

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 2, February 2020

Gross Domestic Product Prediction Using Machine Learning: A Survey

Ashwini Topre¹, Rajesh Bharati²

Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India¹

Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India²

ABSTRACT: Gross Domestic Product is an total estimate of country's economy, in terms of total productive output. It is of great importance to policy makers and central banks to know its trend and predict possible changes. It is of huge importance to coverage policy makers and central banks to know its trend and predict possible changes. This gives accurate assessment of the future state of the economy – whether it is heading toward expansion or contraction – such that preemptive actions can be taken as appropriate. Reliable forecasting of GDP is a very important tool in the macro- economic toolbox. Unfortunately, national economies are very complicated and it isn't clean to version them accurately, not to mention expect their behavior. GDP forecast are very uncertain and unavoidably short-term.

In this work we plan to use decision tree, bootstrap bagging and random forest. The motive for using these algorithms is that the random forest combines trees with the belief of an ensemble. Thus, the trees are weak learners and the random forest is a strong learner. The runtime of random forest is pretty fast, and they are able to handle an unbalanced and missing data. Bootstrap bagging reduces variance and helps to avoid over fitting. And by through the use of decision tree we can predict the classes for our training data.

In this work we plan to implement a system that will help in GDP prediction. Manually it is very difficult and time consuming, tedious work to analyze data of whole country's economical progress based on total production in country and all investments. We can take help of Machine Learning in this task. This will make our job easy and will give results in much less time.

KEYWORDS: GDP, Machine Learning, SVM, Economic Indicators, per capita GDP.

I. INTRODUCTION

Gross Domestic Product (GDP) is a total estimation of the economy, regarding total productive output. It is very vital to policy makers and central banks of any country to know the country's current trend and predict possible changes. This provides accurate calculation of the future condition of the economy of respective country – whether country is heading toward growth or recession, so that necessary actions can be taken to increase the growth rate. Accurate prediction of GDP is a very critical. Unfortunately, national economies are profoundly complex and it is not easy to model them accurately, let alone predict their behavior. Like weather forecasts, GDP forecasts are inherently uncertain and inevitably short-term.

Gross Domestic Product (GDP) calculates the value of all goods and services produced in a country in a fixed period. GDP is defined as the sum of consumption expenditures, investments, and the trade balance (the difference between export and import). General way to calculate GDP of any country is:

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

Where,

C = Private consumption

I = Gross Investment

G = government investment

X = Exports

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 2, February 2020

M = Imports

Benefits of calculating GDP are as follows:

- Gross domestic product checks the condition of a country's economy.
- It produces the value of all goods and services produced over a specific time period within a country.
- Economists can use GDP to predict whether an economy is growing or experiencing a recession.

II. LITURATURE SURVEY

1. **Accuracy Prediction for Distributed Decision Tree using Machine Learning approach** - In this paper author propose an implementation of distributed decision tree in Spark environment. The implementation performs horizontal data parallelism in the shared-nothing architecture of Spark. It is scalable to large dataset. Research on distributed decision tree effectively established the fact that parallelizing the time consuming tasks in an algorithm and careful design considering the scalability can improve the performance significantly.
2. **Improvement of Random Forest Cascade Regression Algorithm and Its Application in Fatigue Detection** – In this paper author proposed an improved facial feature points detection method based on random forest cascade regression. Aiming at the problem of face feature point localization, the face feature point detection model is first trained through the selected sample library, and then the model obtained by training is used to locate the feature points of the face. The experimental results showed that the improved algorithm that was proposed here can accurately and quickly locate the feature points of the eye and mouth contour, and is good in recognizing the rotation of the face. This work can be used to analyze the driver fatigue state.
3. **An idea of a clustering algorithm using support vector machines based on binary decision tree** - In this paper, author proposed a novel clustering algorithm based on binary decision trees and support vector machines. Furthermore, the idea on which the algorithm was based minimized the dissimilarity within each of the produced clusters, for the simple reason that the most dissimilar clusters should be separated at the very beginning of the tree because they are easy to separate.
4. **Difficulties of Marking Decision Tree Diagrams** - This paper focused on the difficulties of marking decision trees, especially the trees which are based on paper. Author showed that if the decision tree is presented on paper, then it is difficult to see and compare similar questions consistently. The paper provided problems of designing decision trees and details of the problem in marking decision trees; the suggested solutions to those problems are identified. In this paper, the recommendations for the difficulties of marking the decision tree have been identified by the author. One of which is to automate decision trees.
5. **A Comparison between Support Vector Machine (SVM) and Bootstrap Aggregating Technique for Recognizing Bangla Handwritten Characters** – This paper presented a comparison between SVM and Bootstrap Aggregating. There was some problem regarding similar characters. Experiment gave 89.8% accuracy for SVM classification model and 93% for Bootstrap Aggregating model.

III. ALGORITHMIC SURVEY

A. Random Forest

Random forests forests is an method used for classification, regression and other tasks that works by constructing many decision trees and giving the class that is the mode of the classes (in case of classification) or mean prediction (in regression) of the individual trees. Random forests overcome the decision trees' habit of overfitting to their training set.

International Journal of Innovative Research in Science, Engineering and Technology

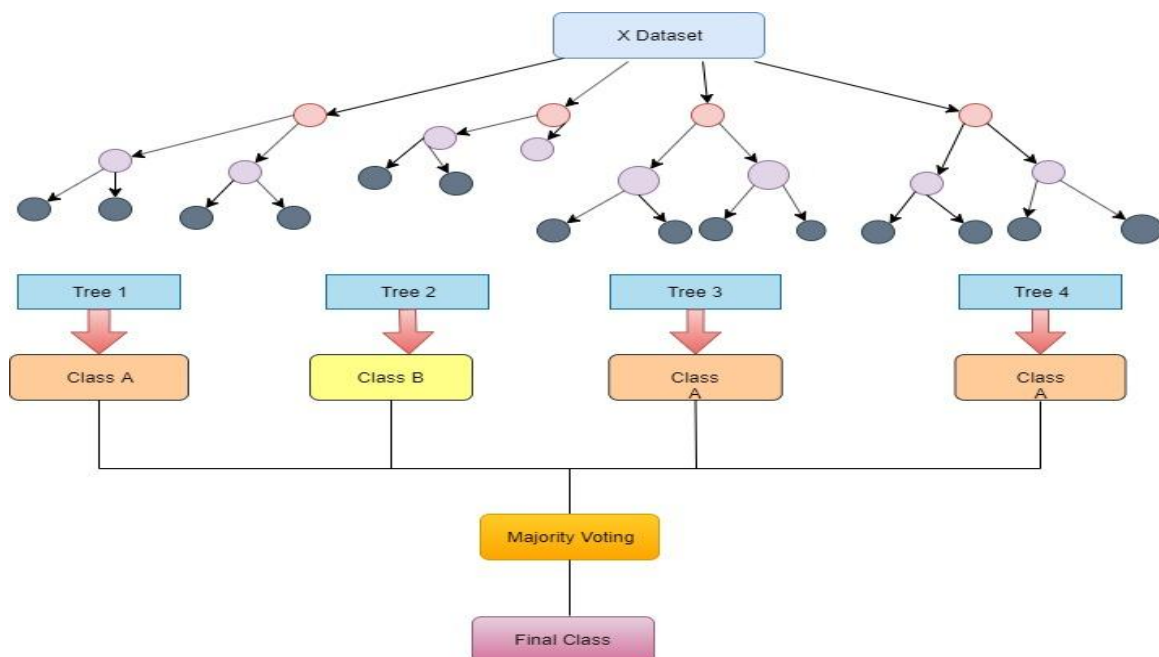
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 2, February 2020

This is a supervised classification algorithm. Task is to create a forest by some way and make it random. The number of trees in the forest and the results it can get are directly proportional: the larger the number of trees gives the more accurate the result. But creating the forest is not the same as constructing the decision with information gain or gain index approach.

The Random Forests Algorithm handles some of the limitations of Decision Trees Algorithm, which is - the accuracy of the outcome decreases when the number of decisions in the tree increases. So, in the Random Forests Algorithm, there are multiple decision trees that represent various statistical probabilities. All of these trees are mapped to a single tree known as the Classification and Regression Trees. In the end, the final prediction for the Random Forests Algorithm is obtained by majority of the results of all the decision trees.



B. Bootstrap

The bootstrap method is a statistical technique for estimating quantities about a population by using estimates from multiple small data samples. Importantly, samples are constructed through drawing observations from a large data sample one at a time and returning them to the data sample after they have been chosen. This allows a given observation to be included in a given small sample more than once. This technique of sampling is called sampling with replacement.

The process for building one sample can be summarized as follows:

1. Choose the size of the sample.
2. While the size of the sample is less than the chosen size
 1. Randomly select an observation from the dataset
 2. Add it to the sample

The bootstrap method can be used to predict a quantity of a population. This is done through continuously taking small samples, calculating the statistic, and taking the average of them. We can summarize this procedure as follows:

The procedure can also be used to check the efficiency of a machine learning model.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 2, February 2020

1. Configuration of the Bootstrap

There are two parameters that must be selected when performing the bootstrap: the size of the sample and the number of repetitions to perform.

2. Sample Size

In machine learning, it is common to use a sample size that is the same as the original dataset. If the dataset is huge and computational efficiency is difficult, smaller samples can be used.

3. Repetitions

The number of repetitions must be large enough to guarantee the meaningful statistics, like mean, standard deviation, and standard error can be calculated on the sample.

Smaller values can be used will further add variance to the statistics calculated on the sample of estimated values.

Generally, the sample of estimates would be as large as possible given the time resources, with hundreds or thousands of repeats.

C. Decision Tree

Decision tree is one of the predictive modelling techniques used in statistics, data mining and machine learning. Decision trees are constructed through a criterion that finds ways to split the data based on various conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are supervised learning method used for both classification and regression tasks.

Trees in which the target variable can take a different set of values are called classification trees. Decision trees in which final variable can take continuous values are called regression trees. Decision tree learning is a method basically used in data mining. The goal of decision tree is create a model that predicts the value of a target variable based on various input variables. Each internal node is one of the input variables; there are edges to children for each of the possible results of that input variable. Each leaf gives a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

A decision tree is a used for classification. Each element is called as a class. A decision tree is a tree in which each internal node is labelled with an input feature. The arrows coming from a node labelled with an input feature are labelled with each of the possible values of the result or the arrow leads to a decision node on a different input feature. Each leaf of the tree is categorized with a class or a probability distribution over the classes, showing that the data set has been classified by the tree into either a specific class, or into a particular probability distribution.

A tree is built by means of splitting the source set, constituting the root node of the tree, into subsets - which represents the children. The splitting is based on a set of splitting rules based on classification features. This method is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is finished when the subset at a node has all the same values of the final variable, or when splitting no longer adds value to the predictions. This process of top-down construction of decision trees is type of a greedy algorithm, and it is the most common strategy for learning decision trees.

IV. PROPOSED SYSTEM

We are here proposing system that will be based on Decision Tree, Random Forest and Boot Strap algorithms and will forecast the GDP of Nation.

We will use data of Agriculture Production, Manufacturing Industry and Service Industry. These are the sectors of main stream that are responsible for the effect on GDP forecasting. We will use multiple algorithms for increasing and improvement of accuracy.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 2, February 2020

V. CONCLUSION

Here we concluded that machine learning which is all new and latest technology which is still in research phase is able to help us to learn and make Forecasting of huge data in terms of time series forecasting. Various Machine Learning algorithms are build and implemented with specific special and dedicated characteristics and properties. For our research Decision Tree, Random Forest and Bootstrap Bagging are provably and effectively satisfying the work of predicting future movement of GDP of country with given data.

REFERENCES

- [1] Siddalingshwar Patil, Umakant Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach" 978-1-5386-9439-8/19/\$31.00 ©2019 IEEE
- [2] Wanshan Peng, Shihua Li, Ze He, Silan Ning, Yuhan Liu, Zhonghua Su, "random forest classification of rice planting area using multitemporal polarimetric radarsat-2 data" 978-1-5386-9154-0/19/\$31.00 ©2019 IEEE
- [3] Juae kim, Young, Joong ko, Jungyun seo, "A Bootstrapping Approach With CRF and Deep Learning Models for Improving the Biomedical Named Entity Recognition in Multi-Domains" 2169-3536 2019 IEEE
- [4] Tao Qunzhu, Zhang Rui, Yan Yufei, Zhang Chengyao, Li Zhijun, "Improvement of Random Forest Cascade Regression Algorithm and Its Application in Fatigue Detection" 978-1-7281-1618-1/19/\$31.00 ©2019 IEEE
- [5] Halima Elaidi, Younes Elhaddar, Zahra Benabbou, Hassan Abbar, "An idea of a clustering algorithm using support vector machines based on binary decision tree" 978-1-5386-4396-9/18/31.00\$ ©2018 IEEE
- [6] Warda Mohamed Al-Hoqani, "Difficulties of Marking Decision Tree Diagrams" Computing Conference 2017 18-20 July 2017 | London, UK
- [7] John Roush, Keith Siopes, Gongzhu Hu, "Predicting Gross Domestic Product Using Autoregressive Models" 978- 1-5090-5756-6/17/\$31.00 ©2017 IEEE SERA 2017, June 7-9, 2017, London, UK
- [8] Warda Mohamed Al-Hoqani, "Difficulties of Marking Decision Tree Diagrams" Computing Conference 2017 18-20 July 2017 | London, UK
- [9] Asish Kumar Ghosh, Shyla Afroge, "A Comparison between Support Vector Machine (SVM) and Bootstrap Aggregating Technique for Recognizing Bangla Handwritten Characters" 978-1-5386-1150-0/17/\$31.00 ©2017 IEEE