preferences, and location (Dipanjan, Satish & Goutam, 2011). These characteristics are relevant to marketing and sales, such as demographics, age, location, nationality, gender, interests, and spending habits (Liu, Liao, Huang, & Liao, 2019). This technique greatly enhances the decision-making process in evaluating new opportunities and enables the retailer to focus on marketing resources, distribution, and creating a unique selling proposition to the most profitable audiences (Yoseph, Ahamed Hassain Malim, Heikkilä, Brezulianu, Geman & Paskhal Rostam, 2020). Incorporating AI methods into marketing to establish market intelligence can potentially create a broad picture of the retailer's existing market (Mageplaza, 2020). According to Lori and Golberg (2018), MI results from AI-based segmentation systems interpreting customer data paired with profile information and demographics.

On the other hand, MBA is the prototypical application of association rule mining. It can be a significant help for marketers to identify with relative certainty the content of the customer's basket based on the types of products they are most likely to purchase in conjunction with other products or more likely to purchase next at a specific time. (Wang & Sun, 2019). MBA's goal is to find common purchase items across the customer's basket, referred to as association analysis. All products have either strong or weak relationships with each other and can exhibit a strong or weak affinity for a period of time. In most cases, the sale of one product is driven by the increase or decrease in the sale of other products. The discovery of this knowledge of association is essential for marketers to estimate the purchasing trending popularity of a certain itemset and drive maximum brand loyalty.

MBA is modeled on association rules mining: IF {itemsets} discovered in data are (Antecedent), THEN {item or set of items} discovered in combination with the (Antecedent) is consequent. MBA uses descriptive and predictive techniques. The descriptive methods rates the association of the product using statistical techniques to derive insights on past data. The predictive methods use supervised learning techniques like regression and classification to mimic products purchased in a sequence to help decision-makers to come up with solutions to problems.

Implementing AI tools like data mining comes with a very expensive price tag and a lengthy implementation period, forcing 65 percent of companies to consider the analytical solution to be very expensive (Coiera, 2019; Power, 2002; Longinus, 2018). AI implementations consist of vendor costs and a number of experts who can answer questions in their area of expertise, making it too expensive for any SME to acquire (Mahoney, 2017).

The Apriori Algorithm

The Apriori algorithm is the first and arguably most influential algorithm for discovering efficient association rules. Proposed by Agrawal and Srikant (1994), it is commonly used to solve ARM problems by determining the more frequently used itemsets in large transactional databases (Akura & Srinivasan, 2005; Szymkowiak, Klimanek & Józefowski, 2018). It is a categorical (non-numeric) algorithm based on using prior knowledge of frequent itemsets properties and in so doing, help businesses enhance sales performance. The algorithm is optimized to support operations in many industries like healthcare, electricity supply, manufacturing, railway safety management, education, retail, finance, and monitoring the quality of web services (McGlothlin & Khan, 2013). Presently, the Apriori algorithm is arguably the ARM algorithm most widely and prominently used, in no small part due to its ubiquity in online commerce, where it is the go-to method for mining small, medium and large databases alike for any and all shopping patterns (Doshi & Joshi, 2018).

Unfortunately, Apriori does come with some notable limitations in terms of computational performance, efficiency, and lack of intelligence. These limitations were a large part of the motivation behind this research study to propose significant improvements upon the computational efficiency and intelligence of the Apriori algorithm.

Firstly, the more complex and numerous the time and space parameters become, the more the performance of Apriori drops. The algorithm needs a new, independent iteration to perform data mining each time a new time/space parameter is added, which massively increases the number of discovered rules, thus slowing performance dramatically. This has generally made it all but impossible for the Apriori algorithm to properly account for the significance of an item to a user or business (Said, Muhammad & Gupta, 2015). Apriori is, thus, very limited in answering complex queries like identifying the important preoperative predictors. This includes a tendency to fail in identifying when an event is unusual and missing potentially infrequent interesting itemsets because they do not meet the support threshold.

Secondly, Apriori is challenged in the area of intelligence: According to Hegland (2003) ARM treats all items in a database equally by considering only the presence and absence of an item within the transaction. This causes the Apriori algorithm to generate itemsets that are redundant or, at best, irrelevant. This deprives database AIs of the intelligence they need to accurately assess customers' buying habits or profitable products. In turn, this causes difficulties in finding rarely occurring events. Furthermore, because of the aforementioned loss of efficiency that the Apriori algorithm suffers from, the more data it needs to process, the less it provides the scalability needed to manage big data tasks (Singh, Garg & Mishra, 2015). This is a weakness not unlike that exhibited by enhanced traditional parallel and distributed algorithms and a particularly crippling one: Considering the constraints related to the analysis of big POS databases, SMEs bring an unambiguous need for high-level knowledge discoveries and scalable algorithmic approaches, as retailers cannot forecast customers' shopping journeys with-out such (Akura & Srinivasan, 2005).

It is true that there is no single program or algorithm that can accomplish all of those mentioned above, but these weaknesses demand a precise investigation because they persist despite the age and the refinements which Apriori has gone through over the years (Akura & Srinivasan, 2005; Ezhilvathani & Raja, 2013). Of greatest concern is that the lack of intelligence in the classic Apriori algorithm remains an unsolved problem despite the many attempts to solve it since ARM's inception (Bansal, Sharma & Goel, 2017; Sharma & Verma, 2014; Moreno, Segrera & López, 2005; Yuan, 2017; Olson & Lauhoff, 2019).

This means that the greater the size of data mining models and the more numerous the data-driven decisions become, the less reliable the insights gained by the Apriori algorithm become. As a result, an organization would take a considerable risk in basing its marketing decision-making on them. If the data that is fed into the algorithm is not cleaned (or is flawed), the results could be far from accurate and potentially misleading to the prediction efforts. In addition, the exponential increase of unstructured customer transactions makes knowledge discovery error-prone and difficult to process.

All told, detecting anomalies early is a must, which leads to the second great imperative of marketing AI: early and automated anomaly detection in real-time. Without such knowledge discovery processes tend to be slow and unreliable. Therefore, it is important to determine if AI-powered anomaly detection can be used in a large POS database to lead to the discovery of hidden knowledge.

Anomaly Detection

Finding anomalies and rare events in the database and the removal of anomalous objects has become an emerging issue in the area of AI. One of the primary focuses of this research is the detection of anomalies, also known as outliers or novelty objects, to identify suspicious transactions in POS database. Anomaly detection is one of the core issues in AI and has gained considerable attention in many industries, like the retail and banking industries (Hawkins, 1980). A data-intensive industry like the retail industry collects enormous volumes of raw POS data daily. But, although the vast majority of this data is correct and complete, a growing number of inputs are rare, even exceptional, and deviate from the rest of the data to varying degrees. This type of data is widely known as anomalies and can result in merchandise and financial loss. Supposing that the SME aim to optimize their work efficiency and increase profitability by using POS data, they have to deal with anomalies. Most importantly, dealing with anomalies generates a more accurate analysis. In that case, data quality and cleanliness are of utmost importance.

Barnett & Lewis (1994) defined anomalies as a subset of observations, which appears to be very inconsistent with the rest of the data. Systematically, we see that anomalies are those objects or data points that follow different patterns, deviate from the mean or fall some standard deviations away from the mean. These anomalies seem to be generated by a mechanism or system different from the rest of the observations. Various reasons such as systematic errors, human errors, and, most importantly, fraudulent behavior can generate anomalies. Regardless of the source of the anomaly, the detection can reveal an interesting pattern in the data, which could be faulted in the system or fraudulent behaviour. Every successfully detected anomaly can assist analysts in understanding the root or the causes of anomalous behaviour.

A philosophy argues that anomalies should not be highlighted nor removed from the database, because, as researchers, we should not prejudge the data. In support of this, previous studies have found that when humans have a strong belief, they subconsciously seek out consistent belief information and discard information contradicting their beliefs (Hawkins, 1980; Barnett & Lewis, 1994). In the context of marketing, this means that extreme anomalies can skew data distribution in the direction of the anomalies, which makes it complex to analyze. Small data deviation might be an anomaly. An anomaly may distort the normal object and blur the distinction between anomaly and normal object (Knorr & Tucakov, 2000). In the literature, anomalies have been much debated regarding to what to do with influential or extreme data points (Osborne & Overbay, 2004). Therefore, it is crucial to develop a new data mining technique for anomaly detection to prevent potentially severe loss without compromising the data's accuracy and legitimacy.

To reduce the risk of developing an inaccurate data mining model, we propose a novel anomaly detection technique for the preliminary data pre-processing, using an enhanced normal distribution method to determine incomplete, inaccurate, or those extreme objects that deviate or are exceptionally far from the mainstream data. This detection technique should, in so doing, reduce the risk of developing DM models that are inaccurate. The following sections expound on this research motivation, research problem, research objectives, and research questions.

1.1 Background and Motivation for Research

The general motivation for this research relates to my professional experience with major financial and CRM systems deployments for SME retailers. This experience has stimulated my interest in learning more about customer psychology, purchasing trends, and the forces behind customer shopping behavior.

Certainly, customers' shopping behavior is a broad and complicated study. Nevertheless, it proposes that the right research mix can provide SME retailers with a systematic understanding of customers' shopping motivations. Moreover, I have serious doubts about the effectiveness of traditional marketing strategies supporting the decision-making process in real-world marketing. Customers are not equal, some earn you an amazingly disproportionate amount of money, and some even waste your time. With the last group, you lose money selling anything to them at all. Customers who buy the most expensive brands always fit a particular demographic, and they are noticeably different from everyone else.

Lee & Gupta (2013) stated that customers buy various combinations of products on a single shopping trip. These multi-category decisions result in the formation of customers (shopping basket), which comprises the collection of categories that customer purchases on a specific shopping trip**.** Thus, SME retailers need to understand the dependencies among customers' purchases. Some products have a higher affinity to be sold together. Hence, the retailer can benefit from this affinity if special offers and promotions are developed for these products to target profitable customers. This will allow retailers to redistribute costs towards aspects of more profitable brands.

Modeling customers' shopping journeys through the improvements of association rule mining (ARM) has been the subject of many types of studies. The primary focus in the previous studies has been on the efficiency of the ARM approach on how quickly it can derive association rules, and to a far lesser extent, the measures taken to determine the intelligence of the derived these rules. As a summary, the traditional ARM approach mentioned above suffers from the following limitations or bottlenecks.

1. The generation of a large number of ARM rules with significant rules redundancies makes it very difficult to comprehend. Thus, it significantly reduces the effectiveness of the association rule mining algorithms.

2. Traditional ARM approach ignores infrequent but elite items often missed by this approach.

3. Traditional ARM ignores the customer's purchase behavior in the interaction among different product profitability trends. Therefore, it ultimately leads to the inability to identify the most profitable products and services.

All this adds up to form a solid motivation to take advantage of the dynamics involved in data mining, such as ARM, classification, and clustering techniques on the content of customer shopping behavior. Through these, it should be possible to ensure the methods and the proposed model can be applied to a real-world context. The first general research motivation aims to present a data-driven market intelligence engine for SME retailers. It is expected that our proposed model will help elevate the issues in the classic Apriori algorithm.

To sum up, the specific motivations for this research are as follows.

1 The first motivation is commercial. To enable SME retailers to grow through an effective market intelligence strategy and systematically forecast customer shopping journeys. This will eventually produce a commercial advantage and tangible return on investment (ROI).

2 The second motivation is technical. Develop novel methods in MBA to help improve the computational and AI of the classic Apriori algorithm.

3 The third motivation comes from the marketing argument that infrequent patterns are also interesting in many real-life (Sim, Indrawan, & Srinivasan, 2008). Infrequent patterns are common to keep more interesting patterns by setting a low minimum support threshold. A low threshold comes with many problems in the mining process. No matter how low the assigned value is, it is always possible that some interesting patterns will be filtered out.

Our proposed method is inspired by the fact that there has not been any research to address the intelligence limitation in MBA or offer a viable solution and bridge the gap between market intelligence and ARM. This research is also motivated by the intuitiveness of the normal distribution approach. The objective is to design an efficient anomaly detection approach that does not require users to possess extensive knowledge of statistical computation. Thus, we introduce an enhanced, normal distribution anomaly detection method to specify the degree of anomaly and likelihood of the data object being generated by an abnormal and normal mechanism. The technique explores different levels of potential anomalies in the POS database, with the ability to fully or partially eliminate anomalies. In other words, the client decides whether or not the suspected data object should be removed or not. The model calculates how far a data object is from the mean (average). In this approach, we predict that the accuracy will increase substantially while errors of inference will to drop significantly.

1.2 Research Problem

Customer shopping behavior is unquestionably a fertile area of investigation. Today, customers are experimenting with different lifestyles and are heavily influenced by fashion. For a retailer, it is essential to understand the attitudes and intentions that affect the customer's shopping behavior (Oliver, Rust & Varki, 1997). Among all retailers, SME retailers suffer from a lack of such understanding. They are facing increasing and serious competition from large retailers and mass merchandisers. Also, customer loyalty is a complex phenomenon that does not emerge on its own, and several variables could affect their loyalty. Commitment, trust, and satisfaction are the global mediators of customer relationship loyalty (Palmatier, Dant, Grewal & Evans, 2006).

The retail industry often does not address individual customers based on their purchase behavior. It only looks in the rear-view of customers' historical data on the assumptions of what makes customers similar to one another. This finding has led scholars to emphasize the need for more studies in the relationship marketing field (Kumar & Reinartz, 2019). The introduction of electronic point-of-sale (POS) records increased the use of transactional data in MBA. The retailers soon realized that this tremendous knowledge trapped inside these POS databases only has value if it can be quickly mined and analyzed to produce results.

Currently, pattern mining from big POS databases is a broad research ar-ea covering both computational algorithms and AI. In the context of customer shopping behavior, each association rules algorithm has its pros and cons. However, all the reviewed algorithms focus entirely on performance and on the processes of candidate-set generation (how quickly the algorithm devises the rules). The quality of the derived rules is of much lesser concern in comparison. For big POS databases, this often leads to associated rules being derived in disproportionately large numbers. This common problem of rules explosion generates many redundant and uninteresting associations. These, in turn, necessitate further human analysis. As a result, decision making based on these rules leads to risky action.

It is for the purpose of overcoming this problem that we propose a simple and free PL/SQL based open-source probabilistic data mining model to intelligently automate repetitive, personalized recommendations and data-driven tasks. This MBA model uses novel computational and artificial intelligence methods to perform fewer database scans and select only legitimately interesting rules.

1.3 Research Objectives

Traditional methods of analyzing Point of Sales (POS) databases are challenging to scale and limited only to analyzing a subset of the data. Utilizing modern data mining techniques allows SME retailers process the entire POS database in small fractions and with less reliance on human experts. The research objectives are twofold as follows, general and specific.

The general objective for this research is to propose market intelligence to digitally forecast customer shopping journey and construct a comprehensive outlook on the customer to deduce conclusions for marketing decisions.

As for specific objectives we highlight points by Ceglar & Roddick (2006) that most of the ARM research efforts explicitly focus on improving the algorithmic performance, and into reducing the algorithmic output set and allowing to express constraints on the algorithmic desired results. Also, they point out that Apriori algorithm uses too many parameters and need to be configured before executing the Apriori algorithm. Finally, they state that for a non-expert in data mining, the generated rules are far too many, and most of them have proven to be non-interesting rules and with low comprehensibility.

Consequently, we come to the objectives that aim at tackling the problems highlighted by Ceglar & Roddick (2006).

1. To demonstrate data mining methods by proposing a simple predictive MBA model using an enhanced Apriori algorithm (Åbo) and by explaining how it can be applied in a large retail POS database. As Agrawal & Srikant (1994) point out, progress in technology has made it possible for retail industry to collect massive sales data. However, mining of association rules in large POS database is a challenging task, because it requires an extensive I/O load. Extracting valuable rules is essential for marketing and customer segmentation based on customer purchase patterns. Therefore, it is imperative to have fast algorithm.

2. To make a significant enhancement to the performance of the classic Apriori algorithm by introducing new computational intelligence techniques. The proposed algorithm aims to divide the POS database into a number of non-overlapping ARM partitions. All partitions are scanned and considered one at a time and kept in the main memory. The algorithm applies new segmentation and compression techniques to avoid repeated database scans and generating non-interesting ARM rules.

3. To make a significant enhancement to the intelligence of the classic Apriori algorithm by introducing new AI techniques to automate the prediction of trends and behaviors and understand the factors that influence products' upward and downward trending popularity. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data.

1.4 Research Questions

To reach the specific objectives, we investigate the customer shopping journey and how market intelligence can automate the business decision-making process in the context of small-to-medium-size retailers. The research aims to answer the following questions:

1. Does the Åbo algorithm improve the computational performance and the intelligence of the classic Apriori ARM algorithm?
2. Do small and medium-sized retailers get better market intelligence with the new algorithm?
3. How can we achieve optimal anomaly detection without compromising the accuracy of the data?

1.5 Research Articles

In this section, we describe the contribution of our original publications, building up this research. We present the techniques applied in each article for modeling consumer shopping behavior. Articles are related to one another, and the contribution of each publication is briefly explained and illustrated. The articles below comprise five scientific publications that enclose all thesis contributions. The included articles are the following:

1.5.1 RA1

This article is titled " Segmenting Retail Customers with Enhanced RFM Data using a Hybrid Regression/Clustering Method," identifies how these 'hybrid regression/clustering methods can be used to identify customer shopping behaving in different demographic categories like gender, age, and nationality. Using an enhanced RFM dataset, the model breaks down the SMR market into four meaningful market segments (best, spender, frequent, uncertain). The knowledge gained from the paper adds to existing studies with the following contributions.

The study identifies areas of prior studies to prevent duplication in this area and increases understanding and knowledge of the retail market in general and the SMEs in particular. It finds the characteristics of customer shopping behavior, like VIP customers who are at high risk of shifting their business to another retailer or falling from a higher segment to lower segments and highly profitable products. Moreover, the paper identifies a small gap within a saturated area of studies and provides a novel interpretation of customer shopping behavior. Finally, it contributes to the prior body of knowledge by extending a theoretical model of customer purchase behavior in a retail setting. In the dissertation, this paper contributes to the model of enhancing market basket analysis to account for simplicity to measure customer shopping power, customer segmentation and shopping trends.

1.5.2 RA2

This article is titled "The impact of big data market segmentation using data mining and clustering techniques" can be used to identify customer shopping behavior in different demographic areas. These categories include gender, age, and nationality. Using an enhanced RFM dataset with new variants, the average purchasing power per all transaction (P), the average purchasing power per product (Q) and change in customer buying behavior or trend using customized change rate (T), we show how the model expands the knowledge of customers segments shopping behavior. The knowledge gained from the paper adds to existing studies with the following contributions.

The study proposes these RFM variants, PQT, using a dataset extracted from big POS database. The variants PQT add more to the understanding and motivation of individual customers segment purchase power. As a result, we compute the average price per items per customer per segment. Moreover, the study increases understanding of big data architecture to improve the SMEs' operations and help develop a personalized marketing campaign.

The new variants help SME retailers to distinguish between different customer segments (Gold, Silver, Bronze) and what motivates customer's purchase habits.

For the dissertation, this paper contributes to the model of enhancing market basket analysis to account for understanding the changes in the customer's average shopping spending. It shows how customer purchase behavior is heavily influenced by age, gender, and nationality. It also helps identify k-means++ algorithm as the best algorithm to measure customer purchase power to build new customers' data-driven marketing strategies. We replicated this knowledge and applied it to the products level in our algorithm. This helps develop a smarter ARM algorithm.

1.5.3 RA3

This article is titled "Outliers identification model in point-of-sales data using enhanced normal distribution method" stems from the fact that AI models perform with higher accuracy using clean data. The study proposes an enhanced normal distribution method to generate insights about different types of outliers found in a retail point of sale (POS) dataset. For the end-user to be able to make an informed decision about outliers, the model allows the end-user to eliminate outliers fully or partially or move them to a separate place for further investigation.

This paper contributes to the model of enhancing market basket analysis to account for understanding different characteristics of outliers. It identifies what outliers are, how they came to be in existence, and if they represent legitimate, error, or fraudulent transactions. This understanding helps to construct the data collection process to prevent outliers from being a normal part of the data process, and also, their removal will not affect the integrity of the extracted data. The key success that underpins our study is that the method is able to distinguish outliers from non-outliers with high accuracy. This model has been successfully implemented into real-life POS database. This increases the analysis accuracy in our Åbo algorithm.

1.5.4 RA4

This article is titled "A clustering approach for outliers' detection in a big point-of-sales database." The study proposes a clustering-based approach to improve the quality of outliers' detection.

The model employs performance-controlled assessment using two clustering algorithms, hard clustering K-means, and soft clustering (FCM) Fuzzy C-means. The model uses size for identifying data points with the largest dissimilarities by calculating the distance between each data point. The method assumes that data objects that belong to dense or large clusters are legitimate transactions, whereas data objects belong to sparse or small clusters, and data objects that don't belong to any cluster are considered outliers.

This paper contributes to the model of enhancing market basket analysis to account for identifying areas of prior studies to prevent duplication in this area and increase understanding of the nature and characteristics of the detected outliers in POS dataset, like those data objects that do not belong to any cluster. The knowledge gained from the study adds to the understanding from the previous RA3 article. Moreover, the results show this model is more efficient when compared to the performance of the previous normal distribution method in RA3 study. However, the method identified far less anomalous transactions when inspected by the domain experts, which resulted in more manual investigation. Based on the results from this study, a decision to favor accuracy over performance in our dissertation is the most fitted outliers approach. Finally, the study contributes to the prior body of knowledge by extending a theoretical model of outliers detection in a retail setting.

1.5.5 RA5

This article is titled “A new approach for association rules mining using computational and artificial intelligence.” The study proposes novel ARM methods to develop MBA model to extract market intelligence insights from big POS databases to forecast customer's shopping behavior.

The model incorporates techniques like CI and AI into one algorithm to create market intelligence for SMEs. The study identifies the difference between two critical marketing segmentation factors, time based-segmentation and occasion-based segmentation. Such understanding proved to be very beneficial to our dissertation. The study identifies the patterns in which products are purchased and how the purchase of one product influences the probabilities of another product being purchased in the same basket.

It contributes to the prior body of knowledge by extending the conventional theoretical model of MBA in a retail setting by incorporating clustering and association rules techniques along with customers purchase transactions.

This paper contributes to the model of enhancing market basket analysis to account for testing the productivity of customers' shopping baskets. The apriori algorithm is effectively modified to increase performance efficiency and enhance mining accuracy. Such modification contributed to the extraction of more focused segmentations and accurate association rules insights related to products purchase decisions with age group, gender, and location.

The key success underpinning this study is developing synergies between customer shopping behavior, CI and AI to develop market intelligence for the SME industry. This model has been successfully implemented into real-life POS database.

1.6 Summary of the Articles

Table 1.1 Summary of the Articles

	No.	Year	Article	Key Contribution	*Authors Contribution*
I.	RA1	2018	Segmenting retail customers with an enhanced rfm and a hybrid regression/clustering method. Yoseph, F., & Heikkila, M. (2018, December). Segmenting retail	1) Simplified data mining model (MS) to measure customer purchase behavior power. 2) Market segmentation methods using classification, clustering and regression techniques based on the new	1) Yoseph, F produced material, analysis and text for the article. 2) Heikkila, M produced text for the article, mentored the research process, some editing and language proofreading of the text.

			customers with an enhanced RFM and a hybrid regression/clustering method. In 2018 International Conference on Machine Learning and Data Engineering (iCMLDE) (pp. 108-116). IEEE.	modified RFM variant (C).	
II.	RA2	2020	The impact of big data market segmentation using data mining and clustering techniques. Yoseph, F., Ahamed Hassain Malim, N. H., Heikkilä, M., Brezulianu, A., Geman, O., & Paskhal Rostam, N. A. (2020). The impact of big data market segmentation using data mining and clustering techniques. Journal of Intelligent & Fuzzy Systems, 38(5), 6159-6173.	1) Design and development of the Generic POS Data Warehouse. 2) Marek Segmentation data mining model using Expectation-Maximization (EM) and K-Means++ and new modified best-fit regression algorithms. 3) New RFM variants (P, Q) provided simplicity to measure customer purchase power and changes.	1) Yoseph, F produced material, analysis and text for the article. 2) Ahamed Hassain Malim and, Heikkilä, M, produced text for the article, mentored the research process, some editing and language proofreading of the text. 3) Brezulianu, Geman, O., & Paskhal Rostam provided language proofreading of the text
III.	RA3	2019	Outliers Identification Model in Point-of-Sales Data Using Enhanced	1) Novel outlier identification data mining model using an enhanced normal distribution	1) Yoseph, F produced material, analysis and text for the article.

			Normal Distribution Method. Yoseph, F., Heikkilä, M., & Howard, D. (2019, December). Outliers identification model in point-of-sales data using enhanced normal distribution method. In 2019 International Conference on Machine Learning and Data Engineering (iCMLDE) (pp. 72-78). IEEE.		method. The Data Mining Model can spot different types of outliers from retail (POS) dataset, giving end-users the ability to fully or partially explore/eliminate outliers.	2) Heikkila, M produced text for the article, mentored the research process, some editing and language proofreading of the text. 3) Howard, D helped in benching marking the model's accuracy.
IV.	RA4	2019	A Clustering Approach for Outliers Detection in a Big Point-of-Sales Database. Yoseph, F., & Heikkilä, M. (2019, December). A clustering approach for outliers detection in a big point-of-sales database. In 2019 International Conference on Machine Learning and Data Engineering (iCMLDE) (pp. 65-71). IEEE.	1)	Efficient outlier detection data mining model based using clustering Hard clustering K-Means algorithm and soft clustering algorithm FCM algorithm.	1) Yoseph, F produced material, analysis and text for the article. 2) Heikkila, M produced text for the article, mentored the research process, some editing and language proofreading of the text.

V.	RA5	2020	A new approach for association rules mining using computational and artificial intelligence. Yoseph, F., & Heikkilä, M. (2020). A new approach for association rules mining using computational and artificial intelligence. Journal of Intelligent & Fuzzy Systems, 39(5), 7233-7246.	1)	Significate step to demonstrate the importance of designing and developing synergies incorporating three different and vital areas of research, namely market research, consumer shopping behavior, and Computational and Artificial Intelligence, to forecasting consumer's shopping behavior and develop market intelligence for the SMRs industry.	1) Yoseph, F produced material, analysis and text for the article. 2) Heikkila, M pro-duced text for the ar-ticle, mentor the re-search process, some editing and language proofreading of the text.

1.7 Structure of the Thesis

Chapter 1 provides the background of the study, with the motivation, problematization, research objectives, and research questions. Chapter 2 provides an account of the research choices and methods used in the thesis. Chapter 3 includes literature reviews on general marketing, market intelligence theories, discipline, benefits, and implementation concerning customer shopping behavior. Chapter 4 provides literature reviews on AI theories, practices, and methods implemented in customer shopping behavior. AI subfields like computational intelligence and data mining are discussed in detail. In chapter 5, the proposed methods and techniques employed as part of this research are discussed in detail. The specifics of the MBA model developed in this research are described. Chapter 6 provides the interpretation of the analyses employed throughout this research. Explanatory and interpretative comparative analyses of customer shopping behavior between the two unsupervised learning algorithms (classic Apriori and the proposed Åbo) are presented and discussed. Appropriate market intelligence strategies are suggested. Chapter 7 provides a critical assessment of the research and answers to the research questions. It demonstrates the research contribution, precision, and thoroughness of the main research findings. Future research recommendations, limitations, and potential future research directions are briefly discussed. Figure 1.1 illustrates how this research is organized.

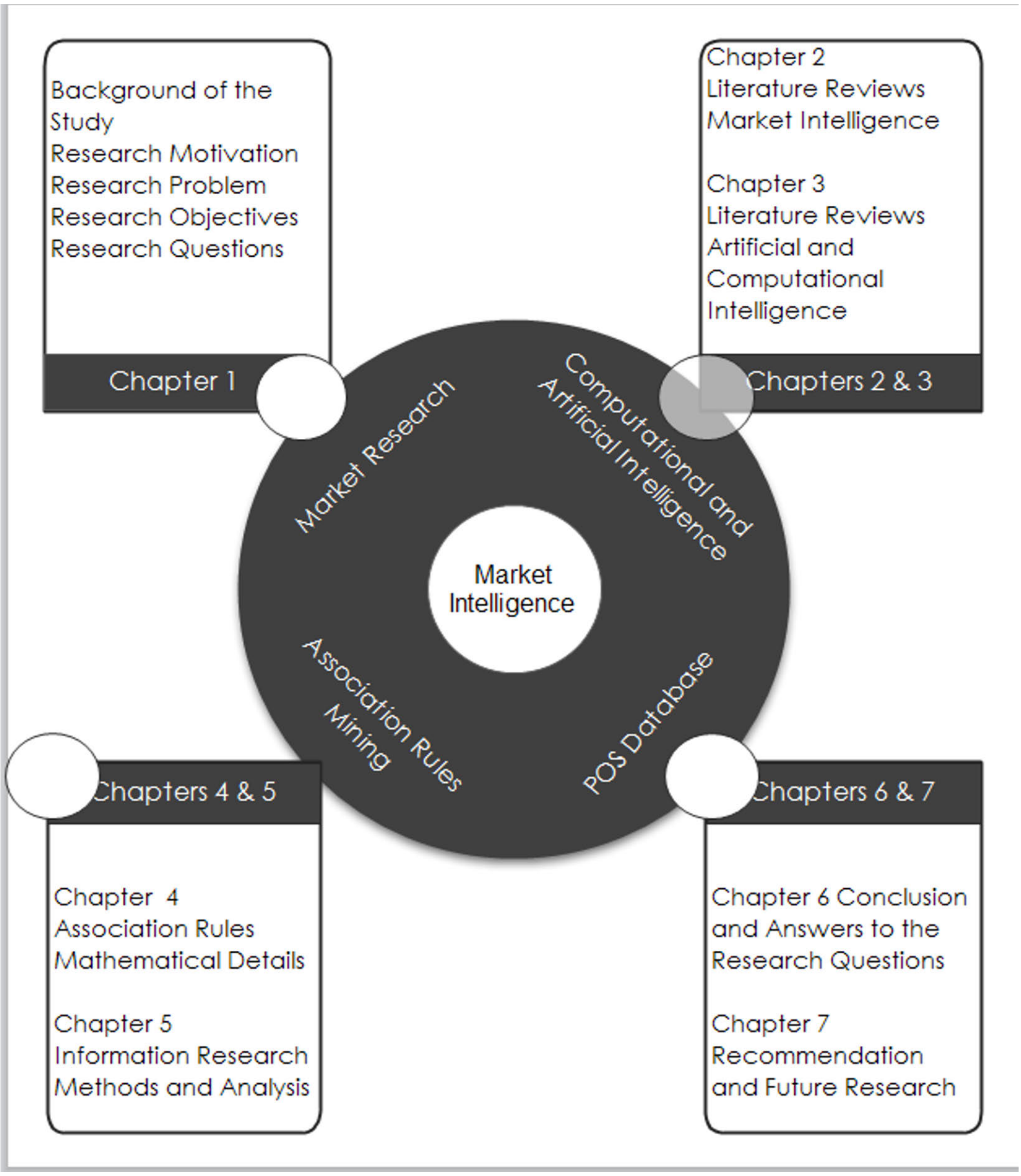


Figure 1.1 Research structure

Chapter 2

Research Approach And Methods

2 Research Approach in Information Systems

Research methodology is a vital element in research to establish a rigorous research structure, the ability to pursue the original research to the general body of knowledge, and present the findings and results in a scholarly way (Wilkinson, 1991). Therefore, research methodology, research philosophy, and research model play a significant role in research.

To provide an understanding of AI, CI, and MI in the context of predicting customer shopping behavior to support the decision making for marketing, this chapter, and the following two chapters describe the three concepts and how they relate to each other. The information systems field has made significant advances in employing advanced statistical modeling techniques to support empirical scientific research (Marcoulides & Saunders, 2006). Analytics involves the use of methodical and iterative techniques to discover, interpret, and analyze meaningful patterns from data (Baltzan & Welsh, 2015). Predictive analytics helps develop new measures, generate new theories, compare existing theories, and assess the relevance of theories (Shmueli & Koppius, 2011). Predictive analytic includes statistical models and empirical methods that are aimed at creating empirical predictions (Shmueli & Koppius, 2011). The analytic plays a significant role in building, testing, and assessing theories. Empirical modeling in information systems has been dominated by statistical explanatory modeling, where statistical inference is used to test hypotheses and evaluate underlying models' explanatory power.

Research philosophy is concerned with the nature, source, and development of the phenomenon and the assumptions on which knowledge should be collected and analyzed (Bryman, 2007). All research aims to advance, refine and expand the body of knowledge, establish facts, and reach a new conclusion using systematic inquiry and disciplined methods. The research design is the strategy researchers often use to answer the research questions, underpinned by philosophy, methodology, and methods (Bryman, 2007).

Grounded theory is the methodology employed in scientific and non-scientific research studies. The theory is rooted back in epistemological objectivity and critical ontological realism (Annells, 1997). This methodology aims to produce or construct an explanatory theory that uncovers a process inherent to the substantive area of inquiry when little is known about a phenomenon.

Quantitative and qualitative data generation and extraction techniques are used in grounded theory research. The theory is grounded in data that has been systematically obtained through social research and analyzed using comparative analysis (Chun. & Francis, 2019). Qualitative grounded theory is a grounded interpretative paradigm which questions the idea that the methods and logic of natural science can be imported into the study of societies (Bryman, 2007).

The research method paradigm does not precede research but rather follows it so that model is grounded on the data generated by the research. Interpretative research has its roots in philosophy, human sciences, and anthropology. The methodology is modeled on the way in which human beings make sense of their subjective reality and attach meaning to it. When following this paradigm, data is gathered and analyzed in a manner consistent with grounded theory (Strauss & Corbin, 1990).

Quantitative grounded theory is grounded in the positivist paradigm that stands for the epistemological assumption that empirical knowledge is based on principles of objectivity, reproducibility, and verifications and is the foundation of all authentic knowledge (Bryman, 2007). The positivist paradigm defines a worldview to research that is grounded in what is known in research methods as the scientific method of investigation (Bryman, 2007). The distinction between interpretivist and positivist theories occurs at the paradigm level, and each method has explicit criteria for data collection, interpretation, and analysis of data. Pure quantitative research primarily relies on the collection and analysis of numerical data such as databases and anything measurable in a systematic way of investigating phenomena and their relationship.

The quantitative method is supported by the scientific paradigm, which leads researchers to regard the world as made up of observable and measurable facts. The results generated by the qualitative research are descriptive rather than predictive. The researcher is able to support and build theories for future quantitative research. In such cases, qualitative research method is used to derive a conclusion to support the hypothesis or theory being studied (Bhat, 2019).

Pure qualitative research methods vary and primarily rely on unstructured or semi-structured methods. Common research methods include non-numerical data such as textual data or visual analysis, interviews, and verbally administered questionnaires.

Quantitative methodology mainly follows the confirmatory scientific approach because it focuses on theory or hypothesis testing with empirical data to see if the hypothesis is supported (Blumberg, Cooper & Schindler, 2008). Where qualitative methodology mainly follows an exploratory scientific approach because it is used to describe what has been seen locally to come with new theories and hypotheses and is commonly used when the phenomenon or topic is little known or unknown. The fundamental difference between quantitative and qualitative is the logic of justification, not the method as a technique. The two research methodologies were generally developed from two different epistemological and ontological perspectives, and each method represents a distinct worldview or paradigm (Silverman, 2004). This distinction has had a remarkable breakthrough in social sciences studies, and the two methods have philosophical roots in the naturalistic and the positivistic philosophies, respectively (Garner, Wagner & Kawulich, 2016).

Contrary to the fundamental divergence in ontology and epistemology, a hybrid method may be needed as a route to increase validity and reduce bias. Many scientific researchers argue that is important to use mixed research of both confirmatory and exploratory methods in one research (Johnson & Onwuegbuzie, 2004). The combination of methods, therefore, treats research design and planning as a case of using methods that best facilitate an understanding of the topic rather than only adherence to a particular approach based on their theoretical underpinnings. Thus, in practice, research may greatly benefit from employing a pragmatic approach that includes both interpretative methods to better understand a topic (Bryman, 2007).

This research follows the grounded quantitative and qualitative theory approach. The quantitative method is derived from the consumer's shopping history stored in (POS) database. The qualitative research material is derived from the knowledge gained by conducting interviews with key marketing and sales managers. This is an essential exploratory approach to understand how the retail business operates and provides a preliminary understanding of consumers' shopping behavior as consumer patterns often change.
Table 2.1 highlights the major differences between qualitative and quantitative research methods.

Table 2.1 The difference between quantitative and qualitative methods

	Methods	Characteristics
1	Quantitative	1) In quantitative research, measurable data is gathered. 2) Use data in the form of numbers, and measurable data is gathered. 3) Requires variables to be predetermined. 4) The data collected through methods such as questionnaires (closed responses), record keeping, and population surveys. 5) The research is conclusive 6) Its methods of conducting quantitative research are structured interviews and observations 7) The purpose of quantitative research is to examine the cause-and-effect relationship between variables.
2	Qualitative	1) A method of inquiry that develops an understanding of human and social sciences to find the way people think and feel. A scientific and empirical research method used to generate numerical data by employing logical, statistical, and mathematical techniques is called quantitative research. 2) Use data in the form of words. 3) Qualitative research is exploratory. 4) Does not require pre-determined variables and can be used for open-ended or exploratory questions. 5) Data is usually collected through methods such as observation, interviews, questionnaires, focus groups, case studies, and document analysis. 6) Qualitative research is conducted to explore and discover ideas used in ongoing processes.

2.1 Research Methods in Predictive Analytics

According to Shmueli & Koppius (2011), predictive analytics had been available for a long time but rarely discussed in the information system discourse, either as a subject or a method. The method was predominantly used to validate empirical models using small datasets mainly gathered from qualitative research like interviews and surveys. The need for predictive analytics is rapidly growing due to the massive surge of unstructured and structured data, and there is the shift from prevailing BI tools to embrace the use of advanced analytics tools (Kotu & Deshpande, 2014). Therefore, the use of predictive analytics has become mainstream to gain a business advantage edge.

In this research, we will not be performing any systematic reviews of prediction techniques (in other discipline's literature). Shmueli & Koppius (2011) provided a model for the process of building and integrating a predictive empirical model in the information system. Figure 2.1 illustrates the 8 proposed steps.

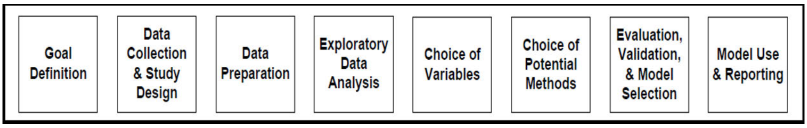


Figure 2.1 Steps in building an empirical model (Shmueli & Koppius, 2011)

The model is not cyclic and considers one iteration to design a predictive analytics model. The observation of the model seems to be more specific to provide guidelines for a single development (justify/build) and evaluation for a predictive empirical model. The first step is goal definition, which defines the purpose of the model design process or criteria constituting a rigorous design. The second step focuses on the quality and relation of the collected data at the time of the prediction used for explanatory modeling. The collected data is listed with its data source, and how it is acquired and if any issue is encountered. The data is then loaded in the target data source and queried to check its completeness. To capture new relationships and sources of information, this step investigates the number of the initial variables, as they are usually large. Each variable needs to be justified based on combining theory and exploratory analysis.

The data preparation step is to limit errors and inaccuracies in data before processing the predictive analytics and preparing the data for the modeling phase. This critical step consists of two sub-steps (missing values and data partitioning). It converts disparate, messy, and missing values in data into a consistent and clean view, such as, e.g., the CRISP-DM model (Shearer et al., 2000), but the development, justification, building, and evaluation phase.

The data preparation step is generally a time-intensive process that includes cleaning and transforming data. In data partitioning, the dataset is randomly partitioned into two parts. Training dataset, which is used to fit the model, and a holdout dataset which is used to evaluate the predictive performance of the final model. The third dataset, known as the validation dataset, is used for model selection. The final model is then selected based upon the validation dataset.

The exploratory data analysis step is selecting and presenting data graphically and numerically. It also involves processing dimensions reduction and handling anomalies. In this step, visualization derives insights and responds to risks more quickly. Visualization empowers analysts to utilize creativity and imagination in their mining for knowledge.

The fourth step is the choice of variables (predictors and response). The predictive model is based on association rather than causation between the (predictors and the response), chosen based on their observable qualities. Building on the understanding of the previous steps, we apply the best fit variables for the predictive analytics model plays a vital role in establishing the relationships in the model and the outcome achieved to help forecast the outcome's future. Prior to modeling predictive analytics model, some variables are more superior to other variables for the purpose of analysis. The choice of variables for causal explanation and representation depends on which variables the analysis employ; many causation treatments will reach a different conclusion about which causal relationships.

In choosing potential methods, choosing an appropriate methodology for the model is a complex and confusing decision. The choice often dictates the right analytical methods underpinning the model used. The type and size of the training dataset and the number of observations in data heavily influence the choice of the methods. A low variance algorithm is highly recommended when the training dataset is sufficiently small and the number of observations is lower as compared to the number of potential features. A high bias-variance algorithm is highly recommended when the training dataset is sufficiently large and the number of observations is higher as compared to the number of potential features (Kinha, Yogita, 2020).

The evaluation, validation and model selection step evaluate the model concerning the business requirements and what to do next. The results generated by the model are evaluated towards the model goals and cost-benefit relation. In this phase, checklists internal benchmarking can be used to assess the model results and create improvement ideas. The step mainly involves evaluating the performance of the predictive model, and the accuracy is measured by applying the chosen method/model to verify the generated predictions. It consists of two steps, model validation and model selection. The model validation is the assessment of the chosen model to compare the performance on the training dataset. The model selection aims to find the right level of model complexity that balances variance and bias to achieve the model's high predictive accuracy. This lifecycle step will improve the model in future iterations.

The final model and reporting step are properly using the results generated by the predictive analytics model to meet the research objectives. The assessment of the method used the accuracy and predictability of model statistical reporting overfitting analysis. The results are used to evaluate the actual precision of the research questions, discover new relationships, and benchmark comparison between the predictive model and more complex models to prove the added complexity and new features provide a useful contribution.

2.2 Approach and Methods in This Research

Research methodology is a systematic process of answering questions to acquire new knowledge.

There are various research methodologies in the field of information systems explored by researchers. Each methodology has its own approach and assumptions in data collection and analysis to achieve the research objectives. A predictive model leads to the discovery of new relationships and previously unknown hidden patterns. Predictive assessment offers ways to assess and evaluate the practical relevance of theories, compare competing theories, and assess the predictability of phenomena.

In this chapter, we discussed the role and representation of predictive analytics in mainstream Information System literature and the process of developing predictive analytics models in information systems. In planning a predictive model, the methodology approach should be based on combining research methods and specific model design. The choice of the methodology should be identified beforehand whether to employ a quantitative, qualitative, or mixed method. However, there will always be trade-offs when choosing the right research methodologies between what is theoretically best regarding the most rigorous research design and what is practically best given the number of constraints.

Finally, the research motivation, questions and conclusions must show how the empirical results generated by the predictive model contribute to the body of knowledge. The research contribution should be in terms of at least one or more of the six roles of predictive analytics discovering new relationships that potentially lead to new theories, contributing to measure development, comparing existing theories, improving existing theoretical models, assessing the predictability of empirical phenomena, and establishing the relevance of existing predictive models (Brodie, Mylopoulos & Schmidt, 2012).

2.3 Summary

Diversity in research methods is considered a major research strength in information systems. Researchers in information systems have employed many different research methods that can at one level be broadly categorized into two research: qualitative and quantitative.

Researchers should have competence in techniques and need to comprehend the theory of knowledge so informed decisions can be made and then reflect critically on the research. Many leading information systems journals were reviewed to understand the state of mixed research methods in the information system. Our reviews show a lack of mixed research methods in the information system. There are no guidelines or standards for evaluating and conducting such mixed research.

Mixed research method attempts to bring together methods from many paradigms. Quantitative research places emphasis on measurement during data collection and analysis. The quantitative method generally follows a natural science theory to establish objective knowledge that exists independently of the values and views of the people involved in the research.

The qualitative research approach follows an interpretative theory of knowledge. The method emphasizes meanings on words rather than distributions of numbers when collecting and analyzing data, primarily used to understand the business model. We seek to apply mixed research methods (Qualitative and Quantitative) in this research. Creswell and Plano Clark (2011) stated that the mixed research approach is designed with its own methods of inquiry and philosophical assumptions to provide the right directions for collecting and analyzing data from multiple data sources in single research.

Fetters (2016) argues that mixed method offers many benefits to approaching complicated research problems as the method integrates post-positivism and interpretivism philosophical frameworks and provides insights knowledge and holistic view into the study phenomena that can not only be fully understood using quantitative or qualitative methods. We use qualitative methods to understand the nature of the business and the employees' perspectives. It is an inductive research approach rather than a deductive approach. The client is exposed to many interviews using structured questions related to the nature and day to day business operations and model analysis according to generated results.

We use quantitative methods in our research in the form of POS database to discover facts, describe patterns and investigate the products' profitability, relationships, co-occurrences, and decline causes.

Chapter 3

Marketing

3 Marketing

The American Marketing Association defines marketing as: "an organizational function and a set of processes, concerned with communicating, creating, and delivering value to customers in ways that benefit the organization and its stakeholders" (Gundlach & Wilkie, 2009). There are several general definitions of marketing when it comes to modern academic research.

(Kotler & Levy, 1969) defined marketing as the performance of business activities directed to guide the flow of goods and services from producer to customer. Later, Kotler broadened the boundaries and marketing concept to include various non-business-related organizations (Kotler, 2019). He states that the essence of marketing is the transaction, defined as the "exchange of values between two parties". How markets create, facilitate and value transactions is of special interest to the discipline of marketing.

According to Bartholomew (2017), the marketing philosophy finds its origin in a business belief that efficient production is the business's main objective and that customer needs are the means of long-term business success. This philosophy has since evolved in a very progressive fashion over time, and nowadays, the line between marketing (as a general topic) and various marketing philosophies are blurred. The philosophy behind establishing the marketing strategy uses data to incorporate and distribute market intelligence throughout business departments and the coordinated implementation of a business's response to various market opportunities (Kotler, Armstrong, Harris & Piercy, 2013).

Morden (2017) argues that basing your business and marketing philosophy on customer satisfaction starts with the discovery of the customer's needs and demands. Thus, it begins with the customer and ends with the customer in a continuous process that aims to fulfil the customer's demands. Therefore, all marketing directs toward customer satisfaction, needs, and wants. Modern marketing seeks profits by creating customer satisfaction and aims to achieve this through an integrated, corporate-wide marketing program (Kotler, Armstrong, Harris & Piercy, 2013).

Marketing approaches have emerged over the years as a manifestation of the different business philosophies aimed at addressing customers at different times. Kotler, Armstrong, Harris, & Piercy, (2013) identify six distinctive marketing philosophies disciplines originating from six historical marketing eras. From the point of view of the central them of the period of time each of the philosophies were developed in, it is natural to present them as a cycle of philosophies as in Figure 3.1.

Figure 3.1 Marketing philosophy eras (Hedin, Hirvensalo & Vaarnas, 2011).

As each section of the circle notifies a step of marketing evolution taking place in the corresponding period, Table 3.1 highlights customers' point of view and key features of the evolution over the years, enabling marketing to become and remain a major influence in market economies.

Table 3.1 Evolution of marketing since 1800s (Kotler, Armstrong, Harris, & Piercy, 2013).

Years dominant	*Era*	*I.Customer's idea of products II.Features*	*Means*
1850s – 1900s	Production	i. Products are available and highly affordable. ii. Improve production and distribu-	Mass Production Low cost
1900s – 1930s	Product	i. Products of greater quality and price ii. Availability does not influence the purchase decision.	High quality Performance Innovation
1930s –1950s	Sales	i. Products that offer the most quality, performance, and innovative features. ii. The right product will sell itself.	Heavy promotional activities
1950s – Today	Marketing	i. Products to be developed according to needs and desires ii. Permeate every department from production to finance to human resources. Customer. All major decisions are based on relevant market considerations.	Integrated marketing
1970s – Today	Societal Marketing	i. Products give benefits to both customers and society. ii. Customer-focused strategies. Consideration the long-term interests of society.	Responsible for social and ethical issues
1990s – Today	Digital Marketing	i. Brand awareness and branding of social media. ii. Search engine marketing and optimization. Digital advertising channels and marketing analytics.	Niche and target marketing

3.1 Market Intelligence

Today the need for new vision and innovation has never been more pronounced for marketers

(Jennifer Polk (Blum, 2020).

In the digital (modern) era, the amount of data available for companies to gather business information is big and grows very fast (ŞERBAN, 2017). One of the important tasks for any marketing department is to know the data they can use for market analyses. Generally, timely market data is the major resource in gaining a competitive advantage with better and faster decisions, internal customer satisfaction and saved time. Hedin, Hirvensalo & Vaarnas (2011) point out that it is equally crucial as the source of new ideas and knowledge. Consequently, marketers need to reinvent how they engage with customers by using online and offline marketing channels to create creative, rich, and immersive customer experiences.

When it comes to market intelligence, the focus is set on the implementation of market data for marketing operations. For modern marketing-driven companies, this sets special requirements on business and marketing strategies. Market intelligence as a business process aims at responding to the quest of operationalizing market data (Jenster & Søilen, 2009). This is why digital marketing in many ways revolves around data and the analysis of this data used to optimize marketing strategies to attain a competitive advantage. Better data insights lead to better marketing campaigns and better sales. From the vantage point of data analysis, AI methods help make sense of this data (Bhor, Koul, Malviya & Mundra, 2018). From the process perspective (Hedin, Hirvensalo & Vaarnas, 2011), using data to keep track of the competition and the state of the industry is an integral part of operating any business. When combining these, by gathering, maintaining, and using market information, we come to Market Intelligence (MI).

Unemyr & Wass (2018) describe MI as the holistic or systematic collection and analysis of internal and external business data, such as the size of the market and number of vendors, that is then used to make decisions on the viability of the current business and markets. Hall (2020) adds that MI data at its core add knowledge about competitors, products and customer shopping trends or behaviors. MI has been regarded to hold exceptional future opportunities. It is considered an essential tool in digital marketing to map the right kind of promotional tools for specific market offerings. The continuous growth of data available boosts the use of MI in digital marketing. In this way, MI collects valuable information about the needs and preferences of customers, which is the pillar in helping businesses expand their digital marketing insights (Hall, 2020).

Kotler, Armstrong, Harris & Piercy (2013) point out that, by its nature, MI is the systematic collection and analysis of publicly available information about customers, competitors, and developments in the market. In recent years, the practice of collecting MI has expanded to include data and business analytics that can help companies improve their business model and projections (Unemyr & Wass, 2018). Franco, Magrinho & Silva (2011) add that MI aims at the relevance of data collected to promote marketing efforts. The objective is to paint an accurate portrait of the prospective market the company wishes to enter, with information about customers, challenges and growth potential for new products and services to help the company in its marketing decisions.

The term market intelligence relates closely to approaches like competitive intelligence for business, knowledge networks, opportunity identification, data mining for market knowledge, and news gathering for business. Because companies typically apply each of these terms in the context of their internal functions and particular data source, the practical ways of working with them differ considerably. To distinguish how MI differs from other approaches, the following summary outlines specific features of these practices (Table 3.2) found in literature sources.

Table 3.2 Market intelligence literature reviews key theories and practices

No.	Approach	Feature	Source
1	Market Intelligence	Umbrella term used within to encompass the external and internal collection of information. The importance of MI for effective business decision-making.	(Chung, Huang, Jin, & Sternquist, 2011). (Sivaramakrishnan, Delbaere, Zhang & Bruning, 2010).
2	Competitive Intelligence	The information the organization should attempt to collect and analyze effectively Described as the external sourcing of information for decision-making, outlined by the most important benefits of competitive intelligence practices: 1. Provide an early warning system for competitive threats.	(Franco, Magrinho & Silva, 2011). (Xinping, 2011).

		2. Improve the quality of strategic planning processes. 3. Enable faster responses to changes in the market.	
	Opportunity Identification	Entrepreneurial ability and mindset as opposed to merely possessing managerial capabilities. Ability to assimilate market information and translate it into beneficial actions for the business.	(Gruber, MacMillan & Thompson, 2012). (Angeli & Grimaldi, 2010).
3	Data mining for market knowledge	The acquisition and storage of data in a readable and accessible format.	(Liu, Cao & He, 2011).
4	Social Media	Use of web-based technologies to create interactive platforms to communicate and share, create, discuss, and change practice social media provides conflicting forces when it comes to market intelligence in that it is seen as inexpensive. Integrating social media into market intelligence allows the organization to develop competitive intelligence, monitor upstream trends, decode market signals, and Reinforce risk surveillance.	(Fournier & Avery, 2011). (Hanna, Rohm & Crittenden, 2011). (Kaplan & Haenlein, 2010).

A general understanding from this review is that the ability to gather and analyze data is critical for effective MI for any business, irrespectively of its size. However, these sources also outline the lack of clarity in the definition of Market Intelligence.

3.1.1 Generation of Market Knowledge

"*Market Intelligence is the continuous, cyclical process that starts from defining decision-makers information demands and ends at the stage of delivering contents that respond to those demands*."

Hedin, Hirvensalo & Vaarnas, (2011)

As the quote above outlines, MI process should be anchored to the existing corporate processes, such as strategic planning, sales, marketing, or product management. They also divide six phases that follow each other in the MI *research cycle* (as in Figure 3.2). In the first phase, the needs analysis, marketing defines the purpose and the scope for the task to drive the process and focus resources correctly. To collect as good data coverage as possible, yet do it efficiently, is the point of interest in the second phase. The pinpoint is in secondary information sources, meaning public data sources that are usually readily available and cost-efficient to use.

From here, the process covers the primary research sources, i.e., data collected based on the specific intelligence needs aimed at answering the most significant questions and problems. The analysis and delivery of the results occur in phases four and five. As the process is cyclical, there is room for feedback loops to verify if the data is sufficient for any conclusions to be drawn and to validate if the conclusions have business value. After these checks, the final sixth phase aims to produce well-informed decisions. It is quite easy to see how it resembles many other data analysis processes; for example, the CRISP-DM process frequently used in data mining and business analytics has similar features (Shearer et al., 2000).

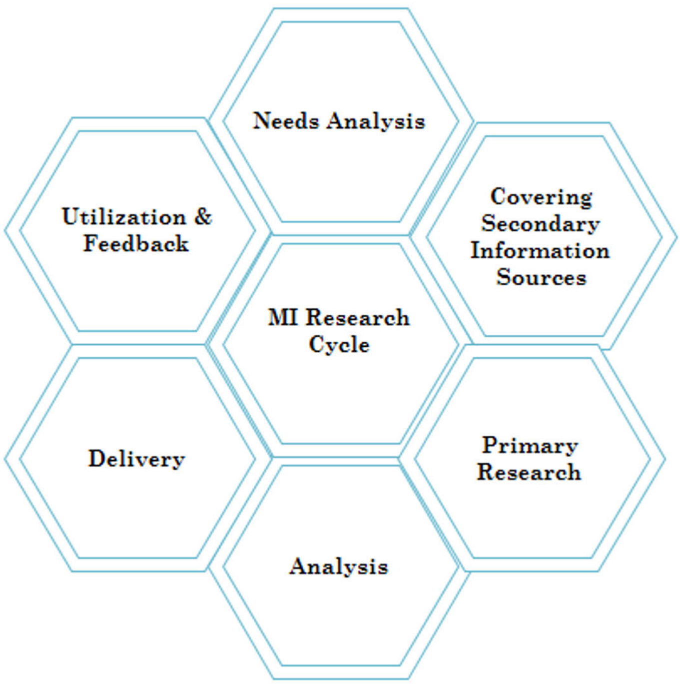


Figure 3.2 MI research cycle (Hedin, Hirvensalo & Vaarnas, 2011).

Unemyr & Wass (2018) present a few general criteria that any organization can use to evaluate the performance of MI concerning the future purchasing potential of customers and get feedback on current or new products and services and what triggers the customers' enthusiasm.

It is a central criterion to understand why and how customers choose and how brands can better align their marketing efforts to gain a competitive advantage. Both products and markets provide specific data usable for developing marketing approaches. To find product intelligence involves taking a deep dive into the various brand products and how these products stack up within the market. To do this, companies typically speak to customers, poll target audiences, or engage with them in surveys. As a result, organizations can better understand their products' differentiators and competitive advantages. Data used for market understanding revolves around examining the marketplaces populated by customers, either regular or prospective. Understanding the areas where target audiences are most active can help companies identify the right combination of media strategies, important sensitive points, and communication channels and use them to see where and how products can fit the markets.

As summary, Hedin, Hirvensalo, Vaarnas, (2011), point out how MI helps companies enter a new market, establish stronger brands and answer complex questions, like

According to Hedin, Hirvensalo, Vaarnas, (2011), MI also helps companies enter a new market and establish stronger brands, answer complex questions like:

1. Where should the company devote more resources?
2. Which markets should the company try to infiltrate next?
3. Are there patterns to what our best customers buy?
4. What products could be cross-marketed to existing customers?
5. Into what demographic segments can the company push new or existing products?

3.1.2 Market Knowledge for Small and Medium-Sized Retailers

The retail industry is arguably in a massive state of business transformation and fluctuation. It attempts to cope with the rapid-changing customer shopping habits caused by social media and online shopping (Jones & Runyan, 2013). While the underlying principles of marketing and MI are equally applicable to large and small firms alike, lack of sophisticated marketing is problematic for smaller firms (Romano & Ratnatunga, 1995).

Hedin, Hirvensalo & Vaarnas (2011) present how a study conducted by the European Commission to investigate how SME retailers gather data for MI, revealed many critical insights. Firstly, there is a gap between theory and practice of market intelligence gathering within SMEs. There is only little insight into how market intelligence for data gathering should be carried out. They write how there is a significant number of lessons found in the literature that have not reached practitioners in the SME industry. Moreover, the SME retailers are typically unfamiliar with the term “Market Intelligence”. Most importantly, SME retailers typically collect market information only after a particular incident.

Raj, Wong, & Beaumont (2016) more generally stated that SME retailers are still lagging behind in their use of AI and MI. On the other hand, Chaudhuri (2011) noted that today it is difficult to find a successful business that has not leveraged MI. The majority of information systems adopting MI are in large multinational enterprises. In order to survive in various conditions, SME retailers need to monitor their business and use all their resources efficiently, also those related to information. Deng & Luo (2010) point out that MI plays an increasingly significant role in an organization’s strategic decision-making. According to Peltoniemi & Vuori (2008), MI enables an organization to acquire relevant knowledge about the plans and actions of competitors on which management can base their decisions.

Prior (2007) states that MI is a factor for ensuring critical success as well as being the key to competitiveness for SME retailers. They should adopt MI processes to help reduce uncertainty-related decisions to the extent that they can make much more efficient choices regarding process improvements, cost reductions, product-mix choices, and new product introductions. Johns & Van Doren (2010) argue that MI fulfils a strategic role in businesses by providing quality insights, increasing general awareness, and improving opportunity and threat identification. Therefore, the benefits of MI acquisition for the SME retail industry far outweigh the costs. They outline four major benefits that can be achieved by adopting MI: 1. Market differentiation, 2. Cohesive marketing communication plans, 3. Pre-selling and cross-selling of an idea to the target market, and 4. building credibility with their customers.

According to Lueg & Lu (2013), market intelligence enhances budgetary efficiency as it increases simplicity, transparency, and friendliness. These factors are essential in improving data validation and thus increase SMEs' budget efficiency. Pranjic (2011) added more benefits gained from adopting Market Intelligence. 1. MI can detect profitable market niches, 3. Competitors' weaknesses and strengths, 3. Early warning signals in the event of political instability, 4. New or potential competition and 5. Economic recession signals. In addition, MI enables decoding competitors' planes or threats, improving analytical skills, ensuring faster and more efficient targeted responses to market changes, and identifying critical vulnerability points.

3.2 Analytics in Market Intelligence

"In God, we trust. All others must bring data".

Deming, W. E. (1986).

Business Analytics covers quite large a number of analytical and statistical methods and approaches. The core of analytics is finding the right data to help solve business problems and find business opportunities. In other words, data is the main resource – some call it the oil of analytics and modern decision support (Humby, 2006). As the prevailing approach to support decision-making with data – rather than with experience, authority or intuition – business analytics finds its roots in the sciences of operations research and management science (Power, Heavin, McDermott, & Daly, 2018).

The field of business analytics is not well-defined by how it uses data methods or statistical tools. The most common classification builds on how BA supports decisions and divides the discipline into descriptive, predictive, and prescriptive modes of analysis and decision support (Davenport & Harris, 2017).

1. The three analytical fields, respectively, try to answer very important questions like What is going on? What is going to happen? and What is the best action we can take? (Davenport & Harris, 2007).
2. Predictive analytics is used to predict the probabilities and likelihoods of market trends and marketing initiatives based on the available data, for example, If/then marketing hypothesis testing and simulations. Descriptive analytics is used to describe product portfolios, customers, and market segments and the relationship between factors. For example, it generates insights to understand the history of brands.
3. Prescriptive analytics is the field of discipline used to solve complex decisions and recommend the next steps/actions based on the analysis across data to prescribe specific marketing strategies. For example, identify a trending product and products mix in a specific region (IBM Center for Applied Insights. 2013).

Marketing data to business analytics comes from many sources of many kinds of data, and multi-sourcing is commonplace (Banasiewicz, 2013). The mechanisms for collecting are competitive intelligence (external business data) and business intelligence (internal business data). External market data is normally gathered by looking at secondary information sources, usually through research or conducted through a continuous monitoring process (Gordon, Perrey & Spillecke, 2013). Internal information for market intelligence is the collection, analysis, and distributing of information that can be sourced internally from marketing operations records, human resource records, client records, staff and owner/manager interaction with clients, as well as their interaction with other stakeholders such as friends, syndicated research reports, books, magazines, journals, etc. (Kotler, Armstrong, Harris & Piercy, 2013).

3.2.1 Market Segmentation (MS)

With the rise of the internet, digital marketing and social media focused marketing has become essential for retailers to better understand their highest-valued customers based on their unique shopping characteristics to maximize their

ROI from marketing campaign instead of pitching the same message to an entire market base (Dahlén & Edenius, 2007).

Today the retail landscape is changing rapidly, and forces like steep recession rapidly reduced customer spending, and at the same time, products choices are proliferating rapidly. Not every customer's shopping preference is the same, not every customer is profitable nor worth retaining, and not every product appeals to every customer. All customers have shopping needs of their own and steps that they go through before deciding to shop.

If the retailer attempts to appeal to all customers broadly, their target audience cannot simply be every customer who might use the product. Hence, the retailer looks for a fit between their competencies and profitability. Segmentation is the process of dividing customers into groups of individuals that are similar in ways relevant to marketing, such as nationality, gender, age, and spending habits. Failure to segment customers and treat each segment differently means that the retailer will be unable to identify unprofitable areas and utilize market intelligence to develop exclusively appealing products and prioritize tailor offerings in the most profitable areas.

An essential element of effective market intelligence is market segmentation (MS) which is based on various criteria, but the common goal is to satisfy the desires and needs of different types of customers (Wedel & Kamakura, 2012). MS is the process of dividing a large heterogeneous mass market of customers into clearly identifiable smaller homogeneous segments. MS divides customers based on certain criteria and similar shopping characteristics that lead to the need for the same products. Segments are made up of potential customers who most likely respond similarly to specific marketing strategies (Kotler & Armstrong, 2013). In doing so, the retailer establishes a clear target for the marketing efforts.

Market segments share common shopping needs, interests, wants and demands. They make planning marketing campaigns easier and help the retailer to focus the marketing efforts on specific profitable market niches instead of targeting the mass market. Segmentation techniques also help marketers to be much more efficient in resources like time and money. Camilleri (2018) stated these identified segments are then targeted with tailored marketing campaigns and clear marketing messages to those segments that are most likely to purchase a service or product. Such messages are referred to as positioning the product or service in the customer's mind to occupy a unique place.

According to Goyat (2011), MS is the way for smaller companies to succeed in big markets, while Marshall & Desborde (2019) stated that if the retailer is to enjoy any level of marketing success, this can be achieved through adeptness and the ability to match its capabilities to the needs and requirements of the marketplace. Central to this matching process is the segmentation of the market. If you fail to segment your customers and treat each group differently, you will be unable to identify underserved areas and utilize this intelligence to develop uniquely appealing products and services. If your competition uses such a strategy, this puts them in a more informed and powerful position. However, effective market segmentation must be dynamic in two ways and always need to be redrawn as it loses its relevance. The first way should concentrate on the customer's shopping behavior, which is rapidly evolving, rather than the customer's personality traits, which frequently persist throughout the customer's life. The second way is the segment is customer shopping behavior being rapidly reshaped by growing economic pressure, advances in technology, and new consumer niches that have their own purchase preferences and demands.

When building market segment, the first step is to establish the basis for segmentation. Various ways of segmenting the market depend on the nature of the industry, brand, and customer data available. For example, it makes sense to put a group of customers together to become more specific, so the retailer might look at female customers living in Helsinki who are interested in an outdoor training suit.

There is a number of *segmentation bases* and *segmentation types* available to build the segmentations. Segmentation bases are the customers characteristics that marketers use to separate an audience into segments, or groups, that can be targeted with customized or specific marketing campaigns such as Demographic, Geographical, Behavioural, and Psychographic bases (Kotler & Keller, 2006). that marketers may use as one or as a combination of more than one to segment a specific market.

According to Kotler & Keller (2006), combinations of different market bases have proven to be more cohesive and comprehensive views of the market and lead to decisions for marketers when identifying better-defined and smaller target groups. There are four most widely used segmentation bases in market segmentation literature (Papadopoulos, Martín, Cleveland & Laroche, 2011).

- Demographic segmentation is used when segmenting the market by size, age, sex and family type.
- Geographical segmentation, customers are classified according to their location population.

Behavioral segmentation, customers are segmented based on their behavioral variables such as occasions, benefits, user status, usage rate, buyer-readiness stage, loyalty status, and attitude (Kotler, 2019; Papadopoulos, Martín, Cleveland & Laroche, 2011).

- Psychographic segmentation, customers are segmented based on characteristics, personality, values, beliefs, lifestyle, attitudes, interests, and social class (Kotler, 2019).

Figure 3.3 Market segmentation bases: (Brotspies & Weinstein, 2019)

Segmentation types are the specific criteria or attributes used for segmentation. Generally, the type of segmentation used depends on the specific business objective. More specifically, they depend on the segmentation criteria used. For example, a marketer can segment customers according to their value. Apart from customer value, socio-demographic features, life-stage information, and behavioural aspects, needs, attitudes and loyalty characteristics work as segmentation types (Kolter, 2005) depending on the specific business objective and target. Different criteria and segmentation methods are appropriate for different situations and business objectives.

As an example, in behavioral segmentation, the customers are grouped by their behavioral and usage characteristics. Behavioural segments can be created with heuristics, such as a set of business rules, but this approach has inherent disadvantages. It can efficiently handle only a few segmentation fields, and its objectivity is questionable as it is based on the personal perceptions of a business expert.

Marketers may use one or a combination of market bases to segment their target market (Kotler, 2019). Demographics or geographic, behavioral, and psychographic features often work as segmentation bases. They can add several layers of

Market Intelligence to what the retailer already knows about customers. In such a case, market knowledge could build on key differentials, such as gender, age, location and spending habits.

3.2.2 Market Basket Analysis (MBA)

Affinity analysis, or association rules analysis, is the study of characteristics or attributes that go together. A frequent pattern is a useful model for extracting the salient features of the data. As mentioned in Chapter 1, it has initially been proposed for analyzing market basket data by Agrawal and Srikant (1994) and typically represents sets of transactions, each transaction containing one or more items.

Market Basket Analysis has attracted considerable interest because the found association rules provide a set of potentially useful information that is both concise and easily understood by the end-users. As Russell (1999) states, using MBA has allowed marketers to build theoretical models of shopping decisions involving items in multiple categories that help them make practical decisions.

As an exploratory data mining method, MBA helps businesses to understand patterns and traits of the business environment. It is mainly used in customers packaged goods (CPG) industry to determine which products are chosen together and how the purchase of one item influences the chances of another item being purchased (Griva, Bardaki, Pramatari & Papakiriakopoulos, 2018). MBA originates from the domain of marketing but is applicable in several other domains as well, such as geophysics, bioinformatics, finance, nuclear science, and geophysics (Fahy & Jobber, 2015).

MBA applies two types of association rules. The most common type of rules, the categorical association rules use simple binary data to find if the item exists in the customer's shopping cart or not; the customer either purchases the item or does not purchase it. The categorical association relates a value of a categorical variable with a value of another categorical variable. The second type of the rules, the *quantitative association rules* apply continuous variables (Aumann & Lindell, 2003), and a quantitative association relates a statistical summary of a continuous variable (mean, median) with a value of a categorical variable (Aguinis, Forcum & Joo, 2013). As MBA dynamically and iteratively creates an expectation based on a set of observations (Silvers, 2011), it helps to verify if business up is still up or down is still down. Szymkowiak, Klimanek & Józefowski, (2018) summarize that the process as consisting of discovery of sets of products (items) present in a large number of transactions (baskets) and that this technique aims to improve the effectiveness of marketing and sales tactics by using the customer purchase data accumulated with the enterprise during sales transactions.

From a theoretical perspective, MBA enables marketers to test theories and develop hypotheses inductively. The hypothesis that customers have a rational model to include associations across different explicit groups of items is testable with the MBA. For example, a sports retailer can test if customers, according to individual preference, tend to purchase sports shoes and water bottles together, despite them belonging to two different product groups, and find out that the products are complementary in terms of interests and activities of the customer. In this case, the MBA analysis can help the retailer's pricing strategy when they know that customers tend to purchase sports shoes and water bottles together, and they may avoid offering these items with a discount or for a price too high. As for a practical shop floor decision, the retailer may position complementary items next to each other or in the same aisle. Increasing the visibility of these items increases the likelihood that customers will find them easily and purchase both instead of only one of them (Russell & Petersen, 2000).

Finding an association between seemingly non-relating objects is of great interest for both the retailers and the scientists interested in discovering hidden but relevant connections in datasets. That is why it is not surprising that using MBA gives some highly relevant insights to both marketing theory development and practice. Cascio & Aguinis (2008) point out that: "*There is a serious disconnect between the knowledge that academics are producing and the knowledge that practitioners are consuming*." However, the use of MBA can produce knowledge that is both relevant and actionable and therefore, has the potential to play an important role in helping bridge the science-practice divide. The promise of the MBA is to lead not only to a substantive theoretical advancement but also to meaningful, practical information systems (Aguinis, Forcum & Joo, 2013).

The ability to keep track of the continuous changes in customers' tastes and predict a customer's shopping behaviors has become vital for retailers (Srivastava, Gupta & Baliyan, 2018). MBA is also one of the best examples of mining association rules and is considered one of the most widely applied techniques over Point-of-sales (POS) transactional data usage by large retailers to uncover associations between items (Wang & Sun, 2019).

As finding association rules is of great practical interest to retailers, there has been a steady and growing demand for efficient methods to search and extract them. Primarily the data mining community has carried out extensive research to develop advanced methods for ARM, resulting in a variety of algorithms that are analyzing data and extracting rules from them (Maimon & Rokach, 2009). Zhang (2002) points out that research work on knowledge discovery in databases (KDD), has helped to develop ARM methods that appeal to retailers with their ability to find essential patterns within point-of-sales databases (Zhang, 2002).

The mining of association rules is a two-step approach: first is the association rule extraction with the ARM algorithm, and the second step is the evaluation of the quality of the rule by the domain expert or using statistical quality measures. For the basic search strategy with the ARM, the retailer can choose between two categories of models: Candidate Generation (CCB) and Pattern Growth (PGB) (Borgelt, 2012).

According to Agrawal & Srikant (1994), MBA rules can be classified into qualitative and quantitative categories. Quantitative rules are defined regarding the type of attributes contained in these relational database objects (tables). Attributes can be either quantitative (salary, weight, etc.), categorical (a type of product, gender, etc.), or Boolean attributes, which can take on one of two options, therefore being a special case of categorical attributes (Griva, Bardaki, Pramatari & Papakiriakopoulos, 2018).

MBA uses logic to measure two events. The first event is an action taken by the company, the *driver object*, and the second is the response to that action, the *correlation object* (Kaur & Kang, 2016). Data from past occurrences of action and response provide the level of affinity. The affinity level is then used to predict future occurrences of responses and actions. MBA can pose an almost infinite array of questions with only three main elements (Itemsets, Object, and Affinity). The depth and simplicity of these elements allow the flexibility that is so essential in MBA (Fahy & Jobber, 2015). Below there is a brief description of each MBA element:

1. Itemset: The systematic interaction with customers is documented by the shopping cart in POS database, in which the customers place the items selected for a transaction.
2. Object: The items in each itemset are called objects. The objects of an itemsets are the products in a shopping cart.
3. Affinity: The probability that one object, such as a product, will occur simultaneously in an itemsets with another object.

3.2.2.1 Found Benefits of Market Basket Analysis (MBA)

The output of MBA analysis is a series of association rules, which enhance the efficiency of the retailer's operational efficiency, marketing strategies, and tactics (Berry & Linoff, 2004). Association rules take the form IF antecedent, THEN consequent; and have few key measures, such as support, and the confidence associated with the rule (Moreno & López, 2005). New algorithms have been developed to predict the itemsets the customer is about to purchase based on a certain group of itemsets that were previously purchased (Szymkowiak, Klimanek & Józefowski, 2018).

Berry & Linoff (2004), as well as Adhikari (2018), note, that association rules are beneficial and actionable for retailers, direct marketers, credit card companies and financial institutions, and so on, dealing with multiple products and services, as they provide companies with the opportunity to dig deeper into the data and continually monitor companies' affinities that achieve key competitive advantages. Fahy & Jobber (2015) outline some major benefits offered by MBA by stating that *"The retailer can use the extracted association rules from MBA to organize store layout and manage shelf space, placing associated products near each other so that customers would not forget to purchase all the desired products on a single shopping trip. Therefore, we can see that this kind of data analysis will help the retailer to determine the optimal store layout and eliminate the guesswork"*.

With MBA, marketers design various promotional marketing strategies, such as providing ideas on product bundling and develop cross-coupon programs where customers are purchasing a product and, at the same time, can get a discount coupon for a product. MBA with temporal components can be very useful to retailers and marketers for selecting cross-selling items. In addition, it is an essential element for designing and creating a recommendation engine (Szymkowiak, Klimanek & Józefowski, 2018).

MBA is applied in many fields of the retail industry to boost sales and increase revenue by identifying the needs and wants of customers. Below are some of their key aspects in which they play a significant role: the following summary outlines each practice (Table 3.3) found in literature sources.

Table 3.3 Market basket analysis practices

No.	Practice	Description	Reference
1	Cross-Selling	The retailer recommends some related products to a customer after buying a product. The cross-selling marking technique is used in a variety of domains.	(Parsian, 2015)
2	Product Placement	Retailers promote products through a non-traditional advertising way by placing complementary and substitute products together. This technique will ultimately influence the customer to purchase both products.	(Kotu & Deshpande, 2014)
3	Affinity Promotion	Promotional events are solely based on associated products. It is an instrumental technique used for preparing and analyzing the data, which comes in the form of a questionnaire.	(Kotu & Deshpande, 2014; Heggde &

			Shainesh, 2018).
4	Fraud Detection	Identify transactions that can be associated with fraud, where data contains credit card usage.	(Kotu & Deshpande, 2014).

3.2.2.2. Market Basket Analysis for Small to Medium Size Retailers (SME)

Nowadays, SME retailers are well aware that they are at a severe disadvantage to big retailers and mass merchandisers that have the financial capabilities and know-how to implement analytical solutions. For SME retailers, access to extensive customer data can be prohibitively expensive. One of the main scopes of MI tools, as well as related Artificial Intelligence (AI) solutions, is to gain a competitive benefit within marketing. The processes containing elements of MI and AI tend to enhance the enterprise performance, lower costs, raise sales, limit flaws and advance data gathering & processing. However, the main challenge that SME retailers face with the AI implementation, to have access to the required technical skills such as data gathering, data engineering, analytics, and machine learning, is a challenge in MI implementation as well (Coiera, 2019).

According to Unemyr & Wass (2018), Many SME retailers do not have enough cash flow to purchase or develop an AI solution. Therefore, digitization of this nature at this point is most prevalent in a large corporation. While large corporations have long ago started implementing new AI solutions to optimize business processes, SME retailers are still standing on the edge, trying to decipher if AI holds any answers for them. But the cost isn't the only barrier depriving SME retailers of acquiring such a tool. SME entrepreneurs sometimes find the whole concept overwhelming because of the lack of data mining expertise and the time resources required to have such a project (Griva & Papakiriakopoulos, 2018).

3.2.3 Integrating Market Intelligence with Database

Chiu & Tavella (2008) stated in their book *Data mining and market intelligence for optimal marketing returns*, that market share in isolation does not provide significant value since the company may be growing its revenue but at the same time losing its market share, while its competitors are growing faster. Insights on the past customer shopping data can often be misleading if the future needs of the same customer differ drastically from their past trends or needs.

When building market intelligence, it is important to have an integrated database system that links together data from marketing, sales, research, customer's transactions, and financial databases that need to be related through some type of identification such as marketing campaign or customer ID, transaction, and date of purchase. Marketers often encounter data quality challenges, like redundant and inconsistent data, missing data and typos. The best way to deal with poor data quality issues is to make sure the key client is fully aware of the imperfection of any data quality issue.

External Market Intelligence consists of data found in secondary information sources, usually through research or conducted through a continuous monitoring process. The goal of sourcing and analyzing published information is to build so called a three-dimensional view of a market (Chen, 2010; Bose, 2008a; Unemyr & Wass, 2018).

The dimensional marketing view reflects market segments, products to market segments facilitated by market channels, and relationships. This al-lows them to build dynamic shopper's profile to differentiate the brand and its services. Three-dimensional views, such as interactive scatter plot visuali-zations, help marketers describe and understand components of customer shopping journey data, answer specific commercial questions, such as: what the market potential is, what the competitors' future plans are likely to be, or what might the prices customers are willing to pay be (Chen, 2010; Bose, 2008b; Unemyr & Wass, 2018).

While much of the market intelligence insights are associated with the collecting information externally, internal market intelligence usually gains a great deal of insight from making better use of the internal information, such as Point of Sales databases or from an area of data known as 'Big Data'. Big data is a combination of unstructured, structured, and semi-structured data gathered by the organization mined for information and used in predictive modeling solutions (James, Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, 2011).

As James, Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, (2022) write, organizations use the big data accumulated in their databases to improve operation, better customer service, gain a potential competitive advantage, and ultimately create a personalized marketing campaign based on specific customer shopping preferences. For internal market information, Kotler, Armstrong, Harris & Piercy, (2013) summarize that the marketer collects, analyses, and distributes information sourced internally from marketing operations records, such as

- sales and client records,
- human resource records,
- occasions when staff and managers interact with clients, as well as
- their interaction with other stakeholders such as friends, syndicated research reports, books, magazines, journals, etc.

3.3 Market Intelligence from Business-to-Consumer (B2C) Relationship

Business-to-consumer marketing (B2C) refers to the transactions and strategies conducted directly between the organization and its customers (Buttle & Maklan, 2018). B2C marketing is one of the most important marketing approaches to date and holds significant business opportunities in the retail industry. Buttle & Maklan (2018) add that retailers can use B2C to increase margins, measure existing brand loyalty and leverage competitive advantage. At the same time, the strategy can increase brand awareness, provide important brand information to consumers, and gather very valuable consumer data to improve and expand business prospects (Johansson, Larsson & Hallin, 2006; Anaza, Kemp & Borders, 2018).

Stormi, Laine, Suomala & Elomaa (2018) present how B2C uses consumers' shopping data to enhance consumer acquisition and retention to produce cross-selling and upselling, to raise brand awareness, boost consumer lifetime value, and to improve consumer satisfaction.

3.4 Scenarios of Market Intelligence

For retail management, it is often useful to verify sales and other key performance measures over time. With time-series analysis, it is possible to extract meaningful knowledge from the shape of the data. When transactions data are analysed with this kind of longitudinal analysis, it is possible to track the products' behavioral changes over time, and when it comes to the association rules algorithms, to overcome the limitations of not accounting for the time aspect of the transactions data. According to Mekala & Srinivasan (2014) a time series database consists of sequences of values that are obtained over a stipulated amount of your time, and describes the behaviour of transactions.

As the retail industry often relies on product trends and up-to-date facts, such as price data, to make key strategic business decisions, adding the time aspect into MBA has a natural appeal. Hasan, Kabir, Shuvro & Das (2022) stated that predicting product sales of large retail companies is a challenging task considering volatile nature of trends, seasonalities, events as well as unknown factors such as market competitions, change in customer's shopping preferences, or unforeseen events such as COVID-19 outbreak.

If most recent transactions, together with older transactions, could be used to determine existing and expected trends, they would provide a rich source of knowledge at hand. However, the older the data used in finding trends, the less it tells about the current product popularity trends. Therefore, our objective is to highlight the importance of the most recent transactions and use them to find out up-to-date product trend information, thereby complementing the MBA model. For retail managers, such information, together with association rules, gives additional intelligence to make better-informed marketing decisions.

Product trends are highlighted by the changes in the profit from one period to the next or by experiencing a surge in the product's popularity, both of which can be found in sales data and contribute strongly to product profitability. In general, there are two types of trends, the upward trend, which leads to the increase in product profitability over time, and the downward trend, leading to a decrease in product profitability. To say it in a retail manager's words, "There are products that have not yielded profit but are likely do it in the future." In other words, managers expect profitability due to a recent change in profitability trending.

On the other hand, an established product that previously generated profit may have had declining popularity recently. This questions its future viability. When looking for a product's trending patterns in the POS data, the retailer can determine both opportunities and potential risks. If the product's trending popularity is declining, the retailer can make timely decisions such as spending more on the product sales efforts, reducing prices or discontinuing. However, if a certain product is selling off the shelves, the retailer should stock inventory accurately across channels.

Hebert, Anderson, Olinsky & Hardin (2014) used time series data mining methods to identify commonalities between sets of time-ordered data from a multi store chain database. The Time series data mining application detects similar time series using a technique known as dynamic time warping (DTW).

Retailers often fail because of their inability to find trending products to sell, due the fact that traditional data mining and statistical methods an are often inappropriate when analyzing data that possess a time factor (Hebert, Anderson, Olinsky & Hardin, 2014).

Mekala & Srinivasan (2014) propose a time series model to stream data patterns in a dynamic shopping mall to improve the prediction method. The model employs techniques from the statistical and data mining to improve prediction accuracy.

Finding one product that sells the best or the characteristics of trending products has become an arduous task, especially when every retailer is trying to follow the same marketing strategy. Several factors influence the demand and lifespan of a product, such as emotional factors, the first impression, impact of design, or the product's visibility. These can be summarized as perceived value-added experienced by the customer. To consider how the customer perceives a product and how satisfied the customer is with the retailer's prices and services is a central concern of any SMR retailer. If the average sales of the product decline continuously, the marketing manager can conclude that the product is falling from a beneficial category to a non-beneficial category, or customers are purchasing from other competitors. In a similar manner, but contrary to the previous case, the marketing manager makes positive conclusions about the product if the sales go up and the product is very profitable, and the manager ranks it accordingly.

The MBA model presented here offers an insight based on empirical data into the products and itemsets' trending profitability, identifies the history of sentiments attached to the product, and provides new kinds of insights into customer's shopping, likes, and dislikes, and thus works as an early warning system of potential popularity issues. With these tools, the retailer can pinpoint areas of success and failure, and marketers can accurately forecast products' future perceptions and maneuver for the future.

3.5 Summary

As business concepts and marketing philosophies have emerged since the 1950s, they have shifted companies' attention from production to marketing, from the product that the firm produces to the product the customers want and from the company to the market environment. Today's assumption is that goods are being made to satisfy rather than to sell. This chapter has highlighted major marketing theories, concepts, implementation, and reasons for businesses to practice modern marketing strategies and use AI's power in marketing.

Market Intelligence (MI) has a vital role in providing market insight, achieving growth, and promoting competitiveness and is often used interchangeably with a competitive advantage, which helps businesses understand and compete successfully in their business environment. As an AI program, MI collects critical information about the market and strategically relevant trends and topics and processes it into knowledge supporting decision-making. We demonstrated how MI should be integrated with retailer's strategic planning to gain a competitive advantage.

The ultimate goal of MI is to provide businesses with the know-how to understand, compete, and grow by implementing an organized MI strategy. As a key intelligence requirement, MI needs to anchor to the existing business or corporate processes, such as marketing and sales, strategic planning, product management, and innovation. There are substantial pieces of evidence that link the customer's shopping journey to AI methods in marketing, simultaneously leading to improved financial performance. We have also discussed the impact on the retailer to use decision intelligence based on MI insights, which is typically critical to the success of any retailer.

This research defines MI as an AI technique to help organizations look into the future and provide actionable insights to influence decision-making. In the next chapter, the AI theories, paradigms, implementation, and key algorithms employed in ARM will be discussed in detail.

Chapter 4

Artificial Intelligence

4 Introducing Computational Tools for Marketing Analysis

Computers have played a central role in marketing for the latest five decades with computers, marketers find decision support for various tasks, including customer relationship management and market segmentation. Kotler (1966) was the first scholar who highlighted the importance of computers in marketing to improve the efficiency and effectiveness of marketing decision-making.

Customer relations management (CRM) software is today one of the major market segments of business software and has an annual turnover of 40.2 billion USD dollars (grandviewresearch.com, 2021) and is expected to reach 145.7 billion USD by 2029 (Fortune business insights, 2020). There is a constant demand for new kinds of tools, new features of CRM as the methods of marketing evolve, and new types of marketing channels and tools become available (Constantinos, Christos & Stafyla, 2003). According to (Buttle, 2004), the rapid growth of CRM can be attributed to (a) economics of customer retention and customer lifetime value, (b) fierce business competition for valuable customers, and (c) advancement in technology.

CRM emerged in business management in the 1970s as a tool focused initially on automating the sales force and marketing services (Buttle, 2004). The technology evolved towards a concept of global customer relationship management whose goal is customer loyalty to improve the organization's results (Guerola-Navarro et al., 2020; Gil-Gomez, Guerola-Navarro, Oltra-Badenes & Lozano-Quilis, 2020). Consequently, CRM is completely aligned to modern customer-centered business management theory (Sin, 2005) because management can use it to analyze and plan sales, marketing, and service strategies, that lead the company to achieve and retain long-term partnerships.

CRM has experienced exponential growth since 2010 in terms of its deployment in all sectors and as a topic of scientific research (Gil-Gomez, Guerola-Navarro, Oltra-Badenes & Lozano-Quilis 2020). From a basic initial approach, the technology has evolved to a global vision of integral management of information about customer knowledge to achieve more effective customer interaction (King & Burgess, 2008). The current major trend in CRM is to establish channels and methods to manage customer-centered information to improve organizational performance and thereby obtain better business results (Gil-Gomez, Guerola-Navarro, Oltra-Badenes & Lozano-Quilis, 2020).

Various tools developed for CRM have proven essential for success in the consumer services and information technology sectors (Guerola-Navarro al, 2020). As CRM typically comprises of technology and strategic knowledge, it is a solution for more accurate data analysis and better business decisions (Krizanova, Gajanova, & Nadanyiova, 2018). According to (Schniederjans, Cao, & Ching Gu, 2012), the capability of CRM tools to profile customers is as important as the product, price, promotion, and place. The most effective CRM tools marketers use tend to simplify and automate many essential marketing tasks that allow marketers to build long-lasting relationships with their customer base. Modern CRM tools include communication tools that offer marketers the ability to track leads on which marketing efforts are generating the most revenue and to segment customers by demographic, geographic and psychographic factors.

4.1 Intelligent Decision Support for Marketing

Artificial intelligence enables machines to execute intellectual tasks related to human minds. Since the dawn of AI, marketers have developed an interest in using advanced AI-based tools for various marketing tasks (Han, Kamber & Pei, 2011). With the help of AI, marketers can process massive amounts of consumer data and fulfill customer expectations.

The various ways in which AI can be used in marketing are explained below in Table 4.1. based on Olson & Levy (2018). Parvatiyar and Sheth (2001) see technological dimensions of AI as ways of promoting cycles of customer management, where every organization should focus on significantly enhancing customer loyalty and sales growth. Today marketing departments deploy AI methods that significantly impact contemporary marketing practices (Sterne, J. (2017). Marketers can use AI methods for key strategic decisions like market segmentation, targeting, and positioning (Sterne, J. (2017).

Table 4.1 AI market methods (Olson & Levy, 2018).

No.	AI Method	Description	Reference
1	Market Segmentation	To slice a market into parts based on the customers unique needs and wants.	Netzer, Lemaire, & Herzenstein (2019) demonstrated that automated text analysis and correspondence analysis can be used for psychographic consumer segmentation.
2	Targeting	To choose the right segment(s) on which to focus the firm's marketing actions.	Drew, Mani, Betz, & Datta (2001) targeted customers using a combination of statistical and data-mining techniques. Simester, Timoshenko, & Zoumpoulis (2020) demonstrated optimizing promotion targeting for new customers using various machine learning methods.
3	Positioning	To bridge product attributes and customer benefits by finding a competitively advantageous position for the product in customers' minds.	Daabes & Kharbat (2017) demonstrate how data mining techniques can be used to distil a customer-based perceptual map, as an alternative to marketer knowledge, from mining customers' perceptions.
4	Association	To discover hidden knowledge related to basket clusters or purchase patterns in a large database	(Chen, Tang, Shen, & Hu, 2005) applied MBA in multiple store retail supermarket environments. The method was proven to be computationally efficient compared with traditional methods.

Most advanced techniques used in marketing today are AI and CI techniques, including data mining. This chapter will introduce the most common market segmentation and MBA techniques and present them in the framework of marketing decision support.

According to Müller (2012), AI is a computer system capable of demonstrating complex living systems, like behaviour, and attempting to understand and build intelligent entities that think and act rationally like humans to solve problems or make rational decisions. Stone & Woodcock (2014) propose a framework to apply AI for marketing decisions, both planning and decision-making. In Table 4.2, Stone & Woodstock (2014) present a few major areas to deploy AI in marketing.

Table 4.2 AI deployment in marketing (Stone & Woodcock, 2014)

No.	Marketing Activities	AI Deployment
1	Customer Service	Assisting in reaching target audiences, e.g., with machine learning.
2	Target Markets and marketing mix	Assessing different potential outcomes based on alternative marketing strategies.
3	Branding	Using evidence from the internet and social media to track and shift brand image. Finding evidence of the reasons causing brand shift and loss of market's share.
4	Competitive strategy	Identifying weak signals of competitors and weaknesses in competitors and own marketing strategy.
5	Direct Marketing	Choosing from or combinations of contact types, channels or content appropriate for different target markets and individuals.
6	Advertising	Designing images and videos to fit individual and market segments in different marketing channels.
7	Sales Promotions	Identifying, choosing, and timing promotions and offers to market segments or individuals.
8	Distribution	Identifying missing channels, optimizing marketing channels, and improving transaction time and streamlining fulfillment.

4.2 Artificial Intelligence

When marketers look for solutions to their analytical problems, they expect the tools they use to be both effective in the sense of results they provide to decision making and efficient in the sense of resources they use to attain these results. Since every customer and every customer relationship is unique, the methods sought after are the ones that can reveal this uniqueness and help marketers to concentrate their marketing efforts correctly with insights from computational analysis. For this purpose, traditional CRM software provides only a few effective tools (Nguyen, Simkin, & Canhoto, 2015).

On the other hand, AI is a branch of computer science that provides a number of software tools to deal with individual tastes and preferences so that they can be applied and implemented in market segmentation and MBA, and deployed as efficient marketing campaigns and individual offerings (McCarthy, 2007). In addition to creating human-like behaving machines, modern AI provides a very powerful platform for solving a wide range of super-complex optimization problems (McCarthy, 2007). Furthermore, AI is not biased by rules, which could be described as conventional wisdom, and preconceived notions the human decision-making process relies on (Scherer, 2016).

The core of AI lies in its algorithms. According to Burgess, (2018), an algorithm is a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate and predict an answer to a problem (Burgess, 2018). Most AI algorithms use probability to find the optimal output to a specific problem or goal from a range of inputs. To perform well, the algorithm has to be fed with lots of data to train the model or system and learn from its past mistakes made during the process (Guida, 2019).

One of the main divisions of AI technologies is the distinction between supervised and unsupervised learning (Burgess, 2018). Today, most AI systems apply supervised learning, which refers to systems trained by using large amounts of known data. On the other hand, unsupervised methods refer to an AI system confronted with an extensive unknown dataset (unlabeled data), where the system can identify clusters of similar data points within the dataset (Burgess, 2018).

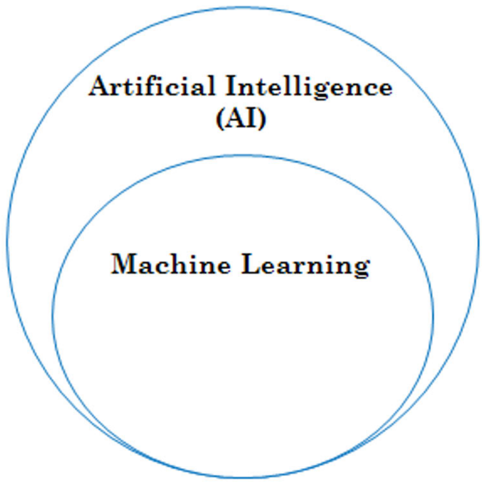


Figure 4.1 The relationship between AI and ML (Zhang & Tang, 2021)

AI is a technology with great potential ML (machine learning) is fundamental for developing AI-based algorithms that can achieve accurate results.

4.2.1 Development of AI

Programs that computationally perform actions of human-like behavior, such as thinking, learning by example, doubting, acting, seeing, and speaking, have the common name AI. Developers of AI have drawn influences from the ways humans and other biological beings accomplish tasks. The roots of this kind of approach to developing technologies lie in the dawn of human history. McCarthy, Minsky, Rochester & Shannon (2006) note that humans have always pursued ways of decreasing manual and mental labour with machines. Originally, this pursuit concentrated on the manual side of labour, but since the 19th century, the advancement of sciences and technologies, such as mathematical logic, automata theory, cybernetics, information theory, computer science, and psychology, has laid the foundation for the development of computing and AI.

According to Simon (1996), computational support is required for a few reasons, not only because of computational effectiveness but also because of boundedness of rationality. Simon further explains how early AI developers were quite optimistic about the development of the branch, but the many aspects of the human way of thinking made the development a rough road to travel (Simon, 1996). Simon argues that humans are limited in their ability to process information and lack the cognitive ability to calculate an optimum solution and recognize possible alternatives. He then proposes bounded rationality as a realistic concept of human decision making, where humans make satisfactory rather than optimal choices. He considers humans to be "simply flawed" because the human capacity to understand the world and solve complex issues is very small compared to the size of real-world issues that require objectively rational human behavior.

The first generation of AI solutions was limited to specific tasks, such as performing problem solving, training computers to mimic basic human reasoning, symbolic methods, multiplication and integrating statistical data analysis (Copeland, 2015). Since the term Artificial Intelligence was presented at the Dartmouth Summer Research Project on Artificial Intelligence in 1956 (McCarthy, Minsky, Rochester & Shannon, 2006), many AI-based systems and decision models have been developed and applied successfully in domains such as Natural Language Processing, Machine Translation, Pattern Recognition, Robotics, and Image Processing.

According to Scherer (2016), advancements in AI technologies over the years will likely increase the frequency of unexpected AI behavior substantially over the next years. This leaves the question of how much risk is involved in driving AI forward without proper supervision and regulation open.

4.2.2 AI as Decision Support for Marketing

There is extensive literature on strategic decisions, and it arises from the need to understand how strategic decisions are made rather than the outcomes of these decisions. However, there is little literature available on the strategic decision-making process in marketing, as most literature covers only general management, specifically financial decision-making related to mergers and acquisitions.

However, the literature about strategic decision-making has not been linked to the use of AI technologies in decision-making (Stone, et, al., 2020).

The distribution line between the roles of computers and humans in management, that used to be within operational management, is now in the strategic management as well. This calls for a wider discussion on the encroachment of AI to business functions, such as marketing. The focus of the discussion is mainly on the ability of AI applications to help humans make sense of high and growing volumes of data, usually referred to as big data (Jarrahi, 2018). Claudé & Comb (2018) state that today, AI is primarily seen as a support to strategic business decisions rather than as a decision-maker. As computational speed and capacity increase and as data available to support decision-making grows, the frontier of substitutability of AI technologies for human decision-making shifts (Stone, et al., 2020). Shrestha & Mahmood (2019) proposed the following alternatives for the process of change:

1. Full Humans to AI delegation, for example, digital advertising, recommender systems, dynamic pricing, and online fraud detection.
2. Hybrid 1: AI to Humans Sequential Decision-Making, for example, recruitment.
3. Hybrid 2: Humans to AI Sequential Decision Making, for example, health monitoring and sports analytics.
4. Aggregated humans – AI decision making, for example, board of directors and senior management.

They see that in strategic decisions, the time it takes to see if a particular approach works, the lack of specificity, possible diversity of interpretations, the fact that speed is less of the essence and relative lack of replicability may mean that aggregated human-AI decision-making is more appropriate. However, they note that Hybrid 1 may be used too.

According to Gordon, J., Perrey, J., & Spillecke, D. (2013), marketing has changed drastically and will continue to change in the future, so the scope for AI applications in the future of strategic marketing decisions making will continue to change in the following five areas:

1. The evolution of marketing analytics towards AI and changes in market research setting.
2. Business-to-business (B2B) marketing.
3. Patterns of expenditure and distribution in consumer markets.
4. Marketing public sector.
5. Data protection and ethical issues.

Maimon & Rokach (2007) state that marketers invest in a variety of innovative AI technologies to become more customer-centric, competitive, and more responsive to the industry and customer demand. In a survey conducted by infosys.com, a pioneer in marketing research, 87% of retailers have deployed some form of AI or automation technology as part of their operations, decision-making processes, and guidance for human decision-makers (Infosys.com, 2017).

This illustrates how vital autonomy in systems and processes is in industries where fast-paced transactions to deliver functional business insights are the core of the business. Automating decision making through AI can help retailers to reach vital decisions like what and when to order and what products to merchandise at the front of the store or on the first page of the retailer's site. Moreover, to decide about individual cross-selling and up-selling opportunities on historical shopping behavior and current basket contents is of central interest of AI in marketing (Stephenson & Bishop, 2005).

4.2.3 Computational Intelligence and Marketing

Chowdhury, Sadek, Ma, Kanhere, & Bhavsar (2006) state that AI methods include two major categories – symbolic AI and computational intelligence, that is CI. Symbolic AI concentrates on developing knowledge-based systems capable of making decisions in a particular domain utilizing knowledge from a human expert. CI consists of fuzzy systems, neural networks, and evolutionary computing methods. The difference in CI from symbolic AI is that the output is generated without using knowledge bases such as rules, frames, or cases.

A major thrust in algorithmic development is the design of algorithmic models to solve increasingly complex performance problems. Modeling of natural intelligence as so-called intelligent systems has achieved considerable successes (Engelbrecht, 2003). Several research disciplines, such as computer science, philosophy, physiology, sociology, and biology contribute to the development of AI (Eberhart & Shi, 2011). Quite often, scholars and scientists disagree about including or excluding specific methods under the umbrella of CI, which has led to different definitions (Engelbrecht, 2003). Although CI has a history of over 59 years, it has only recently gained widespread popularity and somewhat different flavours (Engelbrecht, 2003).

Many research journals with "Computational Intelligence" in their title are being published, the Finnish Julkaisufoorumi lists about 10 of them (jufo.fi). As a scientific discipline, CI is widely recognized as a branch of AI as it studies adaptive mechanisms that enable or facilitate intelligent behavior in complex and changing environments (Eberhart & Shi, 2011). These mechanisms are related to the AI paradigms that exhibit the ability to learn or adapt to new behaviors, to abstract, discover, and associate.

Bezdek (1994) defined CI as a computationally intelligent system that deals only with numerical (low-level) data. Poole (1998), in his book "Computational Intelligence – A Logical Approach," states that the term CI is a typical AI symbol, and AI is the established definition for the field of CI there as it focuses on logic and reasoning. Chen (1999) has a similar definition of CI as the field of studying how to build intelligent agents. Engelbrecht (2011) defines CI as the study of adaptive mechanisms to enable or facilitate intelligent behavior in complex and changing environments.

As there are no effective or common algorithms, either because it is not possible to formulate them, this takes us closer to the definition of CI as a sub-branch of AI and a branch of computer science studying problems for which there are no effective computational algorithms (Eberhart & Shi, 2011). A good part of CI research is concerned with low-level cognitive functions: perception, object recognition, data analysis, discovery of structures in data, control, and simple associations. All methods developed for this type of problem include both supervised and unsupervised learning algorithms (Engelbrecht, 2003). Goldberg & Goldberg (2018) notes that CI is more of a way of thinking about problems and calls for a broader view of the scope of the discipline.

Eberhart & Shi (2011) notes that CI methods are capable of yielding results in a relatively short time. For example, implementing a conventional expert system often takes more than one year and often requires active participation of knowledge experts to build the rule bases and the knowledge base. In contrast, CI systems can often be prototyped in a few weeks to a few months and are implemented using available engineering and computational resources.

A long-term goal for CI is to create cognitive systems that could compete with humans in a large number of areas. So far, this has only proven to be possible in restricted domains, such as processing a large amount of numerical information, recognizing specific patterns, and memorizing numerous details (Duch, 2007). Zhao & Liu (2015) surveyed relevant studies regarding different CI algorithms utilized in unmanned aerial vehicle (UAV) technology planning and the types of environment models and found that CI methods outperform traditional methods. The analysis of the CI studies proved to be useful for identifying key results and trends from UAV technology planning research. Faizollahzadeh, Najafi, Shamshirband, Minaei Bidgoli, Deo & Chau (2018) investigate the state-of-the-art CI methods employed in the hydrogen production industry to identify better performing CI methods in the optimization, prediction, and assessment tasks related to different types of hydrogen production methods. The analysis results provide in-depth knowledge of different hydrogen production methods and modeling techniques. The study identifies CI methods measured by a qualitative analysis of the accuracy of CI in modeling hydrogen production.

Cavalcante, Brasileiro, Souza, Nobrega & Oliveira (2016) conduct a study on CI. The study provides an extensive overview of the most important CI and machine learning studies published from 2009 to 2015. They cover the methods and techniques for clustering and pre-processing financial data to forecast market movements and mine financial text information. The study contributes comprehensive reviews of literature in the CI field, a definition of a systematic procedure to guide the task of building an intelligent trading system and a discussion about the main open challenges and problems in this CI field.

Pedrycz & Chen (2014) conduct a study to examine the possibility that the CI driven sentiment analysis methods can support and enhance the quality of the decision-making process when information to the system is generated from the Web 2.0 economy. No feasible finding was found to confirm that CI methods can support the decision-making process. According to Szymanski, Sarnatowicz & Duch (2007), there are major challenges for the CI community to propose more efficient knowledge representation and retrieval structures using different knowledge representations. They consider employing the probabilistic CI methods to be combined with AI methods to form a hybrid algorithm for a faster and more efficient mining process.

4.3 Data Mining and Knowledge Discovery

Modern information and communication technologies consist of massive amounts of data in databases and other repositories. Transforming the insights about data into knowledge with insights can help companies make better business decisions (Singh, Choudhury & De, 2012). As a subset of AI, data mining is a mathematical engine of the knowledge discovery process (KDD). It transforms the exponential amount of raw customer raw data into actionable and predictable solutions and provides marketers with intelligent recommendations and valuable customer product knowledge (Berry & Linoff, 2004). Consequently, when it comes to explaining how the most successful retailers in the world got to where they are today and how they will continue to dominate the market in the future, many experts and scholars point to their proficiency in data mining (Conroy, Porter, Nanda, Renner & Narula, 2015; Damien, 2019; Wen, Yeh, Tsai, Peng & Shuai, 2018).

Knowledge Discovery in Databases (KDD) is an automatic, exploratory analysis and modeling of large data repositories (Maimon & Rokach, 2009). Data mining plays a central role in the knowledge KDD, as it is used to infer algorithms to explore the data, to develop the knowledge discovery model, and to discover previously unknown patterns. The objective of the developed model is to understand phenomena retrieved from the data, analysis, and prediction (Maimon & Rokach, 2009). The steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data. Data mining techniques have been used in many domains to solve classification, segmentation, association, diagnosis, and prediction problems (Boire, 2017).

According to Tufféry (2010), data mining has been touted as one of the most remarkable technologies to hit the retailing industry this decade, as it is a powerful analytical tool for finding overlooked and useful information. Tufféry adds:" in *the retail industry, data mining is used to provide analyses on product sales and identify naturally occurring clusters of customer shopping behavior, which then form the basis of market segments*."

When we register data into the database, some regularity, similarity, or outlier usually occurs and creates patterns. (Witten Frank, Hall, & Pal, 2016) stated that in market basket analysis, we could see that people who are buying gin are also buying tonic. This is a well-known pattern, but the introduction of the MBA there has provided tools to discover many unknown patterns, e.g., the famous example discovered between beer and diapers. The discovery of this kind of patterns depends on the applied data-mining task, which can be either descriptive or predictive (Witten Frank, Hall, & Pal, 2016).

Descriptive explanatory functions typically describe the relationship between factors. On the other hand, predictive functions and algorithms predict a variable based on the available data. (Witten Frank, Hall, & Pal, 2016).

Summed up the difference between the two approaches with this quote: "*Predictive modeling is all about what is likely to happen, whereas explanatory modeling is all about what can we do about it."* In KDD process, there are two primary data mining goals: verification and discovery, to verify user's hypothesis about the data and to discover unknown patterns automatically. In the next paragraph, we will explain each process goal in detail.

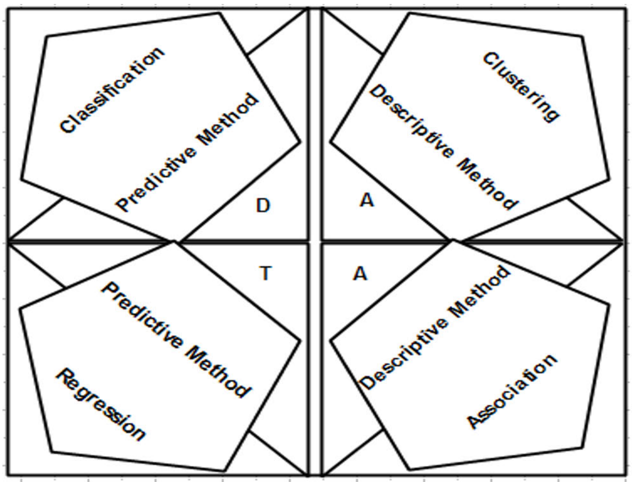


Figure 4. 2 Data mining taxonomy: source (Nguyen, Simkin, & Canhoto, 2015).

The domains of implementing various descriptive and predictive methods have been described by Nguyen, Simkin & Canhoto (2015, see Figure 4.4). They describe how predictive methods aim to discover unknown or potential values of another dataset of interest, while descriptive methods mainly find patterns and come up with new information about the available dataset. As classification algorithms derive models to identify the class of an object based on its attribute, in direct marketing, it can be used to reduce cost. Regression identifies sequences of events based on one or more preceding events, e.g., in stock market regression models.

Descriptive association methods discover connections between items in data set. Association analysis is mainly used in the retail industry, advertising, and direct marketing. Clustering methods identify and describe data objects similar to one another. Similarity exists in various factors, like customer responsiveness to certain actions, purchase behavior and geographical locations. Clustering is popular in the retail, medical, and finance industries.

Marketers can segment markets with data mining, as it is the core process to divide customers into homogeneous and meaningful subgroups with customers-specific attributes and characteristics. As a differentiation marketing tool, it enables marketers to understand customers and build targeted and differentiated strategies (Kotler,2013).

According to Ziafat & Shakeri (2014), organizations traditionally tend to use market segmentation schemes based on demographics and value information, regardless of the industry they operate in. Over the past few decades, they have decided on their marketing activities and developed their new products and services with these segments that have taken shape from such simple business rules. Companies use such unique groups of customers they have identified to target and manage customers more effectively. Among other things, they provide customized product offerings and promotions.

Usually, a market segmentation project starts with the definition of the business objectives and ends with the delivery of differentiated and targeted marketing strategies for the selected segments. Then marketers use market basket analysis to make sense of the company's data and scale up the effectiveness of marketing and sales processes. MBA results provide guidelines in the decisions making processes, for example, in products placement in stores, choosing cross-sell and up-sell categories and in launching targeted marketing campaigns or promotions to customers based on their purchasing behavior (Dolnicar, Grün, & Leisch, 2018).

4.3.1 Data Mining Process and Its Outcomes

Next, we introduce and explain our approach from the vantage point of the data mining process and show how to position data mining within the tiers of KDD. From the point of view of analysis process, we also discuss our approach from the point of view of CRISP-DM (Shearer et al., 2000) as to show how the model for MBA is developed.

As Kumar (2010) pointed out already more than a decade ago, there is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information and knowledge from the rapidly growing volumes of digital data. Such theories and tools are subjects to the emerging field of knowledge discovery in databases (Kimball & Ross, 2011). At an abstract level, the KDD is concerned with the development of methods and techniques for making sense of data. From a more pragmatic vantage point, data mining is a step of the iterative KDD process (Han, Kamber & Pei, 2011).

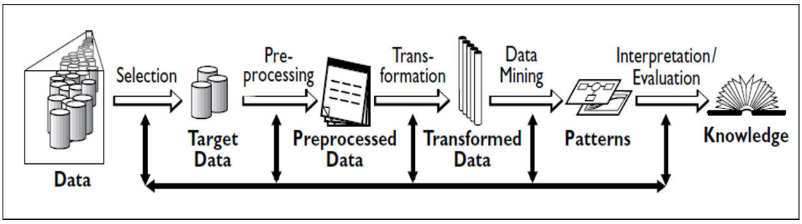


Figure 4.3 Knowledge discovery in database process: (Fayyad & Smyth, 1996)

Data growth results to a great deal from the incorporation of smart systems into corporate information systems. They use various AI methods to analyze exponentially growing data mass, the so-called Big Data. It seems that while data availability increases exponentially, the human processing ability stays almost constant. As the ultimate goal of the KDD is to extract useful knowledge from large databases interactively and repetitively, the process has developed to consist of the few steps described by Fayyad, Piatetsky-Shapiro & Smyth (1996) in Figure 4.3. As the process is repetitive, it means that process might have to move back to the previous step or steps.

The overall goal of the KDD process is to find and interpret patterns from dataset involves repeated steps (illustrated in figure 4.3) to make the decisions of what qualifies as useful knowledge.

Selection of target data is the initial preliminary step that involves creating a target database on which knowledge discovery is to be performed. Once the target data is defined in the source database, then the objectives for the KDD process is determined. The target data is then transferred to a data warehouse database into one unified dataset, and the data discovery is to be performed. This process ensures that data discovery is performed for accurate analytical results.

The objective of the pre-processing step is to improve the reliability of data. The step incorporates data cleansing process to remove of outliers and noisy or irrelevant data from the source data. ETL (Extract-Load-Transformation) tools are the main tools of automation in this step. Data transformation step transforms the pre-processed target data into the form required for mining. During this step, data is organized, often converted from one type to another (i.e., changing nominal to numeric), and new or "derived" attributes are defined. ETL (Extract-Load-Transformation) tools are the main tools in automating this step.

Pattern discovery applies data mining methods, such as machine learning algorithms to extract potentially useful hidden patterns. For example, summarization, classification, regression and clustering are used in this step. The choice of the methods and algorithms relies primarily on the KDD model objectives and on the success of the previous steps. Upon successful completion of the last step, the knowledge discovery, the knowledge activates in the sense that the organization now is able to use the knowledge and measure the effects of the KDD model. The discovered knowledge consolidates, and the user gets the results in a simple and easy to understand representation.

The most comprehensive KDD implementation process model is the CRISP-DM life cycle model (Stirrup & Ramos, 2017). *The Cross-Industry Standard Process for Data Mining* appeared in late 1996 and was built on previous attempts to define knowledge discovery methodologies (Plotnikova, Dumas & Milani, 2020).

CRISP-DM organizes data mining process into six phases, where repeatable approaches are recognized. The phases help organizations understand data mining process and provide a road map to follow while planning and carrying out data mining projects. CRISP-DM is a powerful vendor-independent process for developing data mining projects soundly based on the real-world experience of how scientists conduct data mining projects (Plotnikova, Dumas & Milani, 2020). Based on Stirrup & Ramos (2017), the six phases of CRISP-DM are shown in figure 4.4

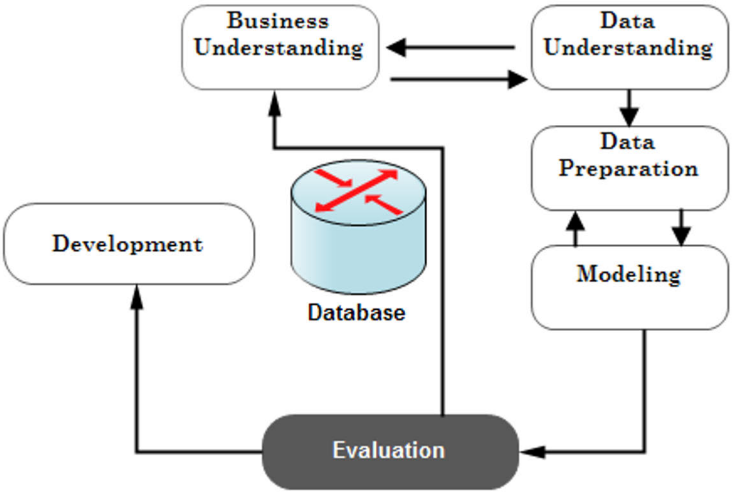


Figure 4.4 - The phases of CRISP-DM: (Stirrup & Ramos, 2017)

The process provides a comprehensive overview of the life cycle of a data-mining project. The generally described phases, their respective tasks, and the relationships between these tasks help users to grasp how the process progresses, what is needed in each phase and how the outcomes of each phase should be understood. Next, the phases are outlined.

The initial phase is Business understanding, and it focuses on understanding the project requirements and objectives from a business perspective. Establishing a clear business understanding is essential to convert this knowledge into data mining and a preliminary plan designed to achieve the project objectives. Data understanding adds to the foundation of Business understanding. It starts with initial data gathering and then identifying the data quality. In this step, project discovers and detects interesting subsets of data to form hypotheses concerning the discovery knowledge.

The data preparation phase involves all activities needed to prepare and construct the final dataset to modeling. Usually, analysts need to repeat this step many times. The tasks in this phase include selection of tables, records and attributes selection, as well as transformation and cleaning of data for the data-mining algorithm. The data preparation phase is likely to take 50% to 80% of the project effort. In the modeling phase, the data scientists develop and assess the data mining model often using various modeling techniques, and their parameters are adjusted to optimal values. Some techniques have specific requirements in the form of data. Therefore, going back to the data preparation phase is often necessary.

The evaluation phase verifies that the data-mining model meets all the business requirements and assures the quality of data analysis. Thorough evaluation and review of the executed steps are essential to proceed to the deployment of the data mining model. A key objective is to determine if some important business issue has not been sufficiently considered. Finally, in the deployment phase, the development of the model and the knowledge gained will need to be presented to the end-user, even if the purpose of the developed model is to increase knowledge. This phase often involves deploying the model within an organization's decision-makers.

4.3.2 Market Segmentation Project with Data Mining

For market segmentation, data mining provides methods and tools that have a natural appeal. Marketers can utilize customer purchase data, and as a result, retrieve customers' purchase behavior and form strategic marketing initiatives. The business objectives of the market segmentation project are usually aimed at fulfilling these very objectives. With differentiated marketing strategies for the segments, the retailer typically gets results where 10 to 30 percent of customers generate 50 to 90 percent of the company's profit. As a data-driven discovery process, data mining provides market segments with similar customer behavior. For example, clustering algorithms that analyze behavioral data, can identify natural groupings of customers and suggest a solution founded on observed data patterns. A properly designed and programmed data mining model can uncover segments with distinct profiles and characteristics and lead to rich segmentation schemes with business meaning and value.

Rostam (2020) proposes data mining for the development of segmentation schemes based on the current or expected value of the customers. Such value-driven segments are necessary to prioritize customer handling and marketing interventions according to the importance and value of each customer. Moreover, since a vital part of a segmentation project is an insight into the derived clusters and an understanding of their meaning, we will also propose ways to profile the clusters and outline their differentiating characteristics.

4.3.3 Data Mining for Anomaly Detection

Anomalies are patterns in data that do not conform to any well-defined notion of normal behavior. Anomaly detection tackles the problem of having and discovering such patterns in data. It is defined as the process of using models to identify behavior that is different from the normal behavior of a system (Chandola, 2009), and it can be applied to any type of data, binary, discrete or continuous, univariate or multivariate.

The terms abnormal behavior, outliers, and anomalies refer to the same concept. Early detection of anomalies in a reliable and robust manner is important to maintain system operations efficiently and safely (Bolton & Hand, 2002). In consumer purchase histories, these methods may help to identify abnormal behavior as fraud. To prevent further unwanted purchases, efficient methods are needed. The difficulty with anomalies is that often they are not clearly differentiated, come in diverse forms, and are adaptive.

Bolton & Hand (2002) list three types of anomaly detection techniques (Hawkins, 1980). 1) point anomalies, 2) collective anomalies, and 3) contextual anomalies. The third type provides an explanation or augmentation of the data when analyzing the first two types (Hawkins, 1980). Aleskerov & Freisleben (1997) state that the choices of anomaly category change the way the data is searched and the type of AI model built for anomaly detection.

In the wake of Covid 19 pandemic outbreak, retailers had to shift their focus from in-person experience to online experience practically overnight due to the unprecedented growth of online shopping and shift in brand loyalty. The sudden shift required ways to accommodate new customer purchase behavior and made anomaly detection in the retail industry more important than ever. As Covid 19 restricted physical movement, customers migrated into virtual world at an unprecedented rate. Retailers needed to go beyond traditional marketing methods of modeling customer purchase behavior.

Chapter 5

Association Rules Mining

5 Market Basket Analysis

In Market Basket Analysis, marketers seek itemsets, *I* in customer shopping baskets and assess the extent to which items frequently co-occur. The data used for the finding applicable rules *D*, comes from the transaction system *P*, where each product of the retailer is listed,

$$P = \{list\ of\ all\ products\} \tag{1}$$

Here, itemset can mean both categories of products explicitly defined by the retailers, for example, product groups, such as running shoes, and, more importantly, categories of goods that might not be obvious at the first glimpse, but after analysis end up having the pattern of being frequently purchased together. Following this, the objective of MBA is to find these implicit, non-obvious associations and build a rationale for the retailers to act upon them.

Data, *D* for MBA consists of all customer baskets, *B* purchased during the analysis period

$$D = \{B_1, \ldots, B_n\} \tag{2}$$

A customer basket B_i has all products the customer buys registered to the point-of-sales (POS) transaction system, for example, an entry of B_1 to POS could look like

$$B_1 = \{p_2, p_3, p_3\} \tag{3}$$

, where p_2 = *pair of running shoes X* and p_3 = *water bottle Y (p_3, p_3 indicates two purchased bottles in the basket)*.

In database systems, products p_k are usually shown by their respective product codes. If, in this simplified example, the product portfolio of the retailer consists of 5 products,

$$P = \{p_1, p_2, p_3, p_4, p_5\} \tag{4}$$

the itemset included in B_1 would be represented by

$$I_1 = \{p_2, p_3\} \tag{5}$$

Where the I_1 consists all of the different products existing in B_1, thus, each I_j constructed from B_i is based on the binomial logic as an answer to the question," Does product p_k exist in B_i?". In a data table, P, B_1 and I_1 would have the following representations.

Table 5.1 Products, basket 1 and itemset 1

Products, P	p_1	p_2	p_3	p_4	p_5
Basket 1, B_1	0	1	2	0	0
Itemset 1, I_1	False	True	True	False	False

Following the day-to-day buyer behavior, the retailer may assume that "*in many cases, buying X leads to buying Y, as well*". With MBA the retailer can test this hypothesis $\{X \Rightarrow Y\}$ with data consisting usually of a very large number of customer baskets *B*.

5.1 Association Rules Mining Process

5.1.1 Rules Extraction

Associations between objects are the primary interest in the rules extraction process. Associations between objects are the primary interest in the rules extraction process. Marketers apply the extracted rules that yielded interesting information based on the co-occurrence of the items. Analysts can apply several strategies to build customer portfolio and adjust sales and marketing strategies for business continuity based on the extracted rules.

Candidate-Generation is a *Breadth-First* search strategy, where the algorithm generates a broad candidate set of ***k+1***-itemsets from point-of-sales data, based on found frequent ***k***-itemsets. In each step of the search (from ***1*** to ***k+1***), the number of occurrences of each unique ***k***-itemset is recorded by its *frequency*. This is then at each step ***k*** compared to a previously given *frequency threshold* to decide if the itemset will be kept for the next step ***k+1,*** or not. At the end of the algorithm, the frequency of any possible itemset satisfying the frequency threshold has been determined. The first and most common algorithm based on CGB strategies is the Apriori algorithm (see a detailed presentation of Apriori in section 5.2.).

Pattern-Growth Growth is a *Depth-First* search strategy that uses a tree-like data structure, the Frequent-Pattern-tree (FP-tree), (Borah, & Nath, 2019). PG algorithms find frequent itemsets starting from ***1***-itemsets (i.e., bottom). They detect sets of items by recursively finding all frequent ***1***-itemsets in the conditional pattern base that the algorithm constructs with node-links represented by FP-tree. The first algorithm based on PGB strategy is the FP-Growth algorithm developed by Han (2000). See Section 5.2.3. for details.

5.1.2. Rules Evaluation

The developers of Apriori, Agrawal & Srikant (1994), present an association rule as an implication expression (co-occurrence) of the form X => Y, where $X \subset I$, and $Y \subset I$ and X and Y are any 2 itemsets, i.e., $X \cap Y = \phi$. Every rule consists of the *antecedent* (*X*) on the left-hand side of the implication and *consequent* (*Y*) on the right hand side. A concrete example based on customer shopping behavior is {Milk, Diaper} → {Coffee}. If you purchase Milk and Diaper, the customer is more likely to purchase Coffee.

There are many metrics to evaluate interesting association rules, like *support, confidence,* and *lift*.

E.g., with Apriori, the strength of the extracted rules must satisfy thresholds set to support, confidence, and lift. As conditional probabilities, they define the predictive quality of the rule, help identify interesting patterns in data and assist practical decision-making (Maimon & Rokach, 2009). To present these metrics, we apply formulations of Borgelt & Kruse (2002).

Support is the rate of occurrence of an itemset in a transaction database (or default popularity of an item). It is calculated by the number of found baskets containing a particular item divided by the total number of transactions.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{No. of\ transactions\ containing\ both\ X\ and\ Y}{No. of\ all\ transactions}$$

Confidence defines the ratio (percentage) of data items that contain Y in the items that contain *X*, i.e., it refers to the likelihood that item *Y* is bought when item *X* is bought. To get confidence, the number of baskets where *X* and *Y* are bought together is divided by the number of baskets with *X*.

$$Confidence\ (\{X\} \rightarrow \{Y\}) = \frac{No.\ of\ transactions\ containing\ X\ and\ Y}{No.\ of\ transactions\ containing\ X}$$

Lift controls for the *support frequency* of consequent while calculating the conditional probability of occurrence of $\{Y\}$ given $\{X\}$. It is a very literal name given to this measure as it describes the rise in the probability of having $\{Y\}$ on the cart with the knowledge of *$\{X\}$* being present over the probability of having $\{Y\}$ on the cart without any knowledge about the presence of $\{X\}$. In other words, it describes how demand of $\{X\}$ lifts (*Lift* > 1) or decreases (*Lift* < 1) the demand of $\{Y\}$ (Kavitha & Subbaiah, 2020).

$$Lift\ (\{X\} \to \{Y\}) = \frac{No.\ of\ transactions\ containing\ both\ X\ and\ Y)|(Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y}$$

5.1.3 Challenges In Association Rules Mining

Analysts apply data mining usually on large or very large databases. One of the challenges for association rule mining is the generation of vary large number of association rules. It is difficult to develop a model or summarize useful information. This has led to vital information being filtered out while the remaining rules may be obvious or already known. However, many researchers have optimized the original algorithm by adding methods such as random sampling (Agrawal, Mannila, Srikant, Toivonen, & Verkamo, 1996) or by declining rules and changing the storing framework (Han, Kamber & Pei, 2011).

According to Borgelt & Kruse (2002), the main problem of association rules induction is that there are so many possible rules. For the product range of a retail store, the variety of products offered, with possibly several thousands of different products, cause millions or even billions of possible association rules. Many association rules cannot be processed or evaluated by inspecting each rule in turn. A widely used approach to efficiently help mitigate this problem, and reduce the number of association rules, is to gradually increase the threshold value of support and confidence until a manageable size of association rules are generated (Borgelt, 2012). However, this may affect the accuracy of the result. This is why other approaches are developed, as well. According to Sethi & Shekar (2018) ARM is shifting focus from generating positive rules to the discovery of negative rules.

Moreno & López (2005) illustrated some of the main drawbacks of the ARM algorithms as:

1. Obtaining non-interesting association rules,
2. Huge number of discovered associated rules and
3. Low algorithm performance.

5.2 Apriori Algorithm and Its Variants

Apriori algorithm employs an iterative approach of the level-wise search, where *k-itemsets* are used to explore the next level *k+1*-itemsets. First, the set of all frequent *1*-itemsets is scanned to count each item in the database. Then each of them is checked for minimum support and minimum confidence requirements, and the resulting set of *1-itemsets* is denoted as L_1. Then this L_1 is used to find L_2, the set of frequent *2-itemsets*, which is used to find L_3, and so on, until no more frequent itemsets can be found (at the level of *k+1-itemsets*). To find each of the levels from L_1 to L_k requires one full scan of the database. To improve the efficiency of the generation of frequent itemsets, *the Apriori property*, that all non-empty subsets of a frequent itemset must be frequent, is used to reduce the search space

The algorithm works in two steps (c.f. Borgelt & Kruse, 2002):

1. Frequent itemsets, sets of items that have at least the minimum support, i.e., at least a given percentage of all transactions, are determined.
2. From the frequent itemsets found in step 1., association rules are generated.

Usually, the first step is the most important because it accounts for the greater part of the processing time.

The Apriori algorithm has a sound theoretical basis, and its focus is on efficiency. In the algorithm, the candidate set of the level C_1 is the set of all transactions T. After it is found, the itemsets with support lower than the minimum support are pruned, which result to L_1, a set of transactions satisfying the minimum support. This is then used to find the candidate set for the next level, C_2. In a similar way, each lower-level set of pruned candidate set, L_{k-1} is used to find the next level candidate set C_k (see figure 4.1 for the pseudocode of the Apriori algorithm).

```
T: Transactional database
Ck: Candidate item set of size k
Lk: Frequent itemset of size k
S: Support
 Apriori (T, s)
   L1 ← {Large 1-item set that appears in more than or equal to s transactions}
   K ← 2
     While Lk-1 ≠ Ø
       For each transaction t in T
         For each candidate c in Ck
           If (c ⊂ t) then
             count [c] ←  count [c] +1
           End If;
         End For;
       End For;
       Lk = Ø
       For each candidate c in Ck     // Prune
         If (count [c] >= S) then
           Lk ← Lk U {c}
         End If;
       End for
       k ← k +1
     End While;
     Return Lk;
End Apriori;
```

Equation 5.1. Apriori pseudocode

In the Apriori algorithm, for each candidate in Ck, the frequency is calculated by scanning the transactional database. After having calculated frequencies for all candidates in a Ck, the frequencies are compared to support, S and candidates with frequencies less than S are excluded. This results in the generation of Lk, and the process follows to k = 2 and further in a similar manner.

5.2.1 Deficiencies and improvements of the Apriori Algorithm

The Apriori algorithm has some flaws like the way it requires repeated scans of the transactional database. According to (Doshi & Joshi 2018), another major limitation with Apriori is the way it handles its calculation. Apriori works from bottom-up, begins from the smallest arrangement of frequent itemset and moves upwards until it achieves a very large itemset.

The Apriori algorithm calculates interesting rules with probabilistic relationships between items in frequent itemsets. As the main limitation of Apriori algorithm is that large frequent itemsets has to undergo many database iterations (Scans). When itemset is too big, memory capacity, and system I/O load consumption is high, causes performance to degrade and slower calculation (Sinthuja, Puviarasan, & Aruna, 2017).

According to Ezhilvathani and Raja (2013), one feasible strategy to improve the efficiency of the Apriori algorithm is to reduce the number of database scans. Also, Borgelt & Kruse (2002) state that efficient algorithms are needed that restrict the search space and check only a subset of all rules, if possible, without missing important rules. Yahya, Hegazy & Ezat (2012) support his by saying that Apriori is computationally expensive. They point out that although Apriori reduces the number of candidate itemsets, the number could still be huge when store inventories are large or when the support threshold is low.

An additional deficiency is that itemsets or items in itemsets have the same importance in Apriori algorithm. The algorithm does not account for differences in real-world itemsets and cannot evaluate if the obtained rules are useful or interesting. Kim and Park (1997) add that the analysis of large inventories involves larger number of item configurations, and the support threshold might have to be lowered to detect certain associations. However, lowering the support threshold might also increase the number of detected uninteresting or useless associations.

These deficiencies call for practical solutions and contributions. According to Zaki (1999), work has been done to reduce the CPU time, memory usage, and I/O computation overhead. Also, several variants of Apriori have been proposed to enhance the efficiency and scalability of Apriori.

Already in their seminal paper, Agrawal and Srikant (1994) suggest several methods, like *Apriori Tid* to reduce the Apriori algorithm's execution time. Apriori Tid applies the same candidate generation function as Apriori. However, it uses a different mechanism to count the support of itemsets after the first pass, by applying encoding of the candidate itemsets used in the previous passes. In later passes k+1, and onwards, the encoding size becomes much smaller than with original Apriori.

The Apriori Tid also introduces a new approach to reduce the memory needed to store transaction identifiers (TIDS). This *diffset* keeps track of only the differences of the Tid of any found k+1 itemset from corresponding k-itemset. A vertical layout is used to keep data scans to one and to record differences in TIDS of a candidate itemsets from its generating frequent itemsets. The diffset approach does not keep the complete TID of each itemset; instead, it keeps the differences between class and its member itemsets. Results are generated on synthetic and real-world datasets (Gouda & Zaki, 2004).

Kim & Park (1997) proposed direct hashing and pruning algorithm that utilizes a hash approach to reduce candidate itemsets. In this approach, the support is counted by mapping the frequent itemsets from the candidate list into buckets (collection of items purchased together), which are then divided according to a specific count system support known as hash table structure.

When new itemsets occur, if the item exists earlier, then the count in the bucket increases. Otherwise, the item starts a new bucket. In the end, the algorithm removes buckets with less than minimum support count from the candidate itemsets.

Zaki (1999) adapted association rule discovery to parallel systems to take advantage of the increased speed and greater storage capacity. For the transition to a distributed processor and memory, he partitioned the database. This technique reduces the quality of candidate itemsets generated from the database.

However, parallel systems algorithm like FP-Growth has a disadvantage that FP-tree becomes too large to be created in memory (Yan, Wang, Wang & Lin, 2011).

Borgelt (2003) used ordered trees, known as "tries," to structure data efficiently in Apriori implementation. In a trie, the data is stored as a set of strings. Every node in the trie contains the same prefix and retrieves any finite ordered set of strings. In this way, the algorithm could reduce the support counting costs.

Said, Muhammad & Gupta (2015) proposed a novel association rules mining scheme after pondering the number of database scans, memory consumption, time, and the interestingness of the rules. The authors extracted association rules of different parameters using Apriori algorithm on medical dataset to heart diseases. The technique is to sort information based on different parameters like gender which extracted some set of rules. To reduce the database scans, the authors used FP-Pattern algorithm that is the more effective structure to extract patterns when the database intensifies.

Most of the ARM algorithms consider a single support value. Setting the support value to a low value generates a large number of useless rules and changing the support value to high leads to the non-selection of some important rules (Agrawal & Srikant, 1994). As an improvement of Apriori, MSapriori considers different support values for every item in the dataset. This has given users the flexibility to express their views and the option to select support according to the nature of data. The support value of a specific rule in MSapriori is the lowest minimum support value between the items in that rule. The generating function of MSapriori is the same as Apriori (Chaudhary, Sharma & Sharma, 2015).

Saurabh Malgaonkar, Sakshi Surve and Tejas Hirave (2010) proposed a method to extract data from large marketing datasets efficiently by using several techniques of ARM. The systems satisfied the objectives of making a more informed decision about product placement, pricing, promotion, and profitability. The model finds associations between patterns and provides insights into the combination of products and services within a customer's basket. Kaur & Kang (2016) present a method to understand the dynamics of the data generation process by examining changes that have taken place in the discovered patterns. Their algorithm tries to capture changing trends of transactions in data. The method applies the basic idea of collaborating Association Rule Miner. It keeps track of the items, which are associated with high confidence.

Szymkowiak, Klimanek & Józefowski (2018) propose an MBA model to describe the theoretical aspects of MBA with an illustrative application based on data from the National Census of Population and Housing in Poland. The proposed model aims at identifying the relationship between legal marital status and actual marital status. Heydari & Yousefli (2017) present an MBA model to allocate parameters as one of the important and effectual factors of the selling rate. The study applies a genetic algorithm (GA) to solve the non-linear binary programming problem and presents a numerical example to illustrate the model. GA is a stochastic search method and belongs to meta-heuristics category uses a randomized choice of operators in search strategy.

To serve the health sector, Keohane, Gambrel, Freed, Stevenson & Buntin (2018) proposed Oaxaca Blinder decomposition model to analyze the sources of per-beneficiary Medicare spending growth between 2007 and 2014. They included demographic characteristics, attributes of Medicare coverage, and chronic conditions. The result of the model was that spending levels for Medicare beneficiaries declined with chronic conditions, which suggest that changing patterns of care use may be moderating spending growth.

Brin, Motwani, Ullman & Tsur (1997) introduced a version Apriori known as dynamic itemsets Counting (DIC). The algorithm splits the database into many partitions by DIC, and each partition is marked with a start point. The support counts of all the generated itemsets so far they are determined and new candidate itemsets are counted dynamically as soon as their subsets become frequent. To reduce the complexity of this dataset, the authors removed all items appearing in over 80% of the dataset transactions, yet still they were only able to mine efficiently at high support.

Han (2000) proposed FP-growth as a new ARM technique to construct a very condensed tree for a transactional database known as a frequent pattern tree (FP-tree). One path in the tree corresponds to one transaction in the database, and the size of the path is the total number of frequent items in the record (transaction) (Layman, 2015). As a tree structure, the FP-tree is useful because there is a lot of sharing between the frequent itemsets. Thus, the FP-tree is smaller in volume than the main database. However, the FP-tree contains all frequent itemsets of the database. The FP-growth constructs the FP-tree in the second scan after the first scan that produces the frequent itemsets with their support values. Unlike Apriori, the algorithm does not generate any candidate rules during the mining process. The major drawback is the large size of the FP-tree that does not fit in the main memory.

Khan, Lee, Lee & Khattak (2010) presented an applied K-means clustering method to classify product baskets and address customers according to their purchasing behavior and power. With proper clustering, retailers can improve their promotional strategies and enhance sales. The technique extracts features such as customers' age, purchasing power, also customer traffic. They use features to research various products and services and find out how well they are selling. This profoundly affects how decisions are made, giving critical insights to automate marketing. In addition, the US Census and the Bureau of Labor Statistics use clustering to determine who gets Federal aid, where assistance programs are targeted, what businesses might move to your community, and how your vote counts in the Electoral College (Sreenivasulu, Viswanadha, & Sambasiva, 2019).

Bansal, Sharma & Goel proposed an improvement to the k-mean clustering algorithm to define the number of clusters automatically and to assign the required clusters to un-clustered points. The proposed improvement aims to improve accuracy and reduce clustering time. They argue that using large datasets with trillions of records, only a little dispersion in the accuracy level will matter a lot and can lead to a disastrous situation if not handled properly. The model result can be more extended to achieve the full accuracy level up to 100%, with very little time and with more quality clusters.

5.3 Research Gap

As this literature review reveals, several attempts have been made to improve ARM and MBA methods. Still, we see that there are problems that have not been solved, especially when it comes to efficient implementation of MBA in small and medium-sized retail companies. To highlight the topics of this thesis, we will now identify a few gaps in research.

Most reviewed ARM models consider only the total monetary value of sold products without accounting for changes in the products sales history and trending patterns. This is why we wish to validate the concept of lifetime trending popularity of products.

This shortcoming highlights the fact that there is a lack of viable solutions of ARM on real-world marketing problems of SMEs. As previous implementations of ARM focus on industries with larger customers-base, small and medium-sized retailers have so far been omitted.

When it comes to market intelligence, the literature review has not provided sufficient insight on how the SMEs can effectively acquire, assimilate, transform, and utilize it. There is a gap between academic understanding and the everyday use of information retrieved by MI.

5.4 Summary

Artificial Intelligence is a fast-growing field of information systems, with increasing interest shown by both marketers and researchers due to the increasing interest in understanding and exploiting the attributes, entities, and relationships in data. In this chapter we focused on methods required by small and medium-sized retailers when they apply market basket analysis. As the review of the literature shows accomplishments have been made, but there is still work to be done. The classic Apriori algorithm gained its popularity because of its intuitive attractiveness and simplicity. The classic version of Apriori has well-known evaluation problems related to support, confidence, and lift. The reviewed literature illustrates these limitations and presents some benefits of the enhanced versions.

Therefore, we here express the urgent demand for an advanced ARM method that accurately predicts customers' shopping journeys, a method to predict the next group of items a customer is about to purchase, based not only on a certain group of items purchased previously but on finding rarely occurring events and popularity trends of brands before they peak.

In the following two chapters, we present the Åbo algorithm and experiments with this MBA tool using a dataset of customer transactions from a point-of-sales database. For association rules mining, we extend the Apriori algorithm with methods providing additional ideas on support with new features of computational intelligence and improved computational performance. We present our new MBA architecture with the improved Åbo algorithm and its implementation process in six steps, as illustrated in Figure 6.1. in Section 6.

Chapter 6

Advanced Association Analysis

6 Advanced Association Analysis

Product associations provide vital information for designing marketing strategies. Considering that a typical retailer sells thousands of products, it is also likely there are thousands of associated product pairs and groups that the retailer may not have recognized. Market Basket Analysis involves a whole set of methodological tools, of which only part is mining association rules with well-known algorithms, such as Apriori, . Typically, the input used for MBA is point-of-sales data on customer transactions.

According to Ansari, (2019), MBA model extracts products and products categories represented in the form of association rules. These associations provide managers, and marketers with different angles on how the products interrelate to each other to develop customized marketing strategies. Nowadays, the MBA is widely applied. Not only the retail industries, such as supermarkets, use it, but other sectors and industries selling multiple products, such as social media, financial institutions, medicine, direct marketers, and many others have implemented it in their processes (Szymkowiak, Klimanek & Józefowski, 2018).

In this chapter, we present new techniques and methods constituting a novel and efficient association rule mining method. We use features of artificial and computational intelligence to augment the knowledge and understanding about customers and products and enhance the retailer's strategic decision-making and tactical marketing capabilities. Our new model of MBA, the Åbo model, aims at helping small and medium-sized enterprises analyze the degree of attraction between items and exploit these relations to implement market intelligence strategies. The model will identify affinities in keystone items and check if there are any patterns in the selection of these items. As the outcome, we expect to have an improved ability to respond to each customer contact and reduce costs due to the optimal allocation of resources, ultimately leading to an increase in revenue.

In this dissertation, we introduce the algorithm used in the Åbo model, that we call the Åbo algorithm. As Åbo is a new variant of the classic Apriori algorithm, we compare it to Apriori with respect to performance and intelligence. There are some fundamental differences in the ways that the two algorithms provide intelligence in decision support and computational performance. We present the architecture of the Åbo model, algorithm and the implementation process. The presentation follows six life-cycle phases, as illustrated in Figure 6.1.

Empirical synthesis comparative analysis is a major method to compare the performance and intelligence of a given algorithm. A single algorithm cannot be applied to all applications due to the complexity for suitable data types of the algorithm. Therefore, selecting data mining algorithm depends on not only the goal of the data mining model, and the compatibility of the dataset.

We present the comparative analysis of two data mining ARM algorithms (Apriori and Åbo) used in retail dataset. A summary of these algorithms with their merits and demerits is discussed in this chapter.

In this research, we use live POS database. The goal is to develop MBA predictive model that offers an intelligent recommendation. CRISP-DM is cost-effective framework involves sequence of events and provides structure for problem solving we might encounter throughout the development of this MBA model.

We start with business understanding to understand the nature of the retailer's transaction lifecycle. For an accurate AI model prediction, data must be explored, gathered, and prepped in away consistent with the business understanding.

Then, we must apply set of AI algorithms to this data, and then MBA model needs to be evaluated for its ability to generalize in real-world.

In practice, several CRISP-DM tasks can be performed/executed in a different sequence, and it is often necessary in an AI development to backtrack to previous task/tasks and repeat.

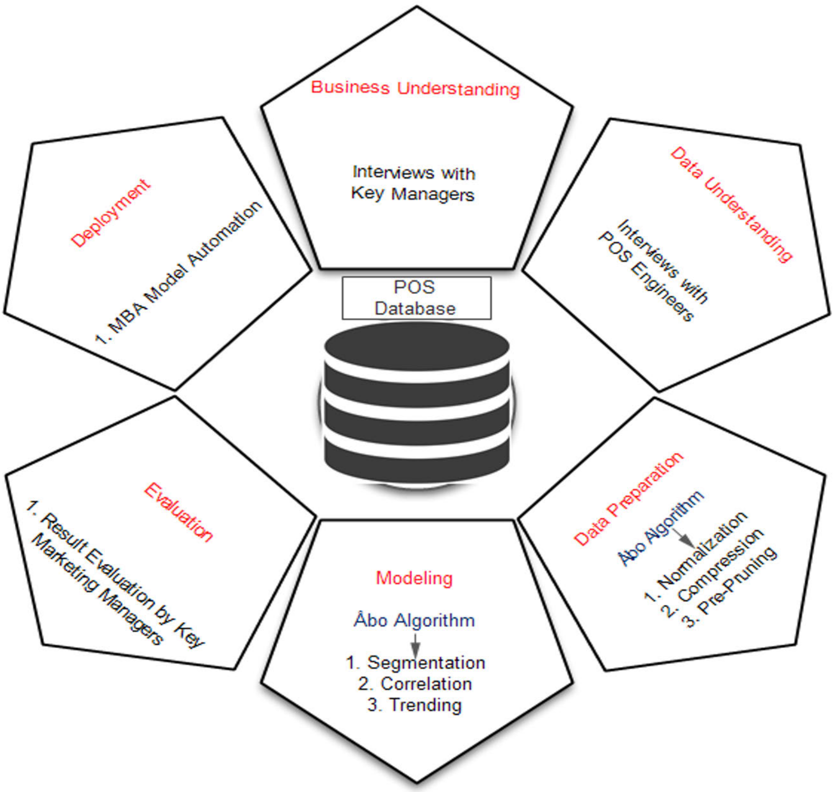


Figure 6.1 The proposed market basket analysis model process

The business understanding in the CRISP-DM life cycle (see section 4.3.1), as well as in this study, focuses on understanding the requirements and objectives of the MBA project. For the work process with the Åbo model, we utilized business expertise to understand business requirements with the Kuwaiti retailer that provided the data and background information about the data. We converted the gathered requirements into the data mining task and sketched a project plan. In the interviews with business experts, we found that the business questions need to be very specific and measurable to get useful questions for data mining.

In the next phase, we wanted to understand the data. Properties and quality of data needed to be verified with respect to data sources and the features of the data set. In this case, we did not apply additional data sources. To understand the properties of data we looked at the metadata, key features, and relationship between different data elements. Specifically, we searched correlations between elements.

After this, we prepared the dataset from the source, the POS database, for modeling. Here we cleaned the data.

The data preparation phase involves many steps. First, we start with the data cleansing to identify if there are any missing data, incomplete or otherwise erroneous within POS data repository, and if these missing data can be substituted or removed. This is essential to improve data quality and ensure more consistent, accurate, and reliable analysis for decision-makers.

We then start with the data construction to derive new attributes from the original POS attributes. Followed by data integration, we run data mapping to match fields from the source POS database to the target POS data warehouse, and ensure it migrated to its destination fields in the way as it was intended.

At this stage, the data is migrated and transformed in the target POS warehouse and is prepared for the next phase (modeling).

The modeling phase consists of three main tasks, 1. model selection and creation, creating a testing plan for the MBA model, and parameters testing and tuning. This phase is tied to data preparation phase because the data preparation influences the model selection and vice versa. This phase may reveal that the selected data doesn’t fit well into the selected algorithm for modeling, and data preparation phase must be revisited. The model selection involves choosing the best-fit algorithm for modeling. Creating a testing plan is generally two steps task (test and training dataset) to adjust the performance and intelligence.

The evaluation phase is where we evaluate the performance of MBA model against the retailer's business goals defined in phase 1. There are three tasks in this phase. 1. Evaluate results: Does the MBA model meets the retailer's business goals defined in phase 1 and reach final decision if the MBA model should be deployed at the retailer premises. 2. Process review is to verify tasks defined are accomplished and then summarize findings. 3. Determine the next steps is based on the preceding three tasks to determine whether we proceed to deploy the MBA model or iterate further.

The deployment phase has four tasks, 1. planning deployment, 2. plan maintenance and monitoring, 3. produce final report, and 4. project review. First, we plan how the MBA model will be deployed. Second, developing a monitoring and maintenance plan to avoid technical issues during the post-operational phase of the MBA model. A final project report is handed over to the retailer (Marketing Department), highlighting all tasks executed and if project goals are met. The project review task assesses all steps executed. Running the preliminary analysis with the point-of-sales database, we can find that the retailer sells diverse products, like clothing, footwear, cosmetics, electronics, and accessories. We carry out further analysis with an example of a two-year-long transaction dataset consisting of customer transaction data (see Table 6.1 for dataset attributes and description).

Figure 6.2 shows the hypothetical POS transaction data from the retailer. There are four transactions, and each transaction consists of a set of items sold to customers by the retailer. The Åbo algorithm is expected to provide a series of association rules where the retailer can infer which items/products the customer purchases together. Each association rule consists of a consequent and antecedent.

TID	Items / Products Purchased (Market Basket)
1	Jeans, Shirt, Belt, Hat
2	Running Shoes, T-Shirt, Socks
3	Perfume, Women Bag
4	Blue Pants, Shirt, School Shoes, Kids Socks

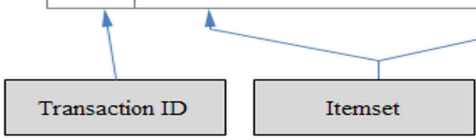


Figure 6.2 Sample of point-of-sale data structure

6.1 Data Exploration and Transformation Practice for Market Basket Analysis

For data mining professionals, the exploration of data is a crucial phase to achieve efficient and accurate data analysis and pattern recognition (Kimball & Ross, 2011). Producing useful databases for data mining tasks requires careful design and artisanship. In their book Database design: know it all, Teorey, Buxton, Fryman, Güting, Hal-pin, Harrington, & Witt (2008) highlight this fact; while the database design builds on scientific, rigorous design principles, it also involves aesthetics and intuition, with the possibility that the designer's subjective biases may inflict the process.

From the retailers' vantage point, a good dataset enables fast decision-making based on relevant data on a factual basis. Being active decision-makers in competitive environments, retailers need to have up-to-date information to stay competitive in dynamic markets. From transformation of point-of-sales (POS) data to data warehouse, this means that POS database structure needs to be transformed into a form that can provide data mining analytical results, yet with the logic of POS design in mind, with tables, indexes, and views.

Clean data is an essential characteristic preserved to meet the organization's goals across a business-ready data pipeline. Clean data is our first step in the data preparation phase to identify and correct messy, raw data. It is defined as the process of analyzing, identifying, and correcting messy raw data. The process involves identifying and fixing errors, rectifying inaccurate data, filling in missing values and determining if all the information is in the right columns and rows. Clean data enables analysis to integrate all relevant data sources to provide a replica of the organization. Therefore, clean data determines the decision-making reliability and expedites the return to sustainable growth (Ilyas & Chu, 2019). Robust data drives the organization's strategic direction.

Our MBA process depends heavily on clean data to learn and systematically discover relevant patterns from data. However, the original POS database of the retailer was designed with many unneeded columns that store empty values and null values. As (Lekberg & Danielsson, 2013) point out, this is quite typical. This is a major design limitation in our POS database. According to Teorey, Buxton, Fryman, Güting, Halpin, Harrington, & Witt (2008), the retrieval of data into a POS database analysis requires looking through hundreds of tables, and each of these may or may not have information for this customer.

Based on our experience, we propose automation of data preparation tasks to provide an explorative summarization of the data to the retailer. In addition, data automation with human intervention remains an option for the retailer when necessary. In our approach, we clean and transform the source data into a highly detailed data warehouse schema that collects and stores a wide variety of large POS databases into tables with a small and constant number of facts and dimensions. This approach will offer retailers the compatibility of quick, consistent, and interactive multi-dimensional analysis to support decision-making. Figure 6.3 illustrates a sample of the POS database used in our experiments after transforming the original into a POS data warehouse schema in a common format

Table 6.1 Original POS dataset attributes description

Column Name	*Description*
Basket_id	Stores the 11 and 12 digits transaction number

Day	Day of the week, Sunday as day 1
Product_Id	Barcode number to track the product
Product_Category	Type of product category
Product_Description	Description of the product
Sales	Sales of each product
Qty	Number of times the product is sold in a transaction
Store_Id	A tag used to track the store name
Retail_Disc	Percentage of discount on the product
Trans_Time	Date when a sale takes place
Week_No	Week number per month
Product Price	Original price before the discount

HOUSEHOLD_KEY	BASKET_ID	DAY	PRODUCT_ID	PRODUCT_CATEGORY	SUB_CATEGORY	PRODUCT_DESCRIPTION	SALES	QTY	STORE_ID
718	26985360571	1	1110793	CLOTHING		MARC BY MARC JACOBS - KNITTOP-DEL2CRNRWSTCH	68.25	2	324
718	26985360571	1	1114050	CLOTHING		SAVAGE CULTURE - SKIRT-80145FA	24	2	324
718	26985360571	1	1130666	Fashion and Beauty	Arabian Fragr...	Dehn al Oudh Abiyad by Afnan Perfumes for Unisex - Eau...	148.5	1	324
718	26985360571	1	1130716	CLOTHING		ICE J - TSHRT-F261/6301	36	1	324
718	26985360571	1	1136199	CLOTHING		ICE ICEBERG - DRESS-G081/1107	176.25	2	324
718	26985360571	1	5567633	SHOES		Lacoste Gazon Bl 1 Cam Loafers for Men - White	261.75	1	324
718	26985360571	1	5591784	ACCESSORIES		Chrysalis Women Brass Tranquility Nature Moss Jade Ba...	66.75	1	324
718	26985360571	1	6551990	FOOD AND BEVERAGE		FRZN WHIPPED TOPPING	28	2	324

Figure 6.3 Sample from the target POS data warehouse

The applied POS data warehouse design comes with performance and storage gain, as we code and store all queries for handling complex operations from the end-users and make them available through stored procedures, functions, and triggers. This means that no further re-processing is required from the data warehouse. The strength of our data warehouse comes from effective entity-centered queries. There is a highly detailed data warehouse schema from the storage side, which collects and stores a wide variety of large POS databases in a small and constant number of facts and dimension tables. Following the method of Tyugu (2007), we transform data into an appropriate format to make patterns discovery easier. Figure 6.3 illustrates our customized ETL process to transfer raw data from the source POS database to the target POS Data warehouse.

In the *extraction* phase, the required POS data for analysis is extracted from the source. Then, raw data (unstructured and structured) and consumer transactions are consolidated into a single data warehouse repository.

In the *transformation* phase, we apply a set of rules to improve data accessibility and quality before loading into the target data warehouse. We also apply data cleansing to check for data inconsistencies, duplicate data, errors, and missing values. These data irregularities are either excluded or replaced with default values. We take into consideration that all databases potentially include some errors and null values. During this phase, potential anomalies will be flagged. To improve mining quality of data, we apply the standardization technique to the extracted source POS database. We transform the data from its original POS format into one unified, generic POS data warehouse presented in a common format with consistent definitions for keys and fields.

Loading phase data is the final step in the ETL. We load the newly extracted and transformed POS data into a new POS data warehouse (full load).

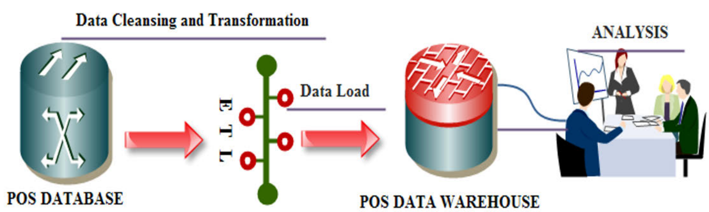


Figure 6.4 POS ETL process from the source pos database to the target pos data warehouse *(Tyugu, 2007)*

We use customized hand-coded ETL pipelines stored as a PL/SQL procedure as the ETL processing method. We base our decision on two reasons. Firstly, we have only one source as a POS database in our data. Secondly, using industry ETL tools can be more complex because of the need to create many tasks to accomplish a simple data transfer, whereas using handwritten stored procedures is just one task. Native stored procedure codes integrate smoothly with any contemporary code management tool.

In addition, the native code allows easy integration with many technologies, and we could automate it with an off-shelf and customized automation tools. To improve the efficiency of the data mining lookups tasks and to handle complex data manipulation, the customized ETL stored procedure uses JOIN and MERGE techniques to bring all the transformed POS transactions over to the data warehouse memory space. The ETL process transforms the data from the source RDBMS POS database and then loads it into a staging POS data warehouse.

A customized stored procedure loads the data from the staging POS Data warehouse into the final data warehouse schema (in our case Oracle). The final DW will be in a common format with consistent definitions for keys and fields. We apply the business rules here in the data load phase. In addition, the missing values are checked and deleted or replaced by default or mean values of each parameter.

We faced some challenges while working with the POS databases in our study, and we have found them to apply to most POS data. POS data contains millions of transactions over a long time span. In many invoices, the total amount shows either null or negative values and transactions that are not linked to any customers on many invoices. This would raise suspicions that there are malicious or fraud involved, and the transactions should be treated accordingly.

It has many fields that are null, obsolete, unnecessary, or, at best, the data is not in a form appropriate for the data mining analysis. Consequently, the POS databases require extensive pre-processing. It needs conversions to make it readable for the retailer. The data format in the obtained POS databases is often unstructured text. It takes extensive effort and time to understand the data structure and processing to transform it into a standardized single unified POS data warehouse to generate accurate and reliable insights. Based on our experience, this is the biggest challenge.

Finally, human judgment plays an important role in understanding the source data, the types of identified anomalies, and most importantly, to verify the accuracy of generated analysis.

6.1.1 Anomaly Detection

The preparation of the database for the exploratory analysis involves the elimination of anomalies. Anomalies in data may indicate experimental, novelty errors, wrong measurement or systematic human input errors (Pearson, 2018). An anomaly may imply existence or absence of critical information. In anomaly detection and analysis, we detect anomalies, investigate them and answer questions on their significance. Depending on their information value we either eliminate, process or label them as legitimate and include them in the analysis.

It is widely known that there is no rigid mathematical definition of anomalies, and the determination of the anomalies is ultimately a subjective human exercise (Adhikari & Karunananda, 2015). Anomalies can be legitimate observations and may represent scientifically important or meaningless aberrations, and hence, we simply do not eliminate the anomalies observation. Regardless of the types of anomalies (as presented in section 4.3.3), it is imperative for the retailer to analyze what anomalies mean.

Many anomaly detection algorithms try to minimize their influence by eliminating them without further verification from the client. As opposed to this, we propose an anomalies detection technique based on an enhanced normal distribution method.

This will enhance the visibility of the anomaly of normal distribution to data in the POS dataset. We believe that the interpretation of anomalies should be a human exercise. The client can benefit from our approach by exploring different levels of potential anomalies that give the client the ability to fully or partially eliminate them.

We use two independent continuous variables in the range (0-1) that give the retailer the ability to partially or fully identify potential anomalies. The closer the variable value is to the extreme 1, the closer it is to the normal distribution shape, and data points are closest to the mean. On the other hand, values near 0 indicate extreme anomalies; the data points are at the furthest point from the mean. Therefore, increasing or decreasing the independent threshold value of the variable results in the detection of less or more extreme anomalies in the extracted data.

What we consider as anomalies are such transactions that in the POS database deviate some standards or are far from the mean. In addition, we consider transactions that will be labelled as potentially fraudulent, i.e. those that deviate in some standard or extreme way. For example, transactions that enter daily outside of the opening time window of 9 am to 9 pm, six days a week, and on Friday, outside of 1 pm to 9 pm. We can also consider the sales peak hours from 2 pm to 6 pm as requiring special attention. Furthermore, some patterns, such as high volumes of transactions made shortly before or after the store is closed, transactions containing low-priced products in high volumes, and transactions made by the same employee in different stores on the same day or within a short period.

The normal distribution (ND) is often utilized for anomaly detection because it fits the behavior of many natural phenomena when most of the dataset aggregates tend to cluster symmetrically around the mean. When data is normally distributed, it follows probabilistic rules. The rationale of using ND is that values following these rules are regular while values that break them are irregular and, according to Schafer (1997), and may indicate anomalies. Our enhanced normal distribution method is illustrated in detail in Research Article 3, and illustrated in Equations (Eqn 1, 2, 3).

The normal distribution has two parameters, the mean and the variance, denoted by μ and σ, as usually (Altman, 1995). The probability density function $Y1$ of a normally distributed variable is in equation (1).

$$Y1 = f(x) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } -\infty < x < \infty \tag{1}$$

$$Y2 = \left(\sigma\sqrt{2\pi}\, Y1\right) = \sigma\sqrt{2\pi} * \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{\sigma\sqrt{2\pi}}{\sigma\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}} = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

$$Y3 = (Y2)^v = \left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right)^v \tag{3}$$

Equation 6.1 Enhanced normal distribution anomalies detection

The normal curve for variable *Y1* value indicates that data in normal distribution shape and coefficient $1/(\sigma\sqrt{2\pi})$ keeps the area under ND curve of value. However, this will cause the mean point value to vary between 1 and 0. While our objective is to keep the mean value point close to 1, we multiply the value of *Y1* by the coefficient to the power of -1 to remove the coefficient effects as shown in Equation *Y2*.

Continuous values of *X* between - ∞ and ∞ so that each conceivable interval of real numbers has a probability other than zero $-\infty \leq V \leq \infty$. Where X is denoted to a single POS transaction, furthermore, we modify the equation to give the user more exposure to potential anomalies by adding variable *(V), V* is an independent variable added to extend the curve horizontally.

Where $0 < V < \infty$. This will extend variable *Y2* value horizontally to make sure that most of the values reside closer to the mean (1). Values that reside far from the mean point are assigned with lower values, such as 0.9, 08, 07, etc.

Variable *Y3* indicates the normal distribution shape in the newly modified equation that removes the anomalies found in the dataset.

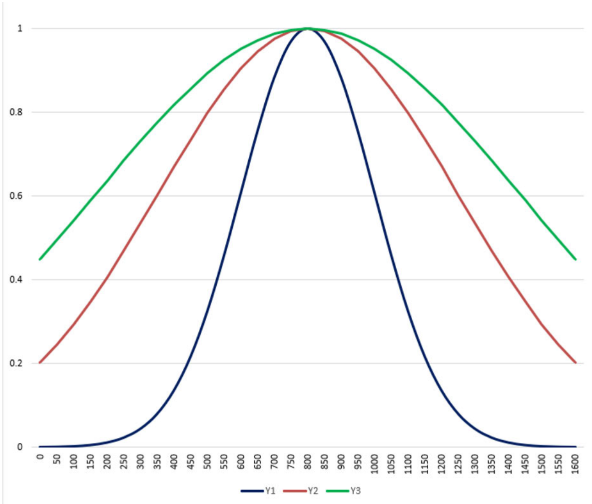


Figure 6.5 Samples POS transactions fall near or far standard deviation from the mean

Figure 6.5 illustrates the shapes of different transactions that lie close to or fall some standard deviations away from the mean, as represented by the normal bell shape, identifying the extent of outliers in the POS dataset by using continuous variables in the range of (0-1). The shapes offer clients a dimensional vision of outliers and potential outliers. The client can further investigate to determine which transactions might be legitimate or potential anomalies, or these transactions are caused by improper measurement and recording errors.

It is important to note that when running anomaly detection, we have spotted thousands of transactions that are not linked to any customers, transactions with null values, or a negative total value. These transactions seem to be against the normal retailer behavior, but they might also be void transactions where the retailer never charged by the customer. However, they should be regarded as potentially fraudulent transactions and investigated by the retailer. These patterns do not conform to the patterns we set for this research. Thus, those transactions were removed as anomalies. Being able to notify the retailer of such transactions proved to be an important side product of this study.

6.2 Market Basket Analysis Implementation

To achieve seamless database integration, we use PL/SQL language for the implementation of the Åbo algorithm. PL/SQL is an Oracle procedural extension of SQL. It brings many advantages to the proposed MBA model implementation.

1. It is highly portable (compatible with all operating systems), which makes it the only language that supports customized error handling.

2. It sends an entire block of statements to the database server at the time, which reduces network traffic considerably.

3. Applications developed using the PL/SQL language can be broken down into smaller modules. For complex applications, PL/SQL allows the developer to break the algorithm down into smaller, manageable, and logically related modules.

4. Oracle provides out-of-the-box tools for the migration of PL/SQL code to other programming languages, JAVA and T-SQL. This is essential in the implementation of MBA to various other industries, like healthcare and banking.

5. It comes with an easy-to-use built-in conversion of codes utility between other applications and other programming languages.

The detailed presentation of our Åbo model is in Research Article 5. Figure 6.6 demonstrates the implementation of the Åbo algorithm as compared to the four first phases of CRISP-DM.

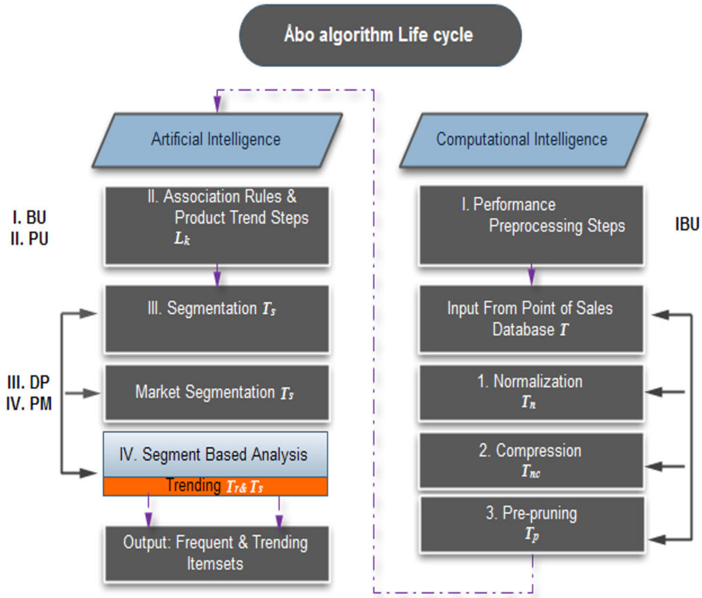


Figure 6.6 Åbo model as compared to the cyclical process of CRISP-DM

The first two phases of the Åbo are qualitative approaches aiming to have complete business clarity. A clear project plan is outlined with a schedule for model completion in each phase. Resources required for the success of the proposed model are requested during these two phases, for example, sample transactional data and meeting with key managers.

To gain an understanding of the business, we address the client with questions about the types of consumer data they have and their technical resources for handling both data and data mining. It is also important to find the possible factors, and if there are any, if the retailer has a contingency plan to mitigate them. The following questions relate to the retailer's business environment data available on that since it is critical for the success of the MBA to address external factors as well. In the practical level, we use semi-structured interviews and questionnaires with marketing and sales managers. For understanding business data, we describe it, perform exploratory analyses and verify data quality.

We also looking for initial insights by conducting interviews and questionnaires with POS engineers. The first two stages closely follow the guidelines of the CRISP-DM.

For preparation and modeling of data, the design and development of the Åbo algorithm is unique. As we aim to improve the Apriori algorithm with intelligent and efficient features, we apply modules of computational intelligence to address the shortcomings of the Apriori. For data preparation, we implement CI to normalize, compress and pre-prune data to reach the following improvements to the Apriori.

a. Avoid rescanning the database.
b. Reduce the size of candidate itemsets.
c. Accelerate both the joining and the pruning processes.

For the modeling , we implement the attenuation factor as a variable to automate the prediction of trend and behavior. With this new computational feature, we aim at understanding the factors that influence the products' upward and downward trending popularity. As we also want to find rarely occurring events and brands with trending popularity before they peak in sales figures, we consider that this feature adds computational intelligence to MBA.

The efficiency of the Åbo algorithm is measured by reducing the number of database scans of $\boldsymbol{k}$-itemsets with new approaches to normalize, compress and pre-prune the data, and hence to reduce the size of the generated dataset. These are implemented as steps 1, 2 and 3 in the preprocessing (Figure 6.6).

The first step to prapare the dataset for the Åbo algorithm is to extract data from the POS dataset *T*, and transform it into a separate data warehouse schema as normalized dataset T_n. Simultaneously with this process, the algorithm converts item identification codes into the new unified 6-digit format.

Then, as the second step of preparation, the algorithm compresses T_n into T_{nc}. As a result, each transaction is now one single horizontal line enabling a horizontal scan to improve scan performance. The algorithm now scans one single record instead of many. Since the average size of a retail POS database is 100 GB, this will significantly decrease the number of database scans.

Then the algorithm prunes T_{nc}. It eliminates itemsets not passing the support and confidence thresholds. After this step we reach dataset T_p, where numerous re-scans of the database and generation of sub-items are no longer needed. This leads to considerable performance improvement compared to the Apriori. When in Apriori, the algorithm scans the entire database in every iteration, the algorithm now scans the database once to generate all frequent itemsets by a single pass over the POS database to count the support of the candidates resulting in the set of 1-itemsets.

Then the extracted analysis results are evaluated and cross-checked with the marketing and sales managers to validate that the MBA model complies to the decision model of the decision-makers. If so, the MBA can deploy to perform systematic knowledge discovery.

6.2.1 Computational Performance and Efficiency

Next, we discuss the results obtained by the application of normalization, compression, and pre-pruning in the Åbo algorithm. We find that these techniques contribute to both efficiency of calculation and computational intelligence and can be regarded as some of the most important contributions in this dissertation. The main efficiency factor is the reduced number of database scans, which leads to the reduction in the size of the dataset. The benefits in computational intelligence partly follow from efficiency.

As for normalization, the reduction in digits into the new unified 6-digit format improves the performance of computation. As Tyugu (2007) points out, the smaller the data types are, the faster the data retrieval is because they occupy less physical space on the hard disk memory and generally require fewer CPU cycles to process.

A sample of the unification of items IDs within the normalization process is presented in Research Article 5.

As for compression, As Singelis (1995) points out that in a typical relational database stores data horizontally in tables, and table is organized into columns. Each column stores one type of data as row (character strings, integer, number). For example, the products table would have columns such as prod number, price, size, color, and vendor. If a customer purchases one shirt, two pairs of pants, one t-shirt, and five pairs of socks. The products table would store each item in a separate row, and for this transaction it will be four rows stored in the database for one single transaction. In our approach, the algorithm places all frequent itemset in the database in a row (or record). The Åbo algorithm now scans one single row instead of many rows. This will significantly enhance the computational performance of Åbo algorithm when dealing with POS database, which stores millions of transactions.

Apriori algorithm is known with its excessive I/O load and occupies much of the memory space due to repeated full database scanning that need storage capacity (Yuan, 2017). The approach uses an iterative method of layer by layer scan to generate high-dimensional frequent itemsets. This approach is only suitable for small transactions database. However, the vast majority of POS databases are Bigdata or data warehouse. Therefore, huge performance bottleneck is expected (Zeng, Xiong, & Li (2018).

Contrary to this approach, we introduce a technique to accelerate the joining and pruning process with two steps.

1. Convert original itemsets from a vertical structure to a horizontal structure.
2. Compress the horizontal itemset to generate an even shorter dataset.

To do this, we combine similar itemsets into one row and then count the number of occurrences for a particular itemset, as demonstrated in Research Article 5.

Our pre-pruning executes a database mapping to avoid repeated re-scanning of the POS database and eliminates the itemsets that have non-frequent subsets. It scans the database only once to create a new dataset. Then it generates a new dataset directly from the previous pass on the same level k -itemsets. Next, new candidate sets are generated by using only the newly generated dataset rather than rescanning the entire original POS database. This step improves the performance and efficiency of the algorithm to generate frequent itemsets and reduces the size of the transactional POS database. Now, the candidate set for the next level k+1 -itemsets is generated by joining, using the same process as in the generation on the lower level k -itemsets. As the approach only uses the new dataset rather than the original POS database, the efficiency of pre-pruning increases in each join

$1. T_n \leftarrow Normalization\ of\ T$	'Normalization of T to reduce database size
$2. T_{nc} \leftarrow Compression\ of\ T_n$	' Compression of T_n to further reduce database size
$3. T_p \leftarrow Pre - pruning\ of\ T_{nc}$	'Elimination of itemsets that do not pass the support threshold
$4. L_k$	'Discovery of frequent items, apriori phase
$5. G$	'Segmentation of the dataset
$6. T_r$	'Discovery of trends in itemsets

Equation 6.2 Variables in the pseudocode of Åbo algorithm

Research article 5 shows how normalization, compression, and pre-pruning reduce the number of candidate k-itemsets and result in smaller and more accurate ARM.

From here, we will next continue to the fourth phase of the Åbo algorithm that generates new candidate ***k***-itemsets using the frequent ***k-1***-itemsets found in the previous iteration. The algorithm eliminates all candidate itemsets whose support count is less than the support threshold. Equation 6.3 presents variables used in this process, while the pseudo-code of the Åbo algorithm is found in Equation 6.4

T: $transactional\ Database$

C_k: $Candidate\ itemset\ of\ size\ k$

$C_{i:}$ $item\ i\ in\ C_k$

$L_{k:}$ $frequent\ itemset\ of\ size\ k$

S: $Support$

CN_c: $number\ of\ occurrences\ of\ itemset\ c$

G: $number\ of\ time\ segments$

g: $time\ segments$

T_g: $all\ transactoins\ in\ time\ segemnt\ G$

t_g: $single\ transaction\ in\ time\ segemnt\ G$

C_g: $total\ count\ of\ itemsets\ in\ time\ segment\ G$

T_r: $itemset\ trend$

T_s: $slope\ of\ the\ itemset\ trend$

att: $Attenuation\ Factor$

Equation 6.3 Variables in the pseudocode of our novel approach

1	*Start Åbo (T_P, S , att, G)*	‘start segmenting the dataset
2	$L_1 \leftarrow (\leftarrow \ \{Large\ 1$	‘count all item sets
3	$k \leftarrow\ 2$	‘set itemset size to 2
4	$While\ L_{k-1} \neq\ \phi$	‘start loop when k-1 is not an empty set
5	$For\ \ (g = 1\ ;\ G\ ;\ g++)$	‘for segment g
6	$T_g\ \leftarrow\ T_p\ for\ \ period\ \ g$	‘all transactions T_g in segment g
7	$For\ each\ transaction\ t_g\ in\ T_g$	‘reduce to transactions in T_g
8	$For\ each\ transcation\ c\ in\ C_k$	‘reduce to transactions in C_k
9	$If\ \left(\ c\ \subseteq t_g\right)\ then$	‘if item set c occurs in transaction t_g
10	$count\left[C_g\right] \leftarrow count\ \left[C_g\right] + 1$	‘add it to the count of item sets in time segment g
11	*End If;*	‘end of If line 9
12	*End For;*	‘end of For line 8
13	*End For;*	‘end of For line 7
14	$T_{rc}\ \leftarrow T_{rc} + T_s\ (count\left[C_g\right], count\ \left[C_g - 1\right])$	‘calculate item set trend T_{rc} with slope T_s (in segment g)
15	$T_{rcatt}\ \ \leftarrow T_{rcatt} + T_R * att_g$	‘multiply with the attenuation factor (in segment g)
16	$R_c\ \leftarrow\ D_{max\,-\,date\,(c)}$	‘calculate R_C, recency of item sets (in segment g)
17	$M_c\ \ \leftarrow \sum cost\ (C_i)$	‘calculate M_C, the monetary cost of item sets (in segment g)
18	$CN_c \leftarrow CN_c$+ count $\left[C_g\right]$ * att	‘calculate the number of itemset *c* occurring (in segment g) and multiply with an attenuation factor
19	*End For;*	end of for line 5
20	*L_k = ϕ*	‘create an empty the set for collecting itemsets of size *k*
21	*For each candidate c in C_k// Prune*	‘final pruning, for each candidate itemset *c* of size k
22	*If (CN_c >= S) then*	‘if itemset support CN_c is larger than support *S*
23	$L_{k:} \leftarrow L_{k:} U\ \{c\}$	‘include itemset *c* into L_k
24	*End If*	‘end of if line 22
25	*End For;*	‘end of for line 21
26	*k ← k +1*	‘add the number of items by 1 for the next loop (to test on line 4)
27	*End While;*	‘end of while line 4
28	*Return L_k*	‘final set of itemsets of size k
29	*End Åbo*	

Equation 6.4 Analysis steps in the Åbo algorithm

6.2.2 Computational and Artificial Intelligence

Now we introduce the time series analysis into the MBA to explain and anticipate the lifetime value of itemsets. Wang, Smith, Hyndman & Alahakoon, (2004) state that a time series is a sequence of observations ordered by time points. One application is to discover items that are commonly bought together across time segments. Finding new, trending products can be challenging for retailers. By the time, the retail manager sources them, the popularity may be over, and the demand has vanished. By knowing popular products before their sales peak (i.e., catching the wave), the retailer can earn money from the growing sales, as every niche has its trends, and increase product cash flow.

Åbo algorithm determines the trending factor that emphasizes the most recent purchase for each itemset. First, we compute a slope for the sales trend from sequential sales data **Cg**, where **g = {1, …, 6}** denotes the time sequence of six months for the itemset C consisting totally of 6 six-month-long periods. Then we introduce the attenuation factor ***att***, used to discount the significance of earlier itemsets and preserve the importance of the most recently realized sales in the trend.

The *attenuation factor*, also known as the natural arithmetical value of signal transmission over long distances, is an exponential function of the path length through the medium, in other words, the strength or intensity of a signal or digits is reduced. This method that generates and validates data and de-noises signal data is widely used in telecommunications, engineering, optic fibers, and ultrasound applications (Tate, 2016). We use the attenuation coefficient as a factor to account for the recency of an itemset, the recency being one of the variables for calculating the product score in segmentation with the RFM-model (McCarty, & Hastak, 2007). In our approach, the longer the time is since the generation of an itemset, the higher the attenuation coefficient is set. In other words, the value of the itemset is discounted based on its historical occurrence to the point of analysis or respectively accentuated based on its recency. For details of using attenuation, we refer to Research article 5.

To calculate the attenuation factor, we first compute itemsets' purchase amounts over the predefined time segments. From here, we can calculate the discounted value of past itemsets purchase slopes. In this way, we preserve the effect of the most recent purchases and emphasize their importance. This enables the retailer to source and spots popular high-demand itemsets. It will also reveal information about infrequent profitable itemsets often missed by the Apriori algorithm. These could potentially contribute and offer the retailer a broader vision of the viability of particular itemsets.

With the attenuation technique, we capture two critical factors in marketing, time-based segments and occasion-based segments. Segmentation with time-based dimensions can be highly effective. For example, some stores work longer hours than others, and some products are sold only at certain times and dates of the year (e.g., Eid Holiday, Month of Ramadan, Christmas, the New year, etc.). Segmentation with occasion-based dimensions is based on the observation that customers tend to behave and think differently on different occasions, and such differences can be beneficial for product segmentation.

6.3 Performance Analysis

In this thesis, the core empirical material consists of the developed MBA model that mines shopping patterns. According to Mowen (2000), customer shopping behavior is the subject of many studies. Most of these studies are, however, from the customers' cognitive perspective. In the implementation of a data mining model questions of performance, both as calculation efficiency and as provided decision support, arise as key areas of inquiry. Consequently, as we have developed an advanced version of a known ARM model, the Åbo model, it is appropriate to compare the performance it to the original, the Apriori model. The results of the performance analysis presented here are based on research that contributed to Research Article 5.

Here, we show the results of implementing and comparing the Apriori and Åbo algorithms. Also, we evaluate the proposed AI techniques and look for answers to what extent the generated data will be efficient for market intelligence.

6.3.1 Algorithmic Performance

This chapter contains analytical comparisons between the Apriori and Åbo algorithms based on the algorithms' execution performance with a large POS database. We use precise operational metrics to assess the health of our POS database and to validate the performance and efficiency of each of the algorithms upon successful execution. We apply six operation metrics to measure resource consumption for the following database categories.

- Speed, as the rate at which the algorithm sends data back to users, measured as execution time.
- Memory consumption, which is the total amount of memory used by the algorithm, measured using execution or run time of each algorithm.
- Storage capability, which is the physical POS dataset size used after the algorithm has completed its mining process.

Table 6.2 shows six operations of metrics used in this research that pertain to the database performance.

Table 6.2 Operation performance metrics Description

Operational metric	Description
Records count	Total number of records of the generated itemsets
Database size	Size of the database
Total IO read (latency)	Time each algorithm I/O takes to complete (Database size * Database scans)
Mining process	Repeated instruction each algorithm takes until all conditions have been met
Execution time	Total time in minutes consumed by each algorithm
Records count	Total number of records of the generated itemsets

The first two methods to increase algorithmic computational performance are the normalization and the compression of the extracted data. As our results show, these methods reduce the time consumed in the generation of candidate itemsets.

The experiment shows that convergence speed and accuracy of the Åbo algorithm significantly enhance the efficiency of all operation metrics, including more efficient and feasible association rules discovery than the classic Apriori algorithm.

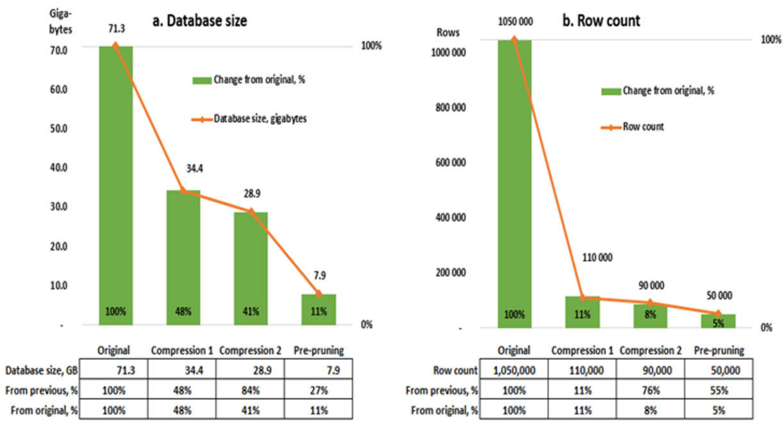


Figure 6.7 Database performance phases (Apriori vs Åbo)

Table 6.3 Operation performance metrics

Matrices	Apriori	Åbo
Database scans	851	784
Records count	1,050,000	50,0000
Database size	71.3 GB	7.9

6.4 Performance of Decision Support

For retail management, it is often useful to verify sales and other key performance measures over time. With time-series data, it is possible to extract meaningful knowledge from the shape of the data. When this kind of longitudinal analysis is used with MBA, it is possible to overcome the limitations of the association rules algorithms that ignore the product's behavioral changes over time.

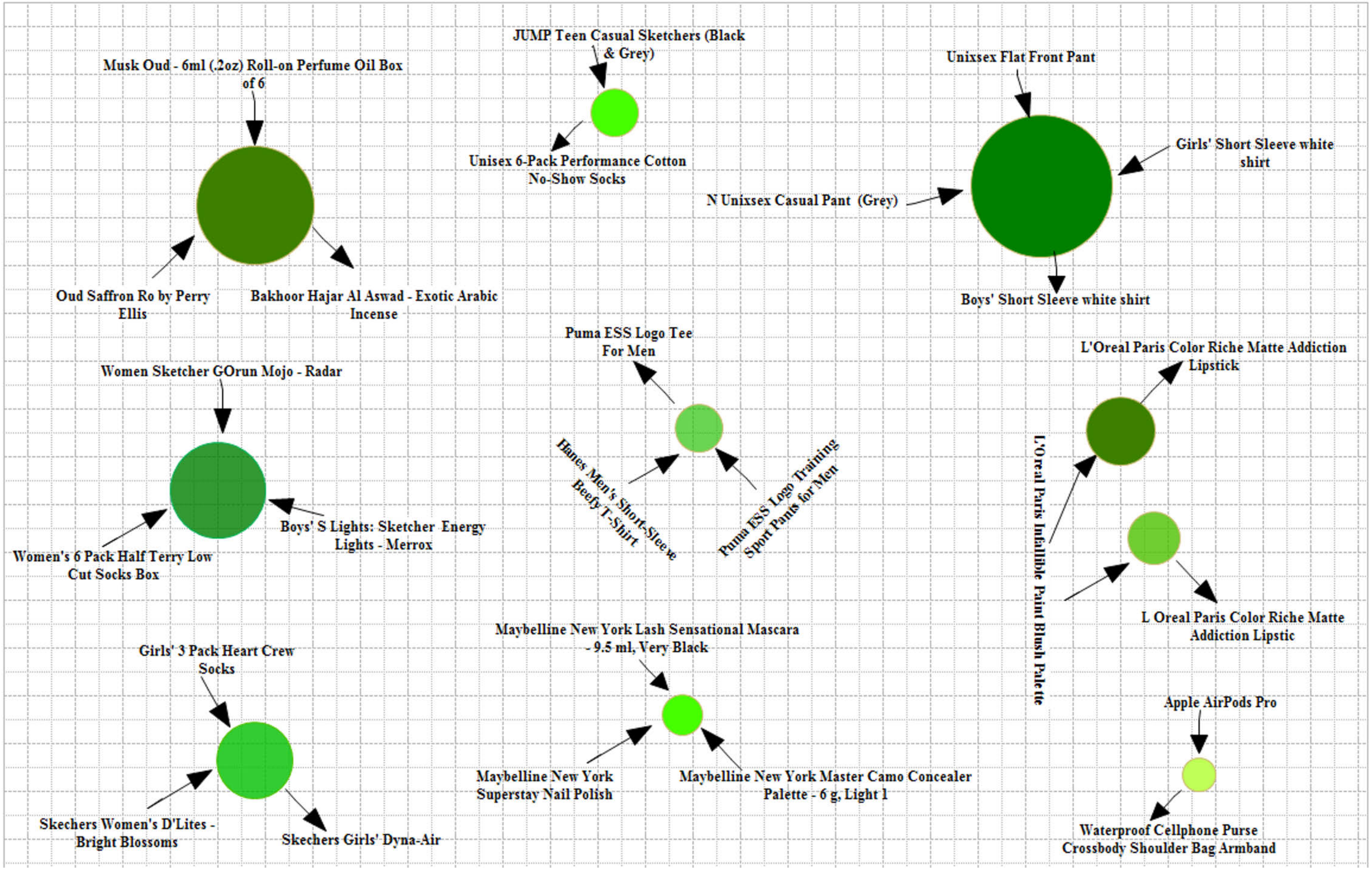


Figure 6.8 The top ten rules in terms of lift

Figure 6.8 The top ten rules in terms of lift

Figure 6.7 shows the top 10 rules in terms of lift. The rules can be increased or/and modified based on the client's preference. It shows how the itemsets cluster around the circle that represents the associated itemsets, while arrows indicate the relationship in the rules. For example, one rule is the purchase of itemset {Women Sketcher GOrun Mojo - Radar and Women's 6 Pack Half Terry Low Cut Socks Box} that is always associated with {Boys' S Lights: Sketcher Energy Lights – Merrox}. Another rule is one of the bestselling itemsets. The {Oud Saffron Ro by Perry Ellis} (Arabian Perfume) is one of the most bestselling itemsets and is always associated with {Bakhoor Hajar AL Aswad - Exotic Arabic Incense}, and also associated with the purchase of {Musk Oud - 6ml (.2oz) Roll-on Perfume Oil Box of 6}.

Another rule is that the purchase of {Apple AirPods Pro} is associated with {Waterproof Cellphone Purse Crossbody Shoulder Bag Armband}. The size of the circle depicts the level of confidence associated with the itemset rule and implies higher support, while the green circle (the color circle) implies a higher lift. Therefore, the larger and the darker the green circle is, the better the rule is, and therefore the products have a higher probability of being purchased together.

We can now conclude that a retailer can project which rules the marketing efforts should focus on to achieve the highest ROI. Another associated rule is that {Unisex Flat Front Pant, N Unisex Casual Pant (Grey}, Girls' Short Sleeve white shirt, Boys' Short Sleeve white shirt} has high support of 0.0097, which means that whenever a customer purchases {Unisex Flat Front Pant}, {N Unisex Casual Pant (Grey}, {Girls' Short Sleeve white shirt} or {Boys' Short Sleeve white shirt}, they also purchase a (School bag}in the same shopping trip.

6.4.1 Finding Trends and Popularity in Itemsets

The main drive to introduce CI into Åbo algorithm stems from the human feature to rely on the ability to visualize the shape of data. Usually, we use descriptive measurements over time to extract knowledge from the shape of data. One of the main limitations of ARM algorithms is that they ignore behavioral changes of products over time and therefore ignore the importance or relevance of changing data. For the retail industry that relies heavily on trends, solid facts and up-to-date data to make key strategic business decisions, the lack of behavioral changes in decision data is not ideal.

For our analysis, we use visual tools to highlight the effect of attenuation in trending. We use three different two-product itemsets[1]. Figures 6.9 and 6.10 show behavior of 2 itemsets, over six-time segments. The blue line represents the original itemset trending, and the green line represents the attenuated trend after the time series attenuation.

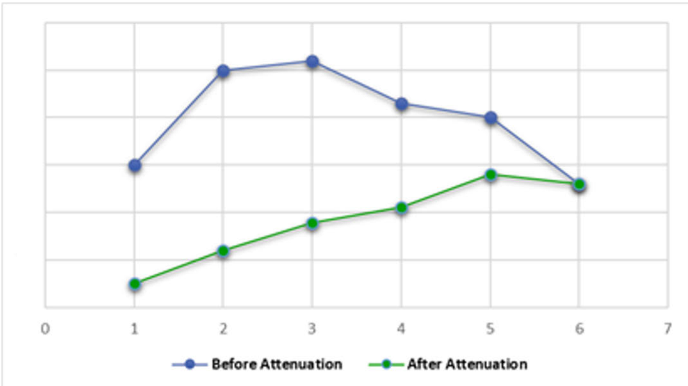


Figure 6.9 Apriori trend attenuation, itemset 1

Figure 6.9 shows itemset 1 trending upwards during the first three periods of the analysis. Consequently, the demand decreased, and itemset popularity declined in later segments, specifically from the fourth time segment onwards. Once the objective facts are extracted, it can be assumed that a separate quantitative analysis of the same data would produce the same results.

The Åbo algorithm, with its attenuated time series, leads to fairly different results even when analyzing the same dataset. The green line of the after attenuated time series shows steady growth, with the itemset trending upwards, up to segment 6 in the green line. Where the degree of the itemset downwards trending is much smaller in the original itemset.

This does not necessarily mean that customers are not actively buying the products of the itemset. Instead, there might be a marketing disconnect between the product and the customer throughout segment 6 or a lack of perceived value from the customer's perspective. Any newly extracted insights will fundamentally influence the marketing perception of this itemset. The retailer could then evaluate the viability of this itemset and perhaps find a way to reverse or slow the most recent downward trend for these products.

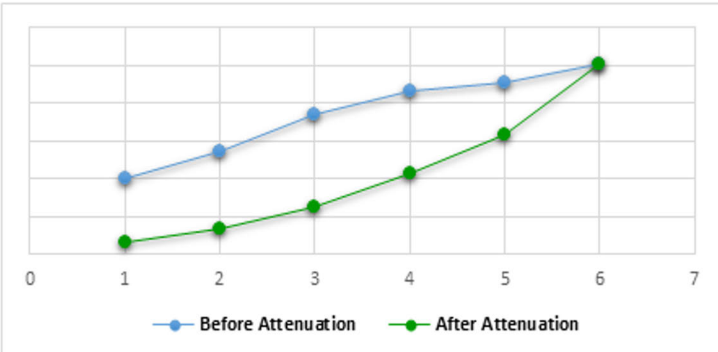


Figure 6.10 Apriori trend attenuation, itemset 2

Error! Reference source not found.6.10 shows the itemset rising in sales t hroughout segments one-five, and its trending popularity shows no sign of slowing down. Would the slight downward trend in the fourth segment suggest that the popularity of the itemset is over? Not necessarily, because the fourth segment also complies with the overall upward trend after attenuation. When we dig further into the features of this itemset, we see that it is a traditional Arabian unisex perfume, usually highly demanded by Arab customers and considered as an exotic and pure expression of Arabian culture.

This itemset tends to be very popular during the holiday seasons, and thus we see the demand has surged during the EID holiday. However, there are spikes of demand throughout the year because of the preferences of Ar-ab customers. Arab Women between the age of 25 and 50 tend to be the largest target audience for this itemset.

The trending popularity of itemsets should encourage the marketing team to promote the itemset in many ways, like creating a collection of similar styles for women's fashion collection.

6.4.2 Products and Services Recommendation Strategies

Effective shopping models that address customers with unique personalized, individually tailored experiences aim at conveying value to customer shopping journeys (Chandra, Verma, Lim, Kumar & Donthu, 2022). Strategies that build on Market Intelligence revolve around having data about customers' personal shopping preferences and interests and considering how often corresponding products exist together in the customer baskets. To design a personalized products recommendation strategy based on customers' shopping preferences, marketers should build their strategies based on factual customer data. Specifically, the customer data should incorporate in-depth customer behavioral data for retailers to be able to provide individualized recommendations.

Systematic knowledge of itemsets popularity will provide in-depth knowledge about the growth of inequality in the distribution of products' profitability. The retailer can then develop Market Intelligence tactics and metrics that best represent the itemset's popularity among customers (Von Reischach, 2009; Yongjie Yang, 2016). This will help the retailer capture customers' shopping behavior patterns and wish lists and help to trigger a sense of urgency among interested customers to boost conversions.

Kotler, Armstrong, Harris & Piercy (2013) summarized the objectives of marketing as.

1. Create product engagement and trial.
2. Maximize market share.
3. Maximize profit while defending market share.
4. Reduce expenditure.

Table 6.3 illustrates some recommendations put together with the client for B2C marketing strategies based on the analysis of the Åbo approach.

Table 6.4 Suggested recommendation strategies

	Segment		Trend	Recommendation Strategy
Itemset 1	Trending Downwards	Trending Upwards	Popularity is Long-lived	Increase Itemset visibility. Products bundling. Marketing these products on Social Networks. Organize contests to promote products with discounts and price matching. Implementing digital marketing with Customer Loyalty Program.
Itemset 2	Trending Upwards	Trending Upwards	Popularity is Long-lived	Increase Itemset visibility. Products bundling. Marketing on these products on Social Networks Organize contests to promote products with discounts and price matching. Implementing digital marketing with Customer loyalty Program. Direct email marketing.

Based on the analysis result of itemset 1 and 2. We have made some recommendations for the retailer to start implementing B2C marketing strategies to help keep itemsets popularity and increase profitability. Our first recommendation is the itemetset visibility.

Itemset 1 trending popularity can be increased by enhancing the visibility of itemsets and present them in more visual appealing to the customer, which will help customers better identify and engage with the itemset. Adding new imagery and graphics will add more personality to the itemset. Also, adapting new target customer's experiences will help capture the interest of both casual and loyal customers.

Our second recommendation is to introduce products bundling strategy and offer complementary products together with the itemset in exchange for an attractive discount. Products bundling strategy is built on customer behavior psychology, as customers always react positively when they are presented with deals, even if they ultimately spend more.

Our third recommendation is to start an eCommerce marketing engagement strategy. Nowadays, the advancement of social media marketing provides a new way to engage with different customers connected to social media platforms. Our fourth recommendation is to adopt a contest marketing strategy that can drive itemset awareness and help engage with bigger local audience.

Our fourth and must-implement recommendation is to implement a digital marketing strategy to harness the power of customer's data. Implementing customer loyalty program is essential to engage with customers digitally. Customer loyalty program will help the retailer create services like customer referral campaigns and offer coupons and gift cards to the customers. Such services will turn happy customers into a stream of revenue.

Happy customers will always praise the experience they received across the retailer’s multiple channels and create positive word of mouth. Another effective way of using digital marketing is using social media influencers to boost itemset awareness. Social media influencers generally have a vast reach and can help promote products across larger audiences.

Since itemset 2 popularity is trending upwards both in the segment and trend categories. We recommend better use of direct email marketing.

Real-time data updates, forms, and user actions, right within the email body. Marketing campaigns and promotions tailored to the customer's shopping preference are sent directly to customers. This approach aims to draw customers' attention, monitor the itemset's success, and improve the relationship between the retailer and customers.

6.5 MBA Analysis

We have here presented the structure of a viable market basket analysis based on the reviewed literature and the developed Åbo model and algorithm. We introduced computational and artificial intelligence methods to complement the Apriori algorithm. In this chapter specifically, the two unsupervised algorithms, the Apriori and the Åbo, experimented with a POS dataset of a small Kuwaiti retailer to forecast customer shopping behavior. We conducted experiments to assess computational performance and efficiency of the algorithms, and the extracted association rules illustrated the outcomes of the developed computational intelligence techniques.

An analysis of the trending popularity over products' lifetime was conducted end presented.

As the second, complementary form of analysis, all the results presented in this chapter were reviewed and cross-validated by the marketing experts we worked with. They provided us with comments and views of market practitioners that confirmed the practical usefulness of the findings in this chapter. As a conclusion of the discussions, we conclude that the key to a successful MBA model is to extract value from the POS dataset and build up an understanding of customers' shopping preferences and needs.

Based on the marketers' response, the proposed MBA model has proven to be scalable and robust in everyday use. The retailer is now able to catch old and new trends in a continuously changing market. Now marketers can systematically implement an iterative process of market knowledge discovery, and the extracted information is readily available. If the retailer is interested in marketing activities such as cross-selling or targeted campaigns, the Åbo model provides useful market knowledge.

A major part of the market knowledge the Åbo algorithm provides lie in the use of products popularity as a "descriptive norm." "Descriptive norm" refers to evidence that informs customers of what typical product behavior is, as opposed to what ought to be, and suggests what is normal in a given situation (Raska, Nichols & Shaw, 2015). The "Descriptive norm" here describes how most customers perceive product's popularity. A piece of knowledge found with descriptive analysis used in a formative way, which lets it be used as a marketing cue, such as the customer intention to make purchase product. As a baseline, it assumes that the customer prefers a product with a reputation of popularity because popularity is perceived as superior quality. This fits past studies on the matter, which have primarily focused on analyzing the market share as a popularity indicator of product on the perceived quality (Dean, 1999). However, it also highlights how rarely there has been an investigation into the effect of time on product perceived quality and popularity in a customer-shopping context. In this chapter, the identified rules and the characteristics of the rules were applied to the individual client. The proposed Åbo model provides valuable insights into the associated itemsets, and we discovered new insights with each rule.

Table 6.5 Excerpt from case data: Top 5 Infrequent Itemsets

No	Rule
1	{Men leather Jacket}, {Men Colored Winter Socks} => {Coffee - Tea Mug}
2	{Boys' S Lights: Sketcher Energy Lights} {Unisex Sweatpant Fleece Workout} => {Green Lazer Pointer Light}
3	{Women Musk Oud}, {Women Suede Pants }, { Small Teddy Bear} => { Legos}
4	{Mens Eco Fleece Sweatpants}, {Electronic Lighter} => {No-brand earbuds}
5	{Barbie Perfums for Girls 2 giftSet: Toilette Spray + Shower Gel}}, {Microsoft Wireless Surface Headphones - Noise Canceling} =>{Polarized Swimming Goggles}

In recent years, there has been an increasing demand for mining infrequent itemset. Cagliero & Garza (2013) proposed an algorithm for discovering infrequent itemsets. However, traditional mining infrequent itemset algorithms still suffer from the scalability, especially if the dataset size is very large.

Table 6.5 shows an example of infrequent itemsets spotted by the Åbo algorithm having different support counts. Our Åbo algorithm uses multiple level supports thresholds to simultaneously discover frequent and potentially infrequent interesting itemsets, whose frequency of occurrence is less than the minimum minsup threshold.

Infrequent patterns are rarely found in big POS databases because they are considered uninteresting and automatically eliminated using the support measure. The retailer can identify positive associations between infrequently purchased items, but slightly missed the minsup threshold or can predict itemset failure. For example, itemset "{Women Musk Oud}, {Women Suede Pants }, { Small Teddy Bear} => { Legos} always see purchasing surge from the beginning of September to December. Itemset "{Men leather Jacket}, {Men Colored Winter Socks} => {Coffee - Tea Mug}" this itemetset is often purchased at the end of April when the itemset is on sale.

Infrequent itemsets are of great interest not because they could be potential candidates for frequent itemset but because they often relate to rare and crucial associations. Some itemsets might relate to valued negative association rules, and such itemsets can be discontinued by the client, or used in cross-selling marketing to bundle it with a trending upward itemset. Also, retailers might spend more time analyzing negative associated itemsets rather than having to read the millions of records of non-profitable products. However, extracting all non-frequent itemsets is out of the scope of this research

Chapter 7

Conclusion And Answers To The Research Questions

7 Introduction

Market Intelligence is the data-oriented insights gained from external and internal data to empower effective and successful marketing strategies around marketing pillars such: product, price, placement, and promotion. Implementation of AI technologies will affect marketing dynamics for many years to come and allow marketers to accommodate changes in consumer purchase expectations and focus their marketing efforts on a higher-level strategic decision to create an impactful message.

'Who are profitable customers?' This is the starting point for many CRM studies pertaining to developing comprehensive models of customer profitability to assess the return on marketing investment. With today's reliance on POS databases, effective extraction of usable marketing knowledge is extremely significant. However, POS databases have gone from being flat to highly structured architectures with a well-defined taxonomy and hierarchy. The number of association rules derived from POS databases has also increased. To address this, we proposed Åbo algorithm as an advanced ARM method to forecast customer shopping behavior.

In the course of developing the proposed Åbo algorithm, we introduced new methods to supplement the classic Apriori algorithm. The Åbo algorithm has shown improvements in two ways, in data utilization, the computational efficiency has improved, and, in knowledge accumulation, the AI methods better utilize product life cycle and trending popularity. The algorithm successfully reduces the number of extracted rules and is more efficient and accurate in extracting interesting rules. The findings found by the implementation of this MBA model in the context of small to medium size retailers lead to logical answers to the research questions.

7.1 Answers to The Research Questions

This research focuses on three research questions.

1. Does the Åbo algorithm improve the computational performance and the intelligence of the classic Apriori ARM algorithm?
2. Do small and medium-size retailers get better market intelligence with the new approach?
3. How can we achieve optimal anomaly detection without compromising the accuracy of the data?

1. Does the Åbo algorithm improve the computational performance and the intelligence of the classic ARM algorithm?

The introduction of artificial and computation intelligence to complement the classical Apriori algorithm shows a considerable increase in both intelligence and computational efficiency. Firstly, from the vantage point of performance, the comparative analysis demonstrates that the time taken to scan database transactions is vastly reduced, and the time used to generate candidate itemsets is considerably decreased. Moreover, the size of the analyzed database decreases.

Secondly, from the vantage point of Artificial Intelligence, Åbo algorithm addresses the time aspect more profoundly. The effect of time on product popularity is logically included in time segments when the MBA model spots the significance of the most recent transactions. The attenuation process enables the retailer to capture trend information with time-discounted weights. For example, itemsets from two years ago should not have the same popularity effect as itemsets data two days ago. Moreover, the analysis results show that Åbo algorithm picked interesting, rare infrequent itemsets by detecting shopping patterns using the time component that the minimum support threshold omits.

As a result, the retailer is now able to understand customer shopping habits and needs in a better way. Cross-selling recommendations can be used to drive up engagement between less popular infrequent itemsets with similar upward trending itemsets. This strategy may maximize the customer's exposure to the product's recommendations.

2. Do The Small and Medium-Size Retailers Get Better Market Intelligence with the new approach?

Finding Market Intelligence for SME retailing is a problem that retailers often pass by because MI seems to be exclusively destined for large retailers with large budgets. However, this research introduced new Artificial and Computational Intelligence methods to demonstrate that these methods are not outside of the SMEs' practical interests. In this research, marketing analysis and sales predictions that marketing and sales personnel previously executed manually are now automated.

Åbo algorithm has shown computational capabilities to compare millions of isolated POS transactions and extract timely and comprehensive insights to support investment decisions to marketing campaigns. To be able to decide how much and when to invest in a campaign benefits the marketers. The strategic benefits of implementing MI in all areas of marketing mix aid decision support processes of the SMEs.

The results confirm the benefits of the model in the operational level as well, as the unusual co-occurrences among products in various baskets lead to a clear vision of what types of products are of prime interest. The data acquisition by MI had a very positive impact on sales.

The analysis result derived from the POS database proves that a marketing intelligence strategy driven by an AI engine can improve the visibility of the products on products promotion, identify which products might be vulnerable to competitive offers, eliminate time-consuming marketing activities, and discover insights about which uncorrelated products are best sold together. This empowers the retailer to optimize marketing processes and replay different marketing scenarios and projects which rules should the marketing efforts to focus on to achieve the highest ROI, knowledge of what products are hot on a particular day and week, hyper-personalized services, and convenient shopping that all contribute to the area of the marketing.

The proposed MBA model answers a widely debated question raised by Zeithaml (2000) “What demographic and psychographic variables are most effective in characterizing profitability tiers?” It was able to identify profitable demographic segments, and the analysis results revealed Kuwaiti citizens with the highest purchase power. Arabian products are the best-selling products with the highest associations from the product's perspective. Also, younger Kuwaiti females are predominately identified as the best customers. Finally, knowledge gained from demographic data was found exceptionally useful for processing decision-making.

3. How Can We Achieve Optimal Anomaly Detection Without Compromising the Accuracy of The Data?

The process of identifying anomalies remains an essential open research topic and research branch in the field of both anomalies’ detection and data mining. We can confirm that anomalies come in different variants like point anomalies, collective outliers, or contextual outliers. The borderline between anomaly and normal object is a grey area, making anomalies very hard to detect and to distinguish them from normal behaviors.

Implementing a context-sensitive systematic anomaly detection method requires careful parameters tuning to flag true anomalies. However, such a process can easily result in blocking good transactions. Identifying abnormal patterns in high dimensionality data, such as retail POS data, is a challenging task because when the number of features or attributes increases, data size grows proportionally. For business data such as retail databases this is a very common and frequently changing feature. This leads to data sparsity causing data objects (points) to be scattered and isolated, leading to difficulties in anomaly detection. Therefore, we conclude the following facts.

1. Anomaly detection model is computationally expensive in a large dimensional transaction database and requires large resources of human monitoring. Therefore, it is difficult to achieve optimal anomaly detection without compromising accuracy.
2. The proposed method shows how data with anomalies compare to normally distributed data. However, it is imperative to deliberate crucial questions using human judgment (human operator) to signify what anomaly is to achieve optimal anomaly detection.
3. Our proposed anomaly detection model showed performance deficiency in mining large dimensional POS databases. POS database is BIGDATA in nature and contains many systematic errors, NULL values, and human errors. The algorithm labels these transactions as potentially fraudulent transactions with different percentages of fraud depending on how far each transaction is from the mainstream data (average). The algorithm scans every transaction, this requires extensive CPU and MEMORY resources to filter out those transactions that deviate from the normal patterns.

7.2 Contribution

In this research, we proposed a novel approach to MBA using an enhanced Apriori algorithm named the Åbo algorithm. Åbo algorithm has shown superior performance in terms of computational and intelligence performance over the classic Apriori algorithm. Åbo algorithm formulated product popularity trends over time with the new conception of calculating frequent itemsets.

Here we summarize the main research contribution as follows.

1. The proposed MBA model simplifies the implementation of market intelligence and provides viable AI methods to forecast customers'

shopping behavior and identify strong association rules discovered in POS databases. The analytical information gained from data is exceptionally useful for strategic planning and decision-making with MI. With MI in place, the retailer has a systematic vision to develop a cross-selling strategy for products and services to increase revenue and generate ROI from marketing efforts. Moreover, the model enables an efficient way to optimize shelf space. As a result, customers looking for one product might also spot another product commonly purchased in the same shelf or aisle and would be tempted to buy them together.

2. To our knowledge, this doctoral dissertation is the first dissertation to propose novel artificial and computational intelligence methods into the classic Apriori algorithm. From the AI perspective, Åbo algorithm has shown capabilities to strengthen and reinforce the effect of the most recent itemsets purchases while justifying the importance of previous itemsets purchases. From the computational perspective, Åbo algorithm has shown a high degree of computational performance in extracting useful association rules.
3. We show that traditional ARM methods fail to predict customers' shopping behavior. Thus, instead of using conventional statistical methods, as most previous studies did, data-driven exploratory methods were used to extract product knowledge hidden in a real POS data warehouse.
4. We incorporate three different areas of study, market research, consumer shopping behavior, as well as computational and artificial intelligence, to develop market intelligence for the SMRs industry.
5. As Sim, Indrawan & Srinivasan (2008) state, infrequent rules (patterns) are often interesting in many real-life cases, but the Apriori algorithm often fails to extract them, as the extracted rules (patterns) have to satisfy the user-defined thresholds of support, confidence and lift. They do not find any itemsets outside of these thresholds but systematically filter them out. The Åbo algorithm sheds light on these interesting infrequent itemsets by detecting shopping patterns using the customer's shopping basket history (time, basket).

6. We proposed a new anomaly detection method using enhanced normal distribution formula that helps us handle anomalies found in POS databases.

7. Kim & Chung (1997) define product popularity as the extent to which the public widely purchases the product. Historically, the product's price and quality define product popularity, and only to a lesser degree, the consumer sentiment that fosters a feeling of personal connection (Aaker, 1996). However, the influence and interaction between the time of purchase and the brand's popularity have not been investigated before.

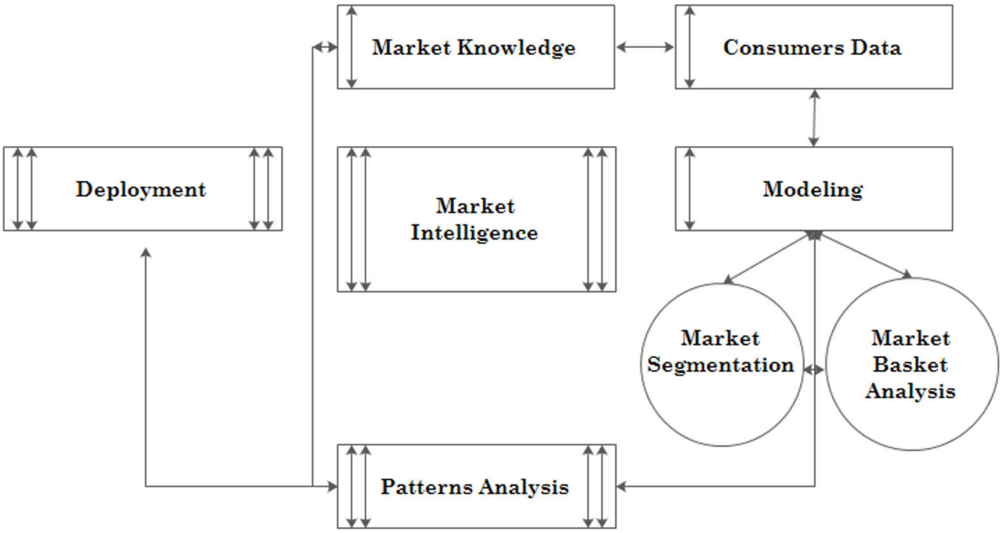
In the research, we investigated whether the time of purchase relates to the brand's popularity and to what extent it affects consumers' shopping behavior. The findings indicate that product popularity is directly affected by the time of purchase in the context of retail, which influences the perceived brand's quality.

We conclude that incorporating various aspects of research in market Intelligence assists SME retailers to leverage MI strategy and drive intelligent responsiveness across processes and operations.

7.3 The Proposed Market Intelligence Process for Data Mining

Based on the findings in this research, we propose a five-phase modeling cycle shown in figure 7.1 that enables SME retailers to apply an iterative MI process.

Figure 7.1 Market Intelligence Implementation Phases for the SME Retailers



Phase 1: Market Knowledge

This quantitative approach outlines the project objectives and requirements in clear technical terms. The goal is to understand the target market and the business goals, like the client's objective and the essential factors and constraints (limitations). This phase also involves understanding problems that the client wants to address that might influence the project's outcome. This phase comprises documentation reading and conducting meetings with the client to ask about relevant context.

Phase 2: Consumer data

Helps to streamline business operations by collecting customers' transactional to help SMEs boost customer service and increase products profitability.

Phase 3: Modeling

This phase defines appropriate modeling and mechanism techniques needed to test the proposed model's quality. During the modeling phase, two unsupervised learning methods are selected. Market segmentation divides the customer base into similar homogeneous groups sharing one or more shopping characteristics. MBA helps the retailer gain deeper insights and understand granular customer shopping behavior such as trends and biases.

Phase 4 Pattern Analysis

This phase assesses the degree to whether the proposed model meets the business objectives set forth. It is a cognitive process used by marketers to distinguish and classify the extracted knowledge/patterns into segmented categories. This phase will allow marketers to understand business dynamics, manage risks, anticipate market shifts, and establish relevance and rigor in the process of decision-making.

Phase 5 Deployment:

The final phase is to bring market intelligence into production. This phase involves the automation of patterns prediction from the MI model.

7.4 Future Research

There are several interesting initiatives arising from this study presented next. The POS database is often complex, containing thousands of irrelevant columns and many incomplete, duplicate, and empty data rows. Anomaly detection can improve the quality of such data with potential outliers, which again influences the quality of the extracted knowledge. According to Davidson & Louis (2012), POS database often ignores normalization, a set of design methods to organize/break down objects (tables and views) into their constituent parts until each object represents and adequately describes what the object represents. To tackle this, we recommend the following research initiatives:

1. Test new algorithms (including fuzzy methods) when training raw POS data to find a precise definition of the customer's profiles.
2. According to product lifetime value (PLV) model, segmented products are expressed in terms of RFM classification. This might help the retailer identify the characteristics of the top associated products and spot the most associated itemsets in terms of their total sales value. If such

knowledge allows retailers to focus marketing messages is worth research.

3. It is also worth research to find if clustering methods help to divide the product assortment into smaller groups based on common product characteristics preceding ARM. This might result in the extraction of more valuable knowledge and application of advanced interactive graphs tools to explore large itemsets of rules.

4. Further research into the Åbo algorithm on other databases needs to be conducted to verify its performance.

5. Further research is needed to justify implementing anomaly detection methods in high-dimensional databases.

6. Finally, this research focused on the SMEs industry. However, the MBA model developed in this research applies to other industries, like Health and Banking, to compare and further validate this research and increase awareness of it in the SME industry.

References

A.P. Engelbrecht, (2003). Computational Intelligence: An Introduction. Wiley.

Adhikari, A. (Ed.). (2018). Strategic Marketing Issues in Emerging Markets. Springer.

Adhikari, A. T. B., & Karunananda, A. S. (2015). Real time anomaly detection of cloud data centers with multi agents swarm. Sri Lanka Association for Artificial Intelligence, 5.

Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. Advances in knowledge discovery and data mining, 12(1), 307-328.

Aguinis, H., Forcum, L. E., & Joo, H. (2013). Using market basket analysis in management research. Journal of Management, 39(7), 1799-1824.

Akçura, M. T., & Srinivasan, K. (2005). Research note: customer intimacy and cross-selling strategy. Management Science, 51(6), 1007-1012.

Aleskerov, E., Freisleben, B., & Rao, B. (1997, March). Cardwatch: A neural network based database mining system for credit card fraud detection. In Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr) (pp. 220-226). IEEE.

Altzan, P., & Welsh, C. (2015). Business driven information systems (Fourth Canadian Edition ed.): McGraw- Kinha, Yogita, (2020). “An Easy Guide to Choose the Right Machine Learning Algorithm.” KDnuggets.

Anaza, N. A., Kemp, E., & Borders, L. A. (2018). Storytelling in Business-to-Business Advertising: An Abstract. In Academy of Marketing Science Annual Conference (pp. 603-603). Springer, Cham.

Angeli, F., & Grimaldi, R. (2010). Leveraging offshoring: the identification of new business opportunities in international settings. Industry and Innovation, 17(4), 393-413.

Aumann, Y., & Lindell, Y. (2003). A statistical theory for quantitative association rules. Journal of Intelligent Information Systems, 20: 255-283.

Bakos, J., Siu, M., Orengo, A., & Kasiri, N. (2020). An analysis of environmental sustainability in small & medium-sized enterprises: Patterns and trends. Business Strategy and the Environment, 29(3), 1285-1296.

Banasiewicz, A. D. (2013). Marketing database analytics: Transforming data for competitive advantage Routledge.

Bansal, A., Sharma, M., & Goel, S. (2017). Improved k-mean clustering algorithm for prediction analysis using classification technique in data mining. International Journal of Computer Applications, 157(6), 0975-8887.

Bartholomew, M. (2017). Adcreep: The Case Against Modern Marketing. Stanford University Press.

Bennett, R. J., & Robson, P. J. A. (2004). The role of boards of directors in small and medium-sized firms. Journal of Small Business and Enterprise Development.

Berry, M. J. A., & Linoff, G. S. (2004). Data mining techniques second edition - for marketing, sales, and customer relationship management. Wiley.

Bezdek, C. (1994). What is computational intelligence? In: Computational Intelligence Imitating Life, pp. 1–12, IEEE Press, New York.

Bhat, A. (2019). Empirical research: Definition, methods, types and examples.

Bhor, H. N., Koul, T., Malviya, R., & Mundra, K. (2018). Digital media marketing using trend analysis on social media. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 1398-1400). IEEE.

Blum, K. (2020, December). Gartner Unveils Top Marketing Predictions for 2021 and Beyond. Retrieved January 16, 2021, from https://www.gartner.com/en/newsroom/press-releases/2020-12-03-gartner-unveils-top-marketing-predictions-for-2021-an.

Blumberg, B., Cooper, D. R., & Schindler, P. S. (2008). Business research methods (Vol. 2). London: McGraw-Hill Higher Education.

Boire, R. (2017). Artificial intelligence (AI), automation, and its impact on data science. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 3571-3574). IEEE.

Borah, A., & Nath, B. (2019). Tree based frequent and rare pattern mining techniques: a comprehensive structural and empirical analysis. SN Applied Sciences, 1(9), 1-18.

Borgelt C, (2012). Frequent itemset mining (Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery,2(6)), pp.437–456.

Borgelt, C. (2003, November). Efficient implementations of apriori and eclat. In FIMI’03: Proceedings of the IEEE ICDM workshop on frequent itemset mining implementations (p. 90).

Borgelt, C., & Kruse, R. (2002). Induction of association rules: Apriori implementation. In Compstat (pp. 395-400). Physica, Heidelberg.

Bose, R (2008a). Competitive intelligence process and tools for intelligence analysis, Industrial Management & Data Systems, Vol. 108, No. 4, pp.510 – 528.

Bose, R (2008b). Customer relationship management: key components.

Brodie, M. L., Mylopoulos, J., & Schmidt, J. W. (Eds.). (2012). On conceptual modelling: Perspectives from artificial intelligence, databases, and programming languages. Springer Science & Business Media.

Brotspies, H., & Weinstein, A. (2019). Rethinking business segmentation: a conceptual model and strategic insights. Journal of Strategic Marketing, 27(2), 164-176.

Bryman A (2007) Barriers to integrating quantitative and qualitative research, Journal of Mixed Methods Research 1 8–22.

Burgess, A. (2018). AI in Action. In The Executive Guide to Artificial Intelligence (pp. 73-89). Palgrave Macmillan, Cham.

Buttle, F. (2004). Customer Relationship Management 1st ed. 30 Corporate Drive,Suite 400,Burlington,MA 01803,USA: Elsevier Ltd.

Buttle, F., & Maklan, S. (2019). Customer relationship management: concepts and technologies. Routledge.

Camilleri, M. A. (2018). Market segmentation, targeting and positioning. In Travel marketing, tourism economics and the airline product. Springer, Cham.

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. Journal of Applied Psychology, 93(5), 1062.

Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., & Oliveira, A. L. (2016). Computational intelligence and financial markets: A survey and future directions. Expert Systems with Applications, 55, 194-211.

Ceglar, A., & Roddick, J. F. (2006). Association mining. ACM Computing Surveys (CSUR), 38(2), 5-es.

Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. Psychology & Marketing.

Chaudhary, R., Sharma, S., & Sharma, V. K. (2015). Improving the performance of MS-Apriori algorithm using dynamic matrix technique and map-reduce framework. Int. J. Innov. Res. Sci. Technol, 2(5), 2349-6010.

Chaudhuri S., U. Dayal, V. Narasayya, (2011). An overview of business intel-ligence technology Communications of the ACM, 54 (8), pp. 88-98.

Chen, HC (2010). Business and market intelligence 2.0, IEEE Intelligent Systems, Vol. 25, No. 1, pp.68-71.

Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). Market basket analysis in a multiple store environment. Decision support systems, 40(2), 339-354.

Chen, Z. (1999). Computational intelligence for decision support. CRC Press.

Chiu, S., & Tavella, D. (2008). Data mining and market intelligence for optimal marketing returns. Routledge.

Chiu, W., Kim, T., & Won, D. (2018). Predicting consumers' intention to purchase sporting goods online: an application of the model of goal-directed behavior. Asia Pacific Journal of Marketing and Logistics, 30(2), 333-351.

Chowdhury, M., Sadek, A., Ma, Y., Kanhere, N., & Bhavsar, P. (2006). Applications of artificial intelligence paradigms to decision support in real-time traffic management. Transportation research record, 1968(1), 92-98.

Chung, J. E., Huang, Y., Jin, B., & Sternquist, B. (2011). The impact of market orientation on Chinese retailers' channel relationships. Journal of Business & Industrial Marketing.

Claudé, M., & Combe, D. (2018). THE ROLES OF ARTIFICIAL INTELLIGENCE AND HUMANS IN DECISION MAKING: TOWARDS AUGMENTED HUMANS?: A focus on knowledge-intensive firms.

Coiera, E. (2019). The price of artificial intelligence. Yearbook of medical informatics, 28(01), 014-015.

Conroy, P., Porter, K., Nanda, R., Renner, B., & Narula, A. (2015). Consumer Product Trends–Navigating 2020.

Constantinos, J. S., Christos, S., & Stafyla, A. (2003). CRM and customer-centric knowledge management: An empirical research. Business Process Management Journal, 9 (5), 617–634. doi:10.1108/14637150310496721.

Copeland, J. (2015). Artificial intelligence: A philosophical introduction. John Wiley & Sons.

Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. (2011). Best practices for mixed methods research in the health sciences. Bethesda (Maryland): National Institutes of Health, 2013, 541-545.

Daabes, A. S. A., & Kharbat, F. F. (2017). Customer-based perceptual map as a marketing intelligence source. International Journal of Economics and Business Research, 13(4), 360–379.

Dahlén, M., & Edenius, M. (2007). When is advertising? Comparing responses to non-traditional and traditional advertising media. Journal of Current Issues & Research in Advertising, 29(1), 33-42.

Davenport T, Harris J (2007). Competing on Analytics: The New Science of Winning(Harvard Business School Press).

Davidson, Louis (2012). "Ten Common Database Design Mistakes." Simple Talk, Simple Talk, 23 Jan. 2019, www.red-gate.com/simple-talk/sql/database-administration/ten-common-database-design-mistakes.

Deng, Z. & Luo, L. (2010). An exploratory discussion of new ways for competitive intelligence on Web2.0. International Federation for Information Processing, 252(1); 597–6.

Dipanjan, D., Satish, G., & Goutam, C. (2011). Comparison of Probabilistic-D and k-Means Clustering in Segment Profiles for B2B Markets. SAS Global Forum.

Dolnicar, S., Grün, B., & Leisch, F. (2018). Market segmentation analysis: Understanding it, doing it, and making it useful (p. 324). Springer Nature.

Doshi, A. J., & Joshi, B. (2018). Comparative analysis of Apriori and Apriori with hashing algorithm.

Drew, J. H., Mani, D. R., Betz, A. L., & Datta, P. (2001). Targeting customers with statistical and data-mining techniques. Journal of Service Research, 3(3), 205–219.

Duch, W. (2007). Towards comprehen-sive foundations of computational in-telligence. In: Duch W,Mandziuk J, Eds, Challenges for Computational Intelli-gence. Springer.

Eberhart, R. C., & Shi, Y. (2011). Computational intelligence: concepts to implementations. Elsevier.

Eggert, A., Ulaga, W., & Schultz, F. (2006). Value creation in the relationship life cycle: A quasi-longitudinal analysis. Industrial Marketing Management, 35(1), 20-27.

Ezhilvathani, A., & Raja, K. (2013). Implementation of parallel apriori algorithm on Hadoop cluster. International Journal of Computer Science and Mobile Computing, 2(4), 513-516.

Fahy, J., & Jobber, D. (2015). Foundations of marketing.

Faizollahzadeh Ardabili, S., Najafi, B., Shamshirband, S., Minaei Bidgoli, B., Deo, R. C., & Chau, K. W. (2018). Computational intelligence approach for modeling hydrogen production: A review. Engineering Applications of Computational Fluid Mechanics, 12(1), 438-458.

Fan, G., Wang, X.L. and Zhu, H.P. (2010). NERI INDEX of Marketization of China's Provinces 2009. Economic Science Press, Beijing.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-37.

Fortune business insights, (2022, April). Https://Www.Fortunebusinessinsights.Com/Customer-Relationship-Management-Crm-Market-103418. Retrieved May 29, 2022.

Fournier, S., & Avery, J. (2011). The uninvited brand. Business horizons, 54(3), 193-207.

Franco, M., Magrinho, A., & Silva, J. R. (2011). Competitive intelligence: a research model tested on Portuguese firms. Business Process Management Journal.

Gal, A. (2010). Competitiveness of small and medium-sized enterprises-a possible analytical framework. HEJ: ECO-100115-A.

Garner, M., Wagner, C., & Kawulich, B. (2016). Quantitative or qualitative: Ontological and epistemological choices in research methods curricula. In Teaching research methods in the social sciences (pp. 81-90). Routledge.

Gil-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R., & Lozano-Quilis, J. A. (2020). Customer relationship management: digital transformation and sustainable business model innovation. Economic Research-Ekonomska Istraživanja, 33(1), 2733-2750.

Goldberg, L. and Goldberg, L., (2018). A Brief History of Artificial Intelligence In Advertising – Econsultancy. [online] Econsultancy. Available at: <https://econsultancy.com/a-brief-history-of-artificial-intelligence-in-advertising/> [Accessed 8 April 2020].

Gordon, J., Perrey, J., & Spillecke, D. (2013). Big data, analytics and the future of marketing and sales. McKinsey: Digital Advantage.

Gouda, K., & Zaki, M. J. (2001, November). Efficiently mining maximal frequent itemsets. In Proceedings 2001 IEEE International Conference on Data Mining (pp. 163-170). IEEE.

Goyat, S. (2011). The basis of market segmentation: a critical review of the literature. European Journal of Business and Management, 3(9), 45-54.

grandviewresearch.com, (2021). Retrieved (2021). Customer relationship management MARKET REPORT. Retrieved February 13, 2021, from https://www.grandviewresearch.com/industry-analysis/customer-relationship-management-crm-market.

Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. Expert Systems with Applications, 100, 1-16.

Gruber, M., MacMillan, I. C., & Thompson, J. D. (2012). From minds to markets: How human capital endowments shape market opportunity identification of technology start-ups. Journal of management, 38(5), 1421-1449.

Guida, T. (2019). Big Data and Machine Learning in Quantitative Investment. John Wiley & Sons.

Gundlach, G. T., & Wilkie, W. L. (2009). The American Marketing Association's new definition of marketing: Perspective and commentary on the 2007 revision. Journal of Public Policy & Marketing, 28(2), 259-264.

Hall, J. (2020). How Artificial Intelligence Is Transforming Digital Marketing. [online] Forbes. Available at: https://www.forbes.com/sites/forbesagencycouncil/2019/08/21/how-artificial-intelligence-is-transforming-digital-marketing/#3ddc7d0121e1 [Accessed 8 Mar. 2020].

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques,|| Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We're all connected: The power of the social media ecosystem. Business horizons, 54(3), 265-273.

Hawkins D (1980). Identification of outliers. Chapman and Hall, London.

Hedin, H., Hirvensalo, I., & Vaarnas, M. (2011). The handbook of market intelligence: understand, compete and grow in global markets. John Wiley & Sons.

Heggde, G., & Shainesh, G. (Eds.). (2018). Social media marketing: Emerging concepts and applications. Springer.

Hegland, M., "Algorithms for Association Rules," (2003). Lecture Notes in Computer Science, Volume 2600, Pages 226.

Heydari, M., & Yousefli, A. (2017). A new optimization model for market basket analysis with allocation considerations: A genetic algorithm solution approach. Management & Marketing, 12(1), 1-11.

Humby, C. (2006). Data is the new oil. Proc.ANA Sr.Marketer's Summit.Evanston, IL, USA,

IBM Center for Applied Insights. (2013). Marketing science: From descriptive to prescriptive. (n.d.). Retrieved January 21, 2021, from https://s3-us-west2.amazonaws.com/itworldcanada/archive/Documents/whitepaper/ITW266E_Marketing_Science.PDF.

Ilyas, I. F., & Chu, X. (2019). Data cleaning. Morgan & Claypool.

infosys.com, (2017). AI: THE PROMISE OF A GREAT FUTURE FOR RETAILERS. [online] Available at: <https://www.infosys.com/human-amplification/documents/retail-ai-perspective.pdf.

James, M., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). The next frontier for innovation, competition, and productivity. Big data.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business horizons, 61(4), 577-586.

Jenster, P. V., & Søilen, K. S. (2009). Market intelligence: Building strategic insight Copenhagen Business School Press DK.

Jobber, D., & Ellis-Chadwick, F. (2012). Principles and practice of marketing (No. 7th). McGraw-Hill Higher Education.

Johansson, M., Larsson, C., & Hallin, I. (2006). Marketing communication in a B2B market: a case study of the store interior industry.

Johns, P. and Van Doren, D.C. (2010), Competitive intelligence in service marketing: A new approach with practical application, Marketing Intelligence and Planning, Vol. 28 No. 5, pp. 551-570.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational researcher, 33(7), 14-26.

Jones, M. A., & Jones, N. A. (2018). Systems and methods of controlling the distribution of products in retail shopping facilities. U.S. Patent Application No. 15/803,343.15/803,343.

Jones, R. P., & Runyan, R. C. (2013). Brand experience and brand implications in a multi-channel setting. The International Review of Retail, Distribution and Consumer Research, 23(3), 265-290.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68.

Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia computer science, 85, 78-85.

Kavitha, M., & Subbaiah, S. (2020). Association rule mining using apriori algorithm for extracting product sales patterns in groceries.

Keohane, L. M., Gambrel, R. J., Freed, S. S., Stevenson, D., & Buntin, M. B. (2018). Understanding trends in Medicare spending, 2007–2014. Health services research, 53(5), 3507-3527.

Khan, A., Lee, S., Lee, Y.-K., & Khattak, A. (2010). Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka. International Journal of Database Theory and Application.

Kim, B. D., & Park, K. (1997). Studying patterns of consumer's grocery shopping trip. Journal of retailing, 73(4), 501-517.

Kim, C. S., & Aggarwal, P. (2016). The customer is king: culture-based unintended consequences of modern marketing. Journal of Consumer Marketing.

Kimball, R., & Ross, M. (2011). The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons.

King, S. F., & Burgess, T. F. (2008). Understanding success and failure in customer relationship management. Industrial Marketing Management, 37(4), 421-431.

Klaus, S. (2018). The Global Competitiveness Report 2018. In World Economic Forum (Vol. 12, p. 2019).

Kotler, P. T. (2019). Marketing management. Pearson UK.

Kotler, P., & Levy, S. J. (1969). Broadening the concept of marketing. Journal of Marketing, 33(1), 10-15.

Kotler, P., Armstrong, G., Harris, L., & Piercy, N. (2013). Principle of Marketing 6th ed., 38-47. Edinburgh Gate, Harlow, England: Pearson Education Limited.

Kotu, V., & Deshpande, B. (2014). Predictive analytics and data mining: concepts and practice with rapidminer. Morgan Kaufmann.

Krizanova, A., Gajanova, L., & Nadanyiova, M. (2018). Design of a CRM level and performance measurement model. Sustainability, 10(7), 2567.

Kumar, A. V. (Ed.). (2010). Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains: Trends and New Domains. IGI Global.

Kumar, V., & Reinartz, W. (2019). Customer Relationship Management: Concept, Strategy, and Tools.

Lambin, J. J. (2013). Changing market relationships in the internet age. Presses univ. de Louvain.

Layman, (2015). Association rules and the apriori algorithm. Layman tutorials in analytics.

Lee, C., Kumar, V., & Gupta, S. (2013). Designing freemium: a model of consumer usage, upgrade, and referral dynamics.

Lekberg, D., & Danielsson, P. (2013). Designing and Implementing Generic database Systems Based on the entity-attribute-value Model.

Leslie, N. (2011). Culture's influence on brand loyalty among culturally diverse consumers in the United States.

Liam Damien. (2019). DATA MINING: Your Ultimate Guide to a Comprehensive Understanding of Data Mining.

Liu, B., Cao, S. G., & He, W. (2011). Distributed data mining for e-business. Information Technology and Management, 12(2), 67-79.

Liu, J., Liao, X., Huang, W., & Liao, X. (2019). Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision. Omega, 83, 1-13.

Longinus, O. (2018). Artificial Intelligence System: Implication for Proper Record Keeping in Microfinance Banks in Nigeria. International Journal of Academic Research in Accounting, Finance and Management Sciences, 8(1), 131-136.

Lueg, R., & Lu, S. (2013). How to improve efficiency in budgeting-The case of business intelligence in SMEs. European Journal of Management, 13(2), 109-120.

Mageplaza. (2020). AI Marketing: What, Why and How to use Artificial Intelligence Marketing. [online] Available at: https://www.mageplaza.com/blog/ai-marketing-what-why-how.html [Accessed 8 Mar. 2020].

Mahoney, m. (2017). Chapter ten the cost of artifical intelligence matt mahoney. Philosophy of Mind: Contemporary Perspectives, 232.

Maimon, O., & Rokach, L. (2009). Introduction to knowledge discovery and data mining. In Data mining and knowledge discovery handbook (pp. 1-15). Springer, Boston, MA.

Maimon, O., & Rokach, L. (Eds.). (2007). Soft computing for knowledge discovery and data mining. Springer Science & Business Media.

Malgaonkar, S., Surve, S., & Hirave, T. (2013, January). Use of mining techniques to improve the effectiveness of marketing and sales. In 2013 International Conference on Advances in Technology and Engineering (ICATE) (pp. 1-5). IEEE.

Marcoulides, G. A., & Saunders, C. (2006). Editor's comments: PLS: a silver bullet?. MIS quarterly, iii-ix.

Marketing-schools.org, (2012). Traditional marketing. Available at: http://www.marketing.com

Marshall, K. P., & Desborde, R. (2019). A Proposal for Fine Arts Market Segmentation based on Functions of Art and Consumer Benefits Sought.

McCarthy, J. (2007). What Is Artificial Intelligence? Stanford Univ. Stanford, CA 94305.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI Magazine, 27(4), 12.

McGlothlin, J. P., & Khan, L. (2013). Managing evolving code sets and integration of multiple data sources in health care analytics. In Proceedings of the 2013 international workshop on Data management & analytics for healthcare (pp. 9-14).

Morden, T. (2017). Principles of management. Routledge.

Moreno, M. N., Segrera, S., & López, V. F. (2005). Association Rules: Problems, solutions and new applications. Actas del III Taller Nacional de Minería de Datos y Aprendizaje, Tamida, 317-323.

Mowen, J. C. (2000). The 3M model of motivation and personality: Theory and empirical applications to consumer behavior. Springer Science & Business Media.

Müller, V. C. (Ed.). (2012). Philosophy and theory of artificial intelligence (Vol. 5). Springer Science & Business Media.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. Journal of Marketing Research, 56(6), 960–980.

Nguyen, B., Simkin, L., & Canhoto, A. I. (Eds.). (2015). The Dark Side of CRM: Customers, Relationships and Management. Routledge.

Nwankwo, S., & Gbadamosi, A. (Eds.). (2010). Entrepreneurship marketing: principles and practice of SME marketing. Routledge.

Oliver, R. L., Rust, R. T., & Varki, S. (1997). Customer delight: foundations, findings, and managerial insight. Journal of Retailing, 73(3), 311-336.

Olson, C., & Levy, J. (2018). Transforming marketing with artificial intelligence. Applied Marketing Analytics, 3(4), 291-297.

Olson, D. L., & Lauhoff, G. (2019). Descriptive data mining. In Descriptive Data Mining (pp. 129-130). Springer, Singapor

Omar, O. E. & P. Fraser (2010). The role of small and medium enterprise retailing in Britain.

Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors influencing the effectiveness of relationship marketing: a meta-analysis. Journal of marketing, 70(4), 136-153.

Paolanti, M., Liciotti, D., Pietrini, R., Mancini, A., & Frontoni, E. (2018). Modelling and forecasting customer navigation in intelligent retail environments. Journal of Intelligent & Robotic Systems, 91(2), 165-180.

Papadopoulos, N., Martín, O. M., Cleveland, M., & Laroche, M. (2011). Identity, demographics, and consumer behaviors. International Marketing Review.

Parsian, M. (2015). Data Algorithms: Recipes for Scaling Up with Hadoop and Spark. " O'Reilly Media, Inc.

Parvatiyar, A., & Sheth, J. N. (2001). Customer relationship management: Emerging practice, process, and discipline. Journal of Economic & Social Research, 3(2).

Pearson, R. K. (2018). Exploratory data analysis using R. Chapman and Hall/CRC.

Pedrycz, W., & Chen, S. M. (Eds.). (2014). Information granularity, big data, and computational intelligence (Vol. 8). Springer.

Peltoniemi, M. & Vuori, E. (2008). Competitive intelligence as a driver of co-evolution within an enterprise population. Journal of Competitive Intelligence and Management, 4(3), 50 – 6.

Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: a systematic literature review. PeerJ Computer Science, 6, e267.

Poole D., (1998). A. Mackworth and R. Goebel. Computational Intelligence – A Logical Approach.Oxford University Press, New York.

Power, D. J. (2002). Decision support systems: concepts and resources for managers, Greenwood Publishing Group.

Power, D. J., Heavin, C., McDermott, J., & Daly, M. (2018). Defining business analytics: An empirical approach. Journal of Business Analytics, 1(1), 40-53.

Pranjic, G. (2011). Influence of business and competitive intelligence on making right business decisions, Business Review, Vol. 1 No. 1, pp. 271-288.

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.

Raj, R., Wong, S. H., & Beaumont, A. J. (2016). Business intelligence solution for an SME: A case study.

Raja, B., Pamina, J., Madhavan, P., & Kumar, A. S. (2019). Market Behavior Analysis Using Descriptive Approach. Available at SSRN 3330017.

Raska, D., Nichols, B. S., & Shaw, D. (2015). When descriptive norm cues fail as persuasion agents in green supermarket advertising. Journal of Promotion Management, 21(6), 721-738.

Romano, C., & Ratnatunga, J. (1995). The role of marketing: its impact on small enterprise research. European journal of marketing.

Rostam, N. A (2020). The impact of big data market segmentation using data mining and clustering techniques. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-15.

Russell, g. J., & Petersen, A. (2000). Analysis of cross category dependence in market basket selection. Journal of Retailing, 76: 367-392.

Said, I. U., Muhammad, J. M., & Gupta, M. K. (2015). Intelligent Heart Disease Prediction System by Applying the Apriori Algorithm. International Journal, 5(9).

Scanzoni J. (1979). "Social Exchange and Behavioral Interdependence." New York: Academic Press, Inc.

Schafer, J. L. (1997). Analysis of incomplete multivariate data. CRC press.

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harv. JL & Tech., 29, 353.

Schniederjans, M. J., Cao, Q., & Ching Gu, V. (2012). An operations management perspective on adopting customer-relations management (CRM) software. International Journal of Production Research, 50(14), 3974-3987.

Sethi, R., & Shekar, B. (2018). Issues in negative association rule mining with business analytics perspectives. DHARANA-Bhavan's International Journal of Business, 11(2), 13-20.

Sharma, N., & Verma, C. K. (2014). An improved apriori-based algorithm for association rule mining.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. J Data Warehousing, 5, 13-22.

Shewhart, W. A., & Deming, W. E. (1986). Statistical method from the viewpoint of quality control. Courier Corporation.

Silverman, D. (2004). Research and social theory. Researching society and culture, 47-58.

Silvers, F. (2011). Data Warehouse Designs: Achieving ROI with Market Basket Analysis and Time Variance. CRC Press.

Sim, A. T. H., Indrawan, M., & Srinivasan, B. (2008, February). Mining infrequent and interesting rules from transaction records. In 7th WSEAS Int. Conf. on AI, Knowledge Engineering and Databases (AIKED 2008) (pp. 515-520).

Simester, D., Timoshenko, A., & Zoumpoulis, S. I. (2020). Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. Management Science, 66(6), 2495–2522.

Simon, H. A. (1996). The sciences of the artificial MIT press.

Singelis, T. M., Triandis, H. C., Bhawuk, D. P., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. Cross-cultural research, 29(3), 240-275.

Singh, D. P., Choudhury, J. P., & De, M. (2012). A comparative study to select a soft computing model for knowledge discovery in data mining. International Journal of Artificial Intelligence and Knowledge Discovery, 2(2), 6-19.

Singh, S., Garg, R., & Mishra, P. K. (2015). Performance analysis of apriori algorithm with different data structures on hadoop cluster. arXiv preprint arXiv:1511.07017.

Sivaramakrishnan, S., Delbaere, M., Zhang, D., & Bruning, E. (2012). Critical success factors and outcomes of market knowledge management: a conceptual model and empirical evidence. In Conceptual Models and Outcomes of Advancing Knowledge Management: New Technologies (pp. 165-185). IGI Global.

Srivastava, N., Gupta, K., & Baliyan, N. (2018). Improved Market Basket Analysis with Utility Mining.

Sterne, J. (2017). Artificial intelligence for marketing: practical applications. John Wiley & Sons.

Stirrup, J., & Ramos, R. O. (2017). Advanced Analytics with R and Tableau. Packt Publishing Ltd.

Stone, M. D., & Woodcock, N. D. (2014). Interactive, direct and digital marketing: A future that depends on better use of business intelligence. Journal of research in interactive marketing.

Stone, M., Aravopoulou, E., Ekinci, Y., Evans, G., Hobbs, M., Labib, A., ... & Machtynger, L. (2020). Artificial intelligence (AI) in strategic marketing decision-making: a research agenda. The Bottom Line, 33(2), 183-200.

Stormi, K., Laine, T., Suomala, P., & Elomaa, T. (2017). Forecasting sales in industrial services: Modeling business potential with installed base information. Journal of Service Management.

Strauss, A., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. NewPark, CA: Sage Publications, Inc.

Stubseid, S., & Arandjelovic, O. (2018). Machine learning based prediction of consumer purchasing decisions: the evidence and its significance. In Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence.

Szymanski, J., Sarnatowicz, T., & Duch, W. (2007). Towards avatars with artificial minds: Role of semantic memory. Journal of Ubiquitous Computing and Intelligence.

Szymkowiak, M., Klimanek, T., & Józefowski, T. (2018). Applying Market Basket Analysis to Official Statistical Data. Econometrics, 22(1), 39-57.

Tengberg, J. C. F. (2013). Inter-organizational information sharing of customer data in retail (Doctoral dissertation, Massachusetts Institute of Technology).

Teorey, T. J., Buxton, S., Fryman, L., Güting, R. H., Halpin, T., Harrington, J. L., ... & Witt, G. (2008). Database design: know it all. Morgan Kaufmann.

Tufféry, S. (2011). Data mining and statistics for decision making, John Wiley & Sons.

Turner, R., & Ledwith, A. (2018). Project Management in Small to Medium-Sized Enterprises: Fitting the

Practices to the Needs of the Firm to Deliver Benefit. Journal of Small Business Management, 56(3), 475-493.

Tyugu, E., & Tyugu, Ė. K. (2007). Algorithms and architectures of artificial intelligence (Vol. 159). IOS Press.

Unemyr, M., & Wass, M. (2018). Data-Driven Marketing with Artificial Intelligence: Harness the Power of Predictive Marketing and Machine Learning. Independently published.

Von Reischach, F., Guinard, D., Michahelles, F., & Fleisch, E. (2009, March). A mobile product recommendation system interacting with tagged products. In 2009 IEEE international conference on pervasive computing and communications (pp. 1-6). IEEE.

Walsh, M. F., & Lipinski, J. (2009). The role of the marketing function in small and medium sized enterprises. Journal of small business and enterprise development.

Wang, L., & Sun, J. (2019). Market Basket Analysis based on Apriori and CART.

Wang, X., Smith, K. A., Hyndman, R., & Alahakoon, D. (2004). A scalable method for time series clustering. Proc. Unrefereed Res. Papers, 1.

Wedel, M., & Kamakura, W. A. (2012). Market segmentation: Conceptual and methodological foundations (Vol. 8). Springer Science & Business Media.

Wen, Y. T., Yeh, P. W., Tsai, T. H., Peng, W. C., & Shuai, H. H. (2018). Customer purchase behavior prediction from payment datasets. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 628-636). ACM.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Xuan, L., Xiaoyan, W., & Zhuang, Z. (2002). Personalized intelligent recommendation service based on Association Rule Mining [J]. Computer engineering and application, 38(11), 200-204.

Yahya, O., Hegazy, O., & Ezat, E. (2012). An efficient implementation of A-Priori algorithm based on Hadoop-Mapreduce model. International sjournal of Reviews in Computing, 12.

Yongjie Yang, Pengfei Yang,Qun Duan., & Lina Han, (2016). Intelligent recommendation algorithm of goods based on association rules [J]. Modern computer, 10 (04), pp. 25-27.

Yoseph, F., Ahamed Hassain Malim, N. H., Heikkilä, M., Brezulianu, A., Geman, O., & Paskhal Rostam, N. A. (2020). The impact of big data market segmentation using data mining and clustering techniques. Journal of Intelligent & Fuzzy Systems, 38(5), 6159-6173.

Yuan, X. (2017). An improved Apriori algorithm for mining association rules. In AIP conference proceedings (Vol. 1820, No. 1, p. 080005). AIP Publishing LLC.

Zaki M.J. (1999). Parallel and distributed association mining: A survey (IEEE Concurrency, Special Issue on Parallel Mechanisms for Data Mining, 7(4)), pp.14-25.

Zeng, M., Xiong, Q., & Li, K. (2018, May). Design and implementation of an improved apriori data mining algorithm. In 8th International Conference on Social Network, Communication and Education (SNCE 2018) (pp. 291-296). Atlantis Press.

Zeithaml, V. A. (2000). Service quality, profitability, and the economic worth of customers: what we know and what we need to learn. Journal of the academy of marketing science, 28(1), 67-85.

Zhan, Y., Tan, K. H., & Huo, B. (2019). Bridging customer knowledge to innovative product development: a data mining approach. International Journal of Production Research, 57(20), 6335-6350.

Zhang, C., & Zhang, S. (Eds.). (2002). Association rule mining: models and algorithms. Berlin, Heidelberg: Springer Berlin Heidelberg.

Zhao, B. G., & Liu, Y. (2015). An efficient Bittable Based frequent itemsets mining algorithm. Journal of Shandong University, (5), 23-29.

Ziafat, H., & Shakeri, M. (2014). Using data mining techniques in customer segmentation. Journal of Engineering Research and Applications, 4(9), 70-79.

Cagliero, L., & Garza, P. (2013). Infrequent weighted itemset mining using frequent pattern growth. IEEE transactions on knowledge and data engineering, 26(4), 903-915.

Hebert, D., Anderson, B., Olinsky, A., & Hardin, J. M. (2014). Time series data mining: a retail application. International journal of business analytics (IJBAN), 1(4), 51-68.

Mekala, P., & Srinivasan, B. (2014). Time series data prediction on shopping mall. Int. J. Res. Comput. Appl. Robot, 2(8), 92-97.

Hasan, M. R., Kabir, M. A., Shuvro, R. A., & Das, P. (2022). A Comparative Study on Forecasting of Retail Sales. arXiv preprint arXiv:2203.06848.

ISBN 978-952-12-4258-8