International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I6.138 ISSN: 1308-5581 Vol 14, Issue 05 2022

PANDEMIC IMPACT OF FINANCIAL PREDICTION USING HIERARCHICAL CLUSTER ANALYSIS AND BOOSTED RANDOM FOREST

Mr.Gangireddy Prabhakar
Reddy
Asst. Professor
Dept. of CSE
St. Peter's Engineering College
prabhakar.sp17@gmail.com

Kalpana Kumari
IV B. Tech
Dept. of CSE
St. Peter's Engineering College
Hyderabad ,TS ,India
kalpanakri44@gmail.com

Nayini Sai Chethan Reddy
IV B. Tech
Dept. of CSE
St. Peter's Engineering College
Hyderabad ,TS ,India
chethan4331@gmail.com

Keerthi Tejaswini
IV B. Tech
Dept. of CSE
St. Peter's Engineering College
Hyderabad, TS ,India
keerthitejaswini25@gmail.com

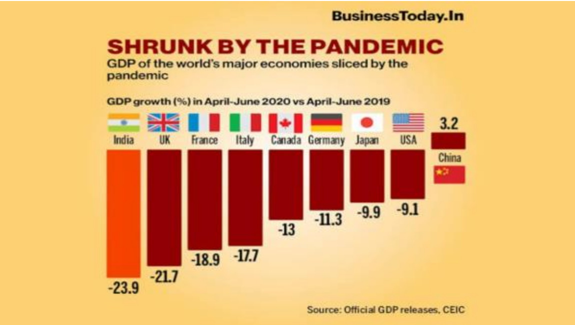
M. Sai Kumar
IV B. Tech
Dept. of CSE
St. Pete's Engineering College
Hyderabad ,TS ,India
Saivadderaj143@gmail.com

ABSTRACT- Economic inequality has historically been significantly impacted by pandemics. The pandemic COVID-19's global spread has drastically altered the political, social, economic, religious, and financial systems of the entire planet. The world's top 10 economies—the US, China, Japan, Germany, the UK, France, India, Italy, Brazil, and Canada—are all at risk of going bankrupt. Global economic growth is being significantly impacted by the pollution crisis. Global trade may potentially experience a 13–32% fall, depending on the depth and breadth of the global economic downturn. This study examines how COVID-19 has affected stock prices and economic growth. Principal component and hierarchical cluster analyses, as well as a customised Random Forest model enhanced by the Ada Boost algorithm, were used to examine the data sets. Even on unbalanced datasets, the suggested approach Boosted yields reliable predictions.

Keywords:
Pandemic, Social , Economic, Financial, States, Boosting Algorithm, Data Analysis, Random Forest, GDP Forecast.

I. INTRODUCTION

In Wuhan, China, COVID-19 made its initial appearance in December 2019 and ha Numerous countries around the world have enacted varied degrees of blockade and closure measures during the viral pandemic, including major commercial closures and stringent restrictions on the flow of people and products. The analysis above demonstrates that COVID-19 has had a terrible effect on the global economy. For instance, the COVID-19 outbreak in India has primarily disrupted the nation's economy. The ministry of Statistics reports that India's growth decreased to 3.1% in the final quarter of the fiscal year 2020. A significant influence on both society and daily life there. The new corona virus is now a widespread pandemic disease as a result of its ongoing propagation. By June 20, 2022, the new corona virus had infected more than 544 million individuals, killed more than 6.35 million of those victims, and seriously harmed both human lives and material wealth.


SHRUNK BY THE PANDEMIC
GDP of the world's major economies sliced by the pandemic

Source: Official GDP releases, CEIC

This study examines how COVID-19 has affected stock prices and economic growth. Principal component and hierarchical cluster analyses, as well as a customised Random Forest model enhanced by the Ada Boost algorithm, were used to examine the data sets. Even on unbalanced datasets, the suggested approach Boosted yields reliable predictions.

Therefore, it is crucial to understand how to use current technology to forecast COVID-19's future effects on the economy and assess how such effects affect economic development. Numerous sectors have employed machine learning algorithms to address some real-world issues as a result of the vast data accumulation and the quick advancement of computer technology. Therefore, utilising pertinent data from 170 nations and areas collected during the COVID-19 era, this article's goal is to forecast the GDP of different countries and regions. In order to obtain a more precise approach, we are using hierarchical cluster analysis and the Random Forest Model Boosted by Ada boost algorithm in this research. The findings demonstrate that the proposed model outperforms all existing models in calculating the GDP of each nation or region.

In order to forecast GDP values, we also run experiments utilising the KAGGLE competitive data set, which includes GDP-related data from 170 nations and regions. In this work, the data is processed, models are created, and Python is used extensively.

In order to guide future efforts, the characteristics that have the biggest impact on the economy during the epidemic era are identified by completing a visual analysis of the correlation between different parameters and GDP.

The methodology's accuracy in predicting GDP has been demonstrated by experimental data.

The following can be used to summarize the contributions of our paper:

• To create a new prediction model in this study that can forecast and assess changes in GDP in various countries or areas, we employed machine learning techniques such Random Forest, ADABoost, and hierarchical cluster analysis. The model can accurately predict GDP value using several data facets.

• The MSE value is used to gauge how well the connected algorithm can predict the GDP of other nations. Our suggested approach surpasses every single algorithm, including random forest and XG Boost, according to the results.

Analyzed are the main elements that impacted a nation or region during the pandemic period. Data analysis also demonstrates the relationship between the data, supplying a data foundation for upcoming work and the economic recovery.

• The COVID-19 risk and vulnerability index:

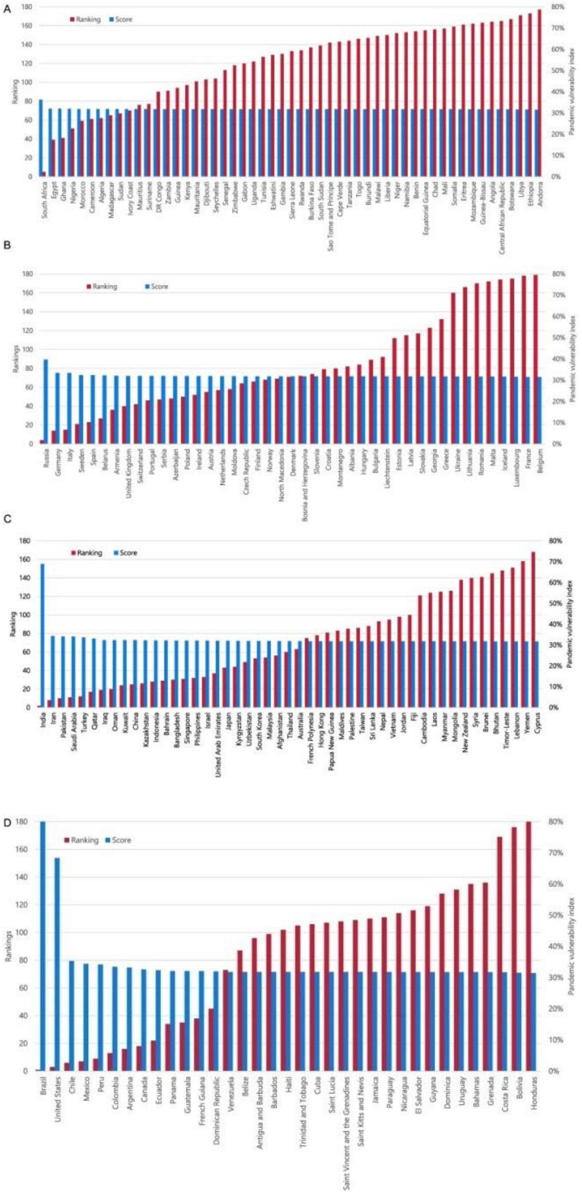
A. Africa

B. Europe

C. Asia and Oceania

D. Americans

- Ranking
- Vulnerability Index Score



II. FEATURE ENGINEERING

The source, make-up, and data preprocessing procedures of the data collection are principally covered in this part. Information about COVID-19's effects on the global economy was taken from the KAGGLE competition platform for the data set utilized in this essay. The global economic data set consists of information on 170 nations and regions over 10 different metrics.

Each day, the pertinent information for each nation or region is saved in one of them. First, we address the dataset's missing values and outliers, then remove the three columns of null values.

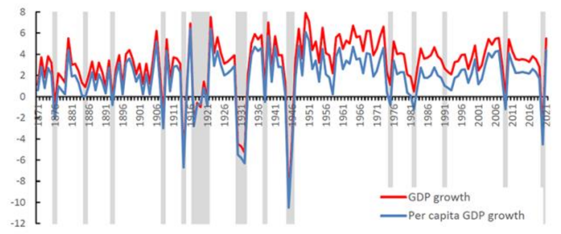


Figure 1. Global GDP growth in a historical perspective

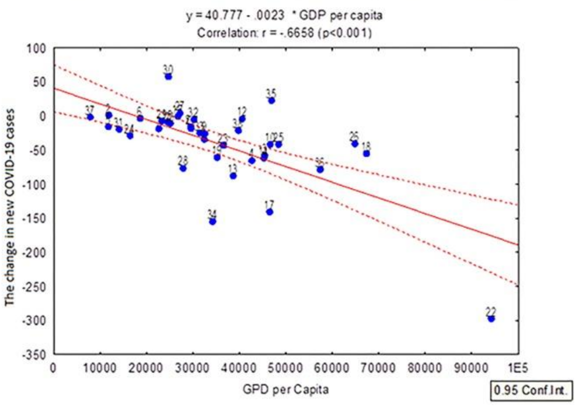
Sources: Bolt et al. (2018), Kose, Sugawara, and Terrones (2019, 2020), and IMF-WEO Apr-2021. Shaded areas refer to global recessions.

This section describes the data set's origin, composition, and data preprocessing procedures in great detail. Information about COVID-19's effects on the global economy was taken from the KAGGLE competition platform for the data set utilized in this essay. Data from 170 countries and regions are included in the global economic data set, which covers 10 main factors. Each day, the pertinent information for each nation or region is saved in one of them. After dealing with the missing values and outliers in the dataset, we remove the three columns of null values.

The results would be significantly impacted by the fact that the dataset's various fields have highly varied data densities.The data in the data set are standardized before the prediction in order to reduce error. The aforementioned data sets are then processed using feature engineering. This study examines how COVID-19 affects stock prices and economic growth.

The data sets were analyzed using principal component and hierarchical cluster analysis, as well as a tailored Random Forest model improved by the Ada Boost technique. The suggested method, Boosted, produces accurate predictions even on unbalanced datasets.

The month, day of the week, weekend day, etc. are examples of temporal aspects that can be extracted from data after it has been preprocessed and compressed into smaller floating-point integers. The statistical features for a certain time period are then used to retrieve the characteristics of the lag, the number of infection and death cases, the characteristics of the maximum value, the characteristics of the minimum value, and the characteristics of the median value. The next stage is to use the rfecv method to choose features, and with each training cycle, certain unimportant traits are eliminated.



The change in new COVID-19 cases per million population plotted against GDP per capita

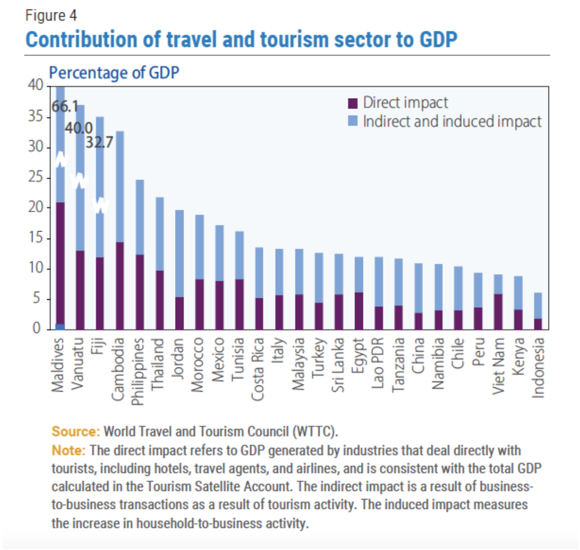
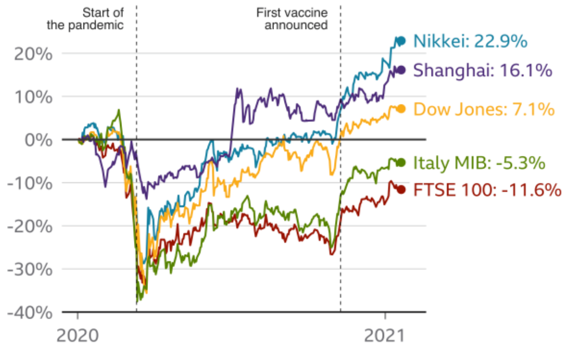


Figure 4
Contribution of travel and tourism sector to GDP

Source: World Travel and Tourism Council (WTTC).
Note: The direct impact refers to GDP generated by industries that deal directly with tourists, including hotels, travel agents, and airlines, and is consistent with the total GDP calculated in the Tourism Satellite Account. The indirect impact is a result of business-to-business transactions as a result of tourism activity. The induced impact measures the increase in household-to-business activity.

III. ALGORITHM MODELS

The impact of coronavirus on stock markets since the start of the outbreak



Source: Bloomberg, 24 January 2021, 00:01 GMT BBC

The Random Forest Method, for starters, is one of the most popular and effective supervised learning techniques based on decision trees. The key concept is to sample the same data set again in order to create N sub datasets with the same amount of data as the original data set. Each sub dataset's decision tree is built, and the ultimate anticipated result is obtained by casting a vote for each sub dataset's associated result. Ada Boost and XG Boost are two boosting algorithms used in machine learning. The AdaBoost method's main idea is to repeatedly train the same training data to generate different weak classifiers.

The sample weights are then recreated in order to include the aforementioned weak classifiers into a strong classifier.

Following are the steps needed to solve an Ada Boost algorithm regression problem.

$D(1)=(w11,w12,w13,\ldots..w1m)$

$W1i=1/m;$

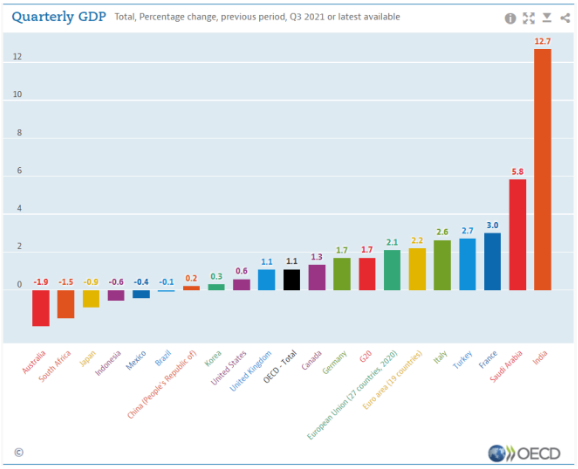
where m is the data volume of the sample set and D (1) is the initial sample weight.

The sample set with weight Dk is then used to train the data to create the weak learner Gk (Xi).

$eki=(Yi-Gk(Xi))2/E2$,

the square error for each sample, where Ek is the maximum error on the training data.

IV. EXPERIMENTS



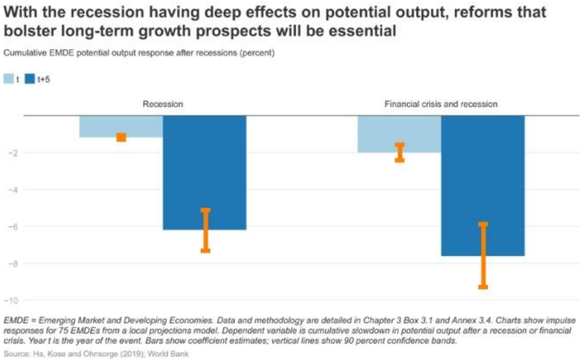
The experimental findings and analzses are presented in this section. To compare the characteristics of Mean square error (MSE) and assess the effectiveness of various algorithm models for GDP forecasting.

A lower MSE value enhances the model's predictive ability.

The following is the MSE calculation formula:

MSE=1/M (m=t) M (Ym-Ytm) 2

where M is the data volume and Ym and Ytm, respectively, represent the actual and forecast GDP levels.



With the recession having deep effects on potential output, reforms that bolster long-term growth prospects will be essential

Cumulative EMDE potential output response after recessions (percent)

EMDE = Emerging Market and Developing Economies. Data and methodology are detailed in Chapter 3 Box 3.1 and Annex 3.4. Charts show impulse responses for 75 EMDEs from a local projections model. Dependent variable is cumulative slowdown in potential output after a recession or financial crisis. Year t is the year of the event. Bars show coefficient estimates; vertical lines show 90 percent confidence bands.

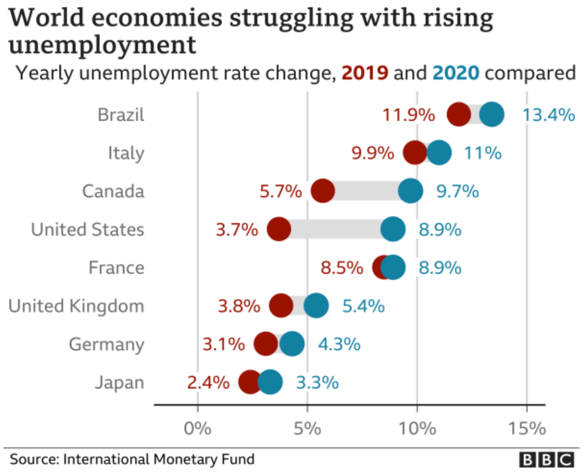
Source: Ha, Kose and Ohnsorge (2019); World Bank

Table 1 displays the experimental findings for our connected model and the competing models.

The connected model has the smallest MSE value, which is obvious.

The results of various models on the economic dataset used to forecast GDP are displayed in Table 1.

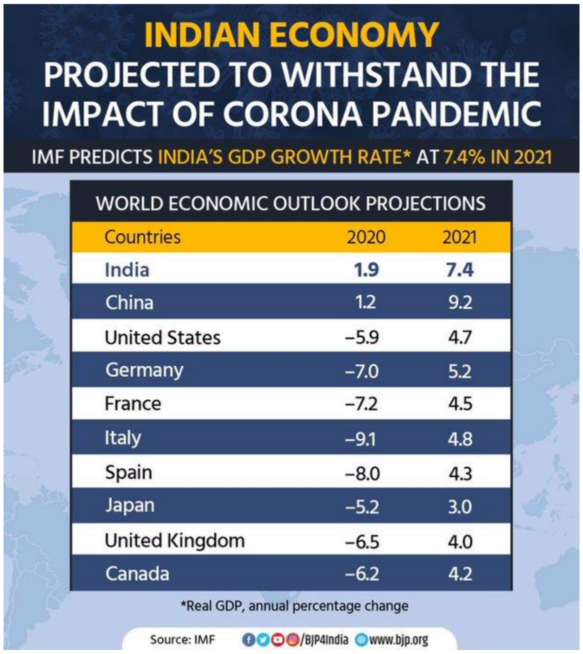
Models	**MSE**
Coupled	6.975
XGBOOST	6.495
AdaBoost	8.650
Random Forest	6.180



World economies struggling with rising unemployment

Yearly unemployment rate change, 2019 and 2020 compared

Source: International Monetary Fund



INDIAN ECONOMY
PROJECTED TO WITHSTAND THE IMPACT OF CORONA PANDEMIC

IMF PREDICTS INDIA'S GDP GROWTH RATE* AT 7.4% IN 2021

WORLD ECONOMIC OUTLOOK PROJECTIONS

Countries	2020	2021
India	1.9	7.4
China	1.2	9.2
United States	–5.9	4.7
Germany	–7.0	5.2
France	–7.2	4.5
Italy	–9.1	4.8
Spain	–8.0	4.3
Japan	–5.2	3.0
United Kingdom	–6.5	4.0
Canada	–6.2	4.2

*Real GDP, annual percentage change

Source: IMF /BJP4India www.bjp.org

V. CONCLUSIONS

In this study, the GDP of various nations and areas was predicted using a linked technique. We combine the Ada Boost and Random Forest algorithms to analyse the economic dataset from the Kaggle competition website. Outcomes indicate that our method may greatly lower the MSE value of the prediction results, demonstrating that the proposed algorithm outperforms all alternatives. We also looked at the relationship between characteristics and GDP figures. Additionally, we receive some sage advice that will help us continue to increase the algorithm's accuracy and use it in a way that will decrease COVID-19's future financial impact.

V. REFERENCES

[1]WHO Coronavirus (COVID-19) Dashboard [EB]. https://covid19.who.int/table. 2020.

[2] Jiang Wan-sheng Wang Bei-bei. Reasons and Inspirations of China's Effective Prevention and Control of COVID-19 in 2020[J].

[3] Mohammad Z-K, Abdollah R-C, Reza R, Meysam A, Taha B.M.J. Ouarda. Machine Learning and Water Economy: a New Approach to Predicting Dams Water Sales Revenue [J]. Water Resources Management. 2020. 1-19.

[4] Ji C, Kuo N F, Peng S L. Potential Impacts of the SARS Outbreak on Taiwan's Economy[J]. Asian Economic Papers, 2004, 3(1):84-99.

[5] Vaishya R, Javaid M, Khan I H, et al. Artificial Intelligence (AI) applications for COVID-19 pandemic[J]. Diabetes and Metabolic Syndrome Clinical Research and Reviews, 2020, 14(4).

[6] Md S H, MM Rahaman. The Post COVID-19 Global Economy: An Econometric Analysis[J]. IOSR Journal of Economics and Finance, 2021, 12(6):22-43.

[7] Oppel S, Huettmann F. Using a Random Forest Model and Public Data to Predict the Distribution of Prey for Marine Wildlife Management[J]. Springer Japan, 2010.

[8] Fang Kuang-Nan. Random forest combination forecasting theory and its application in Finance [M]. Xiamen University Press, 2012.

[9] Abidin A, Darmawan M F, Osman M Z, et al. Adaboost-multilayer perceptron to predict the student's performance in software engineering[J]. Bulletin of Electrical Engineering and Informatics, 2019, 8(4).

[10] Deng, A., Zhang, H., Wang, W., Zhang, J., & Wang, B.. (2020). "Developing computational model to predict protein-protein interaction sites based on the xgboost algorithm", International Journal of Molecular ences, 21(7), 2274.

[11] Yunxin Liang, Jiyu Wu, Wei Wang, Yujun Cao, Biliang Zhong, Zhenkun Chen, and Zhenzhang Li, "Product marketing prediction based on XGboost and LightGBM algorithm", AIPR '19: Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern RecognitionAugust 2019, Pages 150–153.

[12] Xia, Z., Xue, S., Wu, L. et al. "ForeXGBoost: passenger car sales prediction based on XGBoost", Distrib Parallel Databases

[13] Xu ZG. Complex production process prediction model based on EMDXGBOOST-RLSE. Proceedings of 2017 9th International Conference on Modelling, Identification and Control (ICMIC 2017). Kunming, China. 2017. 8.

[14] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., . LightGBM: A highly efficient gradient boosting decision tree. Adv. Neur. Infor. Process. Sys.2017, 30, 3146–3154

[15] Zhu J, Arbor A, Hastie T . Multi-class AdaBoost[J]. Statistics & Its Interface, 2006, 2(3): 349-360.