|  |  |
| --- | --- |
| Team number | 2020090060007 |
| Problem Chosen: | B |

Analysis of the Effect of the US Election On Economy

2020 APMCM summary sheet

Republican candidate Donald Trump and Democratic counterpart Joe Biden are running for president this year. Since Biden and Trump have totally different political tendencies, we analyzes the two candidates’ political guidelines and others various factors based on the macroeconomic model (Y = C+I+G+NX) to forecast the economic changes in the coming year.

In data preparation, we use the closing index of stock market to simulate the daily GDP, the stocks related to the consumption companies to simulate the consumption, daily trading volumes of stocks to simulate the investment and the government budgets to simulate the government expenditure. We find the two candidates' debate draft and political guideline to extract two candidates’ political orientation. Here, we use PCA (Principal Component Analysis) to integrate the effect from one feature and use TF-IDF to extract the key words and analyze the political tendencies by reading the text that contains the key words.

In model selection, we test the present data by using XGBoosting, Linear Regression and LSTM (Long-Short-Term-Memory). Since the data is highly cyclical, we finally choose the LSTM to predict the American economy in the coming year. To forecast the future economy, we firstly use Adaboost with historical true data’s moving average and weight calculated by the trend from the last election in 2015 to forecast the factors’ value in the next year. Then we use SIRD to forecast the covid-19 situation in the next year and AHP to calculate the weights of two candidates in these factors. After finishing the prediction of the features’ future value, we input these data to forecast the future GDP by using selected LSTM model and get the results of the American economy.

For China, the Chinese daily GDP has an increasing linear trend year by year, so we test the present data using Linear Regression and find it fit better in Chinese situation. In China data preparation, we use the same method as we do in preparing the data for America. To observe the impact of election on China's economy, we input the forecast American economy’s data as a variable to regress and predict the Chinese future economy.

*Key words:* LSTM, Adaboot, Linear Regress PCA

**Content**

[1. Introduction 4](#_Toc57591129)

[1.1 Background of the problem 4](#_Toc57591130)

[1.2 Our work 4](#_Toc57591131)

[2. Analysis of the Whole Problem 4](#_Toc57591132)

[2.1 Basic Assumptions 4](#_Toc57591133)

[2.2 Data Pre-processing 4](#_Toc57591134)

[2.3 Symbol Definition 5](#_Toc57591135)

[3. Solution of the Requirement One 5](#_Toc57591136)

[3.1 Analysis of the Requirement 5](#_Toc57591137)

[3.2 Data Explanation 5](#_Toc57591138)

[3.2.1 Consumption 5](#_Toc57591139)

[3.2.2 Investment 8](#_Toc57591140)

[3.2.3 Government Purchase 8](#_Toc57591141)

[3.2.4 Net Export 8](#_Toc57591142)

[3.2.5 COVID-19 9](#_Toc57591143)

[3.2.6 Daily GDP 9](#_Toc57591144)

[3.5 LSTM 10](#_Toc57591145)

[3.5.1 Choosing of the Model 10](#_Toc57591146)

[3.5.2 Predict the Data 12](#_Toc57591147)

[3.5.3 The Result 23](#_Toc57591148)

[4. Solution of the Requirement Two 24](#_Toc57591149)

[4.1 Analysis of the Requirement 24](#_Toc57591150)

[4.2 Data Explanation 24](#_Toc57591151)

[4.2.1 Consumption 24](#_Toc57591152)

[4.2.2 Investment 25](#_Toc57591153)

[4.2.3 Government Purchase 25](#_Toc57591154)

[4.2.4 Net Export 25](#_Toc57591155)

[4.2.5 American Economy 25](#_Toc57591156)

[4.2.6 Daily GDP 25](#_Toc57591157)

[4.3 Regression Analysis 26](#_Toc57591158)

[4.3.1 Choosing of the Model 26](#_Toc57591159)

[4.3.2 Testing of the Model 26](#_Toc57591160)

[4.3.3 Predict the Data 26](#_Toc57591161)

[4.4 The Result 27](#_Toc57591162)

[5. The Letter of Suggestions 27](#_Toc57591163)

[6. Strengths and Weaknesses 30](#_Toc57591164)

[7. References 30](#_Toc57591165)

[8. Appendixes 31](#_Toc57591166)

# 1. Introduction

## 1.1 Background of the problem

Every four years the US holds the presidential election, while this year is special. Due to the effect of COVID-19 and other changes in economic situations, the result of this year's election was very close and the election was very anxious. The political leanings of the two campaigns differ in ways that affect the economy at home and abroad. Our team is required to collect the candidate’s policy propositions, policy guidelines and relevant data in different fields, and calculate their impact on the domestic and foreign economic situation.

## 1.2 Our work

The meaning of the four-sector economic cycle model of macroeconomics is the three-sector economy plus the foreign sector, so the four-sector economy includes consumption, investment, government purchase and net export. We used different data from the above four sectors to analyze the impact on GDP by substituting the political leanings of the candidates.

# 2. Analysis of the Whole Problem

## 2.1 Basic Assumptions

To simplify the following calculation, the assumptions are list as follows:

1. The stock market can reflect GDP, which are positively correlated.

2. The four-sector economic cycle model of macroeconomics is established.

3. The COVID-19 epidemic in China has been controlled, thus not affecting the economy, while the COVID-19 epidemic in the United States has not been controlled and still has a negative impact on the economy.

## 2.2 Data Pre-processing

* Missing Value Processing

Our main data comes from the stock market. The stock market is closed on the weekend, so we will analyze the data after the weekend.

For the missing value of the working day, we take the average value of this data as the population.

* Invalid Value Processing

We removed extreme occurrences of consumption data to avoid affecting the forecast.

## 2.3 Symbol Definition

|  |  |
| --- | --- |
| **Symbol** | **Meaning** |
| C | Consumption |
| I | Investment |
| G | Government Purchase |
| NX | Net Export |
| GDP | Gross Domestic Product |

Table 1

# 3. Solution of the Requirement One

## 3.1 Analysis of the Requirement

For Requirement One, we combine the speeches of two candidates and analyze them by extracting key words from the speeches for the four dimensions of Consumption, Investment, Government Purchase, and Net Export. At the same time, the LSTM algorithm in machine learning will be introduced to predict and analyze the GDP trends of different leaders.

## 3.2 Data Explanation

### 3.2.1 Consumption

In macroeconomics, Consumption refers to the total expenditure of consumer goods consumed by one person or one country in a certain period. CPI and consumer stocks are selected to analyze the consumption.

* CPI: The Consumer Price Index (CPI) is an index that measures the purchase prices of a selected basket of consumer goods, and the weight of each commodity is determined based on its share of urban consumer spending between 1982 and 1984. It is the index of price change that reflects the product and service price statistics that are related to people's life, and it is usually used as an important index to observe the level of inflation.
* Consumer stocks: The change of the consumer stocks can reflect a certain consumer degree of positive reaction. We selected the five relatively representative consumer stocks in different areas, including Coca-Cola, Procter & gamble, PepsiCo, Wal-Mart and Philip Morris. We used python to crawl the daily stock index from 2015 to 2020 to analyze consumer stocks situation changes.

#### 3.2.1.1 PCA

We use principal component analysis (PCA) to integrate each evaluation index into an overall index, that is, to comprehensively utilize the performance to reflect the original information.

Principal component analysis (PCA) is performed by taking correlated data from multiple indicators (e.g., X1, X2… XN) and using spatial downscaling of high dimensional variables, where the components reflect most of the original variable information that is not duplicated. Mathematically, the original N indicators are usually linearly combined to form a new composite indicator. The most widely used method is to use the variance of the first composite (F1) selected, the larger the variance, the more information it contains. In the analysis of Consumption-related data, we selected principal components with a cumulative contribution greater than 80%.

For the U.S. consumption data, we have six variables, the five consumer sector stock indices and the residential consumption index, with a total of 1,478 pieces of data.

**1. Standardize the data, calculate the mean value and standard deviation of the column and conduct z-score standardization.**

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |

So the standardized data is:

|  |  |
| --- | --- |
|  | (3) |

The normalized matrix is

|  |  |
| --- | --- |
|  | (4) |

**2. Calculate the eigenvalue and eigenvector of R**

|  |  |
| --- | --- |
|  | (5) |

r refers to

|  |  |
| --- | --- |
|  | (5) |

**3. Calculate the contribution rate and cumulative contribution rate of principal components**

Contribution Rate :

|  |  |
| --- | --- |
|  | (5) |

Cumulative Contribution:

|  |  |
| --- | --- |
|  | (5) |

The following table describes the eigenvalues of the correlation coefficient matrix, the corresponding eigenvectors and contributions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Eigenvalues | a1 | a2 | a3 | a4 | a5 |
| KO | 0.421 | 0.795 | -0.238 | 0.333 | -0.156 |
| PG | 0.453 | -0.072 | 0.673 | -0.181 | -0.552 |
| PEP | 0.467 | 0.145 | 0.094 | -0.559 | 0.663 |
| WMT | 0.453 | -0.428 | 0.116 | 0.693 | 0.343 |
| CPI | 0.440 | -0.399 | -0.685 | -0.252 | -0.338 |
| Eigenvalues Value | 4.343 | 0.332 | 0.202 | 0.082 | 0.0413 |
| Contribution Rate | 0.869 | 0.067 | 0.040 | 0.016 | 0.008 |
| Cumulative Contribution | 0.869 | 0.935 | 0.975 | 0.992 | 1.000 |

Table the Coefficient Matrix

It can be seen from the table that the cumulative contribution rate of the first principal component reaches 86.9%, so we choose the first to represent the consumption.

|  |  |
| --- | --- |
|  | (6) |

We use python to calculate the reduced dimensional data, which is the U.S. Consumption Index. (Data is attached to the supporting material)

### 3.2.2 Investment

In the United States, the stock market can reflect investment situation. There are three indexes that can reflect the trends and tendencies of investment, which are the NASDAQ, Dow Jones and S&P 500. The Dow Jones has been established for the longest time and is relatively more stable.

### 3.2.3 Government Purchase

The Government Purchases refers to purchase goods and services from all levels of Government spending, it is a part of Government spending. In order to analyze the changes in government purchases in recent years, we selected the federal government financial budget from the Open data of the U.S. Bureau of Commerce.

### 3.2.4 Net Export

NX which refers to the amount of net exports, can be reflect the Balance of trade(the Balance of imports and exports measured in local currency in a certain period of time).

The Balance of trade, also known as net exports, is the difference between a country's imports and exports measured in its own currency over a period of time and forms part of the Current account. The trade account exports more than the import is called "surplus", "surplus" or "surplus"; Otherwise, it is called "deficit" or "deficit" or "excess". Our open data set of the US Bureau of Commerce intercepted the data from 2015 to November 26, 2020 for analysis.

### 3.2.5 COVID-19

The cumulative number of infected people in the United States is now in the tens of millions, and the number of deaths is rising steadily to more than 200,000, with a clear upward trend, with varying degrees of negative impact on all aspects of politics and economics. The two leaders have different attitudes toward the epidemic, so data on the epidemic should be considered.

### 3.2.6 Daily GDP

GDP (Gross Domestic Product) is a good indicator to reflect the economic operation, so we take it as the dependent variable to simulate and forecast. We used the daily closing price of NASDAQ index to simulate the daily GDP according to the following formula:

|  |  |
| --- | --- |
|  | (7) |

In which  represents the closing number, and represents the monthly GDP.

We conducted a correlation analysis of all the indicators and obtained the following correlation matrix, which can justify the selection of indicators.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **C** | **I** | **G** | **NX** | **new\_cases** | **new\_deaths** | **GDP** |
| **C** | 1.00 | 0.47 | 0.93 | -0.15 | 0.37 | 0.38 | 0.43 |
| **I** | 0.47 | 1.00 | 0.51 | -0.30 | 0.14 | 0.26 | 0.18 |
| **G** | 0.93 | 0.51 | 1.00 | -0.16 | 0.28 | 0.33 | 0.46 |
| **NX** | -0.15 | -0.30 | -0.16 | 1.00 | -0.22 | -0.35 | 0.10 |
| **new\_cases** | 0.37 | 0.14 | 0.28 | -0.22 | 1.00 | 0.68 | -0.29 |
| **new\_deaths** | 0.38 | 0.26 | 0.33 | -0.35 | 0.68 | 1.00 | -0.46 |
| **GDP** | 0.43 | 0.18 | 0.46 | 0.10 | -0.29 | -0.46 | 1.00 |

Table the Correlation Matrix,

## 3.5 LSTM

### 3.5.1 Choosing of the Model

We used algorithm of Stacking integration learning, GBXT algorithm, LGB algorithm, and LSTM algorithm to fit and predict, in which THE LSTM algorithm has the best fitting effect.

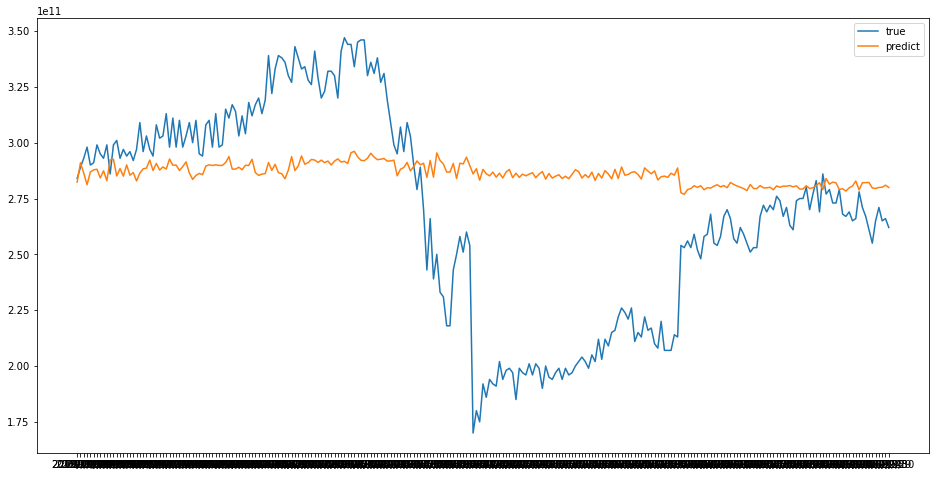


Figure Boosting Fitting

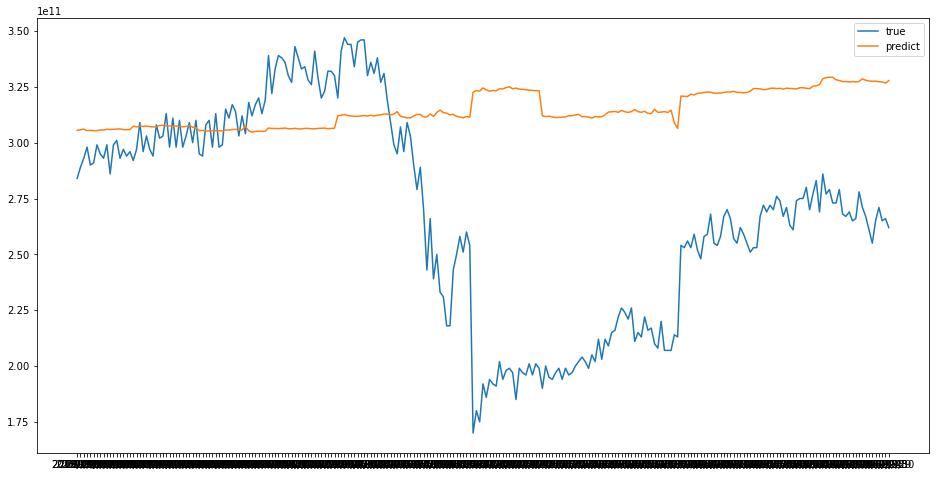


Figure Linear Fitting

LSTM neural network model is a special cyclic neural network model (RNN). One of the key points of RNN is that it can connect the previous information to the current task, that is, it can memorize the previous information and apply it to the calculation of the current output, and analyze and predict the time series data. LSTM mainly solves problems such as GRADIENT disappearance and gradient explosion of RNN, and has its own memory and can make accurate prediction. As a very powerful prediction tool, LSTM is specially designed to solve long-term problems.

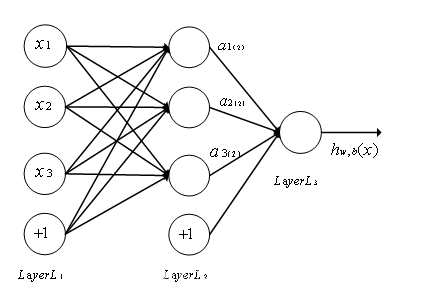


Figure the Explanation of LSTM

#### 3.5.1.1 Testing of the Model

**1.** **The transformation of the raw data**

First, due to different dimensions, we normalized the raw data by Max-min according to the following formula:

|  |  |
| --- | --- |
|  | (8) |

We set the first 80% of the data as a training set, which is a total of 1,200 pieces, and the rest as a test set to evaluate the accuracy of the model.

**2. Selection of Eigenvalues**

According to the initial definition, we had six variables, and we divided them into two categories for simulation testing. The first includes consumption, investment, government purchase and net export (NX), which are the elements of the four-sector cycle in macroeconomics. The second category added the average number of deaths per day and the average number of cases per day to the epidemic data based on the first category. Compare which data classification is more appropriate.

**3. LSTM Neural Network Training**

We used Python Keras package for training. With two hidden layers, the epoch trained three times, the batch\_size (number of steps per training) was 8, the time step was 1, the loss function was mean square error, the activation function of LSTM module was tanh, and Adam optimizer was used to optimize and output the result.

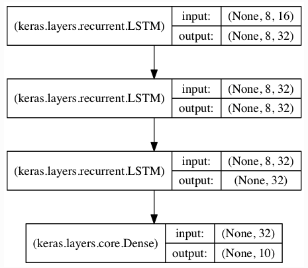


Figure the Step of LSTM

**4. The Result of Training**

Two different training models can be obtained from two different choice. The data are inverse normalized to make an image. When training for the first time, the loss is about 0.036, and as the number of training sessions increases, the loss shows a decreasing trend, and when training more than three times, the loss value is stable at about 0.006. In order to prevent the occurrence of over-fitting of the model, the team set the training period as three cycles.

For the existing data, our fit of the second category accuracy is as high as 86%, so we decided to use the LSTM as the main model to predict the future GDP.

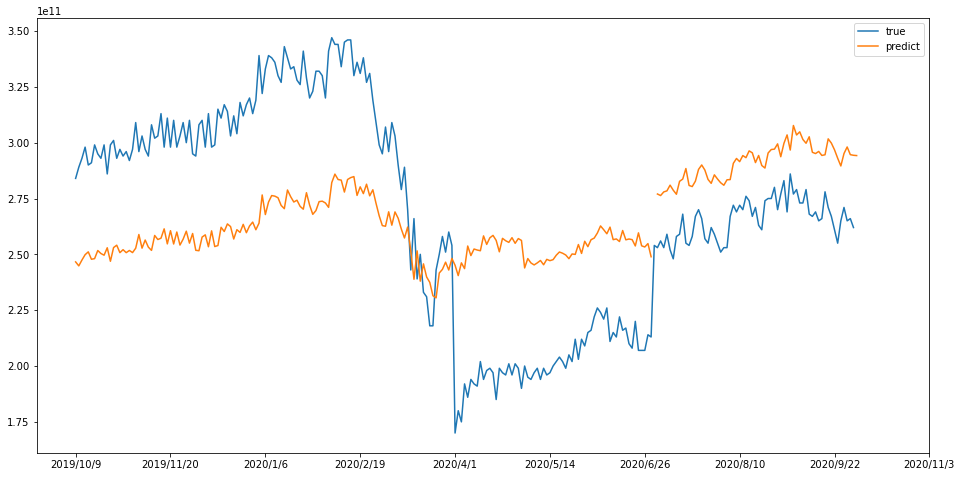
****

Figure the Result of Fitting

### 3.5.2 Predict the Data

In the analysis of the indicators for predicting the U.S., we use different methods for different indicators. First, we use the AHP model to analyze the impact of different candidates' political platforms on various aspects. For the covid-19 data, we use the SIRD epidemic model to predict different infection and cure rates for different candidate attitudes, and AdaBoost combined with the 60-day moving average for the other variables.

#### 3.5.2.1 Text Analysis

* **TP\_IDF**

Term frequency (TF) refers to the number of times a given word appears in the document, while inverse document frequency (IDF) is a measure of the general importance of the word.

In order to present the key words in the debates and speeches of Trump and Biden, the following steps are taken：

First, searching materials such as policy guidelines, speeches and campaign debates from both Trump’s and Biden’s campaign websites. Using python package ‘jieba’ to separate the words of two candidates.

Secondly, creating a thesaurus based on the separated words, combined with common political and economic terms.

Third, we counted the term frequency (TF) and inverse document frequency (IDF) of the two corpus separately, and obtained the TF\_IDF weight ranking, which will be used as a reference for selecting the AHP criterion layer in the following steps.

Term frequency (TF) refers to the number of times a given word appears in the document, while inverse document frequency (IDF) is a measure of the general importance of the word.

|  |  |  |
| --- | --- | --- |
| C:\Users\LENOVO\AppData\Local\Temp\ksohtml23988\wps11.jpg | (9) | |
| C:\Users\LENOVO\AppData\Local\Temp\ksohtml23988\wps12.jpg | | (10) |
| C:\Users\LENOVO\AppData\Local\Temp\ksohtml23988\wps13.jpg | | (11) |

D: Total number of words in the document

Dw: the number of a specific word in a text

We collected a large number of Biden and Trump speeches and debate transcripts and counted their word frequency, as shown in the following word cloud:

|  |  |
| --- | --- |
| C:\Users\LENOVO\AppData\Local\Temp\WeChat Files\8016f11a8974b5b4e015688225b5896.png | **C:\Users\LENOVO\AppData\Local\Temp\WeChat Files\8016f11a8974b5b4e015688225b5896.png** |

Figure the Word Cloud

* **Analytic Hierarchy Process(AHP)**

For each variable (C, G, I, NX), three representative words were selected to analyze the tendencies of the two candidates in the four directions using AHP.

The main characteristic of AHP is to transform the human judgment to the comparison of the importance between several factors by establishing the hierarchical structure, thus transforming the qualitative judgment which is difficult to quantify into the comparison of the operational importance. We will take Trump's Consumption related attitude as an example to explain the establishment and solution of AHP model.

**1. Analyze the relationship between the factors in the system and establish the hierarchy of the system.**

The hierarchical model is divided into three levels. The top level is the target level, that is, the variables of the four departments of economics selected by us; the middle level is the criterion level, that is, the vocabularies selected by us; and the bottom level is the program level, which contains the positive and negative degree of vocabularies.

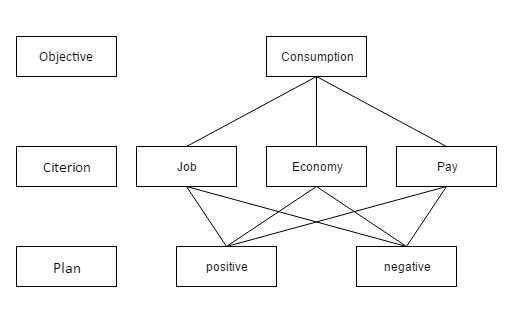


Figure the AHP Model

We judge relative importance according to the following table:

|  |  |
| --- | --- |
| Score | Meaning |
| 1 | The Same Importance |
| 3 | Slightly More Important |
| 5 | Significantly More Important |
| 2、4 | The mean of the Two Importance |
| Reciprocal | If the scale of A and B is 3, then B and A are 1/3 of the scale. |

Table Relative Importance

We construct judgment matrix O-C, and let the three indexes of C layer compare in pairs

|  |  |  |  |
| --- | --- | --- | --- |
| O | **Job** | **Economy** | **Pay** |
| Job | 1 | 3 | 5 |
| Economy | 1/3 | 1 | 1 |
| Pay | 1/5 | 1 | 1 |

Table

We calculate the maximum eigenvalue, and calculate Ci and CR with the following formula

|  |  |
| --- | --- |
|  | (12) |
|  | (13) |

The CI value of the matrix is 0.0146, and the RI value of the matrix is 0.58. Therefore, the CR value of the matrix is 0.0252 less than 0.1, which can pass the consistency test

Therefore, we use the eigenvalue method to calculate its eigenvectors, and get the normalized weight vectors as follows

The corresponding feature vector is: 

After normalization, the weight vector is:

Repeat the above steps to calculate the eigenvalues and eigenvectors of each C-P matrix to obtain the following consistency matrix. ,and all of which pass consistency test.

|  |  |  |
| --- | --- | --- |
| **C1** | **positive** | **negative** |
| **positive** | 1 | 5 |
| **negative** | 1/5 | 1 |

Table

|  |  |  |
| --- | --- | --- |
| **C2** | **positive** | **negative** |
| **positive** | 1 | 3 |
| **negative** | 1/3 | 1 |

Table

|  |  |  |
| --- | --- | --- |
| **C3** | **positive** | **negative** |
| **positive** | 1 | 3 |
| **negative** | 1/3 | 1 |

Table

The table below shows the statistical score results of final Consumption related two different attitudes

|  |  |  |  |
| --- | --- | --- | --- |
|  | **The Weight Indicator** | **Positive** | **Negative** |
| **Job** | 0.6586 | 0.8333 | 0.1667 |
| **Economy** | 0.1852 | 0.75 | 0.25 |
| **Pay** | 0.1562 | 0.75 | 0.25 |

Table the Different Attitude

Therefore, the calculated PC score is 0.80486138, and the NC score is 0.19513862

Repeating the above steps, we separately calculated the positive and negative ratings for each of the two categories of leaders, as shown in the table below

|  |  |  |
| --- | --- | --- |
|  | **Trump** | **Biden** |
| **PC** | 0.80486138 | 0.79580667 |
| **NC** | 0.19513862 | 0.20419333 |
| **PI** | 0.8030621 | 0.75872151 |
| **NI** | 0.1969379 | 0.24127849 |
| **PG** | 0.8030621 | 0.80392009 |
| **NG** | 0.1969379 | 0.19607991 |
| **PNX** | 0.79843895 | 0.8030621 |
| **NNX** | 0.20156105 | 0.1969379 |

Table

For each category of positive trends, we normalize them using the following formula, which is used to predict economic trends in each category of weighted calculations.

|  |  |
| --- | --- |
|  | (13) |

|  |  |  |
| --- | --- | --- |
|  | Trump | Biden |
| PC | 1.005656832 | 0.994343168 |
| PI | 1.028390995 | 0.971609005 |
| PG | 0.999466086 | 1.000533914 |
| PNX | 0.997113239 | 1.002886761 |

Table

The final formula of GDP direction is as follows:

|  |  |
| --- | --- |
|  | (14) |

#### 3.5.2.2 SIRD

Unlike that in China, the COVID-19 epidemic in the United States is not yet under control and will have long-term impacts. Therefore, we will consider the impact of the epidemic on the economy and add it as an independent variable.

Basic assumptions as follows

|  |  |
| --- | --- |
| **Symbol** | **Explanation** |
|  | Percentage of current susceptible population |
|  | Percentage of the population currently infected |
|  | Percentage of population currently cured |
|  | Percentage of the population currently dying |
|  | Number of effective contacts |
|  | Daily Contact Rate |
|  | Daily Cure Rate |
|  | Daily Dead Rate |

Table

Carriers and incubation periods are not considered so the following relationships can be listed for all groups.

|  |  |
| --- | --- |
|  | (14） |

Therefore, the daily contact rate has the following expression (i.e., the basic regeneration rate)



According to the above relationship we can list the difference equation for each group.

For the infected group

|  |  |
| --- | --- |
|  | (15) |

For the safe group

|  |  |
| --- | --- |
|  | (16) |

For the recovery group:

|  |  |
| --- | --- |
|  | (17) |

For the dead group:

|  |  |
| --- | --- |
|  | (18) |

Since the value is of a small period of time , the difference equations are converted into differential equations.

|  |  |
| --- | --- |
|  | (19) |

|  |  |
| --- | --- |
|  | (20) |

|  |  |
| --- | --- |
|  | (21) |

|  |  |
| --- | --- |
|  | (22) |

With the analysis of epidemic in the United States and the census, and with the population of 303 million, we can calculate the current infection rate, recovery rate and death rate to forecast the evolution of the epidemic. We calculated the average infection rate from the beginning of the epidemic to the present as the benchmark, then added the positive and negative factors of the policy, and used MATLAB to predict the development of the epidemic in the next first year.

Biden's policy includes the management of the COVID-19 outbreak, which will increase health care spending on the epidemic. Therefore, it has a positive impact on the control of the epidemic situation, the following graph is the forecast of the development of the covid-19 under the Biden’s policy

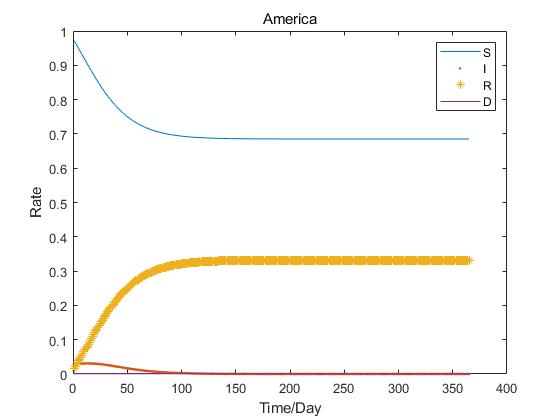


Figure the Predict of COVID-19 under Biden’s Policy

Trump has a negative attitude towards the epidemic, adopts herd immunization, and intends to abolish universal health insurance, so that fewer medical resources will be used for the increase of the epidemic infection rate and the reduction of the cure rate, thus predicting the development trend of the epidemic under Trump's leadership.

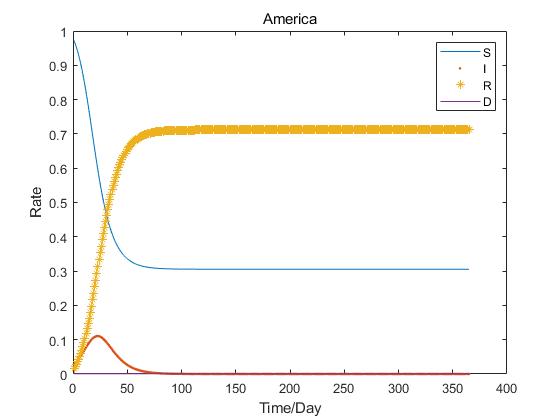


Figure the Predict of COVID-19 under Trump’s Policy

#### 3.5.2.3 Ada boost and 60 average line

* **Adaboost**

The AdaBoost algorithm is a boosting method that takes several weak classifiers and combines them into a strong classifier. In the AdaBoost algorithm, the weights of the weak classifiers are adjusted after each training round, and the weight of the misclassified points in the previous training round will increase.

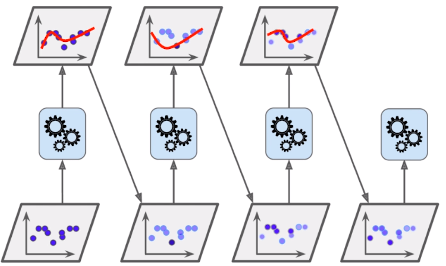
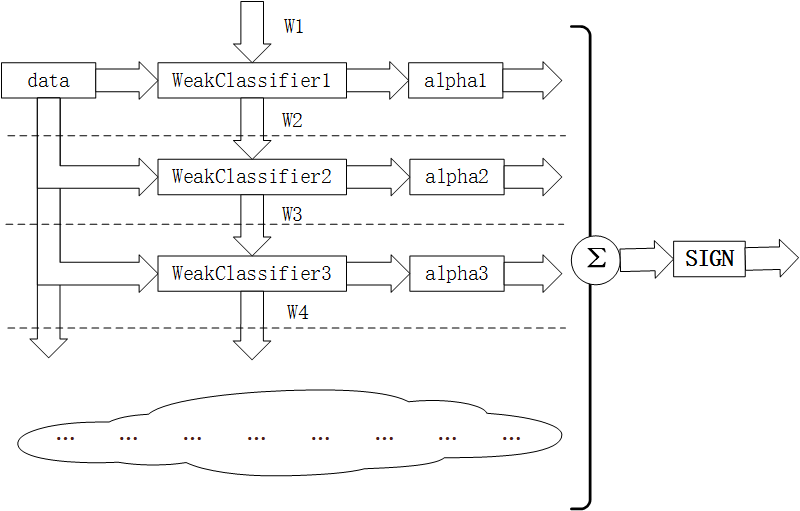


Figure the Explanation of Adaboost

Since the relationship between several classifiers in Ada boost is such that the N classifier is more likely to split the data that the N-1 classifier does not split, it can’t guarantee that the data previously split will be split at the same time. So in Adaboost, each weak classifier has its own most important point of interest, and each weak classifier is only interested in a part of the whole data set, so they must be combined together in order to be more effective. The weights are based on the classification error rate of the weak classifiers. The general rule is that the lower the error rate of a weak classifier, the higher the weight.



Figure

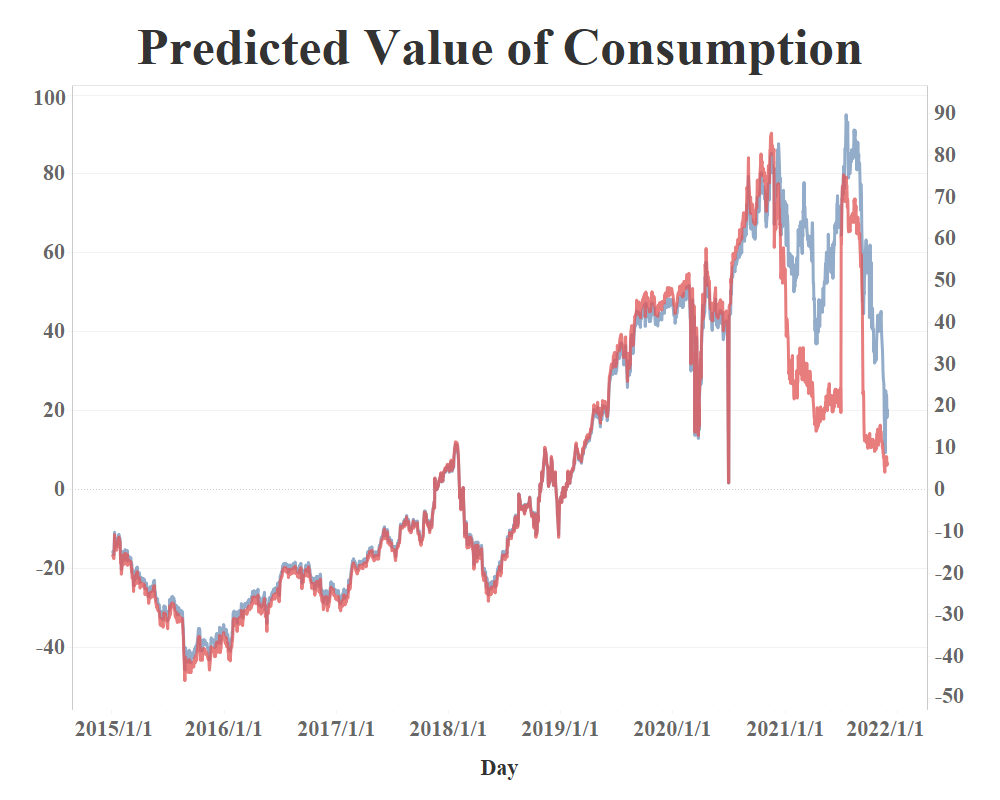
* **60-Day Moving Average**

The moving average is the average closing price of a stock for the previous fixed days, and its significance lies in the fact that it reflects the long-term or short-term changing trend . 60-day moving average is generally a medium- to long-term trend, and the 60-day moving average is the average closing price of the last two months, which is important for the later trend of a stock.

Therefore, we use the 60-day moving average in combination with Ada Boost to make rolling forecasts that better reflect the impact of economic factors in the near future.

Because of this prediction of economic development after the election, we use various data from before and after the 2015 election as weights, while adding thresholds to slow its convergence.

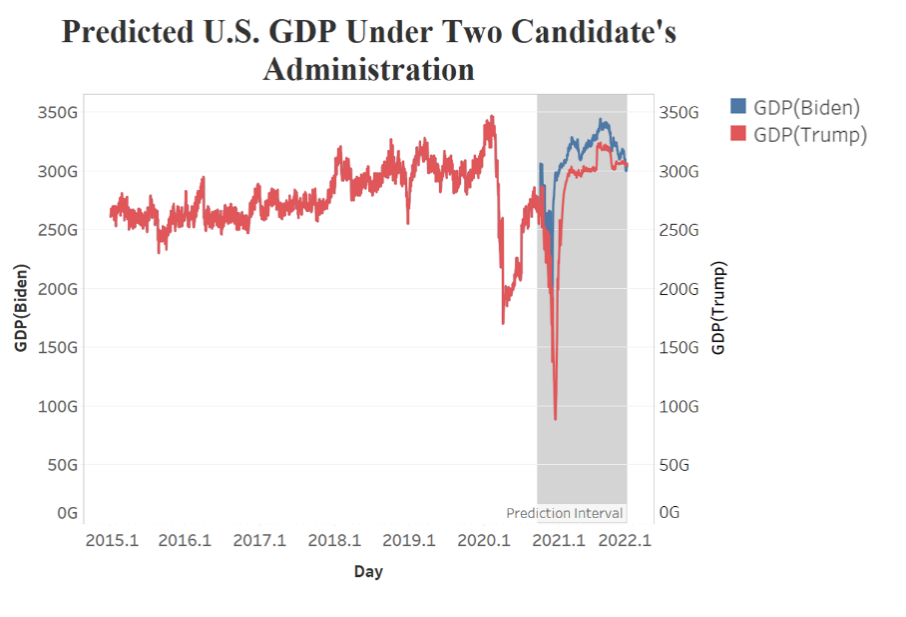
Taking the U.S. consumption data as an example, the following graph is predicted to meet the trend well.



Figure

### 3.5.3 The Result

After predicting each input data, we train the LSTM model by weighting it with political preferences, and then we obtain the following graph of GDP trends under different candidates.



Figure

It is clear from our projections that under Biden's leadership there will be a greater positive impact on GDP growth, while next spring there will be a decline, possibly due to a second outbreak of the epidemic.

# 4. Solution of the Requirement Two

## 4.1 Analysis of the Requirement

For Requirement Two, we used the same indicators from the four sectors of macroeconomics and added the GDP development of the United States under different leaders to get an idea of how China's economic situation would have changed under different leaders.

## 4.2 Data Explanation

### 4.2.1 Consumption

For the Chinese data, we use the CPI, the consumer confidence index and 10 consumption sector index. All of the stocks we chose are representative companies in the consumer sector, covering all aspects of consumer spending.

The consumer confidence Index is an indicator of the strength of consumer confidence. It comprehensively reflects and quantifies consumers' assessment of the current economic situation and their subjective feelings towards the economic prospect, income level, income expectation and consumption psychological state, and predicts the economic trend and consumption trend. It is composed of consumer satisfaction index and consumer expectation index. The former refers to consumers' evaluation of current economic life, while the latter refers to consumers' expectations of future changes in economic life.

For the above indicators, we also use principal component analysis (PCA) to reduce the dimensions, the steps are the same as that in requirement one. More than 80% of the principal component is selected for analysis, and the first component satisfy.

### 4.2.2 Investment

Similarly, we use the daily trading volume of stocks to analyze the amount of investment. The CSI 300 is a financial indicator that reflects the compilation target and operation status of the CSI 300 index, which can be used as an evaluation standard for investment performance. We crawled the DAILY trading volume data of the CSI 300 in the past five years for analysis.

### 4.2.3 Government Purchase

We use the data set of monthly government revenue from Eastmoney

(http://data.eastmoney.com/Eastmoney. com ) ,which can reflect government purchases.

### 4.2.4 Net Export

For the net export data, we calculate from the monthly total of imports and exports.

### 4.2.5 American Economy

Since each country's economy interacts with each other in international trade, the forecasted GDP data from Requirement One is used as a new set of variables.

### 4.2.6 Daily GDP

As for the dependent variable Y, which is the gross production, we also calculate it from the monthly gross GDP, and then used the daily closing price of CSI 300 index to calculate the daily weighted average according to the following formula. The calculation method is as follows:

|  |  |
| --- | --- |
|  | (23) |
| We conducted a correlation analysis of all the indicators and obtained the following correlation matrix, which can justify the selection of indicators: | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | C | I | G | NX | America\_GDP | GDP |
| C | 1.00 | -0.19 | 0.06 | -0.11 | 0.16 | 0.77 |
| I | -0.19 | 1.00 | -0.02 | 0.06 | -0.15 | -0.33 |
| G | 0.06 | -0.02 | 1.00 | 0.27 | -0.25 | 0.10 |
| NX | -0.11 | 0.06 | 0.27 | 1.00 | -0.48 | -0.08 |
| America\_GDP | 0.16 | -0.15 | -0.25 | -0.48 | 1.00 | 0.31 |
| GDP | 0.77 | -0.33 | 0.10 | -0.08 | 0.31 | 1.00 |

Table

## 4.3 Regression Analysis

### 4.3.1 Choosing of the Model

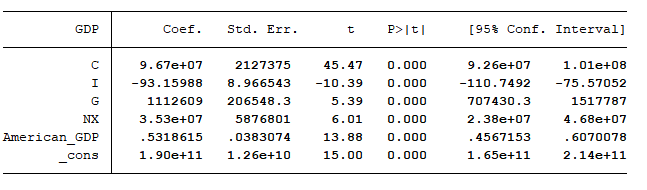
Since the data collected were cross-sectional, we chose a linear regression model. The linear regression model consists of a regression that allows us to obtain the correlation between the independent variable and the dependent variable and the importance of the independent variable.

### 4.3.2 Testing of the Model

We set up the following model to calculate and forecast the subsequent trend:

|  |  |
| --- | --- |
|  | (24) |

Using Least-Squares (OLS), the existing data were tested by Stata and the results are shown in the following table:



Figure

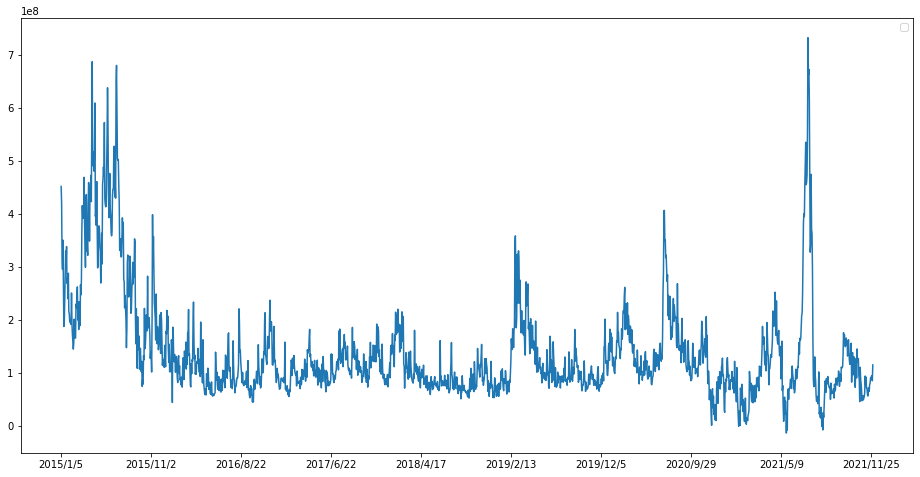
P value less than 0.05 means that the regression coefficient is significantly different from zero at the 95% confidence level, so each variable can reflect well.

### 4.3.3 Predict the Data

As in the same in the requirement one, we use a 120-day forecast period, combining with the Adaboost to make a rolling forecast.

We make an analysis of the Chinese economic situation. Since the economy showed a positive trend after 2015 and a certain decline after that, we will set a maximum threshold and a minimum threshold, with the maximum threshold being the peak in 2019 and the minimum threshold being the fuse during the epidemic to slow down the convergence.

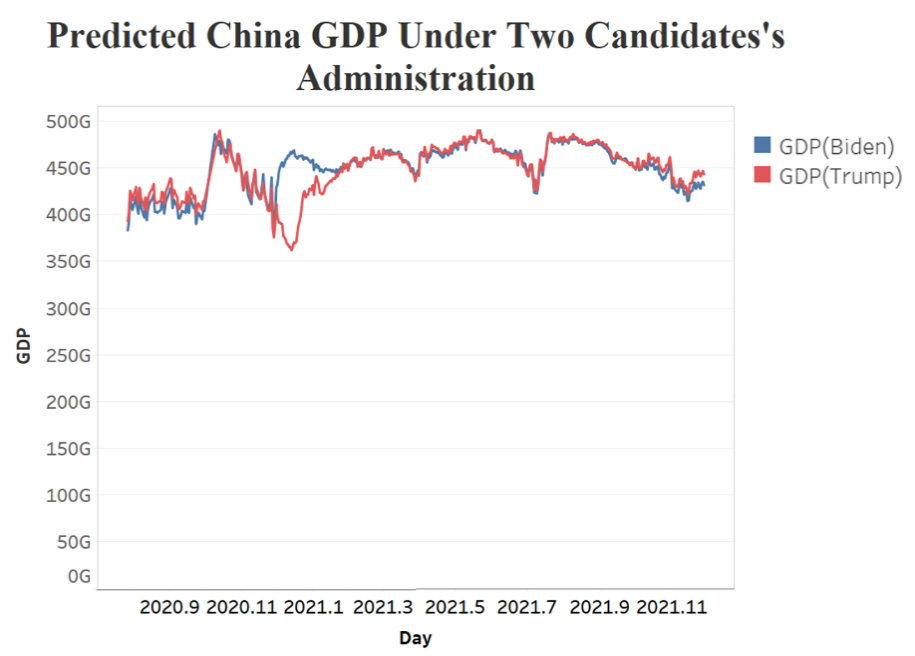
Take investment prediction as an example. As can be seen from the following chart, the fitting can reflect the previous trend.



Figure

## 4.4 The Result

With linear regression and the results of the first question, we make a one-year forecast of China's GDP. The results are shown in the figure below, and it can be seen that under the Biden administration, China's economy has a more positive trend in the near future and will converge in the later part of the year.



Figure

# 5. The Letter of Suggestions

Dear Mr./Ms.,

On behalf of China's Think Tank for Economic Development, I am honored to provide you with some insight into China's economic and political responses to the U.S. election. Two candidates, Joe Biden and Donald Trump will have different impacts on the US economy as well as the international economy due to their political stands and policy guidelines. China, as one of the biggest countries in the world, is also affected by the U.S. election to a certain extent. Therefore, we collected data on the U.S. economy, trade, and consumption, analyzed debate scripts of Biden and Trump, and modeled the impact of the election of the two candidates on the GDP of the U.S. and China, respectively. Here's what we found:

Since Biden took office, the U.S. GDP has steadily increased since the end of 2020, and the U.S. economy has gradually recovered from the impact of the epidemic. Biden will treat COVID-19 scientifically to reduce the impact of the epidemic on the U.S. economy. In terms of fiscal stimulus, Biden has adopted a massive fiscal stimulus, emphasizing large-scale infrastructure (especially in clean energy). This massive infrastructure will provide millions of jobs at home. On foreign trade, the United States will strengthen alliances, lower tax rates, and promote multilateral trade. Since Biden has adopted a milder strategy to contain China's growth, China's GDP trend is more optimistic in the model, but we cannot ignore Biden's tougher stance on China.

As a result, China could make the following adjustments in its policy after Biden takes office.

1.The RCEP (Regional Comprehensive Economic Partnership Agreement) is a trade agreement signed by China, Japan, South Korea, Australia and other countries. After coming to power, Biden is likely to rejoin the TPP and undermine RCEP's landing. In addition, the Biden administration will also negotiate TTIP (Transatlantic Trade and Investment Partnership) with the EU to strengthen U.S.-EU relations. In this scenario, the signing of EU-China investment and trade agreements will be affected by the Biden administration's pressure on the EU. Therefore, China should further increase its openness, speed up the establishment of free trade zones, reach duty-free agreements with many countries, and promote trade in a globalized manner.

2. Strengthen openness, mutual benefit, and attract the pooling of global capital. In the long run, the rise of China will erode the U.S. global profit-seeking space in the long run. When Biden takes office, he is likely to avoid direct conflict with China and at the same time gather allies to form an anti-China coalition. Therefore, China should participate in international affairs with a more open attitude, pay active attention to international trends, and grasp the right to speak out.

3. Cooperate with the United States in various aspects. On the whole, if Biden succeeds in his campaign, it will ease the tension between China and the United States to a certain extent. China can cooperate with the U.S. in the fight against COVID-19 epidemic as well as climate issues, and so on. In terms of the COVID-19, the Democratic Party has put more emphasis on the prevention and control of the epidemic, while the Trump administration has been blamed for the unfavorable outcome of the U.S. epidemic. China can join forces with the U.S. to fight the epidemic to the mutual benefit of both countries. In terms of environmental protection, Biden would rejoin the Paris Agreement and work with China on global climate issues. China can use this opportunity to improve its domestic environmental protection, accelerate the development of related industries.

If Trump is re-elected, the U.S. economy will have a larger decline in the short term and a lower GDP trend in the long term than it would have under Biden’s administration. At the same time, China's GDP will be limited to a certain extent due to Trump's more extreme methods (trade war, financial war) and other means to contain China's development. The model results show that China's GDP declines significantly in early 2021, mainly because of the impact of Trump's presidency. Trump will pursue unilateralism, advocate "America First", and encourage the return of jobs in the manufacturing sector, such as taking back more jobs from China. At the same time, Trump will further restrict China in terms of U.S.-China trade, intellectual property rights, export subsidies, and so on. However, due to Trump's unilateralism around the world, which goes against the trend of globalization, the U.S. appears to have a weakened voice in some areas. China can build on its strengths and gain more opportunities.

In the event that Trump wins the U.S. presidential election, China can focus on the following areas:

1. Build a healthy and sustainable domestic consumer market in China. Trump's term of office will impose further restrictions on the Chinese market. Therefore, the Chinese government will need to step up its efforts to support the transformation and upgrading of consumption, improve the quality of products and services, and cooperate with the new round of supply-side reform in the service industry to address the mismatch between supply and demand in the consumer sector.

2. Maintain a positive and open posture and actively respond to the potential friction through soft diplomacy. For example, even though the U.S. has a high deficit with Chinese goods, service traders have been showing a surplus. U.S. service sector exports to China have created a large number of jobs in the U.S., and U.S. high-tech companies have continued to profit from the Chinese market, which can be a bargaining chip for China to negotiate with the U.S. At the same time, China should continue to increase its service sector exports to China. At the same time, China should continue to improve the development level of the service industry and seek more allies.

3. Adapt to and guide economic globalization and mitigate its negative effects. At a time when Trump is withdrawing from the negotiations of multilateral trade agreements and showing obvious anti-globalization tendencies, China should take this opportunity to show its great power role and accelerate the pace of opening up while pushing for domestic market-oriented reform, and continuously promote the development of multilateral trade. China can establish free trade zones and sign trade agreements with other countries to promote trade.

These are our suggestions to help you in developing your strategy. Hopefully, these suggestions will be of great in making economic strategies and political countermeasures towards U.S.-China relations.

Yours respectfully,

China's Think Tank for Economic Development

# 6. Strengths and Weaknesses

* **Strength：**

Using the 60-day moving average in combination with Ada Boost to make rolling forecasts that better reflect the impact of economic factors in the near future.

* **Weakness：**

Since we only selected data for the last four years, the volume of data does not support long-term forecasts of economic conditions.

When using the stock market to simulate economic changes, it is not possible to predict all of the GDP, only part of it.

# 7. References

[1]Cao Tong,Bai Yanping. Research on air quality prediction by LSTM based on gradient descent optimization[J]. Journal of Shaanxi University of Science and Technology,2020,38(06):159-164.

[2]Xu Quan-Han. U.S. election boosts market risk aversion and international gold prices rise sharply[N]. China Gold News, 2020-11-10(005).

[3]Tang Xiaobin, Dong Manru, Zhang Rui. A Machine Learning LSTM&US Model-based Consumer Confidence Index Prediction Study[J]. Statistical Research,2020,37(07):104-115.

[4]Feng Yuxu,Li Yu-Mei. LSTM neural network based prediction model for CSI 300 index[J]. Mathematical Practice and Understanding,2019,49(07):308-315.

# 

# 8. Appendixes

1.America/LSTM/LSTM code.ipynb

1. **from** sklearn.decomposition **import** PCA
2. **from** sklearn.preprocessing **import** MinMaxScaler
3. **import** pandas as pd
4. **import** numpy as np
5. **import** matplotlib.pyplot as plt
6. data = pd.read\_csv('model data final.csv')
7. data.set\_index('Day',inplace=True)
8. stock = data
9. stock.drop(['Year','Season','Month','GDP Season','Close'],axis=1,inplace=True)
10. """PCA covid\_data"""
11. covid = data[['new\_cases','new\_deaths']]
12. covid.fillna(0,inplace=True)
13. pca = PCA(n\_components=0.9)
14. pca\_data = pca.fit\_transform(covid)
15. pca\_data = pd.DataFrame(pca\_data)
16. pca\_data.index = data.index
17. scaler = MinMaxScaler(feature\_range=(0, 0.8))
18. scaled\_data = scaler.fit\_transform(pca\_data)
19. scaled\_data = pd.DataFrame(scaled\_data)
20. scaled\_data.index = data.index
21. covid\_data = scaled\_data
22. """data prepare"""
23. time\_stamp = 1 # stamp
24. train\_num = 1200
25. train = stock[:train\_num + time\_stamp]
26. test = stock[train\_num - time\_stamp :]
27. scaler = MinMaxScaler(feature\_range=(0, 1))
28. scaled\_data = scaler.fit\_transform(train)
29. x\_train, y\_train = [], []
30. scaled\_data = scaled\_data.tolist()
31. **for** i **in** range(time\_stamp, len(train)):
32. x\_train.append(scaled\_data[i - time\_stamp:i])
33. y\_train.append(scaled\_data[i][-1])
34. x\_train, y\_train = np.array(x\_train), np.array(y\_train)
35. scaled\_data = scaler.fit\_transform(test)
36. scaled\_data = scaled\_data.tolist()
37. x\_test, y\_test = [], []
38. **for** i **in** range(time\_stamp, len(test)):
39. x\_test.append(scaled\_data[i - time\_stamp:i])
40. y\_test.append(scaled\_data[i][-1])
41. **print**(scaled\_data[i - time\_stamp:i])
42. **print**(scaled\_data[i][-1])
43. x\_test , y\_test = np.array(x\_test) , np.array(y\_test)
44. **from** keras.models **import** Sequential
45. **from** keras.layers **import** LSTM
46. **from** keras.layers **import** Dense
47. # param
48. epochs = 3
49. batch\_size = 8
50. model = Sequential() # set\_Sequential
51. # add 2 LSTM model
52. model.add(LSTM(units=100, return\_sequences=True, input\_dim=x\_train.shape[-1], input\_length=x\_train.shape[1]))
53. model.add(LSTM(units=50))
54. model.add(Dense(1)) # set\_Dense
55. model.compile(loss='mean\_squared\_error', optimizer='adam')
56. model.fit(x\_train, y\_train, epochs=epochs, batch\_size=batch\_size, verbose=1) # set\_Optimizer
57. predict\_close\_price = model.predict(x\_test)
58. scaler.fit\_transform(pd.DataFrame(test['GDP\_per\_day'].values))
59. pred = scaler.inverse\_transform(predict\_close\_price)
60. y\_test = scaler.inverse\_transform([y\_test])
61. """Plot the result"""
62. plt.figure(figsize=(16, 8))
63. dict\_data = {
64. 'Predictions': pred.reshape(1,-1)[0] ,# 设置放大参数
65. 'GDP\_per\_day': y\_test[0]
66. }
67. data\_pd = pd.DataFrame(dict\_data)
68. plt.plot(data\_pd['GDP\_per\_day'],label = 'true')
69. plt.plot(data\_pd['Predictions'],label = 'predict')
70. plt.xticks(data\_pd.index[::30],stock[train\_num:].index[::30])
71. plt.legend()
72. plt.show()

2. America/LSTM/Stacking(tesing).ipynb

1. """Using other model to train data"""
2. **import** pandas as pd
3. **import** numpy as np
4. **import** matplotlib.pyplot as plt
5. stock = stock[pd.isnull(stock['GDP\_per\_day']) == False]
6. **from** sklearn.impute **import** SimpleImputer
7. # fill the missing value
8. imputer = SimpleImputer(missing\_values= np.NaN , strategy= 'mean')
9. var\_data = imputer.fit\_transform(stock)
10. var\_data = pd.DataFrame(var\_data)
11. var\_data.columns = stock.columns
12. var\_data.index = stock.index
13. stock = var\_data
14. # other model
15. **from** sklearn.linear\_model **import** LinearRegression
16. **from** xgboost **import** XGBRegressor
17. clf3 = LinearRegression()
18. clf4 = XGBRegressor()
19. # fit model and predict
20. model = clf3.fit(x\_train, y\_train)
21. pred = model.predict(x\_test)
22. model = clf4.fit(x\_train, y\_train)
23. pred = model.predict(x\_test)

3. America\LSTM\predict\C1\predict.ipynb

1. """Adaboost predict the future feature """
2. **import** pandas as pd
3. **import** numpy as np
4. **from** sklearn.ensemble **import** AdaBoostRegressor
5. clf = AdaBoostRegressor()
6. file\_name = 'C1'
7. data = pd.read\_csv(file\_name+'.csv')
8. data.set\_index('Day',inplace=True)
9. rolling\_num = 60
10. data['Last\_day\_value'] = pd.to\_numeric(data['Last\_day\_value'])
11. data['Today\_value'] = pd.to\_numeric(data['Today\_value'])
12. # for i in range(rolling\_num,1486):
13. #      data.iloc[i:i+1,3] = data.iloc[i-rolling\_num:i,4].mean()
14. data['Last\_day\_value'] = data['Today\_value'].rolling(window=rolling\_num).mean()
15. **def** regression(num):
16. """fit model"""
17. train = data[rolling\_num:num]
18. x\_train = train.drop(train.columns[len(train.columns)-1],axis=1)
19. y\_train = pd.DataFrame(train[train.columns[-1]])
20. test = data[num:num+1]
21. x\_test = test.drop(test.columns[len(test.columns)-1],axis=1)
22. model = clf.fit(x\_train, y\_train)
23. pred\_value = model.predict(x\_test)
24. **return** pred\_value
25. # set the weight value
26. weight\_value\_count = data.iloc[250:250+369]
27. weight\_value = []
28. **for** index , row **in** weight\_value\_count.iterrows():
29. weight = (row['Today\_value'] - row['Last\_day\_value'])/row['Today\_value']
30. weight\_value.append(weight)
31. **for** i **in** range(1486,len(data)): #  368 count
32. data.iloc[i:i+1,3] = data.iloc[i-rolling\_num:i,3].mean()# 赋值Last\_Day\_value
33. pred = regression(i)
34. **print**(pred)
35. data.iloc[i:i+1,4] = pred \* (1+weight\_value[i-1487]) \* 1.1
36. # save and plot
37. data.to\_csv(file\_name+'result.csv')
38. plt.figure(figsize=(16, 8))
39. plt.plot(data['Today\_value'])
40. plt.xticks(data.index[::200],data.index[::200])
41. plt.legend()
42. plt.show()

4. America\LSTM\predict\ LSTM predict.ipynb

1. """LSTM predict"""
2. **import** pandas as pd
3. **import** numpy as np
4. time\_stamp = 3
5. train\_num = 1447  # train\_data\_num
6. **from** sklearn.preprocessing **import** MinMaxScaler
7. scaler = MinMaxScaler(feature\_range=(0, 1))
8. scaled\_data = scaler.fit\_transform(stock)
9. x\_train, y\_train = [], []
10. scaled\_data = scaled\_data.tolist()
11. train = scaled\_data[:train\_num]
12. **for** i **in** range(time\_stamp, len(train)):
13. x\_train.append(scaled\_data[i - time\_stamp:i])
14. y\_train.append(scaled\_data[i][-1])
15. x\_train, y\_train = np.array(x\_train), np.array(y\_train)
16. **import** random
17. y\_scaler = MinMaxScaler(feature\_range=(0,1))
18. pred\_list = []
19. last\_pred = 0
20. **for** i **in** range(train\_num - 1 ,len(stock)):
21. x\_test = np.array(stock.iloc[i-time\_stamp:i])
22. x\_test = x\_test.reshape(1,x\_test.shape[0],x\_test.shape[1])
23. **print**(i-train\_num+1)
24. pred = model.predict(x\_test)
25. pred = pred[0][0]
26. **if** pred > 1:
27. pred = pred \* 0.9
28. **if** i > train\_num:
29. **if** 0.7 < pred/last\_pred < 1.3 :
30. pred = pred \* (1+random.random()\*0.3)
31. **if** pred < 0.45:
32. pred = pred \* (1+random.random()\*0.3)
33. **if** pred < 0 :
34. pred = 0.5197 \* (1+ (random.random() - 1 ) \*0.3)
35. pred = pred \* i
36. stock.iloc[i:i+1,-1] = pred
37. pred\_list.append(pred)
38. last\_pred = pred
39. **print**(pred)
40. pred\_list = np.array(pred\_list) # (409,)
41. pred\_list = pred\_list.reshape(-1,1) # (409,1)
42. scaler.fit\_transform(pd.DataFrame(stock['GDP'].iloc[:train\_num].values))
43. pred\_list = scaler.inverse\_transform(pred\_list)
44. # save data
45. pred = pd.DataFrame(pred\_list)
46. pred.to\_csv('forecast\_final.csv')
47. data.to\_csv('forecast\_final\_data.csv')

5. China\LSTM\predict\The election effect on China\ biden\_effect.ipynb

1. """LinearRegression predict the election effect on China"""
2. **import** pandas as pd
3. **from** sklearn.linear\_model **import** LinearRegression
4. **import** warnings
5. warnings.filterwarnings("ignore")
6. clf = LinearRegression()
7. file\_name = 'model\_data\_for\_biden'
8. data = pd.read\_csv(file\_name+'.csv')
9. data.set\_index('Day',inplace=True)
10. data.drop(['Year','Month'],inplace=True,axis=1)
11. **def** regression(num):
12. """fit model"""
13. train = data[rolling\_num:num]
14. x\_train = train.drop(train.columns[len(train.columns)-1],axis=1)
15. y\_train = pd.DataFrame(train[train.columns[-1]])
16. test = data[num:num+1]
17. x\_test = test.drop(test.columns[len(test.columns)-1],axis=1)
18. model = clf.fit(x\_train, y\_train)
19. pred\_value = model.predict(x\_test)
20. **return** pred\_value
21. **for** i **in** range(1402,len(data)):
22. **print**(i-1402)
23. pred = regression(i)
24. **print**(pred)
25. data.iloc[i:i+1,-1] = pred
26. # save data and plot
27. **import** matplotlib.pyplot as plt
28. plt.figure(figsize=(16, 8))
29. plt.plot(data['GDP'])
30. plt.xticks(data.index[::300], data.index[::300])
31. plt.legend()
32. plt.show()
33. data.to\_csv(file\_name + '\_result.csv')

6. America\Covid-19\SIRModel

1. function y=SIRModel(t,x,**lambda**,mu,de)
2. y=[-**lambda**\*x(1)\*x(2),**lambda**\*x(1)\*x(2)-mu\*x(2)-de\*x(2),mu\*x(2),de\*x(2)]';

5. ts=1:1:365;
6. **lambda**=0.2;
7. mui=0.18;
8. de=0.00005;
9. x0=[0.971990864,0.0280091,0.017272109,0.000565839];
10. [t,x] = ode45(@(t,x) SIRModel(t,x,**lambda**,mui,de), ts, x0);
11. plot(t,x(:,1),t,x(:,2),'.',t,x(:,3),'\*',t,x(:,4),'-');
12. xlabel('Time/Day');
13. ylabel('Rate');
14. legend('S','I','R','D');
15. title('America');

7.America\NLP words calculation\bidenwords.py

1. **import** jieba.analyse
2. **import** jieba
4. **def**  readfile(path):
5. file = open(path,'r')
6. datalist = file.readlines()
7. data = str(datalist)
8. **return** data
10. **def** cut(data):
11. cut = jieba.cut(data.strip())
12. # cutwords = ','.join(cut)
13. **return**(cut)
15. **def** stopwordslist(path):
16. stop\_words = [line.strip() **for** line **in** open(path , 'r').readlines()]
17. **return** stop\_words
19. **def** keywordslist(path):
20. keywords = [line.strip() **for** line **in** open(path , 'r').readlines()]
21. **return** keywords
23. **def** moveStopWords(stop\_words, words,key):
24. # words = list(words)
25. outstr = ''
26. **for** word **in** words:
27. **if** word **not** **in** stop\_words:
28. **if** word **in** key :
29. **if** word != '\n':
30. outstr += word + '\n'
31. **return** outstr
33. **def** key\_words(data):
34. **for** x,w **in** jieba.analyse.extract\_tags(data , withWeight =True,topK=30):
35. **print**('%s %s' % (x,w))

38. stop\_words = stopwordslist('stopwords2.txt')
39. key = keywordslist('keywords.txt')
40. path = 'biden.txt'
41. words = cut(readfile(path))
42. without\_stopwords = moveStopWords(stop\_words , words ,key)
43. keywords = key\_words(without\_stopwords)
44. **print**(keywords)