Journal of Financial Risk Management, 2024, 13, 108-129
https://www.scirp.org/journal/jfrm
ISSN Online: 2167-9541
ISSN Print: 2167-9533

Using Machine Learning Models to Predict Economic Recession Caused by COVID-19

Wenxi Xiu

School of Economics and Management, Beijing Jiaotong University, Beijing, China
Email: 22120580@bjtu.edu.cn

How to cite this paper: Xiu, W. X. (2024). Using Machine Learning Models to Predict Economic Recession Caused by COVID-19. *Journal of Financial Risk Management, 13,* 108-129.
https://doi.org/10.4236/jfrm.2024.131005

Received: November 30, 2023
Accepted: February 5, 2024
Published: February 8, 2024

Abstract

Beginning in early 2020, there was a severe global epidemic, which was named COVID-19. Its outbreak has severely disrupted the global economy. This paper attempts to predict the economic fluctuations caused by COVID-19. I collect data about unemployment rate, inflation rate, Producer Price Indices, house price, population, stock market value, treasury bill rate, corporate and government bond yields, net export, savings rate, average price-earnings ratio of the stock market, total credit, loan rate, personal and government consumption expenditures and investment, and so forth from the United States spanning from 1983:1 to 2023:2 and make prediction of the Gross Domestic Product (GDP) using a VAR model, five machine learning models, including gradient boosted regression model, random forest regression model, K-nearest neighbors regression model, linear regression model and support vector machine regression model, along with a deep learning model which is Long Short Term Memory model. The predictive effectiveness of those models is measured by mean absolute error and the results shows that the gradient boosted regression model after hyperparameter optimization, whose error is minimized to about 75, is the best at predicting the US economy. This model also exhibit superior performance in predicting the Italian economy, which to some extent shows its widespread usage.

Keywords

COVID-19, Economic Recession, Machine Learning Models

1. Introduction

An economic recession is a severe, widespread, and prolonged downturn in economic activity that is defined differently in different countries. In the United States, a recession has long been defined by the National Bureau of Economic

Research (NBER) as a significant contraction in economic activity that occurs throughout the economy and lasts for more than a few months.

Global pandemics, as an exogenous shock, can have serious consequences for the economy. In particular, the coronavirus outbreak in early 2020, which was one of the worst global pandemics in recent years. It had a shock that was multicyclical in nature, almost perfectly synchronized within and across countries, and had a catastrophic impact on the economy not only in the foreseeable weeks following the crisis but also over a long period of time (Ludvigson et al., 2020).

The economic effects of COVID-19 can be broadly categorized into supply and demand effects (Padhan & Prabheesh, 2021). The supply effect stems from a reduction in working hours, while the decline in demand stems from a drop in income due to unemployment resulting from the lockdowns. Specifically, the pandemic affects the economy through the following channels: 1) The direct effect of reduced employment. Reduced employment leads to a fall in the demand for capital, which results in a loss of output. 2) The increases in international transaction costs. An increase in the cost of importing and exporting goods and services leads to a decrease in trade and a decrease in productivity. 3) The sharp reduction in travel. Sharp decline in international travel income, which leads to loss of production. 4) The decline in demand for services that require close human contact. Households decline in demand and buy fewer services than before, drastically reducing consumption of goods and services (Maliszewska et al., 2020). In addition, the contraction of foreign direct investment, the real impact of financial shocks, and the fall in oil prices amplify the economic costs associated with COVID-19.

Predicting the macroeconomic impact of COVID-19 in the coming years is critical for formulating appropriate policy responses and guiding firms and households in their decision-making, but it is extremely challenging because the type and magnitude of this economic shock is unprecedented (Huang & Yan, 2023). Therefore, this paper attempts to forecast the US recession due to COVID-19 using a simple vector autoregressive model (VAR), machine learning models, and a deep learning model for historical data on macroeconomic variables, with a view of trying to find a more appropriate forecasting model. Moreover, to validate the applicability of the model, this paper applies a series of models to Italian economic data, revealing that the model performing best in predicting US data continues to perform best in predicting Italian data, confirming the model's applicability.

The rest of the paper proceeds as follows. Section 2, based on the US Google COVID-19 Community Mobility Report, confirms the likelihood that COVID-19 is one of the causes of the US recession, which is a prerequisite for being able to follow up with further research. Section 3 uses a series of models to forecast the US economy and selects the best performing model. Section 4 shows the application of the models to Italian data, verifying that the best model has broad applicability. Section 5 presents the shortcomings of the models. Section 6 is the conclusion.

2. Study Based on Google COVID-19 Community Mobility Reports of US in 2020

Beginning in early 2020, there was a severe global epidemic, which was recognized as a new coronavirus (Severe Acute Respiratory Syndrome Coronavirus 2, or SARS-CoV-2), later renamed Coronavirus Disease-19 or COVID-19 (Qiu et al., 2020).

In order to predict a COVID-19-induced recession in the US, this paper begins with a study based on the 2020 US Google COVID-19 Community Mobility Report to initially determine that this epidemic does indeed affect the economy. The Report uses aggregated, anonymized data to show how busy certain types of areas are, which can also provide some indication of people's propensity to go out or stay at home. Consumption has consistently accounted for 60% to 80% of US GDP since 2000, making it the largest component of GDP. As an example of consumption, in April 2020, the Report shows a significant decrease in the number of people traveling to retail and recreation, which could result in lower consumption. Compared with February 2020, the real personal consumption expenditures in the Fred database fell 16.4%, breaking a record for declines, while the Business Insider declared that retailers had announced plans to close a total of more than 4300 stores in 2020. Retail is an important part of the US economy, and the closure of brick-and-mortar stores is a major shock to the economy as a whole, potentially leading to a recession. And with March to April 2020 coinciding with the first outbreak in the US, if the correlation analysis proves that COVID-19 does indeed correlate with the number of people going to retail and recreation, it will further suggest that COVID-19 could also be a contributing factor to the US recession.

2.1. Backgrounds of the Pandemic in US, 2020

In the United States, the economy has grown rapidly after the economic crisis (see **Figure 1**), with GDP increasing from 6896.561 billion in 1983 to 20386.467 billion in 2023, an increase of almost 300%. However, as can be seen from the **Figure 1**, there is a significant trough in the US economy in 2020, mainly due to the COVID-19 pandemic that had been ravaging the world and its economy since the beginning of the year, and its macroeconomic impact is greater than any catastrophic event that had occurred in the last 40 years (Ludvigson et al., 2020).

The main reason for the decline in GDP was personal consumption expenditures (Chetty et al., 2020), with real personal consumption expenditures plummeting by nearly 20% after the COVID-19 outbreak, as the recurrence and uncertainty of the epidemic made consumers cautious about the economic outlook. Unemployment and reduced incomes during the epidemic forced many households to cut back on spending in response to the uncertainty of the future. At the same time, stay-at-home orders imposed in many US states during the epidemic also restricted travel and socializing, making some consumer scenarios impossible.

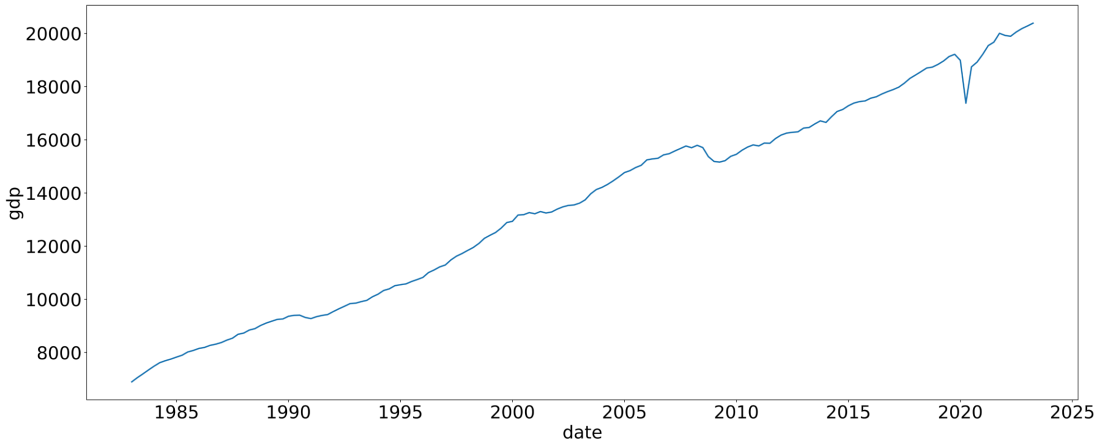

Figure 1. US GDP change over 40 Years. Data sources: U.S. Bureau of economic analysis, real gross domestic product [GDPC1], retrieved from FRED, Federal Reserve Bank of St. Louis.

In fact, social spending activities in the United States began to decrease significantly and rapidly long before the imposition of the lockdown (Farboodi et al., 2020). Chetty et al. (2020) suggest that the higher the rate of COVID infection, the greater the decline in consuming. In this case, 67% of the total decrease in expenditures came from a decrease in service expenditures, while the percentage of cuts in expenditures was higher for higher-income households compared to lower-income households.

The great uncertainty generated by the COVID-19 pandemic (Baker et al., 2020) had also greatly affected investment, both corporate and private. In terms of corporate investment, as macroeconomic uncertainty increased, it increased the financing constraints of listed companies and thus reduced the likelihood of listed companies to increase their investment in fixed assets, financial investment, and innovative investment. In terms of private investment, under the uncertain economic environment, individuals lacked confidence in future economic trends, their tolerance for risk decreased, and they tended to reduce their investments or choose to make short-term investments to reduce risk. This resulted in real gross private domestic investment in the Fred database falling from a peak of 3818.108 billions of chained 2017 dollars in 2019 to a trough of 3145.228 billions of chained 2017 dollars in 2020, a drop of nearly 21 percent.

2.2. Community Mobility Data

The data comes from COVID-19 Community Mobility Reports, published by Google, which is based on smartphone data collected to help public health officials understand how residents everywhere moved during the COVID-19 pandemic.

Each country's data cover six broad categories of mobility information, including retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential. The economic mobility data over time are as **Figure 2** shown. All six curves show a jagged shape, which is mainly influenced by work hours, and people tend to schedule their travel social plans on weekends.

Figure 2. Plot of economic mobility data over time. Data sources: COVID-19 community mobility reports.

In the map, Google uses aggregated, anonymized data to show how busy certain types of areas are in order to predict when business establishments will be most crowded, and also to give some indication of the shifting trends in the propensity of people to go out or stay at home as a result of the epidemic outbreak.

According to the data published by WHO, the number of people infected with COVID-19 in the US jumped from 4 million at the beginning of March to 3493.5 million at the end of April. As a result of the first outbreak, **Figure 2** shows that the number of people going to retail and recreation, grocery and pharmacy, parks, transit stations and workplaces decreased significantly, especially grocery and pharmacy and parks. The reason for this may be that people had a premonition before the outbreak, and it can be seen that the number of people going to these two places increased at the end of March so that people had fulfilled a certain need for relaxation and purchased enough necessities, and changed their way of going out to socialize and relax after the outbreak. In contrast, the proba-

bility of people staying at home increased significantly in the face of the first outbreak, both because 45 states had implemented home confinement orders by April and because people took voluntary measures. The combination of these two factors has reduced social interactions outside the home.

Then, in June 2020, there was another outbreak, which reached 7695.2 million people infected in July, but as people knew more about COVID-19, pharmaceutical company Modena developed the first batch of it as well, people began to lost patience with the Home Rule and disobeyed it, and with the lifting of the residence order, people resumed to retail and recreation, grocery and pharmacy, parks, transit stations and workplaces. As people let their guard down, a third and more dramatic increase in the number of infected people began in October 2020, reaching a peak of 24583.9 million in December. This, coupled with the discovery of new coronavirus variants, Alpha and Delta, in the UK and India in September and October respectively, made people even more fearful of COVID-19, leading to a sharp drop in the number of people going out, especially to parks. It is clear that the spread and development of COVID-19 and the changes in the number of people infected with COVID-19 are influencing people's propensity to go out or stay in their homes to some extent.

2.3. Correlation between Economic Mobility and COVID-19 Data

In order to more accurately see the relationship between the COVID-19 data and economic mobility, I test the correlation between those two in this section using the Pearson correlation coefficient, and the results are shown in Table 1. Where case denotes the number of people US daily COVID-19 cases; retail and recreation denotes the retail and recreation percent change from baseline, and the same for grocery and pharmacy, parks, transit station, workplaces and residential.

Showing from the table, case has a negative impact on the number of people going to retail and recreation, grocery and pharmacy, parks, transit stations and workplaces. The reason for this may be that the COVID-19 is a highly contagious disease that spreads mainly through respiratory droplets, but can also be contracted through general contact, so out of fear of the disease and for their own protection, people are going out less, thus increasing the likelihood of staying at home. In addition to individual will, states led by California had begun to implement homebound orders, and on March 19, 2020, California Governor Gavin Newsom declared a state closure, ordering all 40 million residents to stay home to prevent the spread of coronaviruses. The negative effect of case on the

Table 1. Correlation coefficients for economic mobility and us cases per day.

Pearson	retail and recreation	grocery and pharmacy	parks	transit station	workplaces	residential
case	−0.08	−0.16***	−0.39***	−0.25***	−0.1*	0.06

Note: ***, ** and * indicate significant at 1%, 5% and 10% level respectively.

number of people going to retail and recreation is not significant because Coven and Gupta (2020) found that residents of low-income neighborhoods in New York City are less likely to comply with shelter-in-place activities during non-work hours and they are more likely to frequent retail stores to purchase necessities. Additionally, when essential goods are insufficient, even during severe pandemics, people are compelled to visit retail stores for purchases. Simultaneously, prolonged periods of staying at home can adversely affect people's mental well-being, leading them to be willing to take risks by visiting entertainment venues to fulfill their psychological needs.

Case has a positive impact on the number of people who are homebound, in addition to the reasons mentioned above, Dingel and Neiman (2020) analyzed the feasibility of working from home, and they found that 37% of work can be done from home. Based on this, numerous companies in the US had implemented work-from-home policies. On July 16, 2020, Amazon announced an extension of its work-from-home policy for its employees until January 2021, and Facebook and Google announced policies to extend their work-from-home policies until the end of 2020, which further increases the likelihood that people would be able to work in home.

In summary, the severity of COVID-19 does affect people's propensity to go out or stay at home, providing the premise that this epidemic will have an impact on the economy.

3. Predictive Model

To make a prediction of recession, then a reasonable prediction model must be built. Typically, the initial model obtained from model fitting usually does not work very well, which requires adjusting the hyperparameters to optimize the model. Hyperparameter optimization is the process of determining the correct combination of hyperparameters that maximizes the performance of the model and is one of the more important steps in any machine learning project.

3.1. Data

The research data for this paper mainly come from Federal Reserve Economic Data (FRED) and the World Bank database. FRED is a database maintained by the Research Department of the Federal Reserve Bank of St. Louis and it contains over 820,000 economic data series from 114 regional, national, and international sources. The World Bank database is a repository of global economic and social indicators, including data from more than 1000 countries and regions, covering a wide range of fields such as economy, trade, population, and so on, which can help people better understand and analyze global economic and social changes. These two databases provide sufficient and reliable data support for this paper.

This paper uses quarterly data of the United States from the first quarter of 1983 to the second quarter of 2023. Data includes Gross Domestic Product, unemployment rate, inflation, Producer Price Indices, house price, population,

stock market value, treasury bill rate, corporate and government bond yields, net export, savings rate, average price-earnings ratio of the stock market, total credit, loan rate, personal consumption expenditures and investment, government consumption expenditures and investment.

3.2. Models

The research goal of this paper is to make prediction of economic decline based on evidence from COVID-19 pandemic. A country's economic development tends to be persistent if it is not affected by exogenous factors, such as large-scale diseases, disasters or wars. Clearly, high capital accumulation in the previous year contributes to this year's economic growth. People's psychological expectations also play a role in this. If the previous year's economy was prosperous, people will be more confident that this year's economy will continue to be prosperous, thus expanding consumption and investment and promoting further economic development. At the same time, the country's macroeconomic policies have long-term stability. From this, it can be assumed that the level of the economy in the previous period affects the level of the economy in the following years. In view of this character of the economy, this paper firstly selects vector autoregressive model (VAR), which adopts the form of multi-equation linkage, in which the endogenous variables in each equation of the model regress the lagged terms of all the endogenous independent variables of the model, and then estimate the dynamic relationship of all the endogenous variables. This model is commonly used to predict interconnected time series systems and to analyze the dynamic impact of random perturbations on systems of variables.

Machine learning algorithms are methods that attempt to mine implicit patterns from large amounts of historical data and use them for regression (prediction) or classification. When provided with diverse and complex data, machine learning models outperform simpler time series models, with higher accuracy and simpler variable selection, especially over shorter time horizons. In particular, machine learning models are able to identify turning points in economic movements earlier, suggesting that it can provide more guidance on cyclical fluctuations (Hall, 2018). As being showed in the previous section, the economy experienced a major recession in 2020, after which it recovered and started to grow year after year, and this kind of fluctuation is also applicable to machine learning algorithms. So, this paper further uses gradient boosted regression model, random forest regression model, K-nearest neighbors regression model, linear regression model and support vector machine regression model to predict the economic fluctuations. Among them, the gradient boosted regression model is an integrated learning approach to build a stronger predictive model by combining several weak predictive models. It is based on decision tree model, which progressively improves the accuracy of the predictive model through an iterative approach. Its core idea is that each step tries to fit a negative gradient of the target variable to reduce the loss function of the current model. The random forest

regression model builds multiple unrelated decision trees by randomly extracting samples and features. Each decision tree can produce a prediction result from the extracted samples and features, and the regression prediction result of the whole forest is obtained by averaging the results of all the trees. It is suitable for scenarios that require relatively low data dimensionality and high accuracy at the same time. The K-nearest neighbors regression model is a nonparametric regression model. Its basic idea is to use the eigenvalues and target values of the known samples to find k nearest neighbors by calculating the distance between the samples to be predicted and the known samples, and then the target values of these k nearest neighbors are weighted or simply averaged as the predicted values of the sample to be predicted. The advantages of this model are that it is simple and easy to understand, easy to implement, applicable to data of arbitrary dimensions and nonlinear problems, and it does not require model assumptions and can be obtained directly from the data. The principle of linear regression model is based on data learning methods to obtain a function that predicts the dependent variable by a linear combination of the independent variables. The advantage of this model is its simplicity and high interpretability. The support vector machine regression model is a machine learning algorithm specialized for small sample cases. The algorithm makes a compromise between the complexity of the model and the learning ability. It has strong generalization and accuracy. It ends up with a globally optimal solution, which overcomes the problems of overlearning, underlearning, and local minima of neural networks. Compared to traditional machine learning methods, deep learning methods can automatically extract relevant features and achieve better results when dealing with large amounts of complex data, so this paper also uses Long Short Term Memory (LSTM) model. LSTM algorithm is the most used time series algorithm, which is a special kind of Recurrent Neural Network (RNN), and it is capable of learning long-term dependencies. For traditional RNNs, as the sequence interval lengthens, problems such as gradient explosion or gradient vanishing make the model unstable or simply unable to learn effectively during training. Compared with RNNs, LSTM adds more structure to each unit structure and solves the long-term dependency problem by designing the threshold structure. So, LSTM can have a relatively long short-term memory, which provides better results compared with RNNs and is suitable for processing and predicting important events with relatively long intervals and delays in the time series.

The dependent variable in this paper is Gross Domestic Product (gdp) and the measure uses real GDP to exclude the effect of inflation on it. The unit of the variable is billions of chained 2012 dollars.

The independent variables in this paper are as follows:

- **Unemployment Rate (uneprate).** Unemployment is the idleness of labor resources, and labor resources have immediacy, the labor resources that cannot be used cannot be moved to the next period of use. Thus, the labor force available in the current period of idleness is a permanent waste of this

resource, and a high unemployment rate is not conducive to economic growth.

- **Inflation (infg, infc)**, measured by GDP deflator and Consumer Price Index. Moderate inflation can promote economic growth, but too high inflation can lead to problems such as excessive price increases and loss of purchasing power, which can adversely affect the economy.
- **Producer Price Indices (ppi).** A fall in the Producer Price Indices means that the costs faced by producers have fallen, which may reduce inflationary pressures. However, if inflation is too low or deflation occurs, it can have a negative impact on the economy.
- **House Price (hpi)**, measured by House Price Index. A booming real estate market promotes the development of related industries and employment, leading to an increase in the consuming power of citizens and indirectly contributing to economic growth. Each standard deviation increase in the House Price Index leads to a 0.95 percentage point increase in GDP (Chen & Ranciere, 2019).
- **Population (pop)**, measured by total population. The unit of the variable is thousands. Population growth within a certain range leads to economies of scale that favor economic growth. But population growth beyond a certain range can put enormous pressure on the country's resources, technology, and capital and hinders economic growth.
- **Stock Market Value (stock)**, measured by the Wilshire 5000 Total Market Index. The stock market can have a significant impact on the economy, ranging from capital formation, consumer confidence, wealth effects, profits and employment to monetary policy.
- **Treasury Bill Rate (trrate)**, measured by the 1-Year treasury bill rate. It is closely related to the real estate market, government finances, stock market performance, and business operations, reflecting the market's view of government credit and future economic expectations. In general, the higher the treasury bill rate, the lower the market's trust in the government, and the greater the concern about the future economy.
- **Net Export (netex), Personal Consumption Expenditure (pce) and Investment (gpdi), and Government Consumption Expenditure and Investment (gci).** The unit of netex, pce, gpdi and gci are billions of chained 2012 dollars, billions of dollars, billions of chained 2012 dollars and billions of chained 2017 dollars. For the purpose of the test, the net export data are taken to be the opposite of the data. Consumption, investment and net export are the three main carriages that drive economic growth, and their role in the economy is self-evident.
- **Savings Rate (saverate_per)**, measured by the personal savings rate. Savings rate and economic growth are highly correlated and interact with each other, which is an important factor affecting the potential growth rate of the economy.

- **Price-Earnings Ratio of the Stock Market (pe)**, measured by the MULTPL S & P 500 P/E Ratio. There is a positive correlation between the P/E ratio and gross domestic product, and economic growth is likely to lead to higher profitability of firms, thus pushing up the P/E ratio.
- **Total Credit (tc).** The unit of the variable is billions of dollars. Credit conditions are an important driver of business cycle volatility (Gilchrist & Zakrajšek, 2012; Philippon, 2009). A one standard deviation increase in credit growth is associated with annualized GDP growth of 1.79 percentage points (Chen & Ranciere, 2019).
- **Corporate Bond Yield (ay, by) and Government Bond Yield (sby).** Corporate bond yield is measured by Moody's Seasoned Aaa Corporate Bond Yield and Moody's Seasoned Baa Corporate Bond Yield, and sovereign bond yield is measured by 10-year Government Bond Yield. Corporate and sovereign bond yields are also predictive of macro variables (Chen & Ranciere, 2019). Bond yields affect government borrowing costs, corporate and individual consumption and investment, and asset price volatility and adjustments, among others, and these impacts and risks pose a challenge to the stability and growth of the economy.
- **Loan Rate (lr)**, measured by bank loan rate. The loan rate is strongly positively correlated with future investment growth and negatively correlated with consumption (Chen & Ranciere, 2019) and can be used to make predictions about economic volatility.

Table 2 gives the results of the descriptive statistical analysis of the above variables. Among them, Producer Price Indices, Stock Market Value, Treasury Bill Rate, Total Credit, and Baa Corporate Bond Yield contain missing values, but since the missing values are all less than 10% of the number of observations, they are not excluded, but are instead filled in with the median before forecasting.

3.3. Model Predictions

Based on the above data, this paper makes predictions using VAR model, machine learning methods, and LSTM model in turn. Among them, the mean absolute error (MAE) of the VAR model is 700.3266, the MAE of the optimal model in machine learning methods is 75.7487, and the MAE of the LSTM model is 599.629. The machine learning methods have the best prediction effect, so the focus of this subsection will be on the prediction using the machine learning methods.

After filling in the missing values in the original data with the median, a correlation analysis is first performed, which shows that the three variables most correlated with gdp are the square root of pce, tc and ay and the plotted pairs plot is shown in Figure 3.

Secondly, in this study, each independent variable is transformed into its square root form. After removing collinear variables, 80% of the data is allocated to the training set, with the remaining data is allocated to the test set. The baseline

Table 2. Table of descriptive statistics results for variables.

Variable Name	Number of observations	Mean value	Standard deviation	Minimum value	Maximum value
Gross Domestic Product (gdp)	162	13,565	3819.991	6897	20,386
Unemployment Rate (uneprate)	162	5.984	1.733	3.500	12.967
GDP Deflator (infg)	162	80.900	19.803	48.110	125.620
Consumer Price Index (infc)	162	78.720	22.561	41.290	128.350
Producer Price Indices (ppi)	160	80.590	20.802	52.330	138.040
House Price (hpi)	162	284.400	123.661	114.400	645.299
Population (pop)	162	287,745	32006.410	233,546	335,019
Stock Market Value (stock)	161	52.742	53.389	3.091	227.423
Treasury Bill Rate (trrate)	137	3.741	2.965	0.059	10.711
Net Export (netex)	162	526.630	350.231	50.590	1488.700
Savings Rate (saverate_per)	162	7.170	2.869	2.467	26.267
Price-Earnings Ratio of the Stock Market (pe)	162	17.147	4.593	9.827	46.843
Total Credit (tc)	160	17,412	10013.940	3447	38,832
Aaa Corporate Bond Yield (ay)	162	6.477	2.546	2.233	13.221
Baa Corporate Bond Yield (by)	150	7.006	2.084	3.238	11.382
Government Bond Yield (sby)	162	5.184	2.835	0.650	13.200
Loan Rate (lr)	162	6.470	2.644	3.250	12.992
Personal Consumption Expenditures (pce)	162	8209	4261.458	2185	18,302
Personal Investment (gdpi)	162	2208	829.458	796.300	3892.5
Government Consumption Expenditures and Investment (gci)	162	2784	458.239	1827	3513

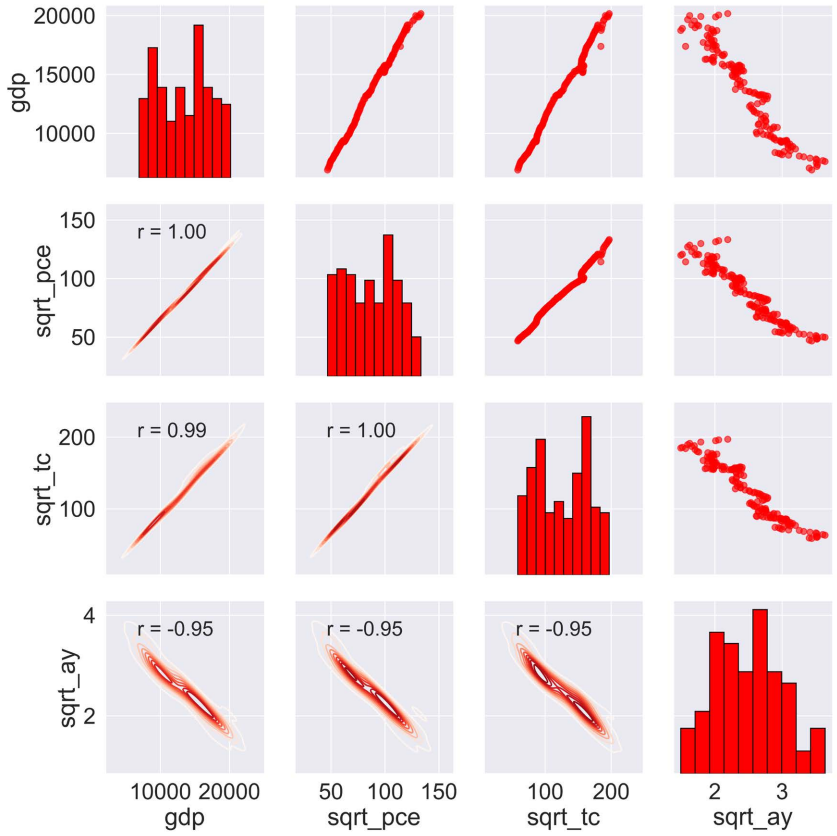

Figure 3. Pairs plot of GDP data.

MAE is constructed as 3348.1324. Further, using five machine learning methods, gradient boosted regression model, random forest regression model, K-nearest neighbors regression model, linear regression model and support vector machine regression model, to forecast the economy. Next, hyperparameter optimization is performed for these five models. The MAEs of the five methods before and after optimization are shown in **Figure 4**. From the figure, it can be observed that the optimized gradient boosted regression model has the lowest MAE, at approximately 75. The support vector machine regression model has the highest MAE, reaching nearly 975, but still significantly lower than the baseline MAE. This suggests that machine learning methods are suitable for predicting economic issues.

Finally, the optimized gradient boosted regression model is explained in depth and it is found that the two variables that play the biggest role in predicting the gdp are infg and infc. Meanwhile, the model's economic forecast graph can be found in **Figure 5**, showing its exceptional predictive capability for the economy.

4. Model Predictions of Italy

The models in the third section are based on US data with good predictions, and in this section the paper searches for some Italian economic data with the aim of seeing if these models are also applicable to Italy.

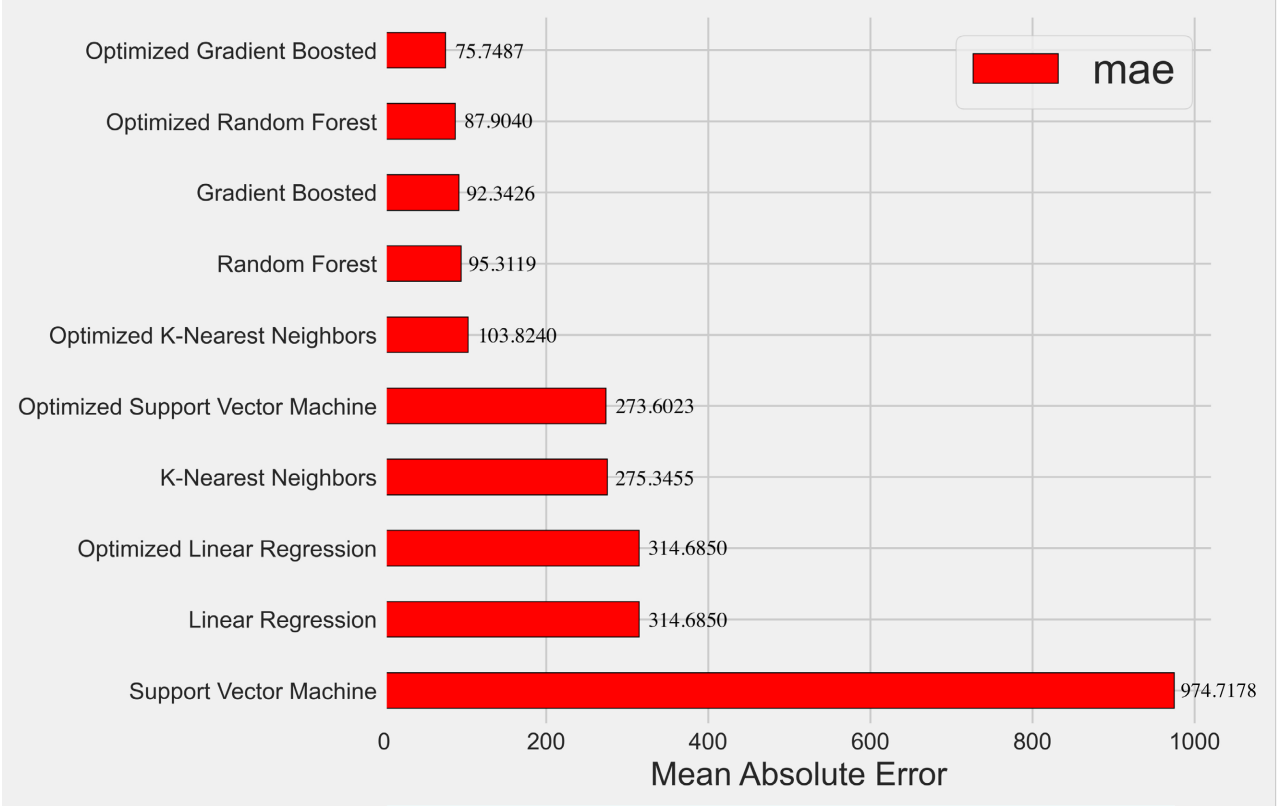

Figure 4. MAEs before and after optimization of machine learning methods.

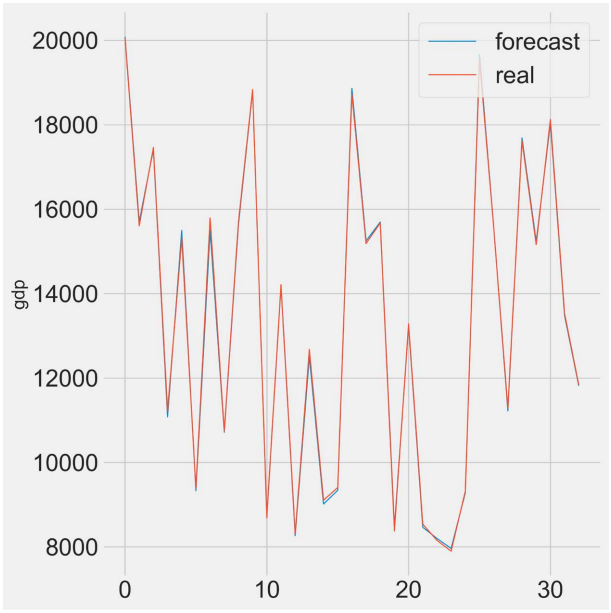

Figure 5. Forecasting the economy using the optimized gradient boosted regression model.

4.1. Data

The Italian data are from FRED, with some of them provided by OECD. The results of collecting data for Italy are not as fruitful as for the US. Firstly, some of

the US data used in the third section do not find their counterparts in Italy. Secondly, a lot of Italian data from 1983 to 1995 are missing. So, from the above two databases, this paper collects quarterly data for Italy from the first quarter of 1996 to the second quarter of 2023 for the study, and the data include GDP, unemployment rate, inflation, Producer Price Indices, house price, population, stock market value, treasury bill rate, government bond yields, net exports, savings rate, total credit, personal consumption expenditures and government consumption expenditures.

4.2. Models

In this section, the economic forecasts for Italy continue to be modeled using the VAR model, five traditional machine learning algorithms, and the deep learning algorithm LSTM.

The dependent variable is Gross Domestic Product (GDP) which is measured by real GDP. The unit of the variable is billions of chained 2010 Euros.

The independent variables are described below, and since them are essentially the same as in the section three, only the specific measures are presented in this section:

- **Unemployment rate (uneprate).**
- **Inflation (infg, infc)**, measured by the GDP deflator and the Consumer Price Index.
- **Producer Price Indices (ppi).**
- **House Prices (hpi)**, measured by nominal House Prices Index.
- **Population (pop)**, measured by population of working age (aged 15 - 64) population. The unit of the variable is thousands.
- **Stock Market Value (stock),** measured by the all-stock prices. The unit of the variable is growth rate previous period.
- **Treasury Bill Rate (trrate)**, measured by the all-treasury bill rate.
- **Net Export (netex), Personal Consumption Expenditure (pce), and Government Consumption Expenditure (gci).** The netex is measured by real net exports of goods and services, whose unit is domestic currency. The pce is measured by private final consumption expenditure, whose unit is billions of chained 2010 Euros. The gci is measured by government final consumption expenditure, whose unit is billions of chained 2010 Euros.
- **Savings Rate (saverate)**, measured by the national savings rate.
- **Total Credit (tc)**, measured by the total private non-financial sector credit. The unit of the variable is billions of dollars.
- **Government Bond Yield (sby)**, measured by 10-year government bond yields.

The results of the descriptive statistical analysis of the above variables are shown in Table 3. Producer Price Indices, Population, Net Export, Savings Rate, and Total Credit contain missing values, but since the missing values are all less than 10% of the number of observations, they are reserved and filled in with the median.

Table 3. Table of descriptive statistics results for variables.

Variable Name	Number of observations	Mean value	Standard deviation	Minimum value	Maximum value
Gross Domestic Product (gdp)	110	395.400	16.967	335.300	426.600
Unemployment Rate (uneprate)	110	9.599	1.824	6.033	12.933
GDP Deflator (infg)	110	91.060	12.346	68.080	113.370
Consumer Price Index (infc)	110	90.370	13.265	67.770	121.600
Producer Price Indices (ppi)	107	92.770	11.981	74.780	127
House Price (hpi)	110	97.610	19.420	59.840	124.430
Population (pop)	101	38,467	473.695	37,189	39,206
Stock Market Value (stock)	110	0.460	3.616	−10.560	11.655
Treasury Bill Rate (trrate)	110	2.224	2.226	−0.510	9.697
Net Export (netex)	109	2729	5870.774	−8723	12,482
Savings Rate (saverate)	108	3.580	2.589	−0.707	7.680
Total Credit (tc)	109	1950.800	714.040	791.200	3026
Government Bond Yield (sby)	110	4.020	1.873	0.637	10.546
Personal Consumption Expenditures (pce)	110	252.200	11.238	215	267
Government Consumption Expenditurest (gci)	110	79.410	3.772	70.400	84.600

4.3. Model Predictions

After that, I use models to predict the above data. Among them, the mean absolute error (MAE) of the optimal model in machine learning methods is the smallest, which is 1267.588, so I'll elaborate further on predictions using machine learning models in this subsection.

Firstly, I fill in the missing values in the original data with the median, and perform a correlation analysis which shows that pce, and the square root of gci

and uneprate are the three most relevant variables to GDP. Figure 6 shows the pairs plot.

Next, I change the data to its square root form, remove collinear features and divide 80% of the data into a training set and the rest of it into a test set. The baseline MAE is 14571.7318, which is very big. To get the smallest MAE, I forecast the Italian economy by using five machine learning methods, gradient boosted regression model, random forest regression model, K-nearest neighbors regression model, linear regression model and support vector machine regression model. And further, I performe hyperparameter optimization for these five models. The MAEs of the initial and optimized methods are shown in Figure 7.

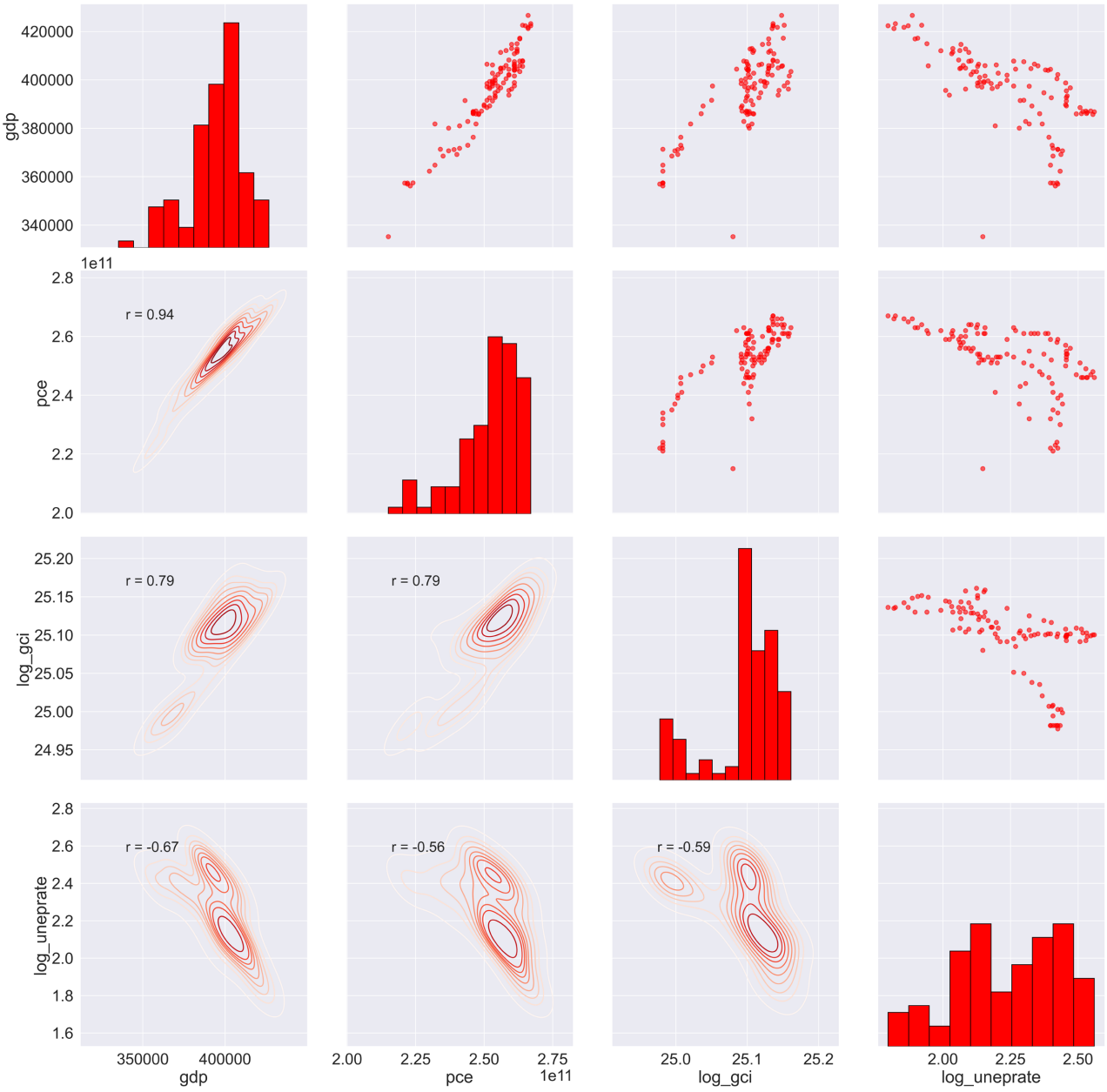

Figure 6. Pairs plot of GDP data.

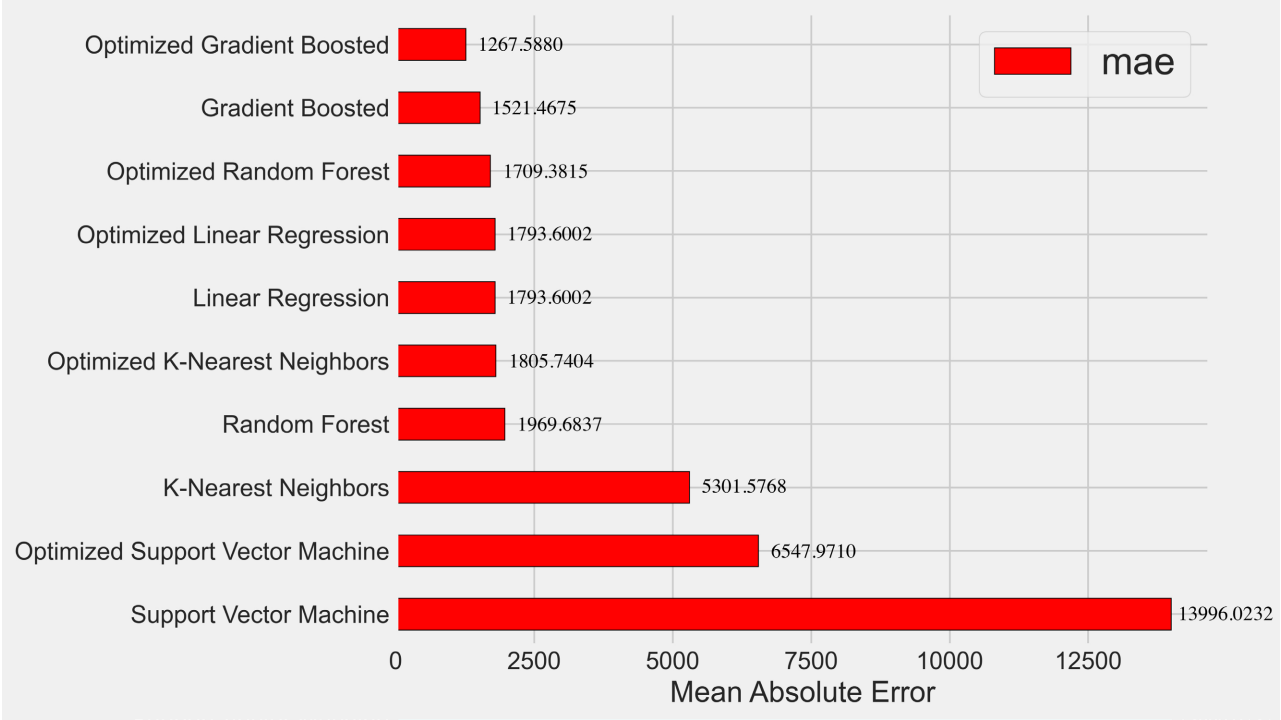

Figure 7. MAEs of the initial and optimized machine learning methods.

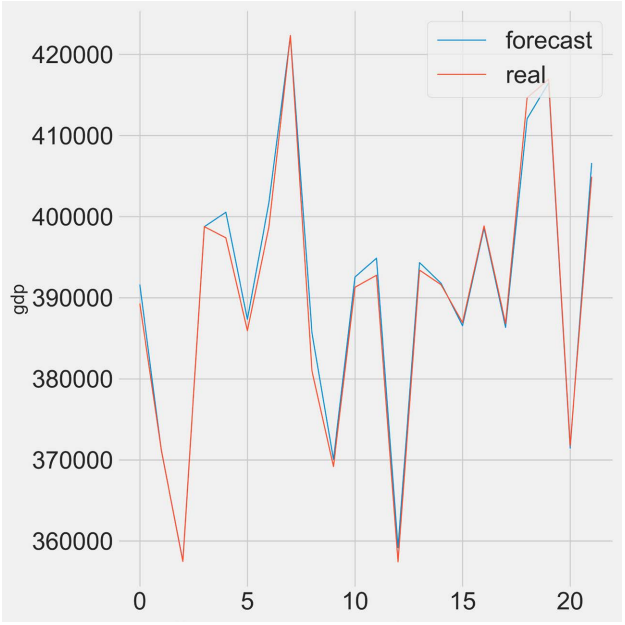

Figure 8. Forecasting the economy using the optimized gradient boosted regression model.

The figure demonstrates that the optimized gradient boosted regression model still exhibits the smallest MAE, at around 1268. On the other hand, the MAE for the support vector machine regression model is still the highest, reaching nearly 13,996. However, it remains significantly lower than the baseline MAE, indicating the suitability of machine learning methods for predicting economic issues.

Finally, I explain the optimized gradient boosted regression model in depth and found that there are four variables play the biggest role in predicting the GDP and they are stock, uneprate, pop and negex. Besides, I plot the model's predictions for economy and it is shown in **Figure 8**. Apparently, the optimized gradient boosting regression model performs well in predicting the Italian data, but the model's prediction error is significantly higher than that for the US data.

5. Limitations of Models

This paper has done reasonably well in forecasting GDP using a range of models, but there are still some limitations in these models.

The VAR model's performance in predicting GDP is average due to its linearity, which sometimes struggles to capture complex nonlinear relationships and dynamic changes within economies. Additionally, external shocks and uncertainties, whether from global or domestic sources, can influence economies beyond the scope of the VAR model's capture. Future modifications or enhancements to the VAR model could involve introducing nonlinear elements, such as utilizing nonlinear functions, lagged squares, or interaction terms, to strengthen its capability. Bayesian methods could also be considered for better handling model parameter uncertainties and providing more accurate forecasts.

The gradient boosted regression model exhibit the best predictive performance among the models considered in this study, yet it also has shortcomings. Its most significant drawback is the lengthy computation time required due to the algorithm's iterative nature in building a robust regression model. Overfitting issues are also observed. To address these, parameter adjustments, such as reducing the number and depth of decision trees to decrease model complexity, and employing feature selection techniques or dimensionality reduction methods like Principal Component Analysis (PCA) can be utilized to improve efficiency and generalization.

Similarly, the random forest regression model, ranking slightly lower than the gradient boosted regression model, faces the challenge of prolonged computation time. As this model constructs multiple decision trees and conducts feature selection and node splitting for each tree, it significantly consumes computational resources. The model's interpretability is also compromised as it generate six-layered decision trees. Solutions for this model can draw from those used for the gradient boosted regression model.

The K-nearest neighbors regression model's results are solely based on the attributes of the nearest points in the test data, making interpretation difficult. To address this issue, a potential approach could involve further utilizing linear regression model for localized predictions, fitting a linear model around each prediction point to provide more straightforward interpretations.

Similar to the VAR model, linear regression model handle linear relationships and may struggle with nonlinear relationships, sensitivity to outliers, and susceptibility to noise in the data, affecting their effectiveness in economic forecasting. To enhance the linear regression model's predictive capability, potential

solutions include expanding the feature space by adding interaction terms, polynomial features, or other economic indicators. Regularization techniques like Ridge Regression or Least Absolute Shrinkage and Selection Operator (LASSO) can also be employed to mitigate sensitivity to outliers and prevent overfitting.

The support vector machine regression model requires adjusting multiple parameters, such as selecting the kernel function and regularization parameters. Poor parameter selection may lead to subpar regression results. Future research in this area could prioritize acquiring more domain knowledge and empirical guidance for parameter tuning. Careful testing and comparisons when selecting kernel functions and regularization parameters should also be considered.

LSTM is sensitive to data quality when dealing with time-series data. Performance can be affected by issues such as outliers or missing values in the dataset, which might reduce the model's effectiveness. Although the missing values are filled using the median, data quality might still be impacted. Additionally, LSTM typically benefits from extensive data, which may be limited in this study. The predictive efficacy of LSTM is also contingent on numerous hyperparameters, and improper adjustment may affect model performance. In future studies, gathering datasets with a broader time span and higher quality could potentially leverage the advantages of LSTM more effectively.

As many other machine learning algorithms also exhibit good predictive accuracy, such as Independently Recurrent Neural Network (IndRNN), Deep Belief Network (DBN), XGBoost, among others. Additionally, ensemble machine learning algorithms show promising performance in forecasting GDP. In the next steps, these models could be employed for economic predictions.

6. Conclusions

The COVID-19 pandemic has caused unprecedented damage to the global economy, making it crucial to forecast the fluctuations in the US economy under this impact.

This paper initially utilizes data from the US Google COVID-19 Community Mobility Reports, finding that the severity of COVID-19 indeed influences people's mobility and tendency to stay at home. The rampant spread of COVID-19 leads to a decrease in visits to retail and recreation, grocery and pharmacy, parks, transit stations, and workplaces, while increasing the probability of people staying at home. This initial finding confirms the relationship between COVID-19 data and economic mobility.

To forecast the economic decline caused by COVID-19 in the United States, this paper collects various macroeconomic data from sources including FRED and the World Bank Database, encompassing GDP, unemployment rate, inflation, PPI, housing price, population, and more. After employing models including VAR model, gradient boosted regression model, random forest regression model, K-nearest neighbors regression model, linear regression model, support vector machine regression model, and LSTM model to forecast the economy, it is observed that the optimized gradient boosted regression model exhibits the

lowest Mean Absolute Error (MAE) and performs the best in prediction. Furthermore, applying these models to Italian data for economic forecasting reveals that the optimized gradient boosted regression model continues to demonstrate superior performance, suggesting its considerable applicability.

While this paper demonstrates relatively good predictive performance of a series of models for the economic recession, these models also have their limitations, which are discussed at the five section of the article.

Although this paper demonstrates relatively good predictive performance of a series of models for the economy, these models also have their limitations. Moreover, the number of models used in this paper is limited, and future research could better optimize and expand the range of research models. However, overall, this paper makes some contributions to the research on economic recession prediction.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Baker, S., Bloom, N., Davis, S., & Terry, S. (2020). *COVID-Induced Economic Uncertainty.* NBER Working Paper No.26983. National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w26983

Chen, S., & Ranciere, R. (2019). Financial Information and Macroeconomic Forecasts. *International Journal of Forecasting, 35,* 1160-1174. https://doi.org/10.1016/j.ijforecast.2019.03.005

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & Team, T. O. I. (2020). *How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data.* NBER Working Paper No. 27431. National Bureau of Economic Research, Cambridge, MA.

Coven, J., & Gupta, A. (2020). *Disparities in Mobility Responses to COVID-19*. NYU Stern Working Paper, 2020.

Dingel, J. I., & Neiman, B. (2020). How Many Jobs Can be Done at Home? *Journal of Public Economics, 189,* Article ID: 104235. https://doi.org/10.1016/j.jpubeco.2020.104235

Farboodi, M., Jarosch, G., & Shimer, R. (2020). *Internal and External Effects of Social Distancing in a Pandemic.* NBER Working Paper No. 27059. National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w27059

Gilchrist, S., & Zakrajšek, E. (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review, 102,* 1692-1720. https://doi.org/10.1257/aer.102.4.1692

Hall, A. S. (2018). Machine Learning Approaches to Macroeconomic Forecasting. *The Federal Reserve Bank of Kansas City Economic Review, 103,* 63-81.

Huang, Y., & Yan, E. (2023) Economic Recession Forecasts Using Machine Learning Models Based on the Evidence from the COVID-19 Pandemic. *Modern Economy, 14,* 899-922. https://doi.org/10.4236/me.2023.147049

Ludvigson, S. C., Ma, S., & Ng, S. (2020). *COVID-19 and the Macroeconomic Effects of Costly Disasters.* NBER Working Papers No.26987. National Bureau of Economic Re-

search, Cambridge, MA. https://doi.org/10.3386/w26987

Maliszewska, M., Mattoo, A., & Van Der Mensbrugghe, D. (2020). *The Potential Impact of COVID*-19 *on GDP and Trade: A Preliminary Assessment.* World Bank Research Working Paper No.9211. https://doi.org/10.1596/1813-9450-9211

Padhan, R., & Prabheesh, K. P. (2021). The Economics of COVID-19 Pandemic: A Survey. *Economic Analysis and Policy, 70,* 220-237. https://doi.org/10.1016/j.eap.2021.02.012

Philippon, T. (2009). The Bond Market's q. *Quarterly Journal of Economics, 124,* 1011-1056. https://doi.org/10.1162/qjec.2009.124.3.1011

Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China. *Journal of Population Economics, 33,* 1127-1172. https://doi.org/10.1007/s00148-020-00778-2