Predicting Impact of COVID-19 on the Global Economy Based on Hybrid Model

Jincheng Zuo[1*], Fenglan Ma[1], Shaun Chen[2]

[1] *Nanjing University*
[1] *Xinjiang Uygur Autonomous Region Development and Reform Commission*
[2] *Shanghai American School*
[*] *Corresponding author. Email: care885@qq.com*

ABSTRACT
The outbreak of COVID-19 has caused hitherto unknown damage to the global economy, resulting in a large number of company failures and staff unemployment. In order to suppress the further spread of the virus, many countries and regions have adopted quarantine policies to varying degrees. This has led to the failure of normal business activities and financial market behaviors. The health crisis caused by the new crown pneumonia has gradually turned into an economic crisis. In this paper, the data provided by the KAGGLE competition platform is used to analyze and predict the impact of COVID-19 on the economy. A mixed prediction model (AdaBoost, Linear Regression, Decision Tree) is constructed to predict the GDP per capita of different countries. The results show that the hybrid model has the highest accuracy. Specifically, the lowest MSE index score of our model is 7.57, and KNN and decision tree algorithm are 2.23 and 2.19 higher than our hybrid model respectively.

Keywords: *economy, COVID-19, hybrid model, AdaBoost, Linear Regression, Decision Tree, MSE*

1. INTRODUCTION

Since the outbreak of COVID-19, it has been spreading all over the world due to its high transmission. In order to suppress novel coronavirus pneumonia, various countries have adopted different measures of isolation and blockade, and developed vaccine against virus antibody, which greatly reduced the mortality of COVID-19 and ensured the safety of people. But at the same time, the lengthening of social distance has led to the failure of normal financial market behavior, business activities, tourism and entertainment, resulting in a large number of company closures and unemployment, which has had unprecedented repercussions for the economic growth of the world. The epidemic has gradually changed from a health crisis to an economic crisis.

To study the law of economic development under the epidemic situation and predict the future economic development trend is an important premise for the government to formulate reasonable and effective policies or measures, so as to ensure the orderly resumption of production and construction and social stability in a country or region. Today, with computer hardware and software progressing rapidly, using machine learning algorithms and big data to explore the consequences brought by the epidemic to economic enhancement has become a hot spot for scholars and relevant personnel in various countries.

Many algorithms have been developed or applied to economic development prediction, such as K-Nearest Neighbor (KNN), random forest, AdaBoost and so on. In order to obtain more accurate prediction results, this paper uses a hybrid algorithm (a new algorithm obtained by integrating linear regression, random forest and AdaBoost algorithms) to predict GDP per Capita, which is a good indicator to measure a country's economy and living standards. The results show that our algorithm has the strongest performance compared with other algorithms, including KNN and Decision Tree.

The structure of this paper goes as follows. We begin by introducing the research background and significance of this paper. In the second part, the current research status quo at home and abroad is introduced systematically. In the third chapter, the related algorithms used in this paper are described. In the fourth part, the experimental process and the experimental structure are shown and discussed in detail. Finally, the conclusion of this paper and the prospect of future work are given.

2.RELATED WORK

Upon the advent of the big data era, machine learning algorithms have gained currency in various industries to solve some nonlinear classification or prediction problems. The GDP prediction under the epidemic situation is a highly nonlinear problem, which will be affected by many different dimensions, such as the country, region, population density, government or local policies and so on. Therefore, it is very appropriate to use machine learning algorithm to forecast the change trend of GDP [1][2].

Many scholars have adopted machine learning algorithms to predict the economic growth [3][4]. For example, Nang et al. [5] applied Support Vector Machines (SVM) to the regression of GDP and obtained some results. Jahn et al. [6] predicted the GDP growth rate of 15 industrialized economies over a 20-year period from 1996 to 2016, using artificial neural network (ANN) algorithm. The results show that the prediction accuracy of the algorithm is better than that of the linear model algorithms. With the continuous development and optimization of machine learning, neural networks have been widely used in prediction, mainly including convolutional neural networks (CNN) [7] and recursive neural networks (RNN) [8][9]. In addition, there are also some other methods applied in the financial field, such as XGBoost [10] and AdaBoost algorithms [11], both achieved high prediction accuracy. In a word, machine learning is applied in economic analysis field abundantly.

In this article, in order to improve the prediction accuracy as much as possible, we couple the three algorithms (linear regression, random forest and AdaBoost algorithms) together to get a new model. And the final experimental results also show that the prediction performance of our hybrid model has been greatly improved.

In this paper, our main work and contributions include:

- We proposed a hybrid model based on Linear Regression, Random Forest and AdaBoost algorithms to predict economic development trend under the COVID-19 epidemic.
- The proposed hybrid model has the best performance compare with other algorithms, including KNN and Decision Tree and provides a reference for a more accurate forecast of GDP.

3.DATASET AND METHODOLOGY

- **Data Set**

In this chapter, we first introduce the data set used in this article.

The data set used in this paper is provided by KAGGLE competition platform, which contains a large number of COVID-19 and economic development data from the end of December 2019 to August 19, 2020 in 210 countries. The data set mainly includes 8 fields, which are country code, date, Human Development Index (HDI), the number of people infected (Tc) and death (Td), the strictness of the policy on the epidemic, the population (Pop), and GDP Per Capita. Then, in order to observe the distribution of data, the relationship between the number of infected and deaths is given as shown in figure.1

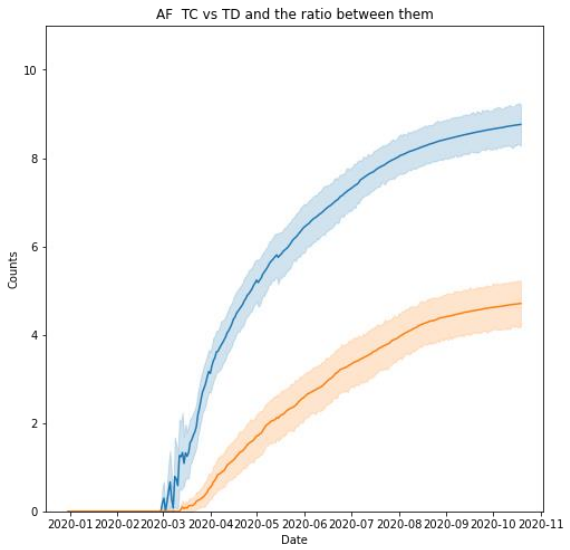

Figure.1 Relationship between infection and death

It can be seen COVID-19 pneumonia is coming or has been effectively controlled with the time going on, the infection and death toll curves are leveling off gradually. In addition, in order to facilitate machine learning for data training, we give the population distribution of different countries as shown in figure.2. Obviously, there is little difference in population distribution in different countries.

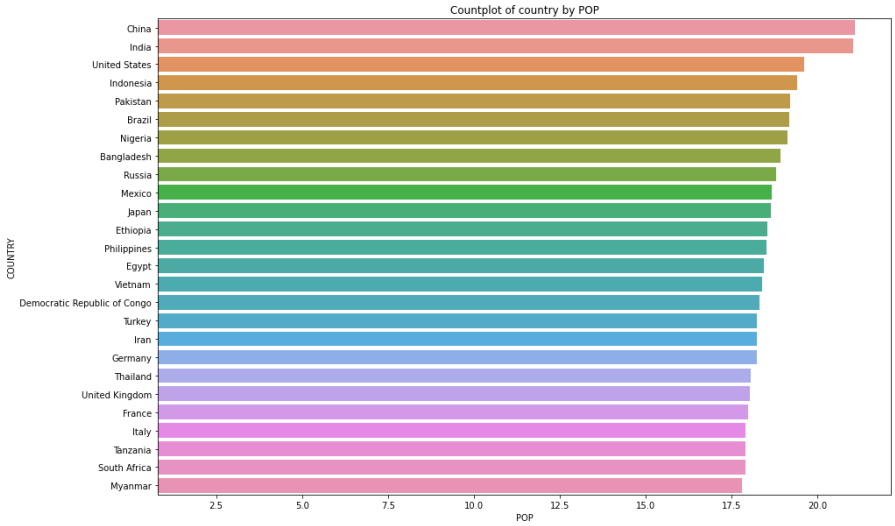

Figure 2. Countplot of country by Pop

- **AdaBoost Regressor**

Then, introduce the algorithm used in this article.

The AdaBoost algorithm is an integrated learning model, which can enhance a weak learner with poor prediction accuracy into a strong learner with higher accuracy. As one of the classic algorithms in the field of data mining, it has the advantages of high precision, high flexibility, relatively simple and no over-fitting [12].

If a sample is correctly categorized during the specific training procedure, its weight will be lowered in the construction of the following training set; Quite the reverse, the weight of a sample point is increased if it is incorrectly categorized. In the meantime, the discourse power in accord with the weak classifier is acquired. Then, the sample set with renewed weights is utilized to train the next classifier, and the whole training course proceeds iteratively. After the training procedure of each weak classifier, the weak classifier with small classification error rate has a greater discourse power, playing a rather momentous part in the eventual classification function, while the weak classifier with large classification error rate has a relatively small discourse power, with a modest influential role in the final classification function.

Assuming that in the m-iteration, the coefficients of the first m-1 and the base learner $G_m(x)$ are fixed, then $f_m(x) = f_{m-1}(x) + \alpha_m G_m(x)$, so that in step m we simply minimize the loss function on the current formula and formula.

And the loss function AdaBoost chose is the exponential loss:

$$L(y, f(x)) = e^{-yf(x)}$$

Then the objective function is to make the loss minimum and get the weight and the weak learners.

$$(\alpha_m, G_m(x)) = \arg\min_{(\alpha,G)} \sum_{i=1}^{N} e^{-y_i f_m(x_i)} = \arg\min_{(\alpha,G)} \sum_{i=1}^{N} e^{-y_i (f_{m-1}(x_i) + \alpha G(x_i))}$$

- **Linear Regression**

Linear regression is the main empirical tool in economics. It is often used to predict consumption expenditure, inventory investment and so on. The model is simpler and more convenient to analyze the multi factor model, but because it is a linear model, it is difficult to accurately analyze and predict the nonlinear model. The commonly used linear regression models mainly include least square method, ridge regression, Lasso regression and so on [13].

- **Decision Tree**

Decision tree algorithm is a very common machine learning algorithm, which belongs to supervised learning algorithm. Similar to the classification process, the algorithm subdivides the relevant data level by level in line with the given rules to define each branch of the decision tree. Generally speaking, the decision tree consists of a root node, a series of internal nodes and the ultimate node. At each bifurcation node of the decision tree, different rules can be selected for further detailed classification, which is the basic idea of the decision tree [14].

- **Hybrid model**

In our experiment, we combine above three base models as shown in figure.3, the weight of each one is 1/3.

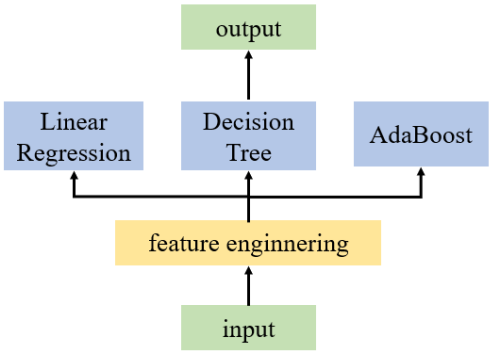

Figure.3 the structure of proposed hybrid model

4.EXPERIMENTS

- **Experimental Results**

It can be seen that the curve of the number of cases and deaths gradually becomes flat, indicating that the epidemic has been effectively controlled. At the same time, the green curve (STI) has begun to decline, which means that as the epidemic slows down, the government's policies have been relatively relaxed, thereby promoting economic recovery and development.

- **Compared Experiments**

Then, in order to quantify a country's economic development level during the epidemic period, we used proposed hybrid algorithms to predict GDP Per Capita. It can be seen that the predicted value is almost consistent with the true value, indicating that the accuracy of the algorithm is good.

To compare our model with other models synthetically, we use the index MSE to evaluate the performance of our algorithm. This index represents the gap between the predicted value and the real value. Generally speaking, the smaller the index, the higher the accuracy. The definition of MSE is shown as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$$

y_i and $\bar{y}_i$ represent the predicted value and label value of the model.

The MSE values calculated according to the prediction results of different algorithms are shown in Table 1. Obviously, our hybrid algorithm has a minimum MSE value of 7.57, which is 2.23 less than KNN and 2.19 less than Decision Tree algorithm. The performance of this algorithm is significantly improved.

Table.1 MSE values of different algorithms

Models	MSE
KNN	9.80
Decision Tree	9.76
Hybrid Model	7.57

5.CONCLUSION

Finally, the main research contents and conclusions of this paper are given. The outbreak of COVID-19 has caused a severe impact on the development of the global economy. Many scholars have studied COVID-19 pneumonia's economic impact on the economy by machine learning. This paper also uses the data set provided by the KAGGLE competition platform to analyze and predict the economic impact of COVID-19. In the data set, the main indexes are human development index, number of deaths, number of infections, etc., and the relationship between these indexes and economic development is analyzed. Next, we obtained a novel prediction model with three models (AdaBoost, KNN and Decision Tree) to predict the GDP Per capita of the COVID-19 epidemic. The results show that our model has the best comprehensive performance. Specifically, the lowest MSE index score of our model is 7.57, and KNN and decision tree algorithm are 2.23 and 2.19 higher than our hybrid model respectively.

In the future, we will continue to optimize our model to obtain more accurate prediction results.

REFERENCES

[1] Martinho V. Impact of Covid-19 on the conver gence of GDP per capita in OECD countries[J]. Re gional Science Policy & Practice.

[2] Tamara N , Muchisha N D , Andriansyah A , et al. Nowcasting Indonesia's GDP Growth Using Machine Learning Algorithms[J]. MPRA Paper, 2020.

[3] Hamed Ghoddusi, Germán G. Creamer, Nima Rafizadeh, Machine learning in energy economics and finance: A review, Energy Economics, Volume 81, 2019, Pages 709-727, ISSN 0140-9883

[4] Burlig, F., Knittel, C., Rapson, D., Reguant, M. and Wolfram, C. (2017). Machine learning from schools about energy efficiency. Working Paper Series.

[5] Xu Haoran, Xu Bo, Xu Kewen. Analysis on application of machine learning in stock forecasting [J]. Computer Engineering and application, 2020, 56(12):19-24.

[6] Study on Prediction and Analysis of GDP Based on Machine Learning[J]. Computer Technology and Development, 2009, 2:227-229, 233.

[7] Nanchang, Gonghe CJ. Research on applying SVM to the regression of GDP[J]. Microcomputer Information, 2007.

[8] Jahn M . Artificial neural network regression models: Predicting GDP growth[J]. HWWI Research Papers, 2018.

[9] Chen XC. Research on Deep Learning Algorithm about Handwriting Based on the Convolutional Neural Network[D]. Zhejiang Gongshang University, 2014.

[10] Longo L, Riccaboni M, Rungi A. A Neural Net work Ensemble Approach for GDP Forecasting[J]. Working Papers, 2021.

[11] Sugiyarto A W , Abadi A M . Prediction of Indonesian Palm Oil Production Using Long Short-Term Memory Recurrent Neural Network (LSTM-RNN)[C]// 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS). IEEE, 2020.

[12] Qureshi S, Chu B M, Demers F S .Forecasting Canadian GDP growth using XGBoost[J]. Carleton Economic Papers, 2020.

[13] Zhan C X. An Empirical Study of the Stock Market Rise and Fall Based on the SVM_AdaBoost Model [D]. South China University of Technology, 2013.

[14] Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from U.S. daily newspapers. Econometrica: Journal of the Econometric Society 78: 35–71.

[15] Goodfellow,I., Bengio,Y., Courville, A. and Bengio, Y. Deep Learning. Cambridge: MIT press. 2016.

[16] Han H , Feng M , Wang W Y . Review of Recent Development in Decision Tree Alg orithm in Data Mining[J]. Application Research of Computers, 2004.