Sebastian Nyholm

MACHINE LEARNING IN MACROECONOMIC FORECASTING

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2022

TIIVISTELMÄ

Nyholm, Sebastian
Koneoppiminen makrotalouden ennustajana
Jyväskylä: Jyväskylän yliopisto, 2022, **32 s.**
Tietojärjestelmätiede, kandidaatintutkielma
Ohjaaja(t): Kuusio, Ari

Dataa on aina ollut saatavilla paljon taloudesta, mutta sen kaiken käyttäminen talouden ennustamisessa on ollut hankalaa. Perinteiset ennustamisen ja arvioinnin mallit eivät ole osoittautuneet olevan kovin tarkkoja makrotalouden ennustamisessa. Modernit koneoppimisen menetelmät ovat osoittautuneet hyviksi monessa eri tilanteessa ja monella eri alalla. Koneoppiminen on vahvimmillaan juuri ennusteiden tekemisessä. Taloutta on aina pyritty ennustamaan ekonometrisillä malleilla, mutta koneoppimisen on huomattu monessa paikassa olevan tarkempi ennusteissaan kuin perinteisemmät mallit. Koneoppimista voidaan käyttää työkaluna ennustamisessa monien eri metodien ja algoritmien kautta, joilla kaikilla on omat vahvuutensa sekä heikkoutensa. Jokaista näistä voidaan käyttää erilaisten ennusteiden tekemisessä juuri niiden vahvuuksien ja heikkouksien perusteella. Ennustaa voi esimerkiksi bruttokansantuotteen kasvua ja pienenemistä, inflaatiota tai velkakirjojen korkoja. Koneoppimisen menetelmien on huomattu olevan tehokkaampia kuin perinteisten aikasarja-analyysien, ja vain tulevaisuus näyttää kuinka tarkasti koneoppimista opitaan hyödyntämään makrotalouden ennustamisessa. Tämä kirjallisuuskatsaus avaa koneoppimista, sekä perehtyy tarkemmin sen eri metodeihin ja kertoo miten koneoppimista ja näitä eri metodeja voidaan käyttää talouden ennustamisessa.

Asiasanat: koneoppiminen, ennustaminen, talous, makrotalous

ABSTRACT

Nyholm, Sebastian
Machine learning in macroeconomic forecasting
Jyväskylä: University of Jyväskylä, 2022, **32 pp.**
Information Systems, Bachelor’s Thesis
Supervisor: Kuusio, Ari

There has always been large amounts of data available of the economy but using all of this to make predictions about the economy has been difficult. Traditional models used in forecasting and in estimates have not proven to be that accurate. The modern methods that machine learning provides have proven to perform well in many different situations and in many different disciplines. Machine learning is at its strongest in making predictions. Econometric models have always tried to forecast the economy, but it has been noted that machine learning is more accurate in its predictions than the more traditional models. Machine learning can be used as a tool in forecasting through many different methods and algorithms which all have their individual strengths and weaknesses. Each of these can be used in making different kinds of predictions based on their strengths and weaknesses. Some good indicators to forecast would be for example the falls and rises of GDP, inflation, or bonds’ interest rates. Machine learning has already been proven to be more efficient than time-series analysis and only the future will tell how well the macroeconomy will be forecasted with machine learning. This literature review explains what machine learning is, familiarizes the reader with different machine learning methods, and explains how machine learning and its methods can be used in economic forecasting.

Keywords: machine learning, forecasting, predicting, economics, macroeconomics

FIGURES

TABLES

TABLE OF CONTENTS

1 INTRODUCTION

Economies struggle to rise from recessions and after crashes households seem to take more debt than they can actually pay back later. Economists have always tried to predict major crashes and recessions, without having much success (Dillow, 2013; Shlaes, 2014). There is an ever-growing amount of data around about the economy. Handling this large data mass has been a problem for macroeconomists long before handling big amounts of data became a well-known problem for other disciplines(Bok, Caratelli, Giannone, Sbordone & Tambalotti, 2018). Machine learning excels exactly at this, predicting something with a lot of data at hand (Varian, 2014, p. 6).

The forecasts made by machine learning could help prepare for economic crises, start preparations against high inflation rates, or even just help households prepare for something that is to come. Machine learning should be a tool that could be found in every economist's toolbox in the future. Machine learning helps make unbiased forecasts on different economic subjects, and it makes them fast. Making quick forecasts gives more time for analyzing how to handle the incoming situation and thereafter actually handling it. Accurate macroeconomic forecasts allow for more rational economic decision-making and are also very crucial to central banks' policy efficacy (Medeiros, Vasconcelos, Veiga & Zilberman, 2021).

The different machine learning methods perform differently and are not equal regarding their forecasting power. Each of the different methods, explained later in this thesis, excel at different scenarios and it should be a common concept to examine and use many of the different methods to see their predictions. This thesis discusses about machine learning in a macroeconomic context, but it will contain a lot of discussion about the different methods machine learning has to offer to economics in its predictions. The center of gravity is on the machine learnings side and the economic forecasting is only an application where machine learnings forecasting capabilities are examined. This way the research is an information systems research and does not branch to other disciplines that much.

1.1 Research questions and the research method

This thesis was conducted as a descriptive literature review so existing literature was used in conducting this study. JYKDOK and Scopus were used as the primary databases for the literature. On top of the literature mentioned before I also used a doctoral thesis conducted by Lauri Nevasalmi in 2020. This doctoral thesis is of very close relation to the subject of this thesis but has its center of gravity on the economic side.

Templier & Paré (2015) state that the point of a literature review such as this is to summarize previously published research on a topic of interest. This way the reader is provided with a comprehensive report on the current state of knowledge in the area under investigation. Baumeister & Leary (1997) note that this serves the scientific field by providing a bridge between a vast assortment of articles on a topic and a reader not having enough time to go through all of it (Templier & Paré, 2015, p. 118).

In search of the literature, I used the advanced search tools provided by the databases. In Scopus, I searched with a list of all the journals I wanted to search from and then followed it up with words to be included in the title. A query string for Scopus would then look something like this:
"EXACTSRCTITLE("European Journal of Information Systems") AND EXACTSRCTITLE("Information Systems Journal") AND TITLE-ABS-KEY(machine AND learning AND forecast*)"
The eight leading information systems journals from the "basket of eight" listing were usually all included in the queries. The query could be modified to include different specification words for the title. These specifications could be adding "prediction" as a substitute for "forecast" or adding a specific economic indicator to predict such as "demand". When researching the different methods, the method name could be added for the title to contain, for example, "Bayesian network". When going through JYKDOK I mainly used the same principles. JYKDOK has its own advanced search, but I used their basic search a few times as well. The main keywords in searches were "machine learning", "forecasting", "macroeconomy", "economy" and "predicting".

The hypernym of machine learning is artificial intelligence, and there is a lot of literature on it in an economic context as well. To keep this thesis within the boundaries of a bachelor's thesis I have specified the focus on the machine learning part. Leaving an important part of machine learning, called deep learning, unrepresented in this thesis was a conscious decision, and done mainly because the boundaries that a bachelor's thesis places. If the reader wants more information about deep learning Goodfellow, Bengio & Courville (2016) wrote a great book about it called "Deep learning". That is a good place to start finding information about it.

I aim to find out how well machine learning can predict different macroeconomic changes, how are machine learning methods comparing to ordinary methods, and eventually why isn't machine learning used more in economic forecasting thus the main research questions are as follows:

- How can machine learning be utilized in macroeconomic forecasting?
- Are machine learning methods better at predicting the macroeconomy compared to those methods used before?
- Why isn't machine learning used more in economic forecasting?

1.2 Structure

This thesis has been divided into four different chapters. The first one being the introduction to the topic. This chapter will explain the motivation behind the study, go through the research question and the method as well as analyze the structure of the thesis.

The second chapter goes more in-depth on machine learning. It defines what machine learning is and also takes a look into a few of the algorithms machine learning uses in its predictions. At the end of the second chapter there will also be a section that discusses machine learning's greatest problems and why it isn't used more.

The third chapter discusses the topic itself, predicting macroeconomics with different machine learning methods. This chapter's main function is to present the findings of the study and compare them slightly to more traditional methods. This chapter will also cover machine learning methods in economics in general and an introduction to nowcasting.

Chapter four is the conclusions chapter, which focuses on briefly wrapping up the whole thesis. This chapter discusses the findings and knowledge of every different chapter provided and summarizes everything the thesis found out.

2 MACHINE LEARNING

This chapter's goal is to give a good understanding of machine learning, what are the main characteristics it has, what are the most commonly used machine learning methods, and finally how reliable is machine learning and what are its biggest problems. First machine learning will be defined according to a few sources and the hypernyms and hyponyms of it are explained. Then the thesis will cover the most used machine learning methods, explain their strengths, and how they are used in the economic context. In the final section, the biggest problems of machine learning are covered, and then the biggest limitations of different methods are covered.

2.1 Defining machine learning

Alpaydin (2020) defines machine learning to being behind the trendiest technologies of today's world. From face recognition to cars that can drive themselves have used machine learning in their pattern recognition. However, these kinds of things are something we humans do unconsciously. We couldn't really define the algorithm we use in recognizing the faces of our close ones and thus couldn't really program a computer program for that without using machine learning. But by analyzing sample face images of a specific person, a learning program captures the pattern and is able to then recognize people (Alpaydin, 2020, p. 3). At this machine learning is probably used the most, pattern recognition. Many different business sectors have found a way to use machine learning to their advantage from manufacturing with demand forecasting to traveling with aircraft scheduling and dynamic pricing. (Vähäkainu & Neittaanmäki, 2018) Based on the perspectives introduced it could be said that there are endless possibilities for machine learning to be utilized in today's world and that we have only scratched the surface of what can really be achieved with it.

Bi, Goodman, Kaminsky & Lessler (2019) define machine learning to be something that aims for the machine to learn by itself without being directly programmed. They also state that in practice experience just usually means fitting to data and thus there is no clear boundary between machine learning and some statistical approaches. Alpaydins' (2020) definition of machine learning is a slightly more complex one, where he defines machine learning to be computer programming to optimize a specific performance criterion using example data or experience. He states that the learning part is the execution of a program to optimize the parameters of a model using past experience. In the end, they are all on the same page on their definitions, one just explained a bit more in-depth.

Sturm et al. (2021) explain that to utilize machine learning in problem solving a learning algorithm is required. According to them all of these algorithms include parameters that need to be tuned according to the problem. They state that the algorithm and the data are needed to train a model to reflect the patterns within the trained data. If new data is achieved, this process can be repeated (Sturm et al., 2021, p. 1584).

Machine learning has a lot of relations to other concepts in the field. Artificial intelligence is considered a hypernym for machine learning. Machine learning itself also has a lot of hyponyms which are usually different methods used in machine learning. Some of these methods are for example decision trees, support vector machines, neural nets, and deep learning(Varian, 2014). Alkharabsheh et al. (2022) also divide the methods into different families. Section 2.2 will go more in-depth on these different methods and algorithms.

In practice machine learning can be used through many different machine learning libraries with open-source code to different algorithms. There are websites that allow machine learning algorithm usage and importing their machine learning algorithms from the website. TensorFlow is a great example of easy machine learning model building and machine learning production library.

2.2 Different algorithms and methods of machine learning

Machine learning can be used through many different algorithms and methods. In this machine learning context methods, algorithms, and techniques as concepts are used interchangeably. The one used in a specific research is up to the writers' preference. On top of the main methods, there are a lot of variations of the more common machine learning methods, so this thesis will not cover all of them. Alkharabsheh et al. (2022) have classified different methods for different families. The methods covered in this thesis are the decision tree (DT), the random forest (RF), the Bayesian algorithm or naïve Bayes (NB), the support vector machine (SVM), and in the end a brief introduction to others that will not be introduced in this thesis more than that. These include for example bagging (BAG), boosting (BST), and k-nearest neighbor (KNN). These different methods have been chosen for their occurrences in the literature regarding economic forecasting.

2.2.1 Decision trees and random forests

Decision trees are probably the most known application of machine learning. These are simple binary trees, which can help the program in making decisions. Varian (2014, p. 6) points out that classification and regression trees often go by the same name of CART. One tree alone is not enough to produce a learning program, but with several trees (a forest) used together and by creating new trees a learning program can be achieved. (Vähäkainu & Neittaanmäki, 2018, p. 14) Figure 1 contains two very simple decision trees. The tree on the left classifies users' actions based on different examples, and the tree on the right makes probabilistic predictions. (Poole & Mackworth, 2010, p. 299)

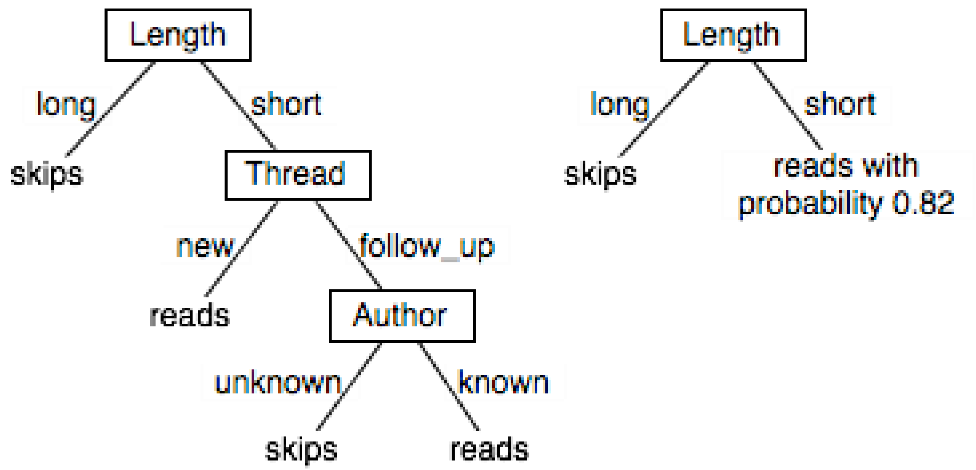

Figure 1. A simple decision tree (Poole & Mackworth, 2010, p. 299).

A decision tree consists of nodes, leaf nodes, branches, and eventually a threshold unit. The data is divided into branches, which start from nodes and lead to leaf nodes. Nodes respond to "if this, then that" circumstances, eventually resulting in a particular outcome. Leaf nodes are basically just nodes without children nodes. (Umar, Fonkam & Prasad 2022)

Random forests are a combination of tree predictors. The generalization error of the whole forest converges as the number of decision trees in the forest grows larger. Other classifiers such as tree strength and the correlation between different decision trees are also things that affect the generalization error. (Breiman, 2001) As mentioned before in chapter 2.1, machine learning is sometimes very close to statistics. Random forests include some similar attributes used in regression models, where the more regressors there are the more accurate the model is. Correlation between different regressors (here the correlation between different decision trees) affects the error and the outcome.

Decision trees get their strength from their simplicity. Learning them is not that hard, which saves time for analyzing the results. They are also quite flexible to nonlinear covariate effects and can include even higher level covariance between covariates. (Bi et al., 2019) This all holds for random forests, as they are a combination of tree predictors.

2.2.2 Bayesian algorithms

In order to get to know what a Bayesian network or a naïve Bayes algorithm is we first have to understand what the Bayes' theorem is. The Bayesian machine learning algorithms are derived and built on top of the Bayes' theorem, so they are of a close relation. The Bayes' theorem is a key concept of Bayesian statistics which will be described briefly next.

Bayesian statistics is often compared to more classical statistics, but the main difference between these two statistic movements is how the population parameters are predicted. More classical statistics assume the population parameters to be unknown constants and when estimated the concept of probability is used whereas Bayesian statistics assume the population parameters to be quantifiable random variables and thus can be described by probability distributions. (Puga, Krzywinski & Altman 2015)

According to Puga et al (2015), the Bayes' theorem is the core of Bayesian statistics. The theorem describes different outcome probabilities using conditional probability. They give a good example of conditional probability where we have two coins, one with the probability of 0.5 to land on tails (and heads for that matter) and a biased coin which has a probability of 0.75 of yielding heads. Conditional probability is basically choosing the coin. The choice of coin affects the toss outcome probability, where heads are now favored because of the biased coin.

Bayesian networks are graphical structural models that encode probabilistic relationships among variables (Heckerman, 1996). Literature regarding Bayesian networks has made major advances in inferring causal relationships from observational data(Binder, Koller, Russell & Kanazawa, 1997; Pearl, 1998; Spirtes, Glymour & Scheines, 2000). Bayesian networks are built of nodes, where there are child and parent nodes (Z. Zheng & Pavlou, 2010). Bayesian network graphs are referred to as DAG which stands for directed acyclic graph(Dharmasena, Domaratzki & Muthukumarana 2021; Zheng & Pavlou, 2010). Zheng & Pavlou (2010) define the directed in this context to mean for a node to have an asymmetric edge over the linked node and the acyclic to reflect the edges not forming a circle. A naïve Bayes algorithm is the same as a Bayesian algorithm, but it just makes the assumption, a naïve one, of predictive variables independence (Bi et al., 2019). Figure 2 is a graph of a generic Bayesian network with five different nodes. A and B nodes being the parents of C and C being a parent of E and D nodes and in return, C being the child of A and B, and E and D being the children of node C.

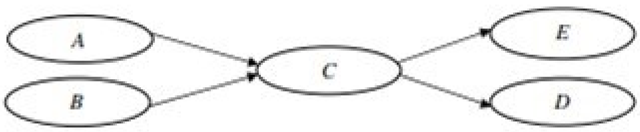

Figure 2. A generic Bayesian network with five nodes (Zheng & Pavlou, 2010, p. 370).

The strengths of the Bayes also come from its simplicity. It has been shown to perform well in the presence of noise, missing data, or even irrelevant features. On top of this the naïve Bayes requires fewer parameters and thus a smaller training set because of its "naïve" assumption. (Bi et al., 2019)

2.2.3 Neural networks

Research about neural networks has been going on since 1943 when Warren Mcculloch and Walter Pitts presented the first model of artificial neurons. Neural networks were originally meant to picture how human brains processed information. (Rojas, 1996) Neural networks used as a machine learning method also go by the name of artificial neural networks. Like people, neural networks learn best from examples, as the biological systems' learning also includes fitting to the synaptic connections between the neurons in the network(Stergiou & Siganos, 2006). Since the research of artificial neurons started, neural networks have been applied to several problems in pattern recognition and are being used as a method in machine learning today. Training neural networks require a lot of experience and experiments and thus some people prefer something more simple, such as the decision tree. (Nilsson, 1998) Decision trees are easy to understand because of their Boolean-like structure, where they only branch in two different directions at a time. Neural networks are also widely in use in nonlinear approximation and when neural networks are applied to forecasting, they can be somewhat of a black-box model, as it is given a set of inputs and learns from them and gives us an output (Butts, Hoest-Madsen & Refsgaard, 2003; Bi et al., 2019). It seems like neural networks are a complex thing to understand as their "hidden" mid-layer between the inputs and outputs isn't visible to the naked eye.

A neural network consists of a big population of neurons interconnected through complex signaling pathways and this structure is used to analyze complex interactions between measurable covariates to eventually predict an outcome (Bi et al., 2019; Vähäkainu & Neittaanmäki, 2018). Neural networks possess layers of these neurons, which are the input layer, hidden layer, and eventually an output layer. In theory, there could possibly be more hidden layers, making the network a more complex one. (Nevasalmi, 2020, pp. 56-57; Vähäkainu & Neittaanmäki, 2018) Every one of these hidden units is a linear combination of the input variables which have a link connecting them (Nevasalmi, 2020, p. 56). Butts et al. (2003) define there to be a weighting value for each input to tell how important a specific input is whereas Bi et al. (2019) and Vähäkainu & Neittaanmäki (2018) and Patil & Subbaraman (2021) state that there is a weighting for each link of the network, even between the hidden layer and the output layer. A weighting is used to define to the neural network how important a specific module of it is, whether a link or an input. When teaching the network the goal is to determine weights for the links that reduce errors (Patil & Subbaraman, 2021).

Figure 3 is a demonstration of a simple neural network with an input layer, one hidden layer, and an output layer. The neurons are portrayed by the grey circles, which are connected by links. It can be seen here how the hidden neurons are actually linear combinations of all the input values, as the links lead from each input to every hidden neuron.

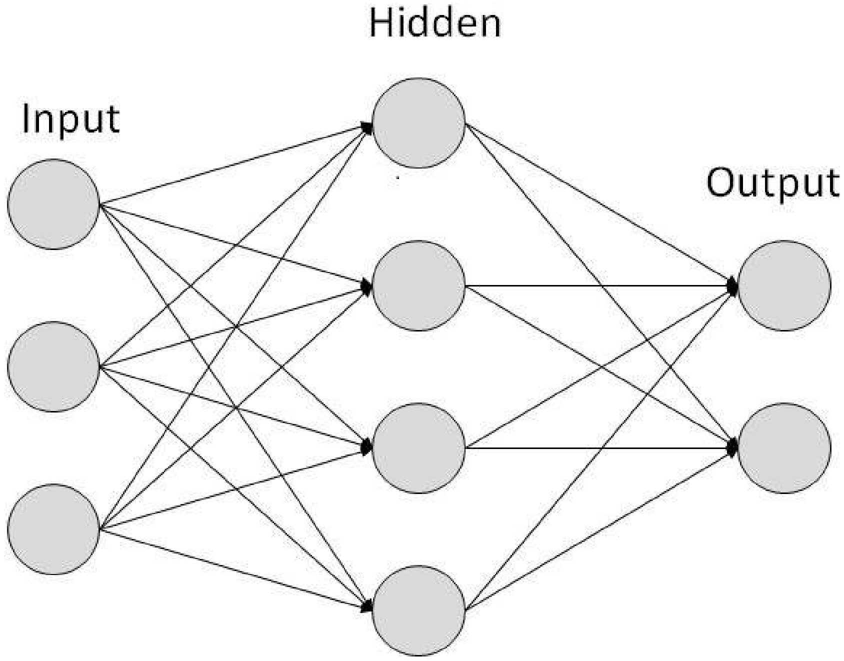

Figure 3. A simple neural network with three layers (Tutorialspoint, 2017).

Neural networks or artificial neural networks -method is probably the one I have seen used the most in research. Bi et al. (2019) define neural networks' strengths to be especially in how it accommodates to variable interactions and nonlinear associations without user specification. It could be said that once an artificial neural network is set it really doesn't care about changes. This could not be said about all machine learning methods.

2.2.4 Support vector machine

Support vector machines were originally presented by Vapnik (1995) (Nevasalmi 2020, p. 58). They are most commonly used in classification and in regression problems (Athey & Imbens, 2019; Bi et al., 2019; Nevasalmi, 2020, p. 58), however Athey and Imbens (2019) state later that support vector machines are traditionally used in classification but have also been extended into regressions. This implicates that they are not traditionally used in regression but have lately also been fitted there. Umar et al. (2022) and Patil & Subbaraman (2021) continue that a support vector machine is actually a binary classifier because it assigns observations into only two different categories. It uses a statistical learning theory to create a consistent estimator for the data.

Bi et al. (2019) start explaining the functionality by defining that support vector machines are used to construct an optimal boundary between observations. Bi et al. (2019) and Umar et al. (2022) call this boundary a hyperplane and Nevasalmi (2020) and Bi et al. (2019) specify there to be a margin that is defined to be the distance from the hyperplane to the nearest observation. Support vector machine training tries to widen the boundary between the two categories as much as possible, and the main purpose of this method is to find the optimum line or decision boundary for categorizing n-dimensional space into different classes so that additional data points can easily be placed into one of these classes (Umar et al., 2022). The bigger the margin or boundary between the different classes is, the easier it is to place these additional data points into a correct category. But what if an observation is not appliable to any of these categories Bi et al. (2019) give a good view on kernel functions and what they do, and this is described briefly after the figure.

Figure 4 is a good example of a support vector machine. $(\omega\ \varphi(x)+ b) = 0$ being the boundary separating the two classes of observations, $(\omega\ \varphi(x)+ b) > 0$, and $(\omega\ \varphi(x)+ b) < 0$ being the support vectors and the area between the support vectors and the boundary being the margin. As stated before the point of the support vector machine is to maximize the area between the two classes.

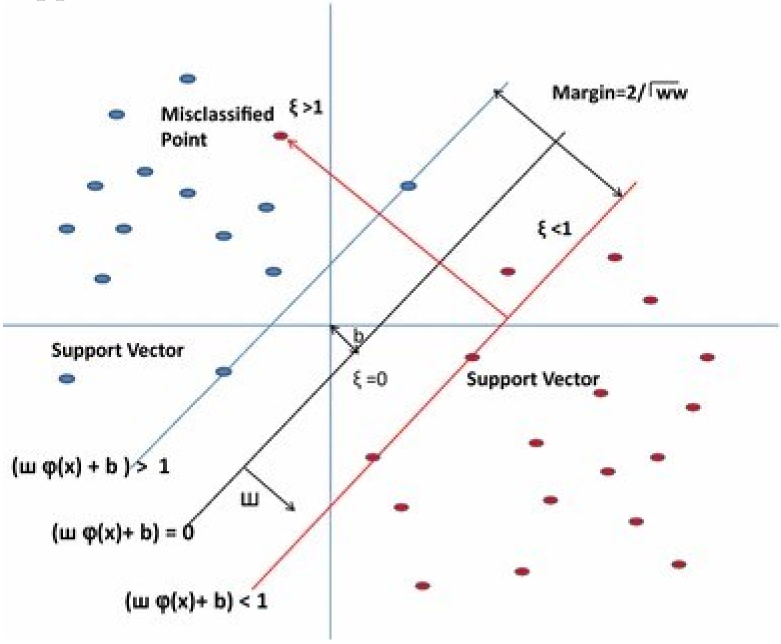

Figure 4. A support vector machine (Chandra & Bedi, 2021, p. 1868).

Many of the observations that support vector machines want to categorize can't yet be separated by a hyperplane. Bi et al. (2019) explain that this can be addressed by using a "kernel function". Support vector machines apply this data transformation tool to the data, and they project the data into a higher-dimensional space where the variables (derived from observations) can now be separated. They state the most popular kernel functions include for example a polynomial kernel, Gaussian kernel, and sigmoid kernel. I will not take a deeper look into these kernels in this thesis but a good place to start could be for example "Survey on SVM and their application in image classification" by Chandra & Bedi (2021) or even "Efficient kernel functions for support vector machine regression model for analog circuits' performance evaluation" by Boolchandani, Ahmed & Sahula (2011). Even though the latter article is not directly of the same subject, it still holds a good introduction to kernel functions and some more research on them.

Support vector machines usually show low misclassification levels and thus are quite reliable. They also scale quite well with higher dimensional data. Support vector machines are also easily interpretable, but some of this interpretability disappears when using a kernel function as it complicates things. (Bi et al., 2019)

2.2.5 Other methods

There are still a lot of methods not introduced in this chapter and thesis. These methods are not any less important in economic forecasting than the ones that were introduced. The methods this thesis covers are chosen for their presence in the literature used in this thesis. Bi et al. (2019) cover a bigger part of machine learning methods or as they say algorithms. That is a good place to look for an introduction to a specific method. Some worthy mentions have to be made to bagging, boosting, and K-nearest neighbors methods. Bagging and boosting belong to an ensemble methods group (Bi et al., 2019) whereas K-nearest neighbors belong to a family of Nearest neighbor methods (Alkharabsheh et al., 2022). Ensemble methods are methods that combine information from many different models and combine their prediction power (Bi et al., 2019).

Bagging is a method that fits the same underlying algorithm to each bootstrapped copy of the original training data. It then creates predictions based on outputs from the resulting models(Bi et al., 2019a; Breiman, 1996). In the end, the quantitative final prediction is made by averaging the predictions made from the copies of the original data, and for a qualitative outcome, the prediction takes either a majority vote or averages the probabilities(Bi et al., 2019). One could say that bagging just creates many copies of the original data set and uses all of these copies to make an aggregated result. Each copy and its result is used in generating the final predictor, and the more copies and possible predictions there are, the more accurate the final aggregated prediction will be.

Boosting is similar to bagging in that it creates copies of the starting data. However, boosting, as its name suggests, boosts the next predictions, and improves them based on the earlier copies and their prediction errors. Different observations are given weights that change based on their classification. (Bi et al., 2019)

K-nearest neighbors method was presented by Evelyn Fix and J.L. Hodges, Jr in 1951 (Nevasalmi, 2020, p. 49). It is a non-parametric method used for classification and regression (Maccarrone, Morelli & Spadaccini, 2021; Zohdi, Rafiee, Kayvanfar & Salamiraad, 2022). The K-nearest neighbors can identify repeated patterns within the time series and are applied to financial time series modeling because of this. The method outperforms many other methods regarding precision with a limited amount of data. (Maccarrone et al., 2021) Zhang (2016) defines the method to work by classifying unlabeled observations into a class with the most similar labeled examples. In a two-dimensional plot, two characteristics can be employed, and the method classifies and places all of the observations in a group fitting that specific observation. Figure 5 illustrates a two-dimensional plot, where fruits, vegetables, and grains are placed. It divides them into groups depending on their sweetness and crunchiness.

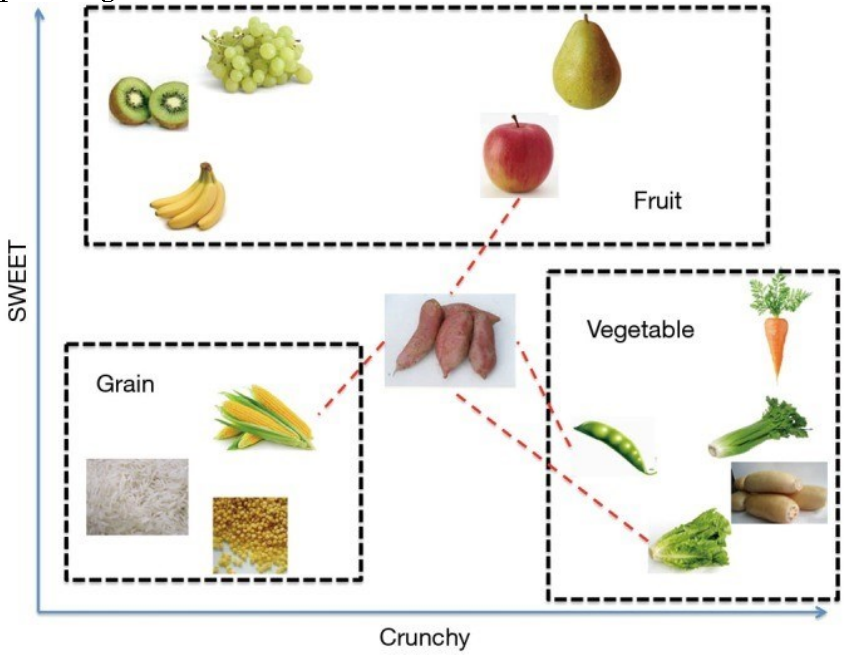

Figure 5. Illustration of how k-nearest neighbors' algorithm works (Zhang, 2016, p. 2).

2.3 Why isn't machine learning used more in economic forecasting?

Machine learning, as new as it is, is rather rarely used in research and researchers still prefer more traditional methods. Several researchers (Kreiner & Duca, 2020; Maccarrone et al., 2021; Soybilgen & Yazgan, 2021) have found out that some machine learning methods do indeed outperform older forecasting strategies. Why is machine learning then not utilized in a bigger scope?

The short answer is that the researchers simply lack the ability to utilize machine learning in their research. This can partly be blamed on the fact that many of the closed-source programming languages are the ones that are gaining popularity. (Bi et al., 2019) As stated before, some machine learning methods do require quite a bit of knowledge on them before they can be used on the problem under research. This can lead to researchers not bothering to learn a whole different method to conduct their research when they can easily do it with methods, they are familiar with. In general, a lack of working knowledge of machine learning algorithms is something that reduces their usage in research (Bi et al., 2019).

Machine learning methods are also quite labor heavy, states Bengio, Courville & Vincent (2013) as well as Athey & Imbens (2019). Bengio et al. (2013) continue that a lot of the actual effort goes into the design of preprocessing pipelines and transforming data. All of this takes time out of doing the actual research. They state that the biggest weakness of machine learning algorithms is their inability to extract and organize the discriminative information from the data. Next, the thesis will go more in-depth on the methods' problems and their limitations in the same order that they were presented.

Decision trees are prone to losing information by categorizing variables where their associations are continuous. Decision trees are also not the most adaptable of methods, because even a small data change or node shift might result in drastically different trees. It should be mentioned that decision trees are also prone to overfitting. Some of these problems can be addressed in a random forest as it connects many decision trees. (Bi et al., 2019)

For the naïve Bayes algorithm, the main weakness comes from its assumption of independence between predictive variables. In reality, this assumption is often incorrect, and thus the base for the whole algorithm disappears. (Bi et al., 2019)

Neural networks can be somewhat of a black box model (Bi et al., 2019; Butts et al., 2003). What this means is that neural networks try to learn something from the inputs and create the hidden middle layer or several of them. Because the layer indeed is hidden, it may be hard to find out the relations between the inputs and the outputs (Bi et al., 2019). Bi et al. (2019) and Athey & Imbens (2019) add that training neural networks may be a complex task and usually require technical expertise and a lot of tuning to work well. Neural networks usually work the best with very large sets of data and features.

Support vector machines' weaknesses mainly come from kernel functions. When it is necessary, experimenting with a set of standard functions is required. This experimenting can be quite time-consuming so once again there will be less time left for everything else. There is also no guarantee that a standard kernel function can work so sometimes even hand-crafted ones are used. (Bi et al., 2019)

It should be noted that when a method has a specific weakness, it can easily be avoided by simply using another method altogether. Some of the methods do have similar weaknesses and limitations, but there should always be one that can avoid a specific one. Varian (2014) also notes that by averaging predictions of several models, a single model can be outperformed.

3 MACHINE LEARNING IN MACROECONOMICS

Machine learning has many usable cases, and it can be used as a tool in many different areas of research. One of these areas is economics, and in this thesis, specifically macroeconomics. Machine learning has a lot to offer when it comes to forecasting the economy and that is what this chapter will cover.

Machine learning is mainly concerned with making predictions when it comes to analyzing data (Varian, 2014). This section will focus on how the economy can be predicted with machine learning and its many methods. First macroeconomics is briefly defined and then there will be a discussion about how different machine learning methods can be utilized in economics. Then there will be a chapter about the research that has been conducted on forecasting macroeconomics with machine learning. The section will also include a chapter about nowcasting which is making predictions of this exact moment, as well as the very near future, and this is a big part of macroeconomic forecasting and knowing what's going on in the macroeconomic scale.

3.1 Macroeconomics

Before getting to know how machine learning can be utilized in forecasting the macroeconomy it is crucial to know what macroeconomics is and what it is not. Tuerck (2014) states that macroeconomics is "the study of the economy as a whole". He explains that microeconomics is the other side of economics which focuses on single economic decision-makers, whereas macroeconomics focuses on interactions between large aggregates such as gross domestic product (GDP) or the market for labor and capital. It is also a macroeconomic concern what policies are needed to make the economy work effectively and what policies are called for when the economy breaks down (Tuerck, 2014, pp. 1-2). Begg, Vernasca, Fischer, Dornbusch & Begg (2014) define macroeconomics quite similarly. They state it stresses broad aggregates, such as the total demand for goods by households and the scope is thus on understanding interrelationships of the big issues that affect the economy. Macroeconomics usually simplifies reality to focus on the key elements of a problem more clearly (Begg et al., 2014, pp. 352-354).

3.2 Nowcasting

Nowcasting is a concept often seen in the field of economic forecasting. It seems to be the combination of the words "now" and "forecasting". Bok (2018) notes that Giannone, Reichlin & Small (2008) were the first ones to build a statistical framework that monitors macroeconomic conditions in real-time. Giannone et al. (2008) add that because all data is in fact released with lag, this means after the things have actually happened, it is important to forecast the things happening right now or in the very near future. Sarduie, Kazemi, Alborzi, Azar & Kermanshah (2020) note that nowcasting is actually the main focus of modern econometric analysis. Because the behavior of economic variables is based on non-linear and complex relationships of data, the use of machine learning techniques has led to a significant improvement in predictions (Sarduie et al., 2020).

Soybilgen & Yazgan (2021) note that in some fields, it is easy to see what exactly is going on today. Weather is a good example, only future weather forecasts are needed as it is obvious what the weather is like outside. They state that in economics, however, we are missing information about the current state. This is because important macroeconomic aggregates can only be measured with a considerable time delay. This is why nowcasting is required. (Soybilgen & Yazgan, 2021)

Without nowcasting, there would be no idea about what is happening on a macro level in our economy. When today's data would be released, maybe two or three months later it would be already old. Nowcasting is something quite important in the field of economic forecasting and the machine learning methods that this thesis introduced can also be used as tools in nowcasting.

3.3 Utilizing machine learning in economics

Athey & Imbens (2019) explain that the traditional approach to econometrics is to specify a target and an estimand. Then a random sample would be taken from the whole population of interest and then using the sample the parameter of interest and other nuisance parameters are estimated by finding the parameter values that best fit the full sample. This is done by using an objective function, for example, the sum of squared errors. The focus is on the quality of the estimators of the target. In contrast, when using machine learning the focus is on developing the algorithm. Then it's the goal of the algorithm to make predictions about some variables given other variables. (Athey & Imbens, 2019) Many econometric models and regressions could be replaced by machine learning. The question just is that is machine learning in a state that is better than the traditional econometric models.

Econometrics is closely related to statistics, and it uses statistics as its base, just applied to economics. Athey & Imbens (2019) state that the machine learning revolution is largely been accepted by the statistics community. However, they add that economics has been a much slower discipline in adopting machine learning as a tool to be used in their field. This may be because of the case that machine learning tools have been particularly successful in big data settings. In reality gathering, big amounts of data for research purposes may be hard.

The chapters coming next, which are focusing on how different machine learning methods are used in economic research, divide the methods to work either in regression problems or classification problems, or even both. Regression analysis is closely related to correlation analysis but regression rather characterizes the law of relation between the variable of interest and the explaining variables and does this not for individual elements but rather for the whole population (Babenko et al., 2021). Babenko et al. (2021) also explain that the point of regression analysis is to find the line that best passes through a set of given points, and this can be done in many different fashions.

Classification is something that tries to group similar objects together by using logistic regression. In classification logistic regression will give us a certain "plane" or specific points of examples. Then by using this plane or example point dividing observations into classes can be done. (Babenko et al., 2021)

Decision trees and random forests are very popular methods for estimating regression functions in economics and especially when out-of-sample prediction power is important. Random forests have also recently been extended to situations where the interest is in causal effects and situations where general economic models' parameters are estimated. (Athey & Imbens, 2019)

Econometrics don't have that many tools to handle situations where out-of-sample predictions are produced so decision trees and random forests predictions in this fashion are welcome. Athey & Imbens (2019) explain the decision trees' and random forests' estimations to work by creating subsamples of the original sample and estimating the regression within the subsamples by averaging. They continue that the main advantages single trees provide in economics is their simplicity. It is easy to explain and interpret results. This can not be said about some of the more complex "black box" methods, where interpreting the results is hard. Random forests share many of the advantages with single decision trees and these include the little tuning they require and quite good out-of-the-box performance and all of this contributes to the results being easy to interpret (Athey & Imbens, 2019).

Decision trees and random forests can easily be modified to tasks that include more classification than regression. In classification, the idea is once again to split the original sample into smaller samples. The main difference in decision trees and random forests regression to classification is the objective function that measures the improvement from a particular split. In classification, this function is called the impurity function, and it estimates how impure different leaves are in the tree. Purity is if all the units in the leaf have the same label, and impurity means the opposite this is if the labels are all different. (Athey & Imbens, 2019)

Neural networks are another flexible approach to estimating regression functions. Neural networks in economics have been found to be very efficient in settings where there is an extremely large number of features in the regression. However, it should be noted that neural networks do require a lot of knowledge and tuning to work well. (Athey & Imbens, 2019)

The network's depth substantially increases the flexibility of the model in practice. In applications, researchers have used neural networks with even ten or more layers and millions of parameters. In these cases, with multiple hidden layers and inside them many hidden neurons, one needs to carefully regularize the parameter estimations. (Athey & Imbens, 2019) Multiple layered neural networks' performance with parameters really shows how flexible neural networks are if used correctly. If one were to possess the skills required to use neural networks flawlessly the possibilities for economic regressions would be endless. This being said, I can't stress enough that multiple hidden layers make the network very complex and require a lot of time and knowledge.

Athey & Imbens (2019) also talk about support vector machines and their kernels in economics, and they state that they are best used for classification problems. They continue that support vector machines can be extended to regression settings but are more naturally introduced in classification. The main idea of support vector machines trying to create a hyperplane with the least classification errors is exactly the same and can be utilized in an economic context as it is.

3.4 Forecasting macroeconomics with machine learning

Predicting macroeconomics is hard, and something that has happened in history is of course no guarantee that it will work the same way in the future. However, history still gives a good overview of what can be expected, and of course, studying the ways things have happened does give some guidelines to how things will behave in the future. There are several studies on how macroeconomics and macroeconomic indicators have been forecasted and this section will cover some studies. Some good examples of macroeconomic indicators are inflation rates, interest rates, and even GDP. These give some insight into what is happening or what will happen in our economy on a broader scale.

Inflation rates have been forecasted a lot with machine learning and of course without it. Medeiros et al. (2019) forecasted inflation rates with the random forest -method and stated that machine learning models with a large number of covariates perform better and give better predictions than the traditional benchmarks. They add that the machine learning method that deserves more attention is in fact the random forest -a method which according to them dominates all other machine learning models.

Medeiros et al. (2019) found out that the gains of using machine learning in inflation forecasting can be as large as 30% in terms of squared errors. A 30% improvement in terms of errors is a huge step up, and this means that forecasting inflation rates will be much more accurate than ever before. Medeiros et al. (2019) also found out that in their forecasts random forests performed better than even their other methods which included bagging and a LASSO method. Interestingly many machine learning methods performed better than the benchmarks they had set, which really gives the idea of what can be achieved with a good understanding of machine learning. Inflation forecasting is important for every individual in the economy, and it allows for more rational decisions when the inflation rates can be known more accurately than before. Machine learning methods allow these forecasts to be their most accurate versions.

GDP is another indicator of the macroeconomy which has been forecasted a lot, and which is under constant monitoring. Soybilgen & Yazgan (2021) forecasted as well as nowcasted the United States GDP using Tree-Based models and dynamic factor models as benchmarks. The machine learning methods used were combinations of a few methods. Bagging and decision trees, Boosting and decision trees, and finally a random forest model. They were compared to benchmark methods of random walk and linear dynamic factor models.

Soybilgen & Yazgan (2021) found out that on average all machine learning methods achieved better performance than the benchmarks. When measured with mean absolute errors or root mean square errors the random forests method performed the best out of all of the methods. However, to take the results into context when measured with mean absolute errors the biggest difference between different machine learning methods was only 0.002 points, while the difference between the random forests method and the best performing benchmark model was 0.022. Over ten times more compared to the difference between the machine learning models. Soybilgen & Yazgan (2021) wanted to highlight that with little information and having done no predictions yet from which to learn from the machine learning models are volatile and actually, a benchmark model was better in the first predictions for the reference quarter. (Soybilgen & Yazgan, 2021) When machine learning is given enough information, it truly outshines the benchmarks.

Soybilgen & Yazgan (2021) explain the CART methods to work in their case by dividing the feature space into regions that minimize an objective function. This is done by recursively splitting the space into two distinct regions and stopping when a certain criterion is reached. Single trees do sometimes make poor and noisy predictions, and that's why random forests were also included in the research.

Interestingly enough some of the simpler methods such as classification and regression trees do produce very accurate predictions when combined, or sometimes even when used as a single tree. Ensemble methods seem to mitigate a lot of the weaknesses of the single methods in practice as well. It would be interesting to compare the simple CART methods and their prediction power to an artificial neural network's prediction power and see does the complexity of the network make any difference. It can be said for sure that random forests introduced by Breiman (2001) are the top of the class when it comes to predictions.

Labor economics is a subject that has been researched a lot, and a part of it is its relation to macroeconomics. A big part of labor economics and thus macroeconomics as well is of course the employment and unemployment rates. Kreiner & Duca (2019) forecasted the unemployment rates using data from the federal reserve. They constructed an artificial neural network and also a model called the LASSO model. The neural network was tested with one two and three hidden layers, but not more because of their computational complexity and run-time. The best configuration according to the errors was the one with the most layers, three. In all three layers, there were 97 neurons. They compared these machine learning methods to the responses of professional forecasters in the known survey of professional forecasters.

The main results in comparing the artificial neural network to the LASSO and the survey was that the artificial neural network performed the best in each period. Especially during the 2007-2012 great recession neural networks really performed well. (Kreiner & Duca, 2020) While artificial neural networks already performed well, it would be interesting to see what could be done with even more hidden layers. Maybe the forecast made by the neural network could be even closer to the actual unemployment rate and the errors of the forecast even smaller. Huang (2009) also notes that artificial neural networks yield better results than ordinary linear models. He also notes the same thing in the economic context that was noted in chapter 2.3 that neural networks require a lot of maintenance, as they can't fit new datasets to the same model when it was trained with different data.

Not too much can be said about the artificial neural network used by Kreiner & Duca (2020). As stated in the second section of this thesis a neural network is a black box model which does not give too much information about its hidden layers. The only things known about the neural networks are the number of the layers inside them, the inputs, and the outputs. It is always important to know how different methods do their predictions, and when the "how" can't be figured out the output of the model has to just be trusted.

Support vector machines according to Huang (2009) have better generalization ability than artificial neural networks. In his study, he forecasts GDP and constructs two models. One based on artificial neural networks, and one based on support vector machines. The main findings in this study are that his support vector machine predicted the GDP much more accurately compared to the artificial neural network model. It reduced the average error rate from about 15% to less than 4%. (Huang, 2009)

Davig & Hall (2019) used a Naïve Bayes model as a recession forecasting tool. They explain that a recession is one of the most significant macroeconomic events, as unemployment rises sharply, and the outputs of the whole economy decline. In their study, they compare the Naïve Bayes models to a logistic regression model and the survey of professional forecasters (also mentioned before). As data, they used the turning point dates from the national bureau of economic research's business cycle dating committee. The main findings of Davig & Hall (2019) were that across many experimental conditions and evaluation criteria the Naïve Bayes model outperforms all of the other models. When forecasting the near future or even nowcasting some models performed slightly better than the Naïve Bayes in certain circumstances. When forecasting a recession in practice it's often more beneficial to get a slight note of it early on, than for the forecast to be extremely accurate, Davig & Hall (2019) note. Usually, the models used in this research all saw a recession coming 3-6 months earlier. The main takeaway is that the most accurate models are not always the ones that see recessions coming at the earliest stages possible, but there doesn't seem to be a way to quantitatively measure this. (Davig & Hall, 2019)

Machine learning offers the field of macroeconomics great improvements in prediction accuracy, but also more work as machine learning does require some knowledge. The future of macroeconomic forecasting will include machine learning in form or another, and it is a tool with great prediction power. Economists working with macroeconomic forecasting should try and adapt machine learning as early as possible to be comfortable using it alongside statistical methods. The problem of machine learning in macroeconomic forecasting requiring multidisciplinary knowledge can be overcome with either there being a machine learning professional that helps people in machine learning usage, or alternatively the forecasters adapting it as a tool of their own.

Having machine learning as a tool used in macroeconomic forecasting will affect the field by providing a time-consuming method to improve the accuracy of the predictions. By using statistical methods, the time used in constructing the method would be used in finding good parameters that best fit the full sample. However, when using machine learning this will be done by the method, but the method has to be trained with the data. It could be said that macroeconomic forecasting will be a trade off between time and prediction accuracy. If a researcher does have the time to orientate to machine learning, better results can be seen in the forecasts. In the long run the time invested in studying machine learning will pay itself back many times over by yielding better results.

4 CONCLUSIONS

This thesis discussed machine learning in forecasting macroeconomics, and the center of gravity was on machine learning and its methods. These different methods have different cases where they work the best, and that applies to the macroeconomic context as well. Research has been done about machine learning and its methods in macroeconomic forecasting and this thesis introduced some of the research. The thesis only included some machine learning methods and not all, as the scope of the thesis had to be narrowed down slightly. This same reason applies to why machine learning was discussed in only the macroeconomic context and not in economics overall.

The reliability of the research has been improved with a familiar research method of a literature review. The validity of the research comes mainly from the qualified references on top of which this thesis was built on. Databases from which the references have been extracted from were mainly JYKDOK as well as Scopus' journal search. Scopus offered several recently written articles about the topic, which have been published in very qualified journals.

While forecasting the macroeconomic indicators machine learning methods were also compared slightly to each other. The main findings in comparing methods in a macroeconomic context, and also while presenting different machine learning methods in chapter 2, is mainly that they all should be used in different cases and one method is not superior to others in all cases. Some methods can be better generalized, as Huang (2009) presented for the support vector machines compared to artificial neural networks. Neural networks seem to be a method that requires the most training and knowledge (Athey & Imbens, 2019; Bi et al., 2019; Kreiner & Duca, 2019). When it comes to machine learning methods it could be useful in some cases to know how it actually does its forecasts. Decision trees are truly simple and easy to read whereas the opposite a neural network contains layers that can't really be read easily (Bi et al., 2019).

Nowcasting is even the main focus of modern econometric analysis (Sarduie et al., 2020), and much of the research presented in this literature review also included nowcasting. Sarduie et al. (2020) explain that the use of machine learning methods has led to a big improvement in predictions, and hence nowcasting now moves from causal inference toward the field of machine learning.

The main question is "How can machine learning be utilized in macroeconomic forecasting?". Machine learning can be utilized by picking a method best fit for that kind of data set and teaching it about the data. Some methods also require some additional tuning. Depending on the method picked teaching can be quite time-consuming, and some methods have other limitations which have to be accounted for. After learning from the earlier data, the method should be able to make predictions based on the data, and the accuracy of the predictions can then be measured with some least errors methods. The methods can be used to predict macroeconomic indicators such as GDP or inflation rates.

The research done in forecasting macroeconomics with machine learning tells quite an unambiguous tale when answering the question "Are machine learning methods better at predicting the macroeconomy compared to those methods used before?". Modern machine learning methods do perform better than the traditional regressions and other traditional models and also better than the survey of professional forecasters. Better performance in the research reviewed in this thesis is defined by smaller prediction errors. There were a few rare situations where machine learning methods did perform worse than the benchmark, but this was mainly due to them being trained only a little and due to only a little data given to the method to learn from (Soybilgen & Yazgan, 2021). Macroeconomic indicators that machine learning forecasted better were inflation, GDP, and unemployment rate. Davig & Hall (2019) point out that measuring what models forecast recessions the earliest rather than what models predict recessions the most accurately could be more beneficial, but there doesn't seem to be a way to quantitatively measure this.

If machine learning indeed is better at forecasting the macroeconomy we get to the third research question. "Why isn't machine learning used more in economic forecasting?" Machine learning is a tool that all economists should have in their box of analyzing tools (Athey & Imbens, 2019), but this would require a lot of multidisciplinary knowledge about economics and about machine learning. This might be the main issue that limits the use of machine learning in economic forecasting. Researchers don't know enough about both machine learning and economics to conduct research using both. Table 1 presents the main findings of the study in a simple table form, where the answers to all research questions can also be found.

Table 1. The main findings of the study.

Research question	**The solution to the research question**
How can machine learning be utilized in macroeconomic forecasting?	A machine learning method has to be picked, tuned, and taught. It can be utilized in predicting indicators such as GDP or inflation rates.
Are machine learning methods better at predicting the macroeconomy compared to those methods used before?	Machine learning methods have proven to be more accurate in their predictions than general regressions or even the survey of professional forecasters.
Why isn't machine learning used more in economic forecasting?	Utilizing machine learning in macroeconomic forecasting requires multidisciplinary knowledge, and thus is not often used.

Future research regarding economic forecasting using machine learning could focus on the methods, more than the problem the method is actually used on. It is important to know which methods work best for which kind of data sets, and why. For now, it seems like a few methods are chosen almost at random because picking one over the other is usually not justified that much, and teaching many methods is usually quite time-consuming so only a few are picked. A better comparison between econometric models and machine learning models could also be conducted by someone more familiar with both of them. A comparison not only comparing the results but rather the process.

REFERENCES

Alkharabsheh, K., Alawadi, S., Kebande, V. R., Crespo, Y., Fernández-Delgado, M., & Taboada, J. A. (2022). A comparison of machine learning algorithms on design smell detection using balanced and imbalanced dataset: A study of God class. *Information and Software Technology*, *143*, 106736. https://doi.org/10.1016/j.infsof.2021.106736

Alpaydin, E. (2020). *Introduction to Machine Learning, fourth edition*. MIT Press.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, *11*(1), 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Babenko, V., Panchyshyn, A., Zomchak, L., Nehrey, M., Artym-Drohomyretska, Z., & Lahotskyi, T. (2021). Classical Machine Learning Methods in Economics Research: Macro and Micro Level Examples. *WSEAS TRANSACTIONS ON BUSINESS AND ECONOMICS*, *18*, 209–217. https://doi.org/10.37394/23207.2021.18.22

Baumeister, R. F., & Leary, M. R. (1997). Writing Narrative Literature Reviews. *Review of General Psychology*, *1*(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311

Begg, D., Vernasca, G., Fischer, S., Dornbusch, R., & Begg, D. (2014). *Economics* (Eleventh edition). McGraw-Hill Education.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239. https://doi.org/10.1093/aje/kwz189

Binder, J., Koller, D., Russell, S., & Kanazawa, K. (1997). Adaptive Probabilistic Networks with Hidden Variables. *Machine Learning*, *29*(2), 213–244. https://doi.org/10.1023/A:1007421730016

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., & Tambalotti, A. (2018). Macroeconomic Nowcasting and Forecasting with Big Data. *Annual Review of Economics*, *10*(1), 615–643. https://doi.org/10.1146/annurev-economics-080217-053214

Boolchandani, D., Ahmed, A., & Sahula, V. (2011). Efficient kernel functions for support vector machine regression model for analog circuits' performance evaluation. *Analog Integrated Circuits and Signal Processing*, *66*(1), 117–128. https://doi.org/10.1007/s10470-010-9476-6

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Butts, M. B., Hoest-Madsen, J., & Refsgaard, J. C. (2003). Hydrologic Forecasting. In R. A. Meyers (Ed.), *Encyclopedia of Physical Science and Technology (Third Edition)*. Academic Press. https://doi.org/10.1016/B0-12-227410-5/00325-2

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in imageclassification. *International Journal of Information Technology*, *13*(5), 1–11. https://doi.org/10.1007/s41870-017-0080-1

Davig, T., & Hall, A. S. (2019). Recession forecasting using Bayesian classification. *International Journal of Forecasting*, *35*(3), 848–867. https://doi.org/10.1016/j.ijforecast.2018.08.005

Dharmasena, I., Domaratzki, M., & Muthukumarana, S. (2021). Modeling mobile apps user behavior using Bayesian networks. *International Journal of Information Technology*, *13*(1). https://doi.org/10.1007/s41870-021-00699-7

Dillow, C. (2013). *Why we can't predict*. https://www.investorschronicle.co.uk/2013/09/03/comment/chris-dillow/why-we-can-t-predict-CBw6Z40EZFbzw09qOKBRyK/article.html

Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique*, *57*(3), 238–247. https://doi.org/10.2307/1403797

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Heckerman, D. (1996). A Tutorial on Learning With Bayesian Networks. *Microsoft Research*.

Huang, X. (2009). Economic forecasting based on chaotic optimized support vector machines. *2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, 124–128. https://doi.org/10.1109/CIMSA.2009.5069931

Kreiner, A., & Duca, J. (2020). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters*, *27*(17), 1434–1437. https://doi.org/10.1080/13504851.2019.1688237

Maccarrone, G., Morelli, G., & Spadaccini, S. (2021). GDP Forecasting: Machine Learning, Linear or Autoregression? *Frontiers in Artificial Intelligence*, *4*. https://www.frontiersin.org/article/10.3389/frai.2021.757864

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). 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

Nevasalmi, L. (2020). *Essays on economic forecasting using machine learning*. https://www.utupub.fi/handle/10024/150668

Nilsson, N. J. (1998). Neural Networks. In *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann. https://doi.org/10.1016/B978-0-08-049945-1.50008-3

Patil, A. R., & Subbaraman, S. (2021). Performance analysis of static hand gesture recognition approaches using artificial neural network, support vector machine and two stream based transfer learning approach. *International Journal of Information Technology*. https://doi.org/10.1007/s41870-021-00831-7

Pearl, J. (1998). *Graphs, Causality, and Structural Equation Models*.

Poole, D. L., & Mackworth, A. K. (2010). *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press.

Puga, J. L., Krzywinski, M., & Altman, N. (2015). Bayes' theorem. *Nature Methods*.

Rojas, R. (1996). The Biological Paradigm. In R. Rojas (Ed.), *Neural Networks: A Systematic Introduction*. Springer. https://doi.org/10.1007/978-3-642-61068-4_1

Sarduie, M. H., Kazemi, M. A., Alborzi, M., Azar, A., & Kermanshah, A. (2020). P-V-L Deep: A Big Data Analytics Solution for Now-casting in Monetary Policy. *Journal of Information Technology Management*, *12*(4). https://doi.org/10.22059/jitm.2020.293071.2429

Shlaes, A. (2014). The Wonks Can't Save Us; Economists have gotten a bad rap for failing to predict downturns like the recent recession. But innovation—Not soothsaying—Is their job. *Wall Street Journal (Online)*, n/a.

Soybilgen, B., & Yazgan, E. (2021). Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors. *Computational Economics*, *57*(1), 387–417. https://doi.org/10.1007/s10614-020-10083-5

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press.

Stergiou, C., & Siganos, D. (2006). Neural Networks. *Imperial College London, Department of Computing.*

Sturm, T., Gerlach, J. P., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organizational Learning. *MIS Quarterly*, *45*(3), 1581–1602. https://doi.org/10.25300/MISQ/2021/16543

Templier, M., & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, *37*, 6-.

Tuerck, D. (2014). *Macroeconomics: Integrating Theory, Policy and Practice for a New Era*. Business Expert Press.

http://ebookcentral.proquest.com/lib/jyvaskyla-ebooks/detail.action?docID=1840750

Tutorialspoint. (2017). *Artificial Intelligence – Neural Networks*. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm

Umar, H. A., Fonkam, M., & Prasad, R. (2022). Towards the sustainability of power utilities in Nigeria: A Bayesian network approach. *International Journal of Information Technology*. https://doi.org/10.1007/s41870-022-00876-2

Vähäkainu, P., & Neittaanmäki, P. (2018). *Tekoäly terveydenhuollossa.* Jyväskylä: Jyväskylän yliopisto.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer. https://link.springer.com/book/10.1007/978-1-4757-2440-0#toc

Varian, H. (2014). Big Data: New Tricks for Econometrics. *American Economic Association*. https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218. https://doi.org/10.21037/atm.2016.03.37

Zheng, Z. (Eric), & Pavlou, P. A. (2010). Research Note: Toward a Causal Interpretation from Observational Data: A New Bayesian Networks Method for Structural Models with Latent Variables. *Information Systems Research*, *21*(2), 365–391.

Zohdi, M., Rafiee, M., Kayvanfar, V., & Salamiraad, A. (2022). Demand forecasting based machine learning algorithms on customer information: An applied approach. *International Journal of Information Technology*. https://doi.org/10.1007/s41870-022-00875-3