Machine Learning in Inflation Prediction for the Finnish Economy

Sukrit Pant

Master's thesis in Governance of Digitalization

Master's Programme

Supervisor: Prof. Jozsef Mezei

Faculty of Social Sciences, Business and Economics, and Law

Åbo Akadmei University

Åbo 2023

Subject: Information Systems
Author: Sukrit Pant (sukrit.pant@abo.fi)
Title: Machine Learning in Inflation Prediction for the Finnish Economy
Language: English
Supervisor: Jozsef Mezei

Abstract:

With growing inflation faced by countries around the world impacting the lives of citizens for each economy, there is a need for a model that can accurately predict inflation for better decision-making. Machine learning techniques have evolved in the past few years, with newer algorithms being invented and improving on previous models.

Most of the research focused on larger economies such as the USA, the UK, Germany, and China. There is a demand for studies that are focused on other economies. Similar studies have not been conducted in Finland. At least those that are available for researchers and economists. The research focuses on improved accuracy of machine learning algorithms compared to traditional econometrics models. A multivariate analysis using a machine learning algorithm is performed to determine the best-performing model for the Finnish economy.

Similarly, the impact of adding further variables in inflation forecasting models is analyzed to understand the change in inflation accuracy. The study suggests that multi-layer perceptron performs the best for both sets of analysis, while cross-validation results in support vector regression for smaller datasets and LASSO for larger datasets. Also, the study reveals that adding more variables impacts the accuracy negatively for most of the algorithms.

Keywords: Machine learning, inflation forecast, inflation prediction

Number of Pages: 74

Date: 10.05.2023

TABLE OF CONTENTS

List of Tables

List of Figures

List of Equations

List of Abbreviation

ANN	Artificial Neural Network
AR	Autoregressive
ARDL	Auto Regressive Distributed Lag
ARMA	Autoregressive Moving Average
CART	Classification and Regression Tree
GP	Gausian Processes
COMPASS	Central Organising Model for Prudential Assessment of Systematic Risk
CPI	Consumer Price Index
GNP	Gross National Product
CSV	Comma Separated Value
E.C.	European Commission
E.U.	European Union
ECB	European Central Bank
OECD	Organisation for Economic Co-operation and Development
EDO	Estimated Dynamic Optimization
GBDT	Gradient Boosting Decision Tree
RMSE	Root Mean-squared Error
MSE	Mean-squared Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
SMAPE	Symmetric Mean Absolute Percentage Error
BVAR	Bayesian Vector Autoregressive Model
BP	Back Propagation
GDP	Gross Domestic Product
GRNN	General Regression Neural Network

GRU-RNN	Gated Recurrent Unit – Recurrent Neural Network
ILO	International Labour Organization
IMF	International Monetary Fund
KNN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
UK	United Kingdom
LASSO	Least Absolute Shrinkage and Selection Operator
MAD	Mean Absolute Deviation
MATLAB	Matrix Laboratory
ML	Machine Learning
MLP	Multi-Layer Perceptron
PCE	Personal Consumption Expenditure
RF	Random Forest
Sklearn	Scikit Learn
STAR	Smooth Threshold Autoregressive
U.S.	United States
USD	United States Dollar
JPY	Japanese Yen
CNY	Chinese Yuan
GBP	British Pound
VAR	Vector Autoregressive Model
SVR	Support Vector Regression
SVM	Support Vector Machine
RBF	Radial Basis Function
XGBoost	Extreme Gradient Boosting

1 INTRODUCTION

1.1 Problem Definition

Inflation can have a significant impact on economies. It can cause prices to rise, leading to higher living costs and lower purchasing power (Oner, 2022). Inflation can also lead to higher interest rates, impacting investment and economic growth. As a result, policymakers must have accurate inflation forecasts to make sound economic decisions.

Forecasting inflation is a crucial part of macroeconomic policymaking, and using machine learning techniques can improve the accuracy of inflation forecasts. First, machine learning is used to identify patterns in historical data that can be used to predict future inflation rates. For example, machine learning can identify relationships between inflation and economic activity or between inflation and changes in the money supply. By understanding these relationships, policymakers can make better-informed decisions about adjusting monetary policy to control inflation. Additionally, machine learning can be used to develop early-warning systems that can alert policymakers to potential inflationary pressures before they develop into full-blown crises. These systems can use data on various economic indicators, such as changes in the prices of essential commodities or the growth rates of critical sectors of the economy. By monitoring these indicators, policymakers can take steps to avert inflationary pressures before they become severe.

With the rise in complexities of the data that comprise different features of the economy, the importance of machine learning is growing. Machine learning will help policymakers understand how various indicators impact inflation. In addition, machine learning can illustrate what variables that do not impact inflation in traditional models could have obscure consequences from the macro-perspective.

Additionally, with the increase in available data, it is likely that making these machine learning models and forecasting can help improve the forecasting and prediction systems. For example, the Bank of Finland (*Suomen Pankki*), uses a Nowcasting system to understand the GDP (Gross Domestic Product) growth in the following quarter (Fornaro & Luomaranta, 2019). Similarly, the Federal Reserve in the United States uses a mixture of various models to forecast the GDP, one being the Estimated Dynamic Optimization (EDO) Model, which has been used since 2006 (Federal Reserve Bank, 2023). Similarly, the Bank of England also uses a Forecasting Platform consisting of different tools named

COMPASS with computing software such as MATLAB to create forecasting models for the following quarter (Burgess et al., 2013). The given examples show that central banks have used different models for the past few years. While research is scarce on the efficacy of economic forecasting systems, the development of the aforementioned forecasting models illustrates the concern of the world's largest economies with predicting the economic future. Advancements in machine learning and neural networks might indicate the need to modernize and update these legacy models, as this contemporary technology offers new opportunities to improve accuracy and efficiency in economic forecasting.

Furthermore, GDP growth prediction has been the focus of researchers and economists in many countries. However, inflation can be a significant hurdle in GDP growth, leading to a need for a better inflation prediction system. Therefore, further research needs to be conducted with rigorous testing of these modern and allegedly superior models for inflation forecasting before incorporating these models in decision-making. While the data available to these central banks are much larger and can probably create a better model for inflation forecasting, working with open data about the economy can help researchers understand how the models work and could be helpful to the field of economics and machine learning as well as the central bank.

1.2 Objective of the Study

The purpose of this study is to understand the various tools used in inflation forecasting and use these to replicate a study based on the Finnish economy. Many studies are focused on countries such as the United States (Aras & Lisboa, 2022; Tattikota & Srinivasan, 2021; Ülke et al., 2016) and the United Kingdom (Chakraborty & Joseph, 2017; Ivan, 2018; Joseph, 2019; Liu et al., 2022). However, in Finland, similar studies are missing that aides in understanding the forecasting accuracy of machine learning models and the importance of variables in machine learning models. Furthermore, while there are models such as Nowcasting that the Bank of Finland implements in its study, they focus on GDP growth rather than inflation. Although, constantly in the past, GDP growth and inflation have had a positive relationship, with terms like stagflation frequently discussed in the media, it is believed that GDP growth has slowed down compared to inflation, changing the dynamics between these macroeconomic indicators.

However, the variables identified in forecasting GDP growth can initiate an understanding of inflation. Therefore, along with previous studies conducted in other countries, this study will try to investigate what these variables might be in the context of Finland. Understanding the variables would help initiate the study and comprehend the impact on the accuracy of forecasting models when additional variables are included. Supposing there is an improvement in the accuracy of the results, these additional variables can be included in the models. In case there are no distinct improvements, the additional variables can be excluded.

Previous research has proven that machine learning is better at predicting inflation than conventional time-series models (Ivan, 2018; Ülke et al., 2016). However, different machine learning models will perform distinctly under various conditions. Therefore, this study attempts to understand whether using a preliminary dataset with few variables might improve results in one machine learning model but using a more extensive dataset might improve a different model. Adding variables will also help with understanding the overall performance of the model and to choose specific models depending on the availability of the data.

The following research questions will guide the research:

1. What is the effectiveness of machine learning algorithms in forecasting inflation?
2. What are the most effective machine learning algorithms in predicting inflation rates in Finland?
3. What are the impact of additional variables and how do they improve machine learning models?

1.3 Structure of the Thesis

The study consists of six chapters. The current chapter provides a background of the topic, the objective of the study, and the research question. The following section of the study will be structured as mentioned below:

Chapter 2 discusses the literature review of the paper, where previous studies will be used to provide information on what studies have been performed in this field worldwide. It will be a mixture of economics-related and machine learning-related studies to supplement the main ideas discussed in this research.

Chapter 3 illustrates the research methods and the various process of research covering the data collection, cleaning, and analysis of the paper. This section explores the various selected machine-learning models and how the machine-learning models were employed in the research.

Chapter 4 covers the information analysis after executing the machine learning model and the performance of the various models. This section will also make a comparison between the various models.

Chapter 5 discusses the analysis in further detail with the vital information that has been received from the study and how it compares to the previous studies discussed in the literature review, along with the contribution of the study, the limitations faced during the study, and how this study can be further elaborated in the future.

Chapter 6 ends the paper with a summary of the key ideas and the conclusion that can be retrieved from this study.

2 LITERATURE REVIEW

The second chapter of the paper will discuss the previous works of literature studying similar topics in a different context. For this chapter, the main discussion is on studies that have been done related to inflation forecasting and studies that cover machine learning. Understanding both pieces of literature is vital for this study as it will explain what the researchers have performed in the past.

2.1 Economic Forecasting and Machine Learning

Machine learning models have improved in the past, and one major use of machine learning has been for predictions. Regression analysis using machine learning models has been one of the major topics of study in economic forecasting. Economic topics such as the stock market index, interest rates, and GDP growth have been vastly studied using traditional and machine learning models. There are many studies that compare these traditional models during the past decade to compare how these models work in different environments.

Economic forecasting has been mostly performed using econometrics models. Regression, Exponential Smoothing, and Autoregressive models have been frequently discussed when studying economic forecasting (Shobana and Umamaheshwari, 2021). Timmermann (2008) evaluated 11 different models to check the predicting ability of traditional and machine learning models when working with stock returns. Prevailing mean, autoregressive model, factor-augmented AR model, exponential smoothing, and double exponential (Holt) smoothing are all linear regression models used in the study. Similarly, Timmermann (2008) also evaluated some non-linear econometric models, two logistic STAR models were used with STAR1 and STAR2 and a single-layer neural network with two hidden units.

Furthermore, the study also had a previous best model and average approach. Both the aforementioned models are performing econometrics study as well as some machine learning for comparative study of economic data. While Shobana and Umamaheshwari (2021) concluded that neural networks were the best-performing models, Timmermann (2008) concluded that economic indicators, as demonstrated in the study by stock return, could not be predicted with reasonable accuracy.

Adding to his conclusion, Timmermann (2008) mentioned that ML techniques used in time series forecasting are likely to over-perform. However, one advantage that traditional econometrics models have over ML models is that traditional econometrics models do not require modeling to understand the non-linear method and can still be helpful.

For the study of sovereign risk, many studies have used machine learning to study these concepts, and ML has proven as a valuable tool. Arakelian et al. (2019) performed regression trees and random forests. Balduzzi et al. (2022) took advantage of the LASSO algorithm for a similar study. Kim et al. (2020) performed an analysis based on recurrent neural networks, support vector regression, long short-term memory, and group methods of handling data. The same study also concludes that ML models outperform traditional econometric models. Collin (2006) and Castellani and Santos (2006) presented a forecast of U.S. Treasury Bonds with an Artificial Neural Network.

In the study of sovereign risk, there is a conclusion that machine learning performs better compared to traditional models. The study evaluates Bayesian model averaging against ML models for support vector regression, elastic net regression, random forests, extreme gradient boosting, and artificial neural networks. The study infers that the ML models are able to comprehend the dynamics of the market for sovereign risk more than the traditional econometrics models. The study concludes that the XGBoost model provided a satisfactory result in their experiment (Belly et al., 2023).

The significant issues with traditional forecasting models have been that they are mostly linear in nature, while economic activities can be a lot more erratic. Studies like Gu et al. (2020) have proven that machine learning models perform much better with a non-linear dataset. As economic predictions consist of non-linear datasets, it can be extrapolated that machine learning will perform well with economic predictions.

2.2 Inflation and Inflation Forecasting

For a country's macroeconomics, inflation is among the biggest concerns, while inflation expectation plays a similar role when understanding inflation. Countries have used inflation targets to control their inflation, but usually, these targets are created based on models which are met with the use of financial and monetary policy. Another aspect related to inflation is that inflation expectations will create inflationary pressure on the economy, which needs to be understood (Gunduz et al., 2020).

Inflation prediction is a necessity when it comes to creating optimal monetary policy. Therefore, central banks use these predictions as there are inflation targets that these

countries need to meet to create stability. Usually, these central banks have been using vector autoregression, Bayesian Vector Autoregression, and Factor Models to predict the CPI of the country (Xie et al., 2007).
Inflation prediction has been in the interest of economists since the start of times as inflation forecasting has been widely studied in studies like Phillips (1958) covering wage inflation and unemployment, along with Samuelson and Solow (1960), who looked at the relationship between inflation and unemployment. In addition, studies like Stock and Watson (1999) and Gordon (1997) have discussed how output gaps and inflation should be considered. Hassani et al. (2013) also elaborated on how GDP and GNP price indices impact inflation.

Stock and Watson (1999) claim that multivariate models perform better in comparison to univariate time-series models.

2.3 Machine Learning Models in Inflation Forecasting

Autoregressive models are some of the most popular methods for inflation forecasting; however, they have the limitation that it works well in a linear model. The problem arises that inflation is non-linear (Binner et al., 2005).

Ülke et al. (2016) compared autoregressive models to other machine learning models like Support Vector Regression, Artificial Neural Network, and K-Nearest Neighbor for inflation forecasting. They concluded that machine learning models performed better compared to univariate or multivariate time-series models.

Previous studies such as Nakamura (2005), Choudhary and Haider (2008), and Binner et al. (2005) have studied the use of the neural network in inflation forecasting, with all of them concluding that neural networks have better accuracy when it comes to inflation forecasting. They also emphasize that neural networks can capture inflation's non-linear features. However, Nakamura (2005), Choudhary and Haider (2008), and Binner et al. (2005) have all used a multi-layer feed-forward neural network to forecast inflation as there has not been a consensus on what models are the best when it comes to the neural network in inflation forecasting.

Ülke et al. (2016) also mention that using machine learning in inflation forecasting is a comparatively new phenomenon. In comparison, there have been macro-economic studies using machine learning, like studies of foreign exchange rates, income forecasts, and stock market volatility. Also, when exploring recent studies, there has been a limit in studies about machine learning and inflation forecasting, with most of them provided by central banks in the form of working papers like Araujo and Gaglianone (2022) at Banco Central Do Brasil and Joseph et al. (2021) at Bank of England.

The nowcasting model is used in Finland to understand the real-time quarterly GDP growth, which updates itself as soon as new data are available to estimate the GDP growth for the upcoming quarter (Fornaro and Luomaranta, 2018). Furthermore, inflation has been considered necessary by economists for a while, which means that a real-time inflation forecasting model is necessary to control the issues.

2.4 Variable for Multivariate analysis in Inflation Forecasting

Multivariate inflation forecasting has many studies that show what might be the possible literature that covers inflation forecasting. For instance, Stock and Watson (2002) use real output and input in the form of an index of industrial production; employment index with the use of unemployment rate, employees on non-agricultural payroll, help-wanted advertising in a newspaper; real retail, manufacturing, and trade sales which include the value of goods; consumption in terms of personal consumption expenditure; housing starts according to the states and authorization of new house; real inventories and inventory-sales ratio; orders and unfilled orders; stock prices; exchange rates; interest rates as interest rates, bond yields, spread rate; money and credit quantity aggregates as money stock, deposits, loans, and securities; price index in terms of producer price index, consumer price index, personal consumption expenditure; average hourly earnings for different sectors; and miscellaneous in terms of exports, imports and trade balance. The authors have analyzed 215 different indexes in their studies under each of the terms that were discussed above.

Similarly, the working paper in Brazil by Araujo and Gaglianone (2022) also evaluated 167 different indexes where they have used index related to inflation, interest rates, money, banking, capital markets, foreign exchange and risk, labor market, industry, sales, energy, climate, public sector as expenses, economic activity, exterior like import price index and export price index, commodities, and global uncertainties.

Additionally, the working paper by Joseph et al. (2021) evaluated 46 indexes using similar variables that were in the previous papers. However, they were not divided into different sections with multiple indexes for each variable. These index for each of the country changes depending on what the different countries register; however, the main economic activities used are similar.

Özgür and Akkoç (2021) also examined 229 different indexes under the heading of commodity price, money market, production, stock market, trade, budget, construction, exchange rates, and balance of payment for their machine learning-related analysis. Rodríguez-Vargas (2020) explores 18 independent variables from a similar heading in his paper. Also, Yang and Guo (2021) applied ten independent variables to forecast CPI with terms like narrow money supply, broad money supply, interbank lending rates, industrial increase, sales of consumer goods, housing boom index, index of real state development investments, housing starts, Shanghai and Shenzhen 300 index, the exchange rate to US dollars.

Finally, when Ülke et al. (2016) analyzed six economic activities with the unemployment rate, index of industrial production, real personal consumption expenditure, employees on non-farm payroll, housing starts, and term spread, all these indexes were in the 215 studied by Stock and Watson (2000).

Regarding multivariate analysis, there is no pre-determined variable for inflation forecasting. With different authors, the need of the paper and the authors' discretion has been the deciding factor related to which independent variables will be part of the analysis.

2.5 Previous Studies

While the studies are limited, and those available are performed in the United States, there have been studies covering how different machine learning models have performed in the past.

Ülke (2016) performed a comparison between univariate autoregressive models, multivariate autoregressive models, and machine learning models to compare the performance of these models to each other and to figure out which model performed the best among the different time horizons that they had selected for the study. They have also used four different target variables with the consumer price index, core consumer price index, personal expenditure consumption, and core personal expenditure consumption for their studies. For the study, they observed that the k-nearest neighbor and artificial neural network were the best-performing models for CPI forecasting. They also conclude that support vector regression outperforms other models regarding core-

PCE forecasting. In addition, machine learning models work better when there are irregularities in the data, while time-series models perform better when data are more stable.

Araujo and Gaglianone (2022) used 167 economic and financial indicators to compare 50 different machine-learning models. When these models were used, they found that ML models invariably beat their benchmark model (ARMA) in many situations where the R-squared value improved with more than two-digit of accuracy.

Chavez-Hurtado and Cortes-Fregoso (2013) performed a similar study to figure out the performance of neural networks to compare with the performance of autoregressive models. In their study, they divided their data into phases where the Mexican economy was in a state of volatility, a transition from volatility to stability, and a state of stability. The study concludes that the neural network performed better compared to the Bank of Mexico model during the state of volatility and state of stability but performed slightly worse during the transition. They find the performance to degrade during the transition due to the change in the situation for which the neural network could not account. They found that the neural network had the capability to anticipate the crisis almost a year before it happened to allow the National Government to modify their policies to avoid the crisis.

In the study by Ahmed et al. (2010), the researchers wanted to find out what models perform the best among multi-layer perceptron, Bayesian neural network, radial basis function neural network, generalized regression neural network, K-nearest neighbor regression, classification and regression trees, support vector regression, and Gaussian processes. From their study, they concluded that multi-layer perceptron and Gaussian processes regression performed the best among their data, while support vector regression performed worse. When compared to Ülke (2016), who found that artificial neural networks performed the best for CPI and multi-layer perceptron being an ANN, their finding is comparable. However, Ahmed et al. (2010) also concluded that support vector machines were better for a classification-related task but performed worse in regression which contradicts Ülke (2016), which found SVR to perform best with core-PCE indicators.

Joseph et al. (2021) used macroeconomic and CPI-based indicators to compare how machine learning models performed. They found that adding macroeconomic indicators did improve the models, but it was insignificant compared to using CPI-based indicators on their own. The authors also discussed how shrinkage models like Ridge or LASSO could help improve the accuracy of the inflation prediction.

Yang and Guo (2021) performed a gated recurrent unit – recurrent neural network (GRU-RNN) for inflation prediction against established autoregressive models. From their experiment, they conclude that these GRU-RNN performs better against autoregressive models and find other suitable deep learning models that can be used to improve the accuracy of inflation prediction.

Özgür and Akkoç (2021) examined the shrinkage model to identify which method performs better than the baseline autoregressive methods. They compare the data from 2007 to 2019 for their forecasting estimate and use five different shrinkage models with ridge, LASSO, adaptive LASSO, Group LASSO, and ElasticNet as the comparative models. They conclude that LASSO and ElasticNet performed best compared to all the other baseline and shrinkage models with root mean squared error of 0.834 and 0.893 with lambda values at 0.065 and 0.710, respectively.

Rodríguez-Vargas (2020) performed a similar machine-learning analysis to forecast Costa Rican inflation. Like other studies, they use the autoregressive model as the base model and compare their chosen machine-learning model. The authors have chosen Univariate K-Nearest Neighbor, KNN with explanatory variable, Extreme Gradient Boost, Random Forest, and Long-Short Term Memory (LSTM). From their analysis, LSTM and univariate KNN were the best-performing models, with random forest and XGBoost performing worse than these models.

The abovementioned studies show that shrinkage models and neural networks have performed well in different economies worldwide, whether in the US and the UK or developing economies like Turkey, China, and Costa Rica.

Study	Author	Location	Subject	Data Set Size	Models Used	Metrics	Best Performer
Explainable inflation forecasts by machine learning models	Aras, Lisboa (2022)	Turkey	Inflation forecasting	26 macroeconomic variables	SVM, MLP, Random Forest, Extremely randomized trees, Adabosst, GBDT, XGBoost	Goodness of Fit (R^2)	Random Forest RMSE = 0.916
Forecasting Costa Rican Inflation with machine learning models	Rodriguez-Vargas (2020)	Costa Rica	Inflation forecasting	19 macroeconomic variables	Univariate KNN, KNN with explanatory variables, XGBoost, Random Forest, LSTM	RMSE	**LSTM**, Univariate KNN, Random Forest and XGBoost RMSE for LSTM = 0.0032
Inflation forecasting in an emerging economy: selecting variables with machine learning algorithm	Özgür and Akkoç (2022)	Turkey	Inflation forecasting	11 macroeconomic variables	ARIMA, Ridge, LASSO, Group LASSO, Adaptive LASSO, ElasticNet LASSO, ElasticNet	RMSE	AdaLASSO RMSE = 0.834

Study	Author	Location	Subject	Data Set Size	Models Used	Metrics	Best Performer
Inflation Prediction Method Based on Deep Learning	Yang and Guo (2021)	China	Inflation forecasting	Ten macroeconomic variables	ARMA, BVAR, BP, GRU-RNN	MSE, MAPE, SMAPE	GRU-RNN MSE = 0.359
A comparison of time series and machine learning models for inflation forecasting empirical evidence from the USA	Ülke et al. (2016)	USA	Inflation forecasting	Six macroeconomic variables	Autoregressive model, Naïve model, ARDL, VAR, KNN, ANN, SVR	RMSE, goodness of fit statistics (R^2)	SVR - Core PCE RMSE = 0.77 ARDL - Core CPI RMSE = 0.62
An Empirical Comparison of Machine Learning Models for Time Series Forecasting	Ahmed et al. (2010)	Egypt	Inflation forecasting	Univariate	MLP, BNN, RBF, GRNN, KNN, CART, SVR, GP	SMAPE	MLP, GP SMAPE = 0.0856
Forecasting UK inflation bottom up	Joseph et al. (2022)	UK	Inflation forecasting	46 macroeconomic variables	Principal Component Analysis, Partial Least Square,	RMSE	Ridge, LASSO RMSE = 0.78

Study	Author	Location	Subject	Data Set Size	Models Used	Metrics	Best Performer
					LASSO, Ridge, ElasticNet, SVM, Random Forest, NN		
Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods	Medeiros et al. (2019)	USA	Inflation forecasting	135 macroeconomic variables	Random Walk, Ridge, LASSO, AdaLASSO, ElasticNet, Target Factors, Factor Boosting, Bagging, Complete Subset Regression, Jackknife Model Averaging, Random Forest	RMSE, MAE, MAD	Random Forest RMSE = 0.70
Machine Learning Methods for Inflation Forecasting in	Araujo and Gaglianone (2022)	Brazil	Inflation forecasting	167 macroeconomic variables	50 models (*See Appendix I*)	RMSE	Shorter Horizon - RNN, RF Longer Horizon -

Study	Author	Location	Subject	Data Set Size	Models Used	Metrics	Best Performer
Brazil: new contenders versus classical models							XGBoost, RF RMSE = 0.055
Forecasting by Machine Learning Techniques and Econometrics: A review	Shobana and Umamaheshwari (2021)	India	Economic Forecasting	Unknown dataset	MLP, Logistic Regression, Decision Tree, KNN, SVR, LASSo, Ridge, ElasticNet	RMSE, MAE, MAPE, WMAPE	MLP, Random Forest
Forecasting sovereign risk in the Euro area via machine learning	Belly et al. (2022)	Euro Area	Sovereign Risk Rate	11 macroeconomic variables, 11 financial market variables, sentiment data from Google search frequencies, six monetary policy variables	SVR, Random Forest, XGBoost, ANN	MSE, RMSE, MAE	XGBoost MSE = 0.044

Table 1Machine Learning Algorithm in Literature

3 RESEARCH METHODOLOGY

This research aims to understand which machine learning model will perform better with a particular data set in Finland and to check whether the machine learning model will perform differently when additional data are appended to the model. The idea of the study is to explore the performance of the model and whether the results that the models produce are similar or different when a new set of variables is added to the original dataset. Various analyses will be made using openly available secondary data from the internet. While this will not cover all the inflation-related aspects, the current research tries to illustrate what can be done in the different phases.

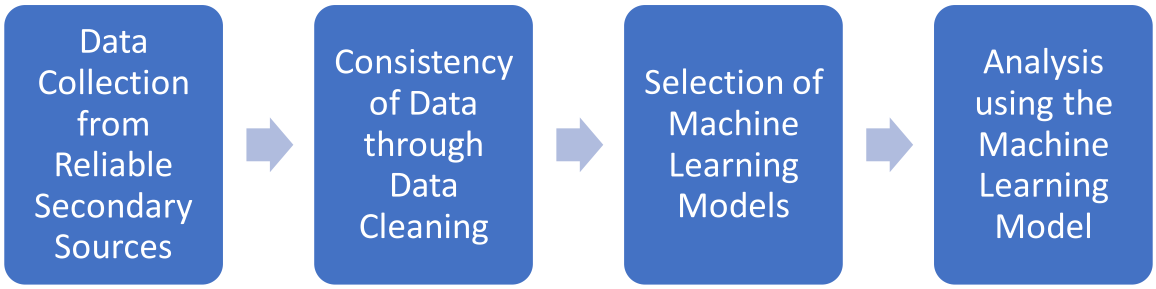

Figure 1 Research Method Flow Diagram

3.1 Research Method

The research conducted is a quantitative analysis research. The research employs various time-series data collected between January 2000 and January 2023, the latest available secondary data. There are 277 monthly data in a time series that can be effortlessly trained and tested with machine-learning models. Much of the data available online is usually recorded annually, so it was important to find data that were recorded monthly to perform the analysis for this study.

The research is also a multivariate analysis that explores various independent variables to determine the target variable, the index determining inflation. To perform the analysis, this research will examine multiple other variables that impact the inflation variable.

Inflation focuses on two sides of the economy, the demand side, which investigates the consumer of goods or services, and the supply side, which examines the supplier of goods and services. A decision to select only the demand side of the economy was made before collecting the data. With the aim of data collection, various data sources had to be explored to locate the data relevant to the analysis. These data are usually recorded by the central bank of the country, the statistical department of the country, the World Bank, and the International Monetary Fund (IMF). In the case of Finland, some data were also available on the European Central Bank (ECB), European Union (E.U.), or European Commission (E.C.) database. In addition, the Organisation for Economic Co-operation and Development (OECD) database was also selected to examine the data available related to Finland.

3.2 Data Selection, Collection, and Cleaning

In machine learning research for inflation prediction, researchers worldwide have chosen various datasets to train and test the model. While some research (Araujo & Gaglianone, 2022; Özgür & Akkoç, 2021; Rodríguez-Vargas, 2020) has used more than 50 variables, whereas other research (Ülke et al., 2016; Yang & Guo, 2021) has been constrained to ten or fewer independent variables for multivariate analysis.

3.2.1 Data Selection

This research will make two separate multivariate analyses. The analyses are called preliminary analysis and supplementary analysis. For the preliminary analysis, five variables were selected, namely the index of production, housing starts, and unemployment which have been used in previous research when using a few variables. At the same time, variables, such as spread rate and personal consumption expenditure, were unavailable for Finland; the yield curve was used as an alternative to the spread rate. Personal consumption expenditure, another inflation index that denotes inflation from the consumption side of the economy, was replaced with the cost-of-living index. The cost-of-living index was chosen because it is another inflation index but is not considered identical to the most common inflation index, the consumer price index (CPI). Finally, the target variable was the most common inflation index, the consumer price index. While previous research also used employees on non-farm payroll as an index that is comparable with the self-employment index in Finland, the variable studying non-farm payroll was removed from the study due to the data being annualized only.

The variables discussed above were used vastly in previous studies, such as Ülke et al. (2016) and Yang and Guo (2021); some indexes used for the supplementary analysis were selected depending on their impact on the consumer price index. The business confidence index is an opinion survey regarding future development that can be used to monitor output growth and anticipate turning points in economic activity.

The producer price index tracks the changes in price at the producer or wholesale level, which is the price change for the suppliers in the economy. At the same time, Import Price Index and Export Price Index determine the change in the price of goods and services imported to and exported from the country. The Basic Price Index of Domestic Goods calculates the price changes of goods produced in the country. OMX Helsinki Index is the stock market index of Finland. Finally, the exchange rate data from Euro to U.S. Dollar (USD), Japanese Yen (JPY), Chinese Yuan (CNY), and British Pound (GBP) were selected from the list of the largest economies that do not use the Euro as their currency.

3.2.2 Data Collection

The data were collected from various sources, with various aggregators used when data were not in the required format. Data available on multiple sources were selected depending on the proximity to Finland. For example, when data were available from a Finnish authority, those were selected first, with data from E.U. entities being prioritized second and data from international agencies selected at the end. However, in some cases, data were selected depending on the data being on the same scale. For instance, when data were available from a Finnish authority with the base year of 2000 and most other data were available with the base year of 2015, the most frequent base year would be selected even though the proximity of the data source was further from Finland.

The data for the index of production were collected from the OECD data bank as those were not available in the record by Statistics Finland or the Bank of Finland. Therefore, the extracted data consider the production in 2015 as the base index, which is 100, and are calculated based on the production variation from the base index.

The data for housing starts were from Statistics Finland. The housing start data employs the index system where 2015 is the reference point with a base index of 100, and monthly variation in housing units from the base year is computed. In addition, residential housing starts were considered for the dataset as residential housing were commonly evaluated in previous studies such as Ülke et al. (2016), and Yang and Guo (2021).

The unemployment data were collected from OECD, which uses the estimates from International Labour Organization (ILO Estimates) for calculating unemployment data. While unemployment data were available in Statistics Finland, the data were not recorded monthly. Therefore, OECD data were selected. The data correspond to the percentage of the eligible population.

The yield data were collected from the Bank of Finland, which records the data for short-term and long-term bond rates. The yield data represents the difference between the short-term and long-term rates for the given month. The Bank of Finland collects the data daily, and the average is used as the aggregator for monthly data.

The cost-of-living index data were collected from Statistics Finland. Unfortunately, the data are available only at the index of 1950 as the base year. Data with other base years were unavailable in the required form in any other data source. Therefore, the data are the change in the cost of living from the base year.

The consumer price index data were collected from OECD even though it was available in Statistics Finland. The reason was that the base year at OECD was 2015, while on the Statistics Finland website, the same information could only be retrieved for the base year 2000 for the time horizon used in the research. Since 2015 was the base year for other indexes, and after matching the data between OECD and Statistics Finland for the shorter time horizon available on their page, it revealed that the values are the same between 2015 and 2023; it was considered the same dataset, just recorded by different entities.

The Business Confidence Index data were collected from European Commission. The business confidence index is calculated as higher or lower than a base value of 50 and records the positive or negative opinion regarding the direction of the economy.

The producer price index, import price index, export price index, and basic price index for domestic goods were compiled from Statistics Finland. These indexes were available with the base year as 2015 at 100. Therefore, the changes were recorded from the base value and the base year.

The Euro exchange rate data were collected from European Central Bank, where the data are recorded. The exchange rates are recorded daily, so the data were aggregated using average to monthly. Therefore, the values show the monthly average daily data for the selected month.

The stock market index OMX Helsinki data were collected from the Bank of Finland database. The database records the index daily, and the average was used as the aggregator to obtain the monthly data for the given time series.

3.2.3 Data Cleaning

The data collected were from various databases. Therefore, some were in ascending order, while others were in descending order. Furthermore, some data were for the selected time horizon, while others were available until February 2023, which is not part of the study. In an Excel sheet, all the required indexes from the downloaded files were collected for the time between January 2000 and January 2023 and sorted from earliest to latest.

Further data preprocessing was done during the analysis phase, which will be discussed in the data analysis section of this chapter. First, all the data were saved as CSV (Comma Separated Value) for easy retrieval using Python. The data were also saved in two files, one containing the preliminary set of five indicators with the consumer price index as the target variable and another file for supplementary analysis containing all the variables mentioned above.

3.3 Model Selection

This research evaluates a few machine learning models used in previous studies to understand which models perform the best for Finnish inflation. The models were selected from previous studies, and various types of models have been selected to check whether various models might have different levels of impact. Machine learning models are available in regression algorithms, tree-based algorithms, shrinkage algorithms, and artificial neural networks. All the equations of the models are models that are covered by previous studies related to machine learning regression models.

3.3.1 Decision Tree

Decision tree is the simplest form of a tree-based algorithm, which can be visualized as a flowchart showing a pathway to the decision. Decision tree is among the most popular and reliable supervised learning models, which is easily understandable to humans (Pathak et al., 2018). As the base form of a tree-based model, it will provide a simple decision tree until the leaf variables reach a point with a single attribute.

This model has been selected to compare to other tree-based models like random forest and extreme gradient boosting. Therefore, the model can be considered the baseline model for this study. Moreover, the decision tree is one of the models Shobana and Umamaheshwari (2021) have selected for their study.

3.3.2 Random Forest

Random forest is among the most popular model utilized in the studies discussed during the literature review; any model which investigates the machine learning model generally consists of the random forest model. Previous studies that applied these models are Aras and Lisboa (2022), Rodriguez-Vargas (2020), Joseph et al. (2022), Medeiros et al. (2019), Araujo and Gaglianone (2022) and Belly et al. (2022). Breiman (2001) proposed the random forest model, which allows for reducing the variance of regression trees. Random forest is also a non-parametric or recursive model that estimates any non-linear function. The model minimizes for the mean absolute error for the available parameters of θ_k where K is the number of terminal nodes.

$$\pi_{t+h} = \sum_{k=1}^{K} c_k I_k\left(x_t; \theta_k\right)$$

Equation 1Random Forest Equation

For $I_k(x_t ; \theta_k)$ is an indicator function using the formula below, where $R_k(\theta_k)$ is the kth region. The regression trees are randomly constructed and aggregated based on bootstrapping represented by B.

$$I_k(x_t; \theta_k) = \begin{cases} 1 \text{ if } x_t \in R_k(\theta_k) \\ 0 \text{ otherwise,} \end{cases}$$

Equation 2 Random Forest Indicator Function

The samples are determined by b, where a tree with K_b reasons is calculated randomly from a subset of original regressors. The forecast is made using the formula,

$$\hat{\pi}_{t+h} = \frac{1}{B}\sum_{b=1}^{B}\left[\sum_{k=1}^{K_b} \hat{c}_{k,b} \mathrm{l}_{k,b}\left(x_t; \hat{\theta}_{k,b}\right)\right]$$

Equation 3Random Forest Sample Forecasting Equation

The major problem with random forest regression trees is that small values change the splits in the data leading to high variance in prediction.

3.3.3 Support Vector Regression

Support vector regression is a model developed under the Support Vector Machine model of classification that estimates a hyperplane that maximizes the separation between data. SVR evaluates the flattest function, which constrains the deviation of the data observed between a fixed amount (Awad & Khanna, 2015). Support vector regression is one of the most popular models being used in machine learning for economic and inflation forecasting; SVR has been used by Aras and Lisboa (2022), Ülke et al. (2016), Ahmed et al. (2010), Joseph et al. (2022), Shobana and Umamaheshwari (2021) and Belly et al. (2022).

The fundamental prediction is made by using a linear function defined by weight vector (w), bias (b), and input vector (x) (Awad & Khanna, 2015).

$$f(x) = w^T x + b$$

Equation 4 Linear Function for Support Vector Regression

The error function uses the following formula with x_m and y_m denoting the mth training input vector and the target output at m = 1 until $m = M$ *(Awad & Khanna, 2015)*.

$$J = \frac{1}{2||w||^2} + C \sum_{m=1}^{M} [\![\; |y_m - f(x_m)_\epsilon]\!]$$

Equation 5 Error Function SVR

In the function, the first term penalizes the complexity of the model while the second term is ϵ-insensitive loss function which is displayed as $|y_m - f(x_m)|_\epsilon = max\{0, |y_m - f(x_m)| - \epsilon\}$. This stops the penalization of errors below the function, which allows for some flexibility for the parameters to move and minimize complexity. The error function is defined with the function below (Awad & Khanna, 2015):

$$f(x) = \sum_{m=1}^{M} [\![(\alpha_m^*]\!] - \alpha_m) x_m^T x + b$$

Equation 6 Support Vector Function

α_m and α_m^* are called the lagrange multiplier, which is also known as a support vector when the values are non-zero (Awad & Khanna, 2015).

3.3.4 Extreme Gradient Boosting

Various researchers have used Extreme Gradient Boosting, and these models have been used various times for economic and inflation forecasting by authors like Rodríguez-Vargas (2020), Li et al. (2022), Belly et al. (2023), Aras and Lisboa (2022).

Extreme gradient boosting is a tree-based machine learning model. Boosting is an ensemble learning model that strengthens the weaker classifiers into a regular decision tree by integrating the model into a more precise model that can control the signal interferences. XGBoost uses various decision trees, which helps it learn regarding the classification of data. Using these multiple decision trees, the model minimizes the error in each iteration and improves the accuracy of the following trees. (Sheridan et al., 2016)

Mathematically, XGBoost can be presented below where the dataset is denoted by $\{x_i, y_i\}$ with p number of samples and q number of features, $\{x_i \in R^q, R^q \to L, y_i \in R, i = 1 \dots . p\}$ where $\hat{y}_i$ is the forecasting value, and the regression tree is denoted by f_{nl}. *SP* denotes the space of tree and *NL* represents the total number of trees.

$$\hat{y}_i = \sum_{nl=1}^{NL} f_{nl}(x_i), f_{nl} \in SP$$

Equation 7 Extreme Gradient Boosting Equation

3.3.5 Least Absolute Shrinkage and Selection Operator

The base of the shrinkage model is built on the concept of $G_h(x_t) = \beta'_h x_t$ *and* $p(\beta_{h,I}; \lambda, w_i)$ being the penalty function where λ is the penalty parameter and has the weight $w_i > 0$. The base model is defined by (Ranstam & Cook, 2018):

$$\widehat{\beta_h} = \arg\min \left[\sum_{t=1}^{T-h} \left(y_{t+h} - \beta'_h x_t\right)^2 + \sum_{i=1}^{n} p\left(\beta_{h,i}; \lambda, w_i\right) \right]$$

Equation 8 LASSO Equation

LASSO model uses the penalty where

$$\sum_{i=1}^{n} p\left(\beta_{h,i}; \lambda, w_i\right) \coloneqq \lambda \sum_{i=1}^{n} \left|\beta_{h,i}\right|$$

Equation 9 Penalty Function

The LASSO model shrinks all irrelevant values to zero, but the model selection can be made under strict conditions.

LASSO model is among the most used shrinkage model in machine learning for inflation forecasting and economic forecasting as applied by Ögzür and Akkoç (2022), Joseph et al. (2022), Medeiros et al. (2019), Araujo and Gaglianone (2022), and Shobana and Umamaheshwari (2021).

3.3.6 Multi-layer Perceptron

Multi-layer perceptron are simple feed-forward neural networks. These models are among the most simple and popular forms of neural networks. The model is stated below:

$$\hat{y} = v_0 + \sum_{j=1}^{NH} v_j g\left(w_j^T x'\right)$$

Equation 10 Multi-layer Perceptron Function

The x' is the input vector (x) augmented with 1, $x = (1, x^T)^T$, w_j is the weight vector for the jth hidden node, $v_0, v_1, \ldots, v_{NH}$ are the weight of the output node and $\hat{y}$ is the network input (Murtagh, 1991).

The complexity of the model can be controlled by changing the *NH* value which is the number of hidden nodes (Ahmed et al., 2010). As one of the basic models of neural networks, MLP still is among the popular models with studies like Aras and Lisboa (2022), Ahmed et al. (2010), and Shobana and Umamaheshwari (2021) have applied these models in their studies.

3.4 Parameter Tuning

Machine learning models require various parameters when they are operated on any programming language (Claesen et al., 2014). Most of the models have predefined default values when they are used in a model for each of the machine learning models. This study has been done with three different sets of parameters. The first round is with the default value for all the parameters as set by the selected libraries in Python. The second round of parameters is tuned using a few parameters that impact each model. These are applied in a grid search that looks for the best accuracy in in-sample prediction, and outputs for the best parameter are used in training and testing with an out-of-sample dataset. Finally, since in-sample and out-of-sample models may have separate accuracy under the same parameters, the parameters are further tuned manually until the results are improved from the default value or until the tuning reaches the default value.

Decision trees were optimized for *max_depth* value, which represents how deep the tree can be; *max_sample_leaf,* which determines how many samples are required to be a leaf node; *criterion* is the function that is used to measure the quality of the split and *splitter* is the strategy used at the node to make the split.

Random forest contains the parameters *max_depth* and *max_sample_leaf,* which is the same as in the decision tree. *N_estimators* is another parameter which is the number of decision trees in the forest. While a decision tree is a single decision tree, a random forest is a collection of various decision trees. *Max_features,* as the name suggests investigates and controls the maximum number of features when making a split and prevents overfitting.

LASSO models consider the parameters of *alpha* that controls the strength for the penalty, *fit_intercept,* which decides whether an intercept is fit to keep or not, and *max_iter,* where the number of iterations algorithm runs through and normalize checks whether the input variables need normalization.

Support vector regression contains the parameters *C*, *kernel,* and *epsilon* that were optimized. *C* value is the regularization parameter that controls the trade-off between the complexity of the model and training error. *Kernel* function transforms the data into higher-dimensional space, and *epsilon* controls the tolerance of error in training data; it is the margin of error that the predictions are allowed to make.

XGBoost library contains the extreme gradient boosting algorithm; the *learning_rate* is the shrinkage rate for each iteration of boosting. *Max_depth* is the size of the decision tree, and *min_child_weight* is the hyperparameter that sets the minimum weight required for a child node to be created when a decision tree is being built.

Multi-layer perceptron required *max_iter* to be changed since the network must converge to provide accurate information. *Hidden_layers* determines the number of hidden layers in the model, *activation* is the function used for each neuron, *solver* is the optimization algorithm that changes the weights and biases in the network, *learning_rate* determines the learning rate of the optimizer and *alpha* which is the regularization parameter penalizing the network such as in LASSO.

3.5 Evaluation Metrics

Machine learning models are evaluated with various tools. As a regression based machine learning mode, there are few tools such as RMSE, MSE, R-squared that are performed to understand the accuracy of the algorithms (Botchkarev, 2019). For this study, the evaluation metrics used for these models are mean-squared error and R-squared values. These two calculations are done between the predicted value from an out-of-sample variable and compare it with the actual value. This evaluation provides clear metrics to compare the various models selected in the previous section. Various evaluation metrics have been used throughout the various studies root-mean-squared error is the most popular model, along with the mean absolute error, mean absolute percentage error, symmetric mean absolute percentage error, the goodness of fit (R-squared), mean squared error and weighted mean absolute percentage error.

3.5.1 Mean Squared Error

Mean-squared error, as the name suggests, is the average of the squares of the errors; it looks at the square of the difference between a predicted value and an actual value. As a squared error, the value will always be positive and non-zero (Botchkarev, 2019). Machine learning uses mean-squared error as an empirical risk minimization. The value being of MSE approaches zero as the measure of error decreases. Mean squared error has been used as an evaluation metric in inflation and economic forecasting in studies such as Yang and Guo (2021), and Belly et al. (2022).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Equation 11 Mean-squared error

3.5.2 Goodness of Fit (R-squared)

The R-squared value is also known as the coefficient of determination, and the coefficient illustrates the goodness of fit for a model. R-squared approximates how the line of regression can predict the actual data (Colin Cameron & Windmeijer, 1997). R-squared value of one means that all the actual data passes through the regression line, while a zero value means that the actual data are very far away from the regression line. A value for r-squared provides that the data is close to the regression line and can predict a value with a narrow range of errors. R-squared as a goodness-of-fit metric has been applied in studies such as Ahmed et al. (2010), and Aras and Lisboa (2022).

$$r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Equation 12 R-squared

3.6 Software and Library

The software used for the machine learning model is Python 3.9.7, which was used inside the Jupyter Notebook extension for Visual Studio Code. The model used the sci-kit learn library, also known as the *sklearn* library, for most of the machine learning model along with the *xgboost* library for the extreme gradient boosting model; pandas library was used to load the data from CSV. From *model_selection* in *sklearn*, the *train_test_split* was imported to create a randomized training and test set for the study. Furthermore, *GridSearchCV* was also used for parameter tuning. From *sklearn.preprocessing*, a *MinMaxScaler* was also imported, which scales the various values in the dataset to values between zero and one to scale the data. For accuracy checks, *mean_squared_error*, *mean_absolute_error,* and *r2_score* library were imported for mean squared error, mean absolute error, and r-squared calculation. From *sklearn.tree* the *DecisionTreeRegressor* was imported for decision tree models, *sklearn.ensemble* was used to import *RandomForestRegressor*, *sklearn.linear* was used for *LASSO*, *sklearn.svm* for *SVR*, *sklearn.neural* network for *MLPRegressor* and *XGBoost* library was used for *XGBRegressor*.

4 RESULTS

For data consistency, the data were preprocessed using a scaler which would scale the data between the value of zero and one depending on the highest and lowest value in the column. However, the target variable was not scaled but recorded in the original scale. Following the preprocessing, the analysis was performed with Python in Jupyter Notebook, and the data were stored in a table in Excel.

In this chapter, the data will be presented along with the analysis of the information that was retrieved from this data.

4.1 Descriptive

Initially, there is a need to understand the variables that are computed during the analysis. Then, the correlation between the variables can provide exciting insights resulting in a better understanding of the variables and the impact of the variables on the results that followed in the analysis.

With the two sets of analysis, namely preliminary and supplemental analysis, the datasets were also provided the same name. While the first five variables of the yield curve, housing starts, production index, cost-of-living index, and unemployment, were used in the preliminary analysis. Nine other variables were added to the supplemental dataset, and the correlation of all these variables is represented in Figure 2, which illustrates the correlation as a heatmap. When investigating each variable, Table 2 emphasizes the correlation score of all the independent and target variables.

As Boivin and Ng (2006) mentioned, adding variables in a machine learning model might have various impacts. The study looks to test the same concept with the provided dataset, and the correlation among the various variables in the dataset might impact the prediction accuracy of the models during further analysis.

Some interesting insights from these are that while the cost-of-living index was the only highly correlated variable with the consumer price index used in the preliminary analysis, the supplemental dataset represents the basic price of domestic goods, which is also similarly correlated to the model. Moreover, the exchange rate between British Pound also has a high degree of correlation along with the import price index and producer price index.

Only two of the 14 independent variables are closely correlated with the consumer price index, with those being the cost-of-living index and the basic price of domestic goods, which are both above 0.9, with the cost-of-living index being more closely related among the two variables.

The producer price index, import price index, and exchange rate of British pounds were the other variables highly correlated with the consumer price index, with all these values being higher than 0.8 but lower than 0.9.

Unemployment and Business Confidence were the variables negatively correlated to the consumer price index and had an inverse relation with the variable.

Independent Variables	**Correlation with Consumer Price Index**
Yield Curve	0.083586
Production Index	0.299384
Housing Starts	0.206322
Cost of Living Index	**0.999982**
Unemployment	-0.410423
Business Confidence	-0.157732
Producer Price Index	***0.822723***
Export Price Index	0.380404
Import Price Index	***0.833128***
Basic Price Index for Domestic Goods	**0.936232**
USD	***0.089669***
JPY	0.078685
GBP	***0.803228***
CNY	-0.529123
OMX Helsinki	0.164265

Table 2 Correlation between variables

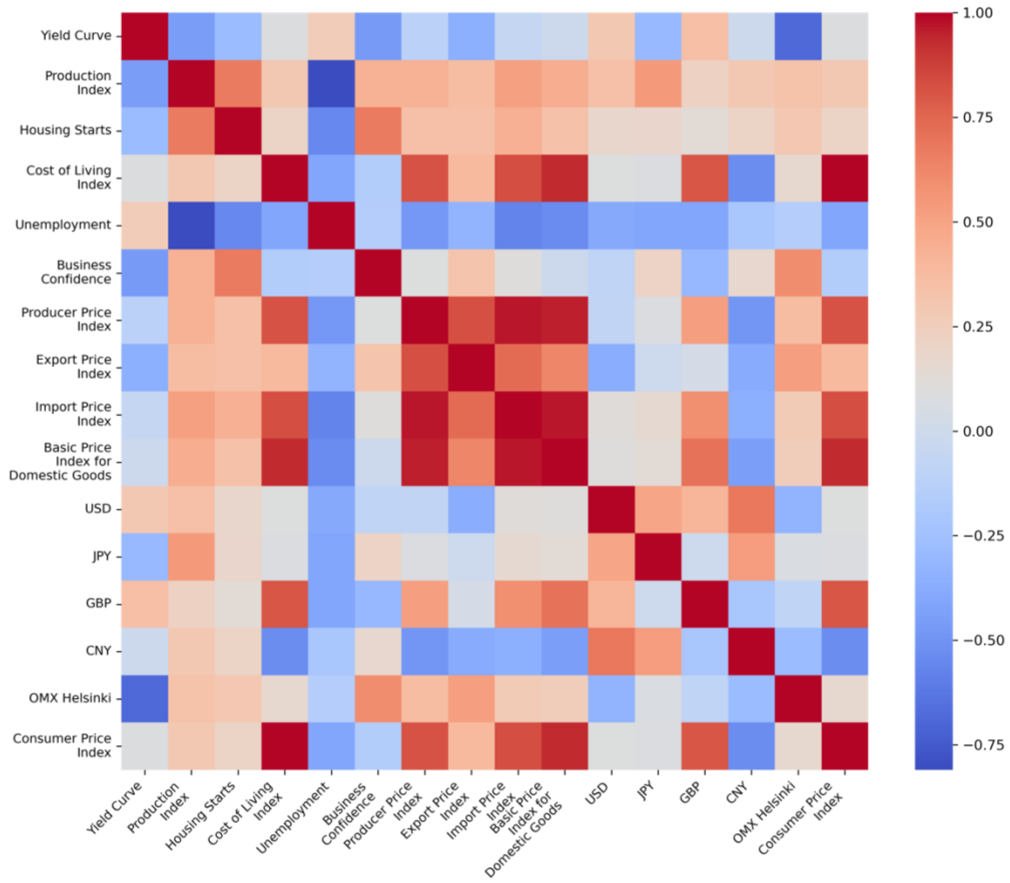

Figure 2 Correlation among variables in the Supplemental dataset

Due to the small dataset, it is quite clear that the cost-of-living index might represent a higher feature importance as opposed to another index. This feature importance is also represented in the table below.

Feature importance represents the independent variables that contribute most to the performance of a particular machine learning algorithm. With different machine learning algorithms, the impact of variables might be different from one algorithm to another, resulting in different levels of performance among the models.

When investigating the feature importance for variables selected for preliminary analysis, it is evident that for all of the models, the cost-of-living index represents the most valuable variable, while other variables do not have any considerable impact in models like decision tree, random forest, LASSO and SVR where the impact of the variable alone is above 90% which is representative of the correlation of the variables to the target variable.

The feature importance for XGBoost and MLP were mixed and provided more interesting insights. For example, while the cost-of-living index was still a large portion of the variable with 44.01% and 56.08%, respectively, these algorithms estimate the impact of other variables to provide the final results rather than a correlation-based analysis.

Variables / Algorithms	Decision Tree	Random Forest	LASSO	SVR	XGBoost	MLP
Yield Curve	0.13 %	0.10 %	0.00 %	0.03 %	19.64 %	20.81 %
Production Index	8.17 %	4.03 %	0.00 %	0.06 %	15.32 %	11.72 %
Housing Start	0.08 %	0.40 %	0.00 %	0.04 %	15.32 %	8.80 %
Cost of Living Index	91.60 %	95.36 %	100.00 %	99.80 %	44.01 %	56.08 %
Unemployment	0.01 %	0.10 %	0.00 %	0.07 %	5.71 %	2.59 %

Table 3Feature importance for variables of preliminary analysis in machine learning algorithms

The following table represents the feature importance when the supplemental analysis was performed. Some changes that are visible in the analysis are that except LASSO, where all the other variables were shrunk to zero, decision tree, random forest, and SVR are less impacted by the cost-of-living index in comparison to the preliminary analysis. All these methods have less than 90% feature importance for the cost-of-living index, and some indices, such as the import price index and the basic price index for domestic goods, are some variables that were able to show more impact on the target variable.

The most noteworthy observation was with the impact of the yield curve on XGBoost and MLP, where both these algorithms find the yield curve as the second most significant feature even though the value is not highly correlated to the target variables.

Variables / Algorithms	Decision Tree	Random Forest	LASSO	SVR	XGBoost	MLP
Yield Curve	0.01 %	0.05 %	0.00 %	0.16 %	15.47 %	7.80 %
Production Index	0.01 %	1.44 %	0.00 %	0.44 %	8.56 %	4.75 %
Housing Start	0.13 %	0.21 %	0.00 %	0.43 %	3.04 %	7.02 %
Cost of Living Index	82.14 %	87.46 %	100.00 %	67.45 %	32.32 %	40.36 %
Unemployment	0.00 %	0.05 %	0.00 %	0.40 %	4.01 %	6.13 %
Business Confidence	0.56 %	0.23 %	0.00 %	0.02 %	3.87 %	3.54 %
Producer Price Index	0.02 %	0.94 %	0.00 %	4.42 %	1.52 %	0.71 %
Export Price Index	0.00 %	1.74 %	0.00 %	5.39 %	1.93 %	1.27 %
Import Price Index	8.17 %	1.39 %	0.00 %	5.38 %	5.25 %	2.55 %
Basic Price Index for Domestic Goods	8.49 %	5.54 %	0.00 %	13.44 %	3.45 %	3.27 %
USD	0.01 %	0.07 %	0.00 %	1.54 %	4.28 %	2.98 %
JPY	0.06 %	0.33 %	0.00 %	0.44 %	5.11 %	4.00 %
GBP	0.00 %	0.20 %	0.00 %	0.29 %	4.83 %	4.00 %
CNY	0.39 %	0.09 %	0.00 %	0.09 %	2.07 %	7.18 %
OMX Helsinki	0.02 %	0.26 %	0.00 %	0.10 %	4.28 %	4.45 %

Table 4 Feature importance for variables of supplemental analysis in machine learning algorithms

4.2 Preliminary Analysis

The preliminary analysis will be performed using the five variables described in the previous chapter. The variables selected for this stage were the index of production, housing starts, unemployment, yield, and cost of living index to predict CPI. The following information was received using the six selected models and the given accuracy parameters.

Method	Initial dataset (Unoptimized)	Initial dataset (Optimized)	Change (%)
Decision Tree	**0.178285714**	0.178285714	0.000 %
Random Forest	0.297251371	0.255316921	14.107 %
LASSO	22.59192551	0.011009331	99.951 %
SVR	10.67622944	***0.007052807***	99.934 %
XG Boost	0.273810202	0.106795189	60.997 %
MLP Regressor	30.13603715	**0.005138005**	99.983 %

Table 5Mean Squared Error for Preliminary Dataset

Method	Initial dataset (Unoptimized)	Initial dataset (Optimized)	Change (%)
Decision Tree	**0.998901976**	0.998901976	0.000 %
Random Forest	0.998169292	0.998427557	0.026 %
LASSO	0.860861142	0.999932196	16.155 %
SVR	0.934247377	***0.999956563***	7.033 %
XG Boost	0.998314395	0.999342271	0.103 %
MLP Regressor	0.814398565	**0.999968356**	22.786 %

Table 6 R-Squared for Preliminary Dataset

Examining the unoptimized dataset where no parameters were adjusted and the default value for each method was used. It was evident that LASSO, SVR, and MLP, models that are commonly used in inflation forecasting, had the most significant errors, with all these models showing an underwhelming performance

Decision tree, which is one of the least sophisticated algorithms, performed the best with the least error at 0.1782 for mean squared error and an R-squared value of 0.9989.

Parameter tuning was done using the Grid Search CV tool from the sklearn library in Python for each machine-learning algorithm. The parameters were initially fed to the algorithm as is from the Grid Search results and checked for improvements from the unoptimized model. In a situation where the algorithm performed worse than the default value, manual tuning was performed until the models improved or the default value was used, as discussed in the previous sections.

The results from the decision tree were unchanged when the parameter tuning was performed. The best result was with the default value; changing the default values resulted in worse performance.

After tuning parameters and rejecting null values for parameters, the optimized random forest improved slightly from the unoptimized results. While the mean squared decreased from 0.2972 to 0.25531, the R-squared value increased from 0.9981 to 0.9984. As a result, there was a gain in performance for the random forest model by 14.11% in mean-squared error and 0.026% in R-Squared.

Similar to the optimized random forest model, other models also improved post optimization where the best was apparent in MLP, which improved from 0.8143 to 0.9999, resulting in a performance improvement of 22.786%. Furthermore, in their R-squared value, SVR and LASSO models also improved by 16.16% and 7.03%, respectively, while the improvement in MSE was from 22.59 to 0.011 for LASSO and 10.68 to 0.007 for SVR.

MLP and SVR were the best-performing indicators, with the mean-squared error of 0.0051 and 0.0071, respectively. While the performance of these algorithms unoptimized is mediocre, their performance after the necessary optimization improved vastly, with both these algorithms illustrating the least variance in their accuracy.

With the preliminary analysis, the best-performing models were MLP and SVR for inflation prediction using machine learning.

While none of the models can be outright rejected as unusable due to low mean-squared error and very high R-squared value. Since the MLP Regressor only had a 0.0051 mean-squared error, we can confirm that the difference between the test and predicted variables from the method is low at 0.005%. In contrast, with the high R-squared value of 0.999968, the model can predict 99.9968% of the variance between the five independent variables and the target variable of inflation (CPI).

From this preliminary analysis, it can be inferred that MLP, which is a neural network, performed the best for inflation forecasting using the dataset selected for this study.

4.3 Supplementary Analysis

The machine learning models presented a reasonable accuracy after the parameter tuning with the preliminary set of variables. Based on these results, it can be recommended that machine learning can be performed using only those datasets. However, some studies, such as Joseph et al. (2021), utilize more variables for similar studies. For example, Joseph et al. (2021) performed research with 167 variables. Although the current study is not as extensive, the study wants to examine how the accuracy changes when variables are added to the model. Furthermore, previous studies have discussed that dominant data can become dominated in a more extensive dataset when variables are added (Boivin & Ng, 2006). Therefore, this study will explore the abovementioned effect on the available dataset.

A supplementary analysis with additional variables needs to be conducted. For the supplementary analysis, the additional variables were selected based on the impact of the variable on demand-side inflation. Therefore, additional data were included, and the same models were trained and executed again to check the accuracy determined by mean-squared error and r-squared value in this regression analysis. In addition, the analysis was performed unoptimized, with the optimization parameter for the preliminary dataset only, and no grid search was performed to optimize the data for the new dataset further.

Algorithm	**Unoptimized**	**Optimized**	**Change%**
Decision Tree	0.355857143	0.355857143	0 %
Random Forest	**0.172804543**	0.174389192	-1 %
LASSO	22.52281835	***0.011035182***	100 %
SVR	19.68670332	0.029941331	100 %
XG Boost	0.415730046	0.094654209	77 %
MLP Regressor	68.83585284	**0.004281389**	100 %

Table 7 MSE for Supplemental Dataset

Algorithm	Unoptimized	Optimized	Change%
Decision Tree	0.997808352	0.997808352	0.000 %
Random Forest	**0.998935734**	0.998925974	-0.001 %
LASSO	0.861286758	***0.999932037***	16.097 %
SVR	0.878753787	0.999815598	13.777 %
XG Boost	0.997439867	0.999417045	0.198 %
MLP Regressor	0.576054642	**0.999973632**	73.590 %

Table 8 R-Squared for Supplemental Dataset

The most interesting observation from this dataset is that the best algorithm is still MLP; however, the second-best model has changed to LASSO rather than SVR. There will be a further comparison between these two results in the following section. Nevertheless, it was quite interesting that such results were visible for these models.

Investigating the unoptimized dataset, LASSO, SVR, and MLP were the worst-performing algorithms with very high MSE and low R-squared values. For example, while the MSE was over 10 for each algorithm, MLP performed the worst with 68.835. Similarly, the R-squared values were under 0.9, with MLP scoring 0.5761, while other models were at least above 0.85.

However, similar to the preliminary results, parameter optimization impacts these worse-performing algorithms highly, with all these models improving from the worst among the six algorithms to the best three algorithms, with MLP evolving from the worst-performing algorithm with the MSE of 68.835 and R-squared of 0.576 to the best-performing algorithms, with MSE of 0.004 and R-squared of 0.9999. Similarly, LASSO and SVR also improved vastly from MSE of 22.522 and 19.687 to 0.011 and 0.0299.

For the second time, optimized multi-layer perceptron performed the best while LASSO performed the second best; however, the variance between MLP and LASSO in the supplemental analysis is higher compared to the variance between MLP and SVR in the preliminary analysis.

4.4 Comparison of Optimized Algorithms

Optimized algorithms have performed better in both the preliminary and supplemental analyses of the two datasets. Therefore, an analysis is needed to check whether the improvement from preliminary to supplemental analysis has a significant degree of improvement.

Method	Preliminary	Supplemental	Change %
Decision Tree	0.17829	0.35586	-99.599 %
Random Forest	0.25532	0.17439	31.697 %
LASSO	0.01101	***0.01104***	-0.235 %
SVR	***0.00705***	0.02994	-324.531 %
XG Boost	0.10680	0.09465	11.368 %
MLP Regressor	**0.00514**	**0.00428**	16.672 %

Table 9MSE Comparison for Optimized Hyperparameters for preliminary and supplemental dataset

Method	Preliminary	Supplemental	Change %
Decision Tree	0.99890	0.99781	-0.109 %
Random Forest	0.99843	0.99893	0.050 %
LASSO	0.99993	***0.99993***	0.000 %
SVR	***0.99996***	0.99982	-0.014 %
XG Boost	0.99934	0.99942	0.007 %
MLP Regressor	**0.99997**	**0.99997**	0.001 %

Table 10 R-Squared Comparison for Optimized Hyperparameters for preliminary and supplemental dataset

When investigating the information, decision tree, SVR, and LASSO models have degraded when the supplemental analysis was performed with the additional dataset. At the same time, the random forest, XG Boost, and MLP regressor improved significantly with the additional dataset.

When investigating the MSE, SVR was the worst loser which has underperformed the preliminary dataset by 324%, while decision tree model has also underperformed by 99.59% from the preliminary analysis. Finally, while LASSO has underperformed in the supplemental dataset, it has only underperformed by 0.235%.

The numbers make sense more when investigating the R-squared value where only decision tree and SVR have underperformed while the performance for LASSO is unchanged. Decision tree has underperformed in R-squared value by just 0.109% while SVR performed worse by just 0.014. So, the performance based on R-squared is worse for decision trees when compared to MSE.

MSE checks the prediction accuracy of the algorithm while R-squared evaluates how closely the model can predict the regression line built by these algorithms. So, while the accuracy of the model declines highly when these new independent variables are added for the various models, the overall variance of the model slightly improves or does not decline as severely.

With the best-performing model of MLP, it can be deciphered that the addition of these variables does improve the overall accuracy of the model. While the other two better-performing models, as discussed in the previous sections, are not to the same degree.

4.5 Cross-Validation

All the analyses above were performed using a test set and train set, which splits the data using the *train_test_split* function in Scikit learn library in Python. The train and test set have only been split once for the analysis, and all the results are based on only that split of the dataset.

Cross-validation is an evaluation technique that creates a partition of the dataset into several folds of equal sizes. The evaluation is performed by storing one of the datasets as a test set, and the remaining data are used for training the data. Depending on the size of the fold, each fold will be evaluated once as a test set, and the average of the model is the performance of the model.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean MSE	Standard Deviation
Decision Tree	10.81464	16.89786	7.17600	4.50073	59.95309	19.86846	20.46871
Random Forest	8.49662	6.75061	11.26479	0.95284	64.19588	18.33215	23.17904
LASSO	0.03105	0.00342	0.00119	0.00285	0.01286	0.01028	0.01116
SVR	0.02129	0.00192	0.00121	0.00141	0.00192	0.00555	0.00787
XGBoost	7.35764	5.57528	4.85607	0.41213	62.39913	16.12005	23.25227
MLP	0.02076	0.08652	0.00130	0.00149	0.00128	0.02227	0.03299

Table 11Cross-Validation table for preliminary Analysis

The table above displays the mean-squared error received for each algorithm during the various folds. When observing the data, the average mean-squared error is high for the models like decision trees, random forest, and XGBoost. Most of the folds represent significantly high mean-squared errors for all these algorithms. In contrast, 'Fold 4' represents the best-performing model, which has similar results to the analysis made with the split data. According to the cross-validation, the best-performing model is support vector regression, which has performed better than any other model, and MLP, which was considered the best algorithm, has performed worse compared to SVR and LASSO.

After further addition of variables, the table below represents the cross-validation data for the supplemental analysis. Similar to preliminary analysis, MLP has performed worse than SVR and LASSO; however, the model is still better performing compared to a decision tree, random forest, and XGBoost.

The main issue with the MLP algorithm can be clearly visible in 'Fold 2', which performs much worse compared to other folds, which has changed the average result of the data severely. While the other four folds are in hundredth or thousandth decimal value, the value of the second fold is more than one exhibiting an overall worse algorithm.

The overall performance of the LASSO model is notably better compared to SVR, which changes the best model to use when more extensive variables are analyzed for multi-variate forecasting.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean MSE	Standard Deviation
Decision Tree	19.77268	17.02750	56.55436	25.63427	66.88800	37.17536	20.49572
Random Forest	14.76114	2.18564	10.28772	4.61850	65.37109	19.44482	23.37824
LASSO	0.05500	0.00344	0.00120	0.00339	0.01651	0.01591	0.02028
SVR	0.35226	0.86685	0.14021	0.02639	0.12226	0.30159	0.30200
XGBoost	8.10878	5.86686	4.95656	0.47631	61.57337	16.19638	22.82382
MLP	0.02291	5.13963	0.00256	0.00157	0.13903	1.06114	2.03988

Table 12 Cross-validation table for supplemental analysis

5 DISCUSSION

The study has evaluated various works of literature to discuss the past performance of machine learning algorithms and how these algorithms performed compared to the traditional econometric models used to perform similar predictions. Also, the study has collected the necessary data and ran the algorithms to evaluate which models have performed in what manner in the context of Finland and the data available for the country. This chapter will discuss all the various concepts in further detail.

5.1 Comparison to Previous Studies

During the literature review, various studies have been discussed regarding how machine learning models have performed compared to traditional econometric models for economic forecasting. Similarly, there were also discussions of how time-series models used to calculate inflation forecasting perform worse compared to machine learning algorithms. With multiple algorithms used in multiple studies, a few of these algorithms were selected to evaluate similar models in Finland.

Table 1 Machine Learning Algorithm in Literature represents all this information in more detail with the algorithms that have been used by each of these models as well as the country that these models were used. In addition, some information has been extracted from the table to represent how much each of these algorithms has been used in the selected studies and how well a model performed when it was selected.

Algorithm	**Times Used**	**Best Algorithm**	**Percentage**
Random Forest	Six (6)	Four (4)	66.67%
LASSO	Five (5)	Two (2)	40.00%
Support Vector	Seven (7)	One (1)	14.28%
XGBoost	Four (4)	Three (3)	75%
MLP	Four (4)	Two (2)	50%

Table 13 Frequency of Models used and best model selected

Table 9 illustrates five out of the six algorithms that have been used in this study. These algorithms were handpicked due to their frequency in similar multivariate studies. While there were only 11 studies evaluated, few studies have presented more than one model as the best-performing model. Evaluating 11 studies focused on economic and inflation forecasting, the decision tree was the only model which was not selected by the research more than once, and the model was also not considered the best model in the research it was evaluated. For the rest of the five models, support vector machine in the form of support vector regression was the most popular model used in the study; however, it was only considered the best-performing model in one study, which too was for a specific form of inflation that was evaluated in the study (Ülke et al., 2016).

The preliminary dataset used in the study does indicate that support vector regression was the best-performing algorithm among the non-neural network model. While the performance did deteriorate when more variables were added. An interesting observation is that the study by Ülke et al. (2016), that SVR performed the best in the previous study, was also performed with only six independent variables, like the five independent variables used in the current study. This observation might inform that the algorithm for SVR deteriorates when the addition of a newer variable dilutes the dominant variable in the dataset. The same observation was also made through cross-validation, where the SVR performed severely worse for the more extensive set of independent variables but was the best model with fewer independent variables.

The second most popular model evaluated in previous studies was random forest, which was also preferred as the best algorithm by the researchers of those studies due to the performance of the model. In terms of frequency in this limited set of studies, random forest is the most picked model; however, as a ratio, the model performed best every two times out of three when the algorithm is used for evaluation in a study.

As a percentage of being the preferred model when it is selected, XGBoost was considered the best algorithm in 75% of the studies that it was examined. However, the algorithm did not perform as well in the current study with the dataset that was collected for the study. No observable traits were available in the data as the countries of the study, variable size, and field of studies were different from each other, with no overlapping traits visible for the model.

Shrinkage models had been used in many past studies that were reviewed during the literature review, and LASSO was almost unanimously selected model for shrinkage algorithm study. It might be the same reason that LASSO was used five times in the 11 studies that are represented in the table. However, LASSO was considered the best model only twice out of the five studies the algorithm was utilized in.

Finally, MLP was the most common neural network algorithm due to its nature of being the most straightforward neural network. MLP has been utilized by studies in both general economic forecasting and inflation forecasting; however, the studies that have selected MLP have only found that the model performed best 50% of the time.

However, neural networks have been used multiple times across various studies, and neural network was the best model for a majority of the research that they were studied in. From the table, nine of the 11 studies have examined some form of a neural network, and five have mentioned that neural network models have performed better than other machine learning models. At the same time, more studies have compared time-series and traditional econometrics models to neural networks and have regarded neural networks to improve prediction accuracy significantly.

Compared to previous studies, while the overall prediction accuracy for the training and test data is best for a neural network model, on further analysis through cross-validation, the neural network model does not perform as well as the other two models of SVR and LASSO.

For the current study, while LASSO displayed a better accuracy compared to other algorithms examined in the study, all the variables other than the highly correlated variable were ignored and were pushed down to zero, which shows that the model does not consider other variables (Ranstam & Cook, 2018). The prediction is based on a single value that has displayed a significant correlation to the target variable.

Many studies believe that the non-linear nature of inflation makes machine learning algorithms perform better than traditional models and, in extension, neural network models.

5.2 Hyperparameter Optimization

Hyperparameter optimization has displayed a critical role in the analysis. The default value for all the models was evaluated for the unoptimized testing of the models. In contrast, the optimized testing and cross-validation of the models were performed based on the optimized parameter. The table below illustrates the various parameters and the value of the different optimization parameters that were performed using grid search.

Parameter optimization has been crucial in previous studies. As hyperparameters are built to evaluate models, and there have been numerous studies in hyperparameter tuning, machine learning researchers can perform better analyses. Similarly, with the help of a better hyperparameter and the ability to represent the hyperparameter in the table as represented below allows the study to be reproducible in the future and improves the reproducibility of the study (Hutter et al., 2019).

The method used to perform hyperparameter tuning in the current study is Grid Search. Grid Search is a basic method used for hyperparameter optimization; however, the method is usually time-consuming, and the dataset in this study also required continuous parameters, but the limitation of grid search requires that the continuous parameter is predefined (Chavez-Hurtado & Cortes-Fregoso, 2013; Yang & Shami, 2020). In the current study, the decision tree required two continuous parameters of *max_depth* and *min_sample_leaf,* which have been selected depending on the researcher's discretion; however, there are only a few values applied, and manual tuning was required to perform the analysis to check either some values above or below the best parameter were able to improve the performance of the algorithm.

Similar issues were faced with random forests where in addition to *max_depth* and *min_sample_leaf*, the *n_estimator* parameter is another continuous parameter that had to be predefined and did not provide the best result for the models.

Similarly, the value of *alpha* in LASSO algorithms could be better optimized when a continuous optimization parameter could be used as well as the number of iterations (*max_iter*) that the models need to shrink all the irrelevant values to zero also can perform better when a continuous variable has been used. Similar problems are faced by the values for *C*, and *epsilon* in SVR, *learning_rate*, *max_depth,* and *min_child_weight* in XGBoost.

The most complication for hyperparameter optimization is encountered with MLP. A neural network is extremely dependent on the input and output layers in the data. A significant issue for a researcher during hyperparameter optimization for neural networks would be to find the correct number of hidden layers (*hidden_layer*) for grid search. One major issue with grid search in tuning for hidden layers can be the overfitting and underfitting problem, where the value provided as the optimal parameter could come from an overfitted model. When defining the grid search, a discrete set of hyperparameter were provided for all continuous data so that it could be performed with a modest number of parameters tuned for each of the machine learning algorithms that can improve the performance of the algorithm from the unoptimized algorithms.

Algorithm	Library	Evaluated Parameter	Best Parameter (Grid Search)	Best Performing Parameter (After Manual Tuning)
Decision Tree	Scikit Learn	max_depth: [2, 4, 6, 8, 10, 11], min_samples_leaf: [1, 2, 5, 10, 15, 20, 25], criterion: ['mse', 'friedman_mse'], splitter: ['best', 'random']	max_depth = 10, min_sample_leaf=1, criterion = 'MSE', splitter = 'random'	max_depth = None, min_sample_leaf=1, criterion = 'MSE', splitter = 'best'
Random Forest	Scikit Learn	max_depth = [1, 5, 10, 15], min_samples_leaf = [1, 5, 10, 15, 20], n_estimators = [100, 200, 300], max_features= ['auto', 'sqrt', 'log2']	n_estimators=100, max_depth=15, min_samples_leaf=1, max_features='auto'	n_estimators=200, max_depth=10, min_samples_leaf=1, max_features='auto'
LASSO	Scikit Learn	alpha: [0.01, 0.1, 0.5, 1.0, 2.0, 5.0], fit_intercept: [True, False], normalize: [True, False], max_iter: [1000, 5000, 10000]	alpha =0.01, fit_intercept=False, normalize=True, max_iter=5000	alpha = 0.01, fit_intercept=True, normalize=False, max_iter=1000
SVR	Scikit Learn	kernel: ['linear', 'poly', 'rbf', 'sigmoid'], C: [0.1, 1, 2, 3, 4, 5, 10], epsilon: [0.01, 0.1, 1]	kernel = 'linear', C = 10, epsilon=0.01	kernel = 'linear', C = 10, epsilon=0.01

Algorithm	Library	Evaluated Parameter	Best Parameter (Grid Search)	Best Performing Parameter (After Manual Tuning)
XGBoost	XGBoost	learning_rate: [0.005, 0.0005, 0.05, 0.1, 0.15], max_depth: [1, 2, 3, 4, 5], min_child_weight: [1, 2, 3, 4, 5]	learning_rate=0.1, max_depth=5, min_child_weight=1	learning_rate=0.1, max_depth=5, min_child_weight=1
MLP	Scikit Learn	hidden_layer_sizes: [(10,), (50,), (100,), (10,10), (50,50), (100,100)], activation: [‘logistic’, ‘tanh’, ‘relu’], solver: [‘lbfgs’, ‘adam’], learning_rate: [‘constant’, ‘adaptive’], max_iter: [1000, 2000, 3000, 4000, 5000]	max_iter=4000, hidden_layer=(50, 50), activation=relu, solver=lbfgs, learning_rate=constant, alpha=0.01	max_iter=4000, hidden_layer=(50, 50), activation=relu, solver=lbfgs, learning_rate=constant, alpha=0.01

Table 14 Hyperparameter Optimization

5.3 Best Algorithm for the Finnish Economy

Two different datasets were evaluated when evaluating the best algorithm for the Finnish economy. Both data sets concluded that multi-layer perceptron, a neural network, performed the best for the available dataset when evaluated for monthly data.

Multi-layer perceptron, which was also the best-performing algorithm in Ahmed et al. (2010) for their study in Egypt related to inflation forecasting, and Shobana and Umamaheshwari (2021) for their study in India related to economic forecasting concludes that MLP is a neural network that provides a high degree of prediction accuracy. The popularity of neural networks in studies such as Rodriguez-Vargas (2020), Yang and Guo (2021), and Araujo and Gaglianone (2022) also concluded that the prediction accuracy of the neural network was better compared to other machine learning models. These algorithms were compared to other machine learning models such as random forest, XG Boost, SVR, Gaussian Process, LASSO, Ridge, ElasticNet, and Logistic Regression, among many other studies.

In addition to MLP, the prediction accuracy of SVR was also relatively close to the score provided by MLP during the preliminary analysis. For example, SVR displayed an MSE of 0.0071 to the MSE of MLP of 0.0051.

However, on performing cross-validation analysis, the best model for the preliminary analysis was SVR, and the best model for supplemental analysis was LASSO. Therefore, further studies need to be conducted to understand the overall impact and prediction accuracy using cross-validation and the prediction capability of the algorithms through a comparison of actual and predicted values.

Furthermore, this study can be considered a pilot study for the Finnish economy to evaluate machine learning models to forecast inflation in Finland. However, with a larger dataset and a more significant number of machine learning models, as evaluated by Araujo and Gaglianone (2022) for the Brazilian economy, a similar study needs to be conducted in Finland so that a larger dataset with a larger time horizon could be investigated and the best model for each situation identified.

According to the discussion in the previous chapters, the Nowcasting system (Fornaro & Luomaranta, 2019) that forecasts GDP growth in Finland was a starting point for this study to discover the various variables that need to be evaluated for the study related to inflation forecasting. A robust system that can understand inflation forecasting is necessary. It could already be used internally but not exhibited similarly due to the risk of inflation expectation which can be another variable that impacts inflation (Carlson & Parkin, 1975).

5.4 Impact of Additional Variable

When evaluating the impact of additional variables on the accuracy of the machine learning algorithm, no significant changes were visible in the accuracy of these models. However, a grid search was performed when the variables were re-evaluated in the supplemental dataset. Except for XGBoost, all other models performed worse when the grid search results were applied to each algorithm.

The hyperparameters used in the preliminary analysis were reused for the supplemental analysis. In all situations, except XGBoost, the hyperparameter used for preliminary analysis performed better than the hyperparameter provided by Grid Search for supplementary analysis.

With minor improvements with the additional dataset, it cannot be conclusively decided that adding further variables improves the prediction accuracy of the model. However, many studies use many more independent variables compared to the current study; there are requirements for additional studies with all the various economic variables.

Additionally, the cross-validation revealed a better picture of the impact of additional variables on prediction accuracy. For all the algorithms evaluated during the study, each had a worse average mean-squared error when the number of independent variables increased from five to 14 for the study.

The table below represents the average MSE for the cross-validation dataset and illustrates the worse prediction accuracy for each algorithm when variables were added. The train and test set utilized during the primary analysis of each algorithm had a few improvements, such as the random forest and LASSO algorithm for the unoptimized models and random forest, XGBoost, and MLP Regressor for the optimized set of variables. However, the cross-validation does not produce a similar result. Each algorithm performed worse for the more extensive set of independent variables compared to the initial analysis that utilized fewer independent variables to reach a conclusion.

Algorithms	Preliminary	Supplemental
Decision Tree	19.86846	37.17536
Random Forest	18.33215	19.44482
LASSO	0.01028	0.01591
SVR	0.00555	0.30159
XGBoost	16.12005	16.19638
MLP	0.02227	1.06114

Table 15Average mean-squared error for cross-validation

Even though the addition of the variables has resulted in a negligible improvement for the primary analysis and worse performance in cross-validation, there needs to be a realization that the more extensive variables were not as diverse as they were in previous studies, so further analysis needs to be made with a more diverse set of variables.

5.5 Inflation related independent variables

Many of the previous studies evaluated inflation-related variables in inflation forecasting. For example, studies such as Medeiros et al. (2021) studied their data through a FRED dataset, including data for inflation such as the consumer price index for various goods and services such as apparel, transportation, medical care, commodities, durables, services, All items less food, all item less shelter, all items less medical care. Similarly, the study also used a different inflation indicator known as personal consumption expenditure: chain index, durable goods, nondurable goods, and services for inflation forecasting. While no correlation exists between these variables and the target variable revealed in the study, these variables have been used for making forecasts.

Similarly, Araujo and Gaglianone (2022) studied 167 different variables when studying various forecasting models for inflation, and on their dataset, 21 of the independent variables were under the inflation category.

Furthermore, studies such as Ülke et al. (2016) include the real personal consumption expenditure as an independent variable which is another inflation-related variable to forecast the accuracy of inflation prediction in their study.

While the models discussed above are the ones that have used explicit datasets that are related to inflation in their independent variables, there are other studies such as Ahmed et al. (2010); Joseph et al. (2021); and Özgür & Akkoç (2021) where other variables that have an implicit impact on inflation are included which are also highly correlated to the values in the models.

The cost-of-living index and consumer price index are both inflation-related indicators. Therefore, when navigating the website for Statistics Finland, both these indices are stored in the heading of inflation. However, there are differences between these two variables regarding what they consider during the calculation.

The primary function of the cost-of-living index is to minimize the cost of achieving a level of utility or satisfaction and then maximize the marginal utility within their budget (Diewert, 1990). Therefore, each household or individual preferences are under consideration, where they will work on maximizing their living standard under budget constraints. However, the consumer price index is the measurement change of average prices paid by urban consumers for a basket of goods and services (Abraham, 2003). The central bank determines the basket of goods and services and updates it every few years depending on the country and its policy.

Due to the difference in how they are calculated, the cost-of-living and consumer price indexes should show inflation differently. While the general assumption for economists is that the consumer price index and cost-of-living index are different forms of inflation index (Triplett, 2001). However, the values are highly correlated to each other when analyzed for Finland. According to the Official Statistics of Finland (OSF) (n.d.), the consumer price index is defined as the development in prices of products and services purchased by households in Finland and is considered the general measure of inflation. However, in Finland, the cost-of-living index is calculated based on the consumer price index with the base year of 1951:10=100, which is a long-term time series that chains the latest consumer price index. Therefore, the only difference in the cost-of-living index is adjustments of rental agreements on dwellings, business premises, and land, which are tied to the cost-of-living index.

When inflation-related variables are used in inflation prediction, there are various chances of a high degree of correlation between these values, which can result in better accuracy for the model.

6 IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH

6.1 Policy Implication

As previously discussed in the study, the Nowcasting system has been used in Finland to forecast GDP growth, with the Bank of Finland making the data publicly available. While an inflation forecasting system might already exist, it is not publicly available. While it might not be reasonable to provide the information publicly as the Bank of Finland does with GDP growth. It might be necessary to publish public information regarding how this information is being calculated by the Bank of Finland.

Inflation forecasting has been of interest to many economists as well as the general population. This kind of understanding will allow researchers to study the models and help in expanding the models. While the Bank of Finland does not need to mention the exact algorithm they are using for this inflation forecasting-related study, it might be a good practice to provide the information on the data being used at the Bank of Finland to make such predictions.

With the information on the data, it would be insightful for the researchers to evaluate the various models and recommend them to the Bank of Finland and the Finnish government. As discussed in the first chapter, understanding inflation allows for better decision-making for entities such as the Bank of Finland as well as the Ministry of Finance in Finland. With further research being conducted in this field by researchers in various universities and even foreign economists, Finland might be able to control their inflation which would change by the value of one in a whole year in the past few years, with 2022 January only recording 108.8 since a base 99.5 in January 2015 to a base point of 118.0 in January 2023. The inflation from January 2022 to January 2023 has been higher than from January 2015 through January 2022.

6.2 Limitations of the Study and Future Research

There were various limitations that the study faced during the whole process, and many further analyses could be performed related to the same topic. First, the dataset available was collected from publicly available sources; since the data had to be extracted from various sources and some of the data was used as alternative variables, it was quite tricky during the data collection, and the analysis might have increased errors in the model. Also, some of the data were only available in annualized form, and monthly data is usually required to train machine learning models. However, some entities might have collected all the data, primarily entities like the Statistics Department and Central Bank, who could perform a more detailed analysis.

Similarly, as discussed in the previous section, the number of variables had to be limited for the analysis due to the availability of data. Therefore, a future study can look to replicate the studies that used almost all macroeconomic data for multivariate inflation forecasting.

Additionally, the study also displayed that splitting the dataset into a training set and test set, which is used for scoring the prediction accuracy of the algorithms, illustrated a specific result. However, cross-validation of these algorithms presented a different environment where MLP, considered the best model in the training and test split, performed worse in cross-validation. Therefore, future research needs to be conducted to study the phenomenon where cross-validation is rigorously tested for the prediction accuracy of the models.

Similarly, Grid Search was the method that was performed for the optimization of parameter, and the issues with the optimization method has been discussed in the previous chapter. Furthermore, methods such as random search, gradient-based algorithm, Bayesian optimization, and metaheuristic algorithms are other methods for hyperparameter optimization. Therefore, these various forms of hyperparameter optimization could be conducted in future studies for better prediction accuracy.

Furthermore, the current study evaluated multivariate prediction; however, there are machine learning models that are focused on univariate machine learning models, such as ARIMA and SARIMA, which can be tested to perform inflation prediction. These models might illustrate a model which can quickly predict inflation. However, there will be a need to understand the prediction accuracy with the application of these models.

Furthermore, due to the knowledge of the researcher as well as the limitation of time due to the need to graduate, only six machine learning algorithms were utilized in performing the analysis; however, just observing the literature review, there are a lot more algorithms such as Ridge, ElasticNet, LSTM, AdaBoost, Bagging, AdaLASSO, RNN that can be evaluated in future research.

Additionally, the current study followed the studies discussed in the literature, where many of these previous studies contain highly correlated variables employed during the evaluation of prediction accuracy for the machine learning models. Nevertheless, these algorithms show a high degree of prediction accuracy when employed. However, since the current study attempted to examine the same concepts, there were some issues when the cost-of-living index and consumer price index were highly related, even though they are supposed to calculate different forms of inflation. Therefore, a comprehensive study that excludes such inflation indicators must be explored so that future research does not face similar issues. Furthermore, a study needs to be performed where these variables, which are significantly correlated and that cannot be extracted beforehand, need to be excluded from measuring the accuracy of the algorithm.

One remarkable research that can be conducted is through the evaluation of fiscal and monetary policies. Natural language processing can be performed on the fiscal and monetary policies for a ten-year period, where various economic indicators are evaluated depending on the fiscal policy and monetary policy. Economic indicators such as GDP growth, unemployment, and inflation can be evaluated and checked whether these monetary policies and fiscal policies significantly impact these indicators.

Furthermore, for entities like the Finnish government and the central bank, a study can be conducted to evaluate whether a system equivalent to Nowcasting can be created for inflation forecasting for better decision-making in these economies.

7 CONCLUSION

Scholars and researchers have pursued the idea of perfect economic and inflation forecasting for an extended period. Each iteration of mathematical and statistical models has been evaluated to reach this perfect accuracy. However, with the growth in machine learning, economic prediction has found a new understanding, and more researchers are focused on utilizing these newfound models in forecasting inflation. While many studies have been conducted in the field, a large amount of research is focused on the superpowers of the global economy, such as the UK, the USA, China, and Japan. However, every country around the world is yearning to identify a model that can predict their economic indicators rigorously.

The study was conducted to answer three research questions related to using machine learning in inflation forecasting through previous literature, the best machine learning algorithm to predict inflation forecasting in Finland through a multivariate approach, and the impact of additional data in multivariate analysis.

The first research question that evaluates the use of machine learning in inflation forecasting can be concluded with the various previous studies that have marked that machine learning algorithms have outperformed traditional econometrics models in the accuracy of prediction of economic indicators as concluded by Belly et al. (2023), Özgür & Akkoç (2021), Rodríguez-Vargas (2020), Shobana & Umamaheswari (2021), Timmermann (2008), Ülke et al. (2016), and Yang and Guo (2021). This illustrates that machine learning algorithms can provide better accuracy in inflation forecasting, which corroborates the need to evaluate machine learning models for inflation forecasting.

The second research question examines a set of machine learning algorithms utilized in inflation forecasting. The machine learning algorithms were selected based on the frequency of use in previous works of literature, and the analysis was performed on two different multivariate datasets of different amounts of independent variables. Through the initial analysis, MLP displayed the best prediction accuracy. However, cross-validation displayed a different result where SVR and LASSO models performed better for the two datasets.

Finally, the third research question explores the impact of a more extensive number of variables in multivariate analysis. With various multivariate studies in machine learning and inflation forecasting where the number of independent variables varied from singledigit numbers to hundreds, it was necessary to understand and evaluate the impact of larger datasets in inflation forecasting through machine learning models. Previous studies such as Boivin and Ng (2006) mention the concept of variable dilution, where a variable that impacts a machine learning model gets diluted with a larger dataset leading to a worse-performing model when additional variables are provided to the machine learning model. A similar trend is observed in the current study where few models, such as decision tree, random forest, SVR performed worse for larger datasets. Similarly, all the models performed worse for cross-validation, which represents that the larger dataset might not be able to improve the accuracy of machine learning models.

While the study focused on the research question mentioned above, a few more impacts were observed during the study, which were not explicitly questioned but were answered during the study.

Hyperparameter optimization illustrated that machine learning models tend to underperform when the hyperparameters are not correctly tuned. With these optimizations, there were large degrees of improvement in the prediction accuracy of the model. However, there were limitations to the hyperparameter optimization method employed during this research. Therefore, a future study can be conducted where a better hyperparameter optimization method is selected.

Similarly, various previous studies have used an inflation-related variable as an independent variable in inflation forecasting. Due to the high degree of correlation among these variables, while the prediction accuracy of these models has improved, there is a further need to study the inclusion of such variables that can boost the accuracy and whether such variables would be available to researchers and decision-makers when making future predictions. The nowcasting model used for GDP growth does not use another GDP indicator in creating a prediction model. Similarly, there is a need to understand whether a forecasting model for inflation can also operate similarly.

Furthermore, due to inexperience and time limitations, there are also ways to improve these studies. The study can be conducted using a more extensive set of independent variables. Also, more machine learning algorithms and deep learning algorithms could be evaluated in the study to expand the horizon for these studies. A study evaluating the impact of monetary policy and fiscal policy could be conducted where various economic indicators are evaluated with natural language processing and other machine learning algorithms.

One of the most significant limitations faced during this research was the availability of the data. While various data were available related to economic indicators for the Finnish economy, some data were available only in annualized form, while all the data that were used during the study were monthly. Therefore, some entity that has access to these data could perform a better analysis depending on the availability of the data. Furthermore, some of the variables used in the study are alternate variables that differ from generally used variables and might display a lower degree of correlation between the variables resulting in a worse-performing model.

REFERENCES

Abraham, K. G. (2003). Toward a Cost-of-Living Index: Progress and Prospects. *Journal of Economic Perspectives*, *17*(1), 45–58. https://doi.org/10.1257/089533003321164949

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*, *29*(5-6), 594–621. https://doi.org/10.1080/07474938.2010.481556

Arakelian, V., Dellaportas, P., Savona, R., & Vezzoli, M. (2019). Sovereign risk zones in Europe during and after the debt crisis. *Quantitative Finance*, *19*(6), 961–980. https://doi.org/10.1080/14697688.2018.1562197

Aras, S., & Lisboa, P. J. G. (2022). Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*, *207*, 117982. https://doi.org/10.1016/j.eswa.2022.117982

Araujo, G. S., & Gaglianone, W. P. (2022). *Machine Learning Methods for Inflation Forecasting in Brazil: new contenders versus classical models*. Banco Central Do Brasil.

Awad, M., & Khanna, R. (2015). Support Vector Regression. *Efficient Learning Machines*, 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4

Balduzzi, P., Savona, R., & Alessi, L. (2022). Anatomy of a Sovereign Debt Crisis: Machine Learning, Real-Time Macro Fundamentals, and CDS Spreads. *Journal of Financial Econometrics*. https://doi.org/10.1093/jjfinec/nbac021

Belly, G., Boeckelmann, L., Mateo, C., Di Iorio, A., Istrefi, K., Siakoulis, V., & Stalla-Bourdillon, A. (2023). Forecasting sovereign risk in the Euro area via machine learning. *Journal of Forecasting*, *42*(3), 657–684. https://doi.org/10.1002/for.2938

Binner, J. M., Bissoondeeal, R. K., Elger, T., Gazely, A. M., & Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia. *Applied Economics*, *37*, 6. https://doi.org/10.1080/0003684052000343679

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*(1), 169–194. https://doi.org/10.1016/j.jeconom.2005.01.027

Botchkarev, A. (2019). A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 045–076. https://doi.org/10.28945/4184

Burgess, S., Corugedo, E. F., Groth, C., Harrison, R., Monti, F., Theodoridis, K., & Waldron, M. (2013). The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of Models. In *Bank of England*. https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2013/the-boes-forecasting-platform-compass-maps-ease-and-the-suite-of-models.pdf?la=en&hash=C0D385B7FA637B9E0F96AC50B0C2ADEA01206927

Carlson, J. A., & Parkin, M. (1975). Inflation Expectations. *Economica*, *42*(166), 123. https://doi.org/10.2307/2553588

Castellani, M., & Santos, E. (2006). Forecasting Long-Term Government Bond Yields: An Application of Statistical and AI Models. *Working Papers Department of Economics*. https://ideas.repec.org/p/ise/isegwp/wp42006.html

Chakraborty, C., & Joseph, A. (2017). *Machine Learning at Central Banks*. Bank of England.

Chavez-Hurtado, J. L., & Cortes-Fregoso, J. H. (2013). Forecasting Mexican inflation using neural networks. *CONIELECOMP 2013, 23rd International Conference on Electronics, Communications and Computing*.

Claesen, M., Simm, J., Popovic, D., Moreau, Y., & De Moor, B. (2014). Easy Hyperparameter Search Using Optunity. *ArXiv:1412.1114 [Cs]*. https://arxiv.org/abs/1412.1114

Colin Cameron, A., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, *77*(2), 329–342. https://doi.org/10.1016/S0304-4076(96)01818-0

Colin, A. (2006). Fixed income attribution with minimum raw material. *Journal of Performance Management*, *11*, 2. https://eprints.qut.edu.au/225277/

Diewert, W. E. (1990, January 1). *The Theory of the Cost-of-Living Index and the Measurement of Welfare Change* (W. E. Diewert, Ed.). ScienceDirect; Elsevier. https://www.sciencedirect.com/science/article/pii/B978044488108350007X

Federal Reserve Bank. (2023, March 22). *March 22, 2023: FOMC Projections materials, accessible version*. Www.federalreserve.gov. https://www.federalreserve.gov/monetarypolicy/fomcprojtabl20230322.htm

Fornaro, P., & Luomaranta, H. (2019). Nowcasting Finnish real economic activity: a machine learning approach. *Empirical Economics*, *58*(1), 55–71. https://doi.org/10.1007/s00181-019-01809-y

Gordon, R. J. (1997) The time-varying NAIRU and its implications for economic policy. J. Econ. Perspect., 11, 11-32.

Gunduz, S., Yildirim, S., & Durukan, M. B. (2020). An Investigation of the Factors Affecting Inflation Perceptions: A Case Study on Business and Economics Undergraduate Students. *Sosyoekonomi*, 245–263. https://doi.org/10.17233/sosyoekonomi.2020.03.14

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The Model Confidence Set. *Econometrica*, *79*(2), 453–497. https://doi.org/10.3982/ecta5771

Hassani, H., Soofi, A. S., & Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society. Series a (Statistics in Society)*, *176*(3), 743–760. https://www.jstor.org/stable/43965659

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). *Automatic Machine Learning: methods, systems, challenges.* Springer.

Ivan, B. (2018). Inflation Forecasting Using Machine Learning Methods. *Russian Journal of Money and Finance*, *77*(4), 42–59. https://doi.org/10.31477/rjmf.201804.42

J. Binner, R. Bissoondeeal and T. Elger, “A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro divisia”, Applied Economics, vol. 37, No. 6, pp. 665-680, 2005

Joseph, A. (2019). Parametric inference with universal function approximators. In *Bank of England*. Bank of England. https://www.bankofengland.co.uk/working-paper/2019/shapley-regressions-a-framework-for-statistical-inference-on-machine-learning-models

Joseph, A., Kalamara, E., Kapetanios, G., & Potjagailo, G. (2021). Forecasting UK inflation bottom up. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2021.03.005

Kim, W. J., Jung, G., & Choi, S.-Y. (2020). Forecasting CDS Term Structure Based on Nelson–Siegel Model and Machine Learning. *Complexity*, *2020*, 1–23. https://doi.org/10.1155/2020/2518283

Li, Y.-S., Pai, P.-F., & Lin, Y.-L. (2022). Forecasting inflation rates be extreme gradient boosting with the genetic algorithm. *Journal of Ambient Intelligence and Humanized Computing*, *14*(3), 2211–2220. https://doi.org/10.1007/s12652-022-04479-4

Liu, Y., Yang, D., & Zhao, Y. (2022). *Housing Boom and Headline Inflation: Insights from Machine Learning*. International Monetary Fund.

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2019). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, *2*(5-6), 183–197. https://doi.org/10.1016/0925-2312(91)90023-5

Official Statistics of Finland (OSF). (n.d.). *Statistics Finland - Statistics by topic - Consumer price index*. Www.stat.fi. Retrieved May 9, 2023, from https://www.stat.fi/til/khi/meta_en.html

Oner, C. (2022). *Inflation: Prices on the Rise*. International Monetary Fund. https://www.imf.org/external/pubs/ft/fandd/basics/30-inflation.htm

Özgür, Ö., & Akkoç, U. (2021). Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms. *International Journal of Emerging Markets*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ijoem-05-2020-0577

Pathak, S., Mishra, I., & Swetapadma, A. (2018, November 1). *An Assessment of Decision Tree based Classification and Regression Algorithms*. IEEE Xplore. https://doi.org/10.1109/ICICT43934.2018.9034296

Phillips, A. W. (1958) The relationship between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. Economics, 25, 283-299.

Ranstam, J., & Cook, J. A. (2018). LASSO regression. *British Journal of Surgery*, *105*(10), 1348–1348. https://doi.org/10.1002/bjs.10895

Rodríguez-Vargas, A. (2020). Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking*, *1*(1-4), 100012. https://doi.org/10.1016/j.latcb.2020.100012

Samuelson, P. and Solow, R. (1960) Analytical aspects of anti-inflation policy. Am. Econ. Rev., 50, 177-194.

Sheridan, R. P., Wang, W. M., Liaw, A., Ma, J., & Gifford, E. M. (2016). Extreme Gradient Boosting as a Method for Quantitative Structure–Activity Relationships. *Journal of Chemical Information and Modeling*, *56*(12), 2353–2360. https://doi.org/10.1021/acs.jcim.6b00591

Shobana, G., & Umamaheswari, K. (2021). Forecasting by Machine Learning Techniques and Econometrics: A Review. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. https://doi.org/10.1109/icict50816.2021.9358514

Stock, James H. and Watson, Mark W., Forecasting Inflation (March 1999). NBER Working Paper No. w7023, Available at SSRN: https://ssrn.com/abstract=155850

Stock, J. H., & Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Tattikota, S. R., & Srinivasan, N. (2021). *Integration of Econometric Models and Machine Learning - Study on US Inflation and Unemployment* (p. Madras School of Economics) [Working Paper]. https://www.mse.ac.in/wp-content/uploads/2021/10/Working-Paper-206.pdf

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, *24*(1), 1–18. https://doi.org/10.1016/j.ijforecast.2007.07.008

Triplett, J. E. (2001). Should the Cost-of-living Index Provide the Conceptual Framework for a Consumer Price Index? *The Economic Journal*, *111*(472), 311–334. https://doi.org/10.1111/1468-0297.00633

Ülke, V., Sahin, A., & Subasi, A. (2016). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing and Applications*, *30*(5), 1519–1527. https://doi.org/10.1007/s00521-016-2766-x

Xie, H., Zhang, M., & Andreae, P. (2007). Genetic Programming for New Zealand CPI Inflation Prediction. *2007 IEEE Congress on Evolutionary Computation*. https://doi.org/10.1109/cec.2007.4424790

Yang, C., & Guo, S. (2021). Inflation Prediction Method Based on Deep Learning. *Computational Intelligence and Neuroscience*, *2021*, e1071145. https://doi.org/10.1155/2021/1071145

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

APPENDIX

Appendix I

Table representing the models evaluated by Araujo and Gaglianone (2022):

1	Random Walk	26	Hybrid Random Forest – Adalasso
2	Random Walk (Atkeson-Ohanian)	27	Hybrid Random Forest – XGBoost
3	ARMA	28	Inflation Expectations (Breakeven)
4	VAR	29	Inflation Expectations (Focus Survey)
5	Phillips Curve (Backward)	30	Combination 1 (Mean)
6	Phillips Curve (Hybrid)	31	Combination 1 (Median)
7	Factor Model 1	32	Combination 1 (Granger-Ramanathan)
8	Factor Model 2	33	Combination 1 (Constrained Least Squares)
9	Factor Model 3	34	Combination 1 (Complete Subset Regression)
10	Factor Model 4	35	Combination 1 (Adalasso)
11	Elastic Net	36	Combination 1 (Random Forest)
12	LASSO	37	Combination 2 (Mean)
13	Adaptive LASSO (Adalasso)	38	Combination 2 (Median)
14	Ridge Regression	39	Combination 2 (Granger-Ramanathan)
15	Random Forest	40	Combination 2 (Constrained Least Squares)
16	Quantile Random Forest	41	Combination 2 (Complete Subset Regression)
17	XGBoost	42	Combination 2 (Adalasso)

18	Recurrent Neural Network (RNN)	43	Combination 2 (Random Forest)
19	Disaggregated Inflation (ARMA)	44	Combination 3 (Mean)
20	Disaggregated Inflation (Adalasso)	45	Combination 3 (Median)
21	Disaggregated Inflation (Random Forest)	46	Combination 3 (Granger-Ramanathan)
22	Hybrid Adalasso-OLS	47	Combination 3 (Constrained Least Squares)
23	Hybrid Adalasso – Random Forest	48	Combination 3 (Complete Subset Regression)
24	Hybrid Adalass – XGBoost	49	Combination 3 (Adalasso)
25	Hybrid Random Forest – OLS	50	Combination 3 (Random Forest)

Notes: Combination 1 is based on models 1-27. Combination 2 and 3 are based on the superior models of the *model confidence set* of Hansen et al. (2011), considering model 1-27 or 1-29, respectively.