MACHINE LEARNING IN MACROECONOMICS AND GDP FORECASTING

SUBMITTED BY:

Shivyankar Singh Rathore - 2020MT10851

Contact: 9111607980

Email: shivyankarsinghrathore@gmail.com

SUBMITTED TO:

Prof. Amlendu Kumar Dubey (IIT Delhi)

MACHINE LEARNING IN MACROECONOMICS

ABSTRACT:
Although there has always been a wealth of economic data available, forecasting the state of the economy has proven to be challenging. The accuracy of conventional forecasting models has not been very good. The innovative techniques that machine learning offers have excelled across a wide range of contexts and academic specialties. Economic forecasting has always been attempted using econometric models, however it has been recognised that machine learning is more precise than the more conventional models in its forecasts. GDP, inflation, or interest rates are a few reliable forecasting variables. Time-series analysis has been shown to be less effective than machine learning.

INTRODUCTION:
After economic catastrophes, households seem to take on more debt than they can realistically repay in the future. Economies struggle to recover from recessions. Without much success, economists have sought to foresee significant crashes and recessions throughout history. There is an expanding body of information available regarding the economy. Macroeconomists have struggled with how to manage this massive amount of data. This is precisely where machine learning excels, making predictions based on a wealth of data. The predictions provided by machine learning could aid in preventing economic crises, beginning defenses against high inflation rates, or even just assisting people in preparing for the future. In the future, machine learning ought to be a tool in every economist's toolbox. Machine learning facilitates quick, accurate forecasting on a variety of economic topics. Quick forecasting allows you greater time to plan how to manage an upcoming problem before confronting it. Accurate macroeconomic projections are essential to the effectiveness of central banks' policy decisions and enable more logical economic decision-making. In terms of performance and forecasting capability, the various machine learning techniques vary. In this research, machine learning is discussed in a macroeconomic setting. Economic forecasting is merely one application where machine learning's forecasting abilities are assessed.

1. **Research Questions:** 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?

2. **Structure:** There are four distinct chapters in this report. The introduction to the subject comes first. This chapter will go through the research topic, describe the purpose for the study, and analyse the structure. Machine learning is covered in more detail in the second chapter. In addition to defining machine learning, it also examines some of the algorithms. There will also be a section at the end of the second chapter that goes over the

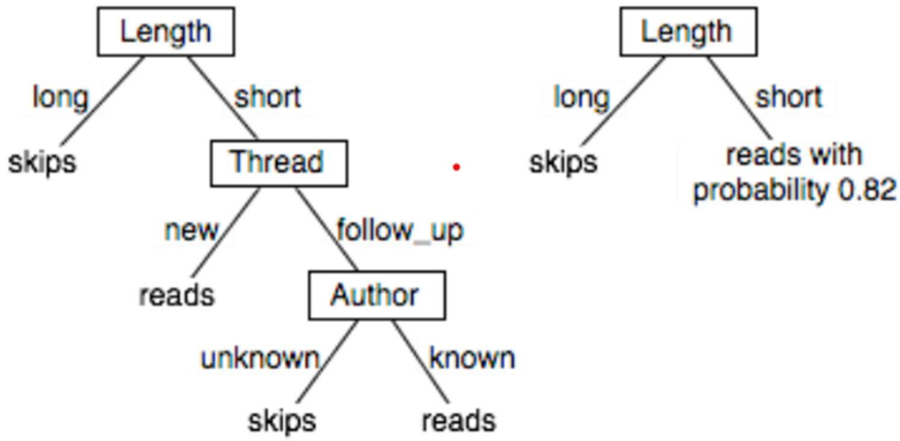
biggest issues with machine learning. The topic itself—predicting macroeconomics using various machine learning techniques—is covered in the third chapter. The primary goal of this chapter is to present the study's findings and provide a little comparison to more conventional approaches. The last chapter is conclusions.

MACHINE LEARNING:

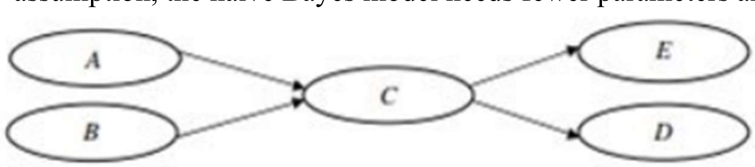
The purpose of this chapter is to provide a thorough overview of machine learning, including its primary characteristics, the most popular machine learning techniques, the reliability of machine learning, and its key drawbacks. First, machine learning will be described in accordance with a few sources, followed by an explanation of its hyponyms and hypernyms. The thesis will next discuss the most popular machine learning techniques, describe their advantages, and describe how to apply them in an economic environment. The key issues with machine learning are discussed in the final part, followed by the biggest drawbacks of the various approaches.

1. **Defining Machine Learning:** Pattern recognition using machine learning has been utilized in anything from face recognition to self-driving cars. But we humans unconsciously engage in these kinds of behaviors. Without employing machine learning, we would not really be able to specify the algorithm we use to recognize the faces of our loved ones and, consequently, we couldn't really programme a computer program for that. However, a learning programme is able to recognize people by identifying patterns in sample face photos of a particular person. Machine learning has been successfully applied in a wide range of industries, from manufacturing with demand forecasting to travel with aircraft scheduling and dynamic pricing. We have only begun to scrape the surface of what ML is truly capable of.

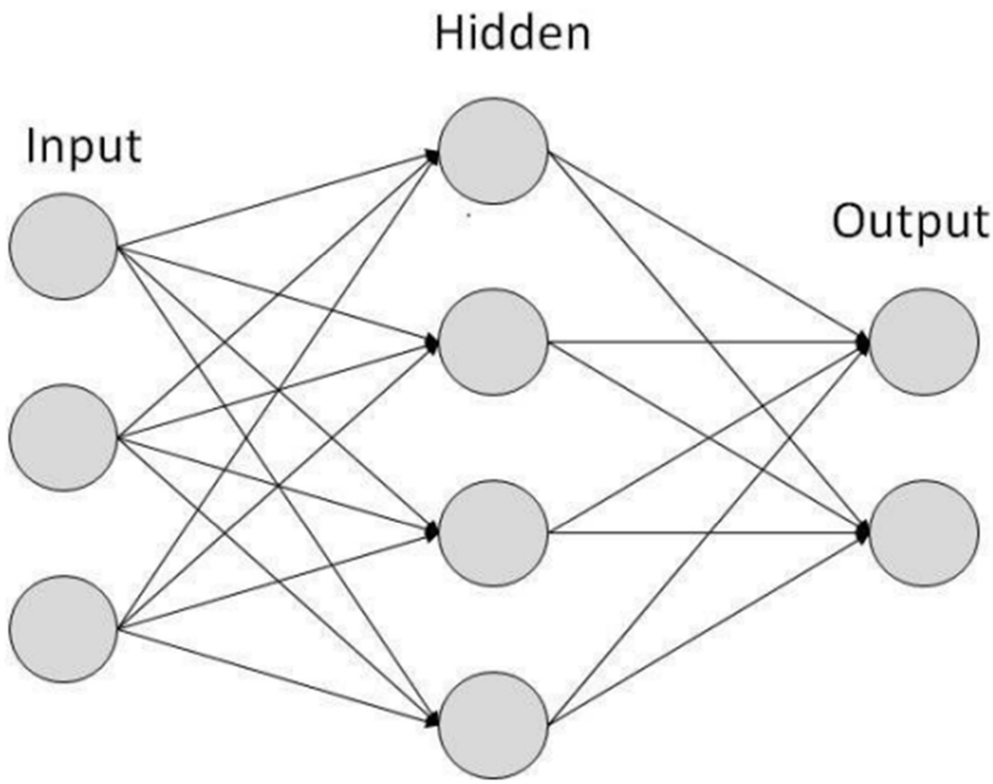
2. **Different Algorithms of ML:** Decision Tree (DT), Random Forest (RF), Bayesian Algorithm or Naive Bayes (NB), Support Vector Machine (SVM), and k-Nearest Neighbor (KNN) are the methodologies mentioned in this study. These several approaches were picked because they frequently appear in the literature on economic forecasting.
    - **Decision Trees and Random Forests:** The most well-known application of machine learning is undoubtedly decision trees. These straightforward binary trees can aid the programme in selecting choices. A learning programme cannot be produced by one tree on its own but can be produced by a forest of trees working together and by planting additional trees. Nodes, leaf nodes, branches, and a threshold unit make up a decision tree. Nodes react to "if this, then that" situations, leading to a certain result in the end. Combining various tree predictors results in random forests. As the number of decision trees in the forest increases, the generalization error of the entire forest converges. Decision trees' power comes from how straightforward they are. They are simple to learn, which frees up time for thorough analysis of the data.



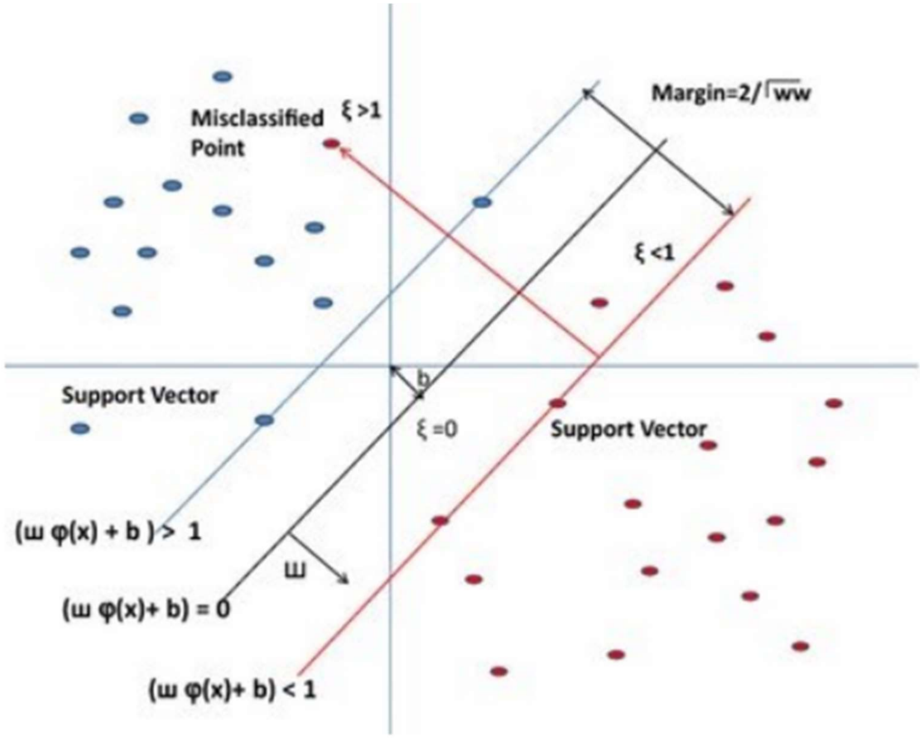
- **Bayesian Algorithms:** While Bayesian statistics imply that the population parameters are measurable random variables and can therefore be characterized by probability, more classical statistics consider that the population parameters are unknown constants and when approximated, the idea of probability is utilized. Graphical structural models called Bayesian networks are used to represent probabilistic interactions between variables. A Bayesian network is composed of nodes, each of which has a parent node and a child node. DAG, which stands for directed acyclic graph, is the term used to describe Bayesian network graphs. A nave assumption about the independence of the predictive variables is all that separates a naive Bayes algorithm from a Bayesian one. The nodes A and B are the parents of C, and C is the parent of the nodes E and D. The simplicity of the Bayes model is also one of its advantages. It has been demonstrated to function successfully when there is noise, missing data, or even irrelevant features. Additionally, because to its "naive" assumption, the naive Bayes model needs fewer parameters and a smaller training set.



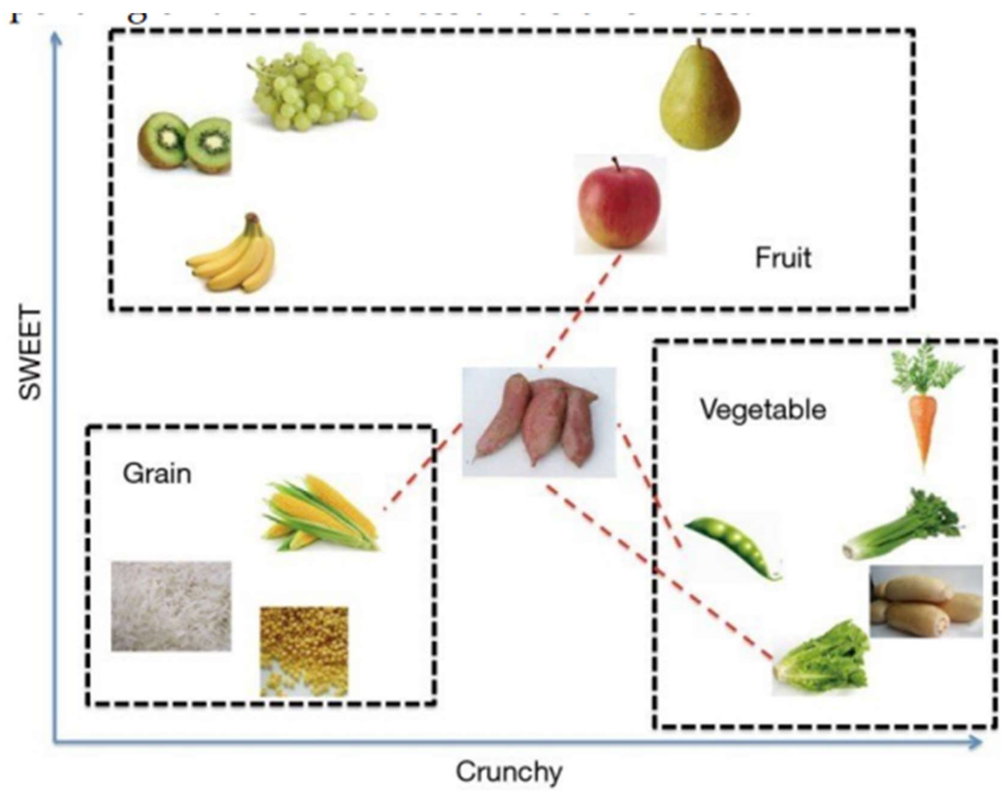
- **Neural Networks:** The original purpose of neural networks was to simulate the information processing of human brains. Like humans, neural networks learn best from examples since learning in biological systems also involves adapting to the synaptic connections between the network's neurons. Many experiments and years of expertise are needed while training neural networks. In nonlinear approximation, neural networks are also frequently used, and when they are used for forecasting, they can resemble a black-box model since they are given a set of inputs, learn from them, and then provide us with an output. Given that the "hidden" mid-layer between the inputs and outputs of neural networks cannot be seen with the human eye, they are difficult to comprehend. A neural network is made up of a large number of neurons connected by intricate signaling channels. These neurons are found in layers inside neural networks, including the input layer, hidden layer, and the output layer. The network could contain additional hidden layers, increasing its complexity. Each of these concealed units is a linear combination of the input variables that are connected by a link. A weighting is used to specify to the neural network the relative importance of each link or input. The objective of educating the network is to choose link weights that minimize mistakes. 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.



- **Support Vector Machine:** They are primarily applied to classification and regression issues. Even though they are not typically utilized in regression, they have recently been fitted there. A support vector machine categorizes observations into just two groups, making it a binary classifier. To provide a reliable estimator for the data, it applies statistical learning theory. A perfect border between observations is built using support vector machines. The main goal of support vector machine training is to find the best line or decision boundary for classifying n-dimensional space into distinct groups so that additional data points can be conveniently added to one of these classes. Support vector machine training attempts to widen the boundary between the two categories as much as possible. There are still many observations that support vector machines cannot yet separate using a hyperplane. The variables (derived from observations) can now be separated in the data after applying this data transformation tool (kernel function) to the data and projecting it into a higher-dimensional space. They claim that the polynomial, Gaussian, and sigmoid kernels are some of the most widely used kernel functions. Support vector machines typically exhibit minimal levels of misclassification and are therefore extremely trustworthy. They also work well with data that has higher dimensions. It is also simple to interpret support vector machines, although some of this interpretability vanishes when using a kernel function because it makes things more difficult.



- **k-Nearest Neighbors:** It is a non-parametric technique for regression and classification. The k-nearest neighbors are used in financial time series modelling because they can spot recurring trends in time series. With a small amount of data, the approach performs better in terms of precision than many other methods. The technique involves grouping unlabeled observations into a class with the labelled instances that share the most similarities. Two features can be used in a two-dimensional plot, and the procedure categorizes and groups all the observations into a group that fits that particular observation.



3. **Why isn't machine learning used more in economic forecasting?** Machine learning is seldom ever employed in research; instead, scientists use more conventional techniques. Several researchers have discovered that some machine learning techniques outperform more established forecasting techniques. Why then isn't machine learning used more widely? The simple explanation is that the researchers are unable to incorporate machine learning into their work. As previously said, some machine learning techniques do call for a significant amount of background knowledge before they can be applied to the study subject. As a result, researchers may decide not to bother learning a completely new method when they can carry out their research just as easily using techniques, they are already familiar with. The design of pre-processing pipelines and data transformation represent a significant portion of the real labour involved in machine learning approaches, which are likewise labour-intensive. All of this eats up time that could be spent conducting the research. The incapacity of machine learning algorithms to extract and organise the discriminative information from the input is one of their main flaws. By categorising variables where their relationships are continuous, decision trees are susceptible to information loss. Additionally, decision trees are not the most flexible of methodologies because even a slight change in the data or a node's position might produce significantly different trees. It should be noted that overfitting can occur with decision trees as well. Given that a random forest connects numerous decision trees, some of these issues can be solved there. The primary flaw in the naive Bayes algorithm is that it assumes independence between the variables that will be used to predict outcomes. In some ways, neural networks resemble a black box model. This indicates that neural networks attempt to learn something from the inputs and build one or more hidden middle layers. It could be challenging to determine the relationships between the inputs and

outputs because the layer is indeed buried. Large data and feature sets are typically the best for neural networks to function. Kernel functions are the main source of problems for support vector machines. Experimenting with a collection of common functions is required when it's necessary. Again, there will be less time for anything else because this testing might take a lot of time. Additionally, there is no guarantee that a standard kernel function will function, thus occasionally even custom ones are utilised. It should be noted that if a method has a particular vulnerability, it can usually be fixed by switching to a different approach entirely. While some of the solutions do share some flaws and restrictions, there should always be one that can overcome a particular one. A single model can be outperformed by averaging the predictions of multiple models.

ML in MACROECONOMICS:

ML is a technology that can be applied to numerous types of study. Economics is one of these disciplines. ML has a lot to offer in terms of economic forecasting. When analyzing data, ML is primarily interested in producing predictions. This section will concentrate on using machine learning to forecast the economy. After a quick definition of macroeconomics, there will be a discussion of the various applications of machine learning in economics. The research on using machine learning to forecast macroeconomics will then be covered in a subsequent chapter. A chapter on nowcasting, which involves making forecasts for both the present and the immediate future, will also be included in this part.

1. **Macroeconomics:** It is important to understand what macroeconomics is before learning how machine learning can be used to forecast the macroeconomy. "The study of the economy as a whole" is macroeconomics. The study of macroeconomics focuses on the interactions between big aggregates like the market for labor and capital or the gross domestic product (GDP). What policies are required to make the economy function well and what policies are required when the economy falters are other macroeconomic concerns. Macroeconomics typically simplifies reality to highlight a problem’s essential components more clearly.

2. **Nowcasting:** Because all data is released with a lag, which implies after events have really occurred, it is crucial to predict events that will occur soon or right now. The primary focus of contemporary econometric analysis is nowcasting. Weather is a fantastic example because it is easy to tell what the weather is like outdoors; therefore, just future weather forecasts are required. Without nowcasting, it would be impossible to understand the macroeconomic trends affecting our economy. If today's data were released two or three months from now, it might already be outdated. In the subject of economic forecasting and machine learning techniques, nowcasting is something that is highly crucial.

3. **Utilizing machine learning in economics:** Establishing an objective and an estimand is the conventional method for conducting economic analysis. The parameter of interest and other nuisance parameters are then estimated using the sample by determining the parameter values that best suit the entire sample. A random sample would then be selected from the entire population. An objective function, such the sum of squared errors, is used for this. The accuracy of the target estimators is the main concern.

When employing machine learning, however, the emphasis is on creating the algorithm. The algorithm's objective is to then generate predictions about some variables based on other factors. Machine learning has the potential to replace many economic models and regressions. Statistics and econometrics are closely related fields. The statistics community has essentially accepted the machine learning revolution. Machine learning has been embraced by the discipline of economics much more slowly than other disciplines. This may be because large data environments have seen success with machine learning techniques. Large data sets may be challenging to collect for research purposes. While correlation analysis and regression are closely connected, regression really characterises the relationship between the relevant variable and the explanatory factors. Finding the line that best passes through a set of specified points is the goal of regression analysis, which can be accomplished in a variety of ways. By utilising logistic regression, classification attempts to group related things together. In classifying data, logistic regression will provide us with a certain "plane" or set of examples. Then, it is possible to categorise observations using this plane or example point. When out-of-sample prediction power is crucial, decision trees and random forests are particularly common methods for estimating regression functions in economics. In terms of economics, single trees are mostly advantageous for their simplicity. Results are simple to interpret and explain. The results are straightforward to interpret because random forests have many benefits with single decision trees, such as the little tuning needed and decent performance right out of the box. It is simple to adapt decision trees and random forests to jobs that need more classification than regression. The objective function that assesses the improvement from a specific split is the primary distinction between decision trees and random forests regression to classification. This function, known as the impurity function in classification, determines how impure each leaf in the tree is. Another adaptable method for calculating regression functions is neural networks. When there are a colossal number of features in the regression, neural networks have been discovered to be quite effective in the field of economics. It should be mentioned, though, that neural networks do need a lot of expertise and fine-tuning to function effectively. The depth of the network significantly boosts the model's adaptability. Researchers have employed neural networks with up to ten layers and millions of parameters in applications. One needs to rigorously regularise the parameter estimations in these situations since there are numerous hidden layers and numerous hidden neurons within them. The performance of multiple-layered neural networks with parameters demonstrates how adaptable neural networks can be when used properly. The potential for economic regressions would be limitless if one had the abilities necessary to employ neural networks flawlessly. The network is extremely complicated and demands a lot of effort and expertise due to its numerous hidden layers. The best application for support vector machines is in classification issues.

4. **Forecasting macroeconomics with machine learning:** Macroeconomic forecasting is challenging because past events may not necessarily indicate how something will turn out in the future. Nevertheless, reviewing past events does provide some pointers for how things will act in the future and provides a solid picture of what may be expected. Macroeconomic indicators include things like GDP, interest rates, and inflation rates. Machine learning has been used extensively to forecast inflation rates, as well as naturally without it. The random forest machine learning method, which

outperforms all other machine learning models, merits more attention. Machine learning can improve inflation forecasting by up to 30% in terms of squared errors. A 30% reduction in errors is a significant improvement, and this indicates that projecting inflation rates will be considerably more precise than in the past. It's interesting to note that several machine learning techniques outperformed the benchmarks they had established, which really illustrates the potential of machine learning. Everybody in the economy needs to be able to predict inflation because it helps people make more informed decisions when the inflation rates are known with more certainty. These forecasts are made possible by machine learning techniques and are the most precise ones. Another macroeconomic metric that has been heavily predicted and is continuously monitored is GDP. Surprisingly, some of the less complex techniques, such the combination of classification and regression trees or even just one tree, can yield extremely accurate predictions. In practise, many of the shortcomings of the single approaches appear to be significantly reduced by ensemble methods. To examine whether the complexity of the network affects prediction accuracy, it would be interesting to compare the prediction accuracy of simple CART approaches to that of artificial neural networks. The relationship between labour economics and macroeconomics is one of the many aspects of the subject that has been extensively investigated. Of course, the employment and unemployment rates play a significant role in labour economics, and consequently macroeconomics as well. Results from artificial neural networks are superior to those from regular linear models. The number of layers, inputs, and outputs of neural networks are the only things that are known about them. Understanding how various approaches make their predictions is always vital, and when the "how" cannot be determined, the model's output must simply be trusted. Artificial neural networks can't generalise as well as support vector machines. As unemployment increases significantly and the outputs of the entire economy decrease, a recession is one of the most important macroeconomic events. The Nave Bayes model outperforms all other models under various experimental settings and evaluation standards. Some models outperformed the Nave Bayes in specific situations when forecasting the near future or even nowcasting. When predicting a recession in real life, it is frequently more advantageous to have a general idea of it at an early stage rather than a forecast that is particularly exact. Macroeconomics can greatly enhance prediction accuracy thanks to machine learning, but there is also more labour involved because machine learning does require some understanding. Macroeconomic forecasting will be impacted by the usage of machine learning because it offers a labour-intensive way to increase prediction accuracy. By employing statistical techniques, the time spent developing the procedure might be put towards identifying the parameters that would best match the entire sample. One could argue that macroeconomic forecasting will include a time-accuracy trade-off. If a researcher does have the time to become familiar with machine learning, the forecasts will perform better.

CONCLUSIONS:

The focus of this report's discussion on machine learning and related techniques for macroeconomic forecasting was on those topics. The macroeconomic setting is also one of the different situations where each of these strategies performs well. Machine learning and its applications to macroeconomic forecasting have been studied. Machine learning techniques were also slightly compared to one another when forecasting the macroeconomic indicators. When presenting several machine learning approaches and comparing them in a macroeconomic context, the major conclusions are that each method should be utilised in a distinct situation and that no one way is necessarily better than another. Some techniques can be more broadly applied. The technique that seems to demand the most training and expertise is neural networks. Knowing how machine learning algorithms generate their predictions may be helpful in some situations. While a neural network has layers that can't be read simply, decision trees are genuinely simple and straightforward to comprehend. In fact, the main goal of contemporary econometric analysis is nowcasting. Nowcasting shifts from causal inference to the field of machine learning because of the significant improvement in predictions made possible by the application of machine learning techniques. How can machine learning be used in macroeconomic forecasting is the main query. By selecting a technique that is most suited for that kind of data collection and educating it about the data, machine learning can be used. Some techniques also need some extra fine-tuning. Teaching can take a lot of time depending on the approach used, and some methods have other restrictions that must be taken into consideration. The approach should be able to produce predictions based on the data after learning from the prior data, and the accuracy of the predictions can then be assessed using several least errors methods. Predictions of macroeconomic indices like the GDP or inflation rates can be made using the approaches. Are machine learning approaches better at predicting the macroeconomy than those employed previously? is a subject that has been addressed through research on macroeconomic forecasting using machine learning. Modern machine learning techniques outperform conventional regressions, other conventional models, as well as surveys of expert forecasters. Machine learning techniques did occasionally outperform the benchmark, but this was mostly because they had received insufficient training. Inflation, GDP, and unemployment rate were macroeconomic metrics that machine learning anticipated more accurately. It could be more useful to measure which models predict recessions the earliest rather than which models do it most accurately. We move on to the third research topic if machine learning really does do better in macroeconomic forecasting. The question is, "Why isn't machine learning used more in economic forecasting?" All economists should have machine learning in their toolbox for analysis, but doing so would need extensive multidisciplinary understanding of both economics and machine learning. This might be the biggest difficulty preventing machine learning from being used for economic forecasting. To conduct research using both machine learning and economics, researchers lack the necessary knowledge in both fields. Future study on economic forecasting using machine learning might concentrate more on the techniques than the actual problems they are applied to. It is crucial to understand why some strategies are effective for particular types of data sets. For the time being, it appears that a few ways are picked nearly at random because teaching various methods typically takes a lot of time and choosing one over the other is typically not justified that much. Someone more knowledgeable with both econometric models and machine learning models could potentially make a better comparison between them. a comparison that focuses more on the process than just the outcomes.

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.

REFERENCES:

1. Sebastian Nyholm. MACHINE LEARNING IN MACROECONOMIC FORECASTING
2. Alpaydin, E. (2020). Introduction to Machine Learning, fourth edition. MIT Press.
3. 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
4. Begg, D., Vernasca, G., Fischer, S., Dornbusch, R., & Begg, D. (2014). Economics (Eleventh edition). McGraw-Hill Education.
5. 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
6. 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. Macroeconomic Nowcasting and Forecasting with Big Data | Annual Review of Economics (annualreviews.org)
7. 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
8. Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/BF00058655
9. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
10. 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

11. 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
12. 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
13. Dillow, C. (2013). Why we can't predict. Why we can't predict - Investors' Chronicle (investorschronicle.co.uk)
14. 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
15. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. The MIT Press.
16. Heckerman, D. (1996). A Tutorial on Learning with Bayesian Networks. Microsoft Research.
17. 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
18. 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
19. 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
20. Nevasalmi, L. (2020). Essays on economic forecasting using machine learning. https://www.utupub.fi/handle/10024/150668
21. 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
22. Puga, J. L., Krzywinski, M., & Altman, N. (2015). Bayes' theorem. Nature Methods.
23. 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
24. 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
25. Stergiou, C., & Siganos, D. (2006). Neural Networks. Imperial College London, Department of Computing.
26. 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
27. Tutorialspoint. (2017). Artificial Intelligence—Neural Networks. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm

28. 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
29. 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
30. Varian, H. (2014). Big Data: New Tricks for Econometrics. American Economic Association. https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3
31. 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
32. 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. Demand forecasting based machine learning algorithms on customer information: an applied approach | SpringerLink

FORECASTING OF INDIAN GDP

INTRODUCTION:

Machine learning can be used to evaluate a vast amount of data to make the best financial decisions. In a flourishing country like India, economic measures play an essential role in commercial and governmental sectors when making judgments. Economic forecasting helps business executives, officials, and investors make better decisions. As a result, making accurate projections about these measures is increasingly critical. In its most fundamental form, econometrics is nothing more than economic statistics. Machine learning is used in economics to accomplish a similar goal, but with massive amounts of data. Keeping this in mind, we examine the conceptual framework and apply it to the forecasting of India's gross domestic product (GDP). Gross domestic product (GDP) is a measure of the total worth of goods and services generated inside a country's borders over a year. Indicators such as GDP growth rate are critical to assessing a country's economic health. GDP can tell us if the economy is in a recession, depression, or boom. The GDP is a comprehensive indicator of the country's total economic output.

$$\mathbf{GDP = C + I + G + (X- M)}$$

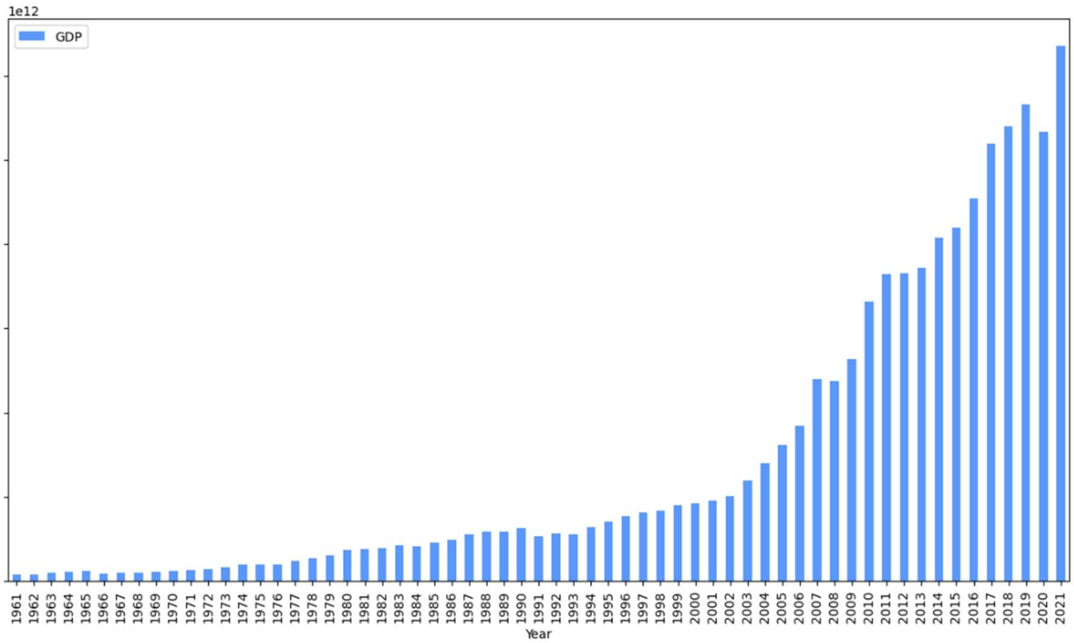
- C = Annual Consumption (Personal Consumer Expenditure)
- I = Gross Private Domestic Investment
- G = Government Spending
- X = Total Amount of Exports
- M = Total Amount of Imports
- (X-M) = Total Net exports

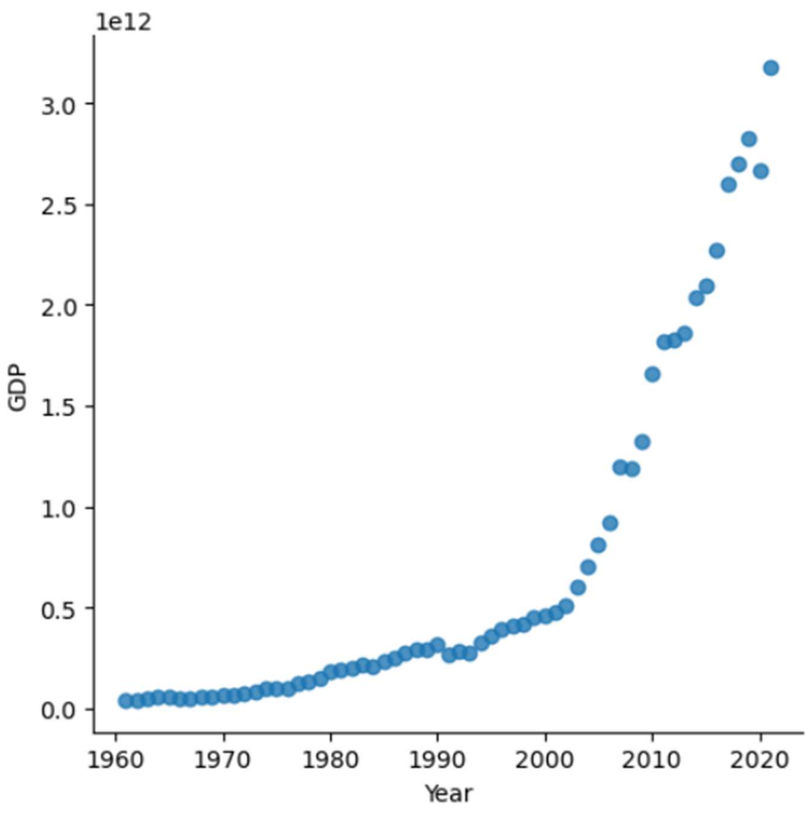
METHODOLOGY:

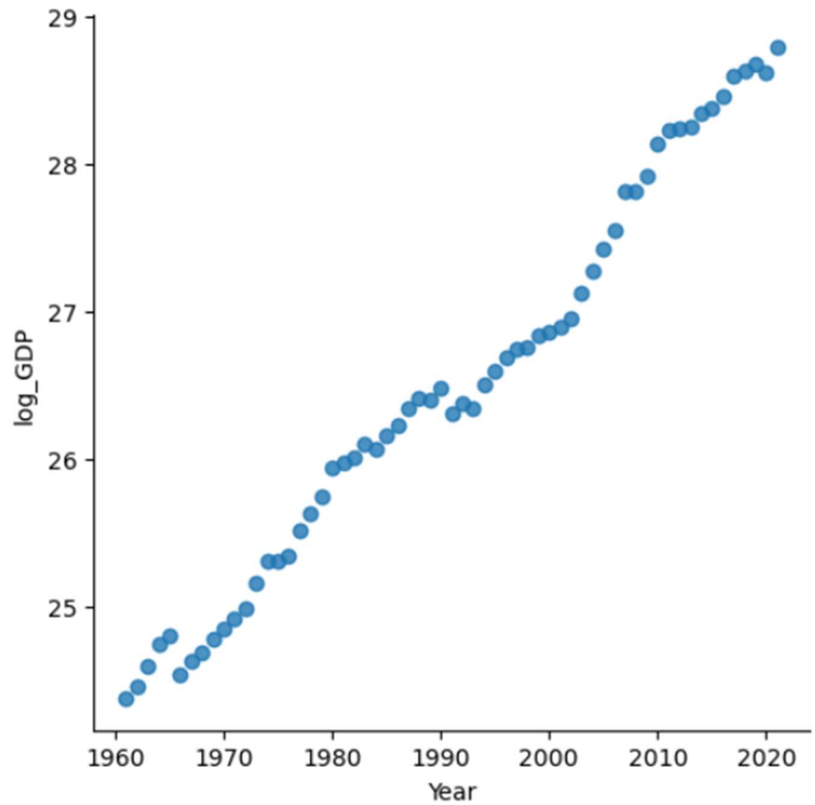
1. **Data Collection:** The first step is to determine what sort of data is needed to address a specific problem and whether or not there are any privacy problems associated with the data. It is the initial stage in the process of developing a machine learning model. To conduct our research, we used World Bank open source data.

2. **Data Preparation:** Data preprocessing is the process of extracting raw data, which is data that has been acquired in the actual world and has been turned into a clean data collection before it is used. A series of procedures were carried out to reduce the data to a manageable amount. Null values have been dropped and filled up throughout our work.

3. **Data Visualization:** In data visualization, we can observe how the data appears and what type of association exists between its many characteristics. It's the quickest approach to verify if the attributes match the output. We've utilized Python packages like Matplotlib and Seaborn to show data in an appealing way.


1e12
GDP
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021
Year


1e12
3.0
2.5
2.0
1.5
1.0
0.5
0.0
GDP
1960
1970
1980
1990
2000
2010
2020
Year



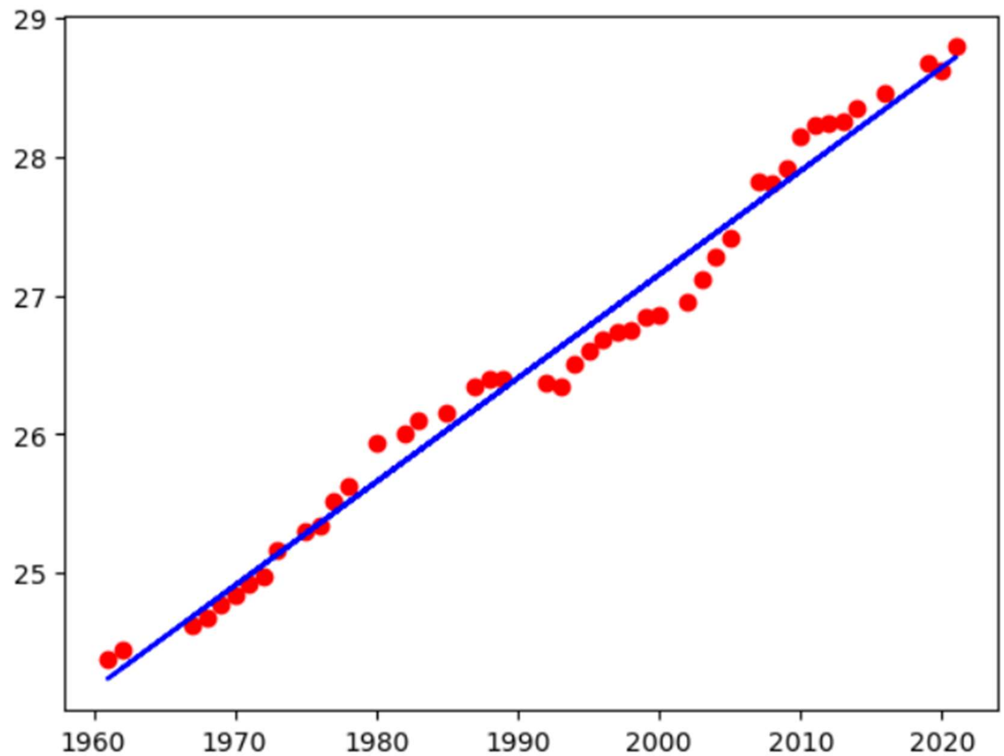
4. **Data Analysis:** For GDP prediction, we have made use of Linear Regression algorithm and Random Forest Regression algorithm.
   - **Linear Regression:** Linear Regression is a supervised learning-based machine learning technique. Variable correlation and forecasting are two of its most common applications. There are a variety of different regression models, each of which has a different number of independent variables and a different correlation between the dependent and independent variables.
   - **Random Forest Regression:** A random forest is a meta estimator that employs aggregating to increase predicted accuracy and mitigate over-fitting by fitting a number of classification decision trees to various subsets of the dataset.

5. **Model Training:** When developing algorithms, model training is the stage of the data science development lifecycle in which researchers attempt to find the optimal mix of weights and biases for the methodology to limit a loss function throughout the predicted range. There are 70% training and 30% testing datasets in our research.

6. **Model Evaluation:** Model evaluation is the method of assessing a machine learning model's progress and identify any flaws. Our model compares predicted data to actual data. Among the metrics we utilized for our assessments were Mean Squared Error (MSE).

7. Results:

Year	Actual GDP	Predicted GDP
2022	3.5e+12 $	3.19e+12 $
2023	---------------------	3.44e+12 $



CONCLUSION:

In this study, we provide a comparison of two machine learning algorithms, Multiple Linear Regression algorithm and Random Forest Regressor for the prediction of GDP of India. The results of our analysis show that the Random Forest Regressor algorithm outperformed the Multiple Linear Regression. In a developing country like India, evaluating economic metrics such as GDP plays an important part in taking decisions, both in private sectors as well as public sectors. In this study, we have used machine learning algorithms to predict one of the key economic metric GDP (in current USD) with the data obtained from the World Bank.

	Linear Regression	Random Forest
Mean Squared Error	0.032	0.013

REFERENCES:

1. Economy of India, Wikipedia
   https://en.wikipedia.org/wiki/Economy_of_India#cite_note-53
2. Yoon, J. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. Comput Econ 57, 247–265 (2021)

3. Padmawar, Vaishnavi & Pawar, Pradnya & Karande, Akshit. (2021). Gross Domestic Product Prediction using Machine Learning. 08. 2395-0056.
4. Bang, James & Sen, Tinni & Basuchoudhary, Atin. (2015). New Tools for Predicting Economic Growth Using Machine Learning: A Guide for Theory and Policy.
5. World Bank Dataset - https://data.worldbank.org/country/IN
6. Government of India Dataset - https://data.gov.in/resources/indian-economy-selected-indicators-2010-11-2016-17
7. Kaggle Dataset - https://www.kaggle.com/paree24/india-gdp-growth-world-bank-1961-to-2017