Team Number :	apmcm2203263
Problem Chosen :	C

2022 APMCM summary sheet

The problem of anomalous increase in global average temperature has led to more and more horrible results, and it has become more and more important to analyze the trend of global average temperature increase and its causes. In this paper, we mainly use ARIMA model and LSTM model(Improved RNN model) for global temperature prediction, and use correlation coefficient analysis to analyze the correlation between average temperature and longitude, dimension and time, and analyze the effect of volcanic eruption on the change of average temperature by comparing the change of average temperature of its nearby cities before and after volcanic eruption.

For question 1, the latest data provided in the annex is only up to 2013, and we are not able to use this dataset to analyze the trend of temperature change over this year. Therefore, our team collected the global average temperature for each month of the last two hundred years to analyze the global trend, and built two models, ARIMA and LSTM, to fit and predict the global average temperature. The RMSE of the LSTM model is significantly lower than that of the former model, and the LSTM model is able to give very accurate predictions of future global temperature changes.

In response to question 2, in which we were able to see a strong correlation between global temperature and time as well as in question 1, we also analyzed the correlation coefficients between global temperature, longitude, and latitude and visualized them in a heat map. In order to analyze the effect of volcanic eruptions on the average temperature, a very good example is the average temperature change in Changchun, China, during 1903, when a volcanic eruption occurred in the Changbai Mountain volcanic area near the city. The temperature in the area dropped significantly in the two years following the eruption, and we have also analyzed how the eruption affected the average temperature. In fact, volcanic eruptions can bring in large amounts of CO2. Therefore, for question c, we consider global carbon emissions as the main cause of global temperature. Finally, a variety of specific measures are listed to environmental global temperature increase trends.

Finally, we also prepared a non-technical report to describe the model developed, to illustrate the severity of the global average temperature increase, and to suggest possible measures to mitigate the global temperature increase.

Keywords：ARIMA;LSTM Recurrent Neural Network; Correlation Coefficient Analysis Method

Content

1. Introduction 1
   1. 1 Background 1
   1. 2 Our Work..2
2. Problem analysis..2
   2. 1 Analysis of Problem 1 2
   2. 2 Analysis of Problem 2 2
3. Model assumption 3
4. Symbol description..3
5. Model building and solution for question 1 3
   5. 1 Data preprocessing..3
   5. 2 Data Clustering & Dimensionality Reduction 4
   5. 3 Building of ARIMA model 4
   5. 4 Building of LSTM model..7
   5. 5 Answers to questions..10
6. Correlation analysis of factors affecting temperature change 13
   6. 1 Global temperature in relation to time, longitude and dimension 13
   6. 2 Impact of natural disasters on global temperature 14
   6. 3 The main causes of global temperature change 15
   6. 4 Measures to mitigate global warming..15
Reference 15
Report..17
Appendix..18

1. Introduction

1. 1 Background

According to the World Meteorological Organization (WMO) Global Climate 2019 report published in 2020, the world is continuing to warm, with 2015-2019 being the five hottest years since the Industrial Revolution and 2010-2019 being the tenth hottest year since the Industrial Revolution, with the rate and magnitude of global warming and its duration The speed and magnitude of global warming and its duration are rare in history. At the same time, global atmospheric concentrations of greenhouse gas emissions continue to increase, with record-breaking global CO_2 emissions of 407.8 ± 0.1 mg/L in 2018, and continuing to rise in 2019.[1]

The heat energy brought by the rising temperature will provide the air and ocean with huge kinetic energy, thus forming large, or even super large typhoons, hurricanes, tsunamis and other disasters. In particular, large amounts of rainfall from typhoons and hurricanes and other disasters can lead to mudslides and landslides, which can seriously threaten traffic safety and the safety of people's lives.

Rising temperatures will not only draw water directly from the ocean, but also from the land, causing widespread droughts in the interior of the country, which will reduce food production, as well as feed production. Food and meat supplies will be scarce, directly threatening the stability of the country.

The icebergs that are melting as temperatures rise are the primary source of fresh water that we depend on for survival. A large portion of our underground freshwater reserves comes from iceberg meltwater. When the temperature balance is normal, the ice circulation system of icebergs, which means that they melt in summer, flows downhill and into the ground, accumulating fresh water for the plains and acting as a filter. The water returns to the mountains in winter as water vapor and re-accumulates snow and ice through heavy snowfall, also a filtering process. The whole cyclic process makes our fresh water guaranteed with a stable balance. Today, on the other hand, global warming has caused the accumulation of ice and snow on icebergs to be much slower than the rate of melting, and even some icebergs are no longer accumulating, which cuts off local fresh water for drinking. This will bring about conflicts and wars due to water shortage.

Rising atmospheric carbon dioxide levels will lead to rising carbon dioxide levels in the oceans, carbonating them, which will kill off a large number of microorganisms. Rising ocean temperatures will also destroy a large number of coral-centered biotic chains. The disappearance of food at the very bottom causes the ocean food chain to start at the bottom and break rapidly upwards and spread beyond the ocean. With no food left, there will be a massive die-off of marine life, and other organisms that feed on marine life. The death of a large number of organisms in the ocean will pollute the ocean and accelerate the death of other organisms; at the same time, a large amount of greenhouse gases will be released, accelerating global warming, forming a vicious circle.

As temperatures rise, invertebrates, especially insects, wake up from hibernation earlier, while long-distance migratory animals that depend on these insects for their livelihoods cannot catch up in time and miss the opportunity to feed, thus dying in large numbers. Insects wake up early, because there is no natural predator, will be recklessly eat large areas of forest and crops. Without forests, the carbon dioxide content will increase, accelerating global warming and forming a vicious circle; without crops, there will be no food for humans.

The massive decrease in the number of bees is also a precursor to the complete collapse of the natural food chain. Without the help of bees to spread pollen, plants will not be able to reproduce. In other words, crops will not be able to reproduce and bear fruit, and humans will have no food. Humans worldwide will face food shortages, and wars over food will become more and more frequent and closer. And there will be less and less food for humans to compete for.

1. 2 Our Work

In this paper, we mainly use ARIMA model and LSTM model for global temperature prediction, and use correlation coefficient analysis to analyze the correlation between average temperature and longitude, dimension and time, and analyze the effect of volcanic eruption on the change of average temperature by comparing the change of average temperature of its nearby cities before and after volcanic eruption. To propose feasible measures to mitigate environmental global warming based on the impact of carbon emissions on global average temperature.

2. Problem analysis

2. 1 Analysis of Problem 1

Question 1 requires a joint modeling analysis using the dataset accompanying the question as well as the dataset collected by our team.

- Problem a needs to analyze the temperature change every 10 years, we can collect the average temperature of all places in March every 10 years and find its average to represent the global average temperature in March of that year.
- Problems b and c require two time series prediction models, and we have built two models, ARIMA and LSTM, to perform the global average temperature prediction.
- Problem d requires a comparison of the accuracy of the two models based on the predicted and fitted results.

2. 2 Analysis of Problem 2

Question 2 requires our team to collect specific information about the occurrence of natural disasters.

- For problem a, our team used correlation coefficient analysis to give specific correlation coefficients for global average temperature, time, longitude, and dimension.
- For questions b and c, we use the information collected on natural disasters to analyze the change in local average temperature before and after the natural disaster to show how the factor affects the global average temperature and

further analyze how the factor affects the global average temperature.

- For question d, our team gives many feasible measures to mitigate global warming.

3. Model assumption

Our team made the following assumptions in building the model based on the analysis of the data set.

1) Any uncertainty in the average temperature less than 0.5 is a plausible value, and data with uncertainty greater than 0.5 are not sufficiently plausible or necessary for analysis. Therefore, our team believes that it is reasonable to directly remove the series of data with uncertainty greater than 0.5.
2) For the data given in the Appendix, our team assumes that the average temperature of all cities in a country in a given year is summed to obtain an average that represents the national average temperature of the country in that year.
3) Similarly, averaging the temperature of each month can represent the average temperature for the whole year.

4. Symbol description

Symbol	Description	Unit
T_y	Time variable	Year
T_m	Time variable	Month
AT	Average temperature	℃
ATI	Increase of Average temperature	℃
Lat	Latitude	°
Lon	Longitude	°
RMSE	Root Mean Square Error	/
F(t)	The average temperature prediction at time t	℃

5. Model building and solution for question 1

5. 1 Data preprocessing

Our team first performed linear interpolation of the null values, but found that there were so many temperature vacancies before 1900 that too much linear interpolation would significantly affect the final resulting annual average temperature, so our team chose to simply discard all the null values. These null values represent only 4.6% of the total data and do not have a significant impact on the overall modeling and analysis.

```
In  [6]: df.isnull().sum()

Out[6]: dt                                0
        AverageTemperature            11002
        AverageTemperatureUncertainty 11002
        City                              0
        Country                           0
        Latitude                          0
        Longitude                         0
        dtype: int64

In  [7]: df.isnull().sum()['AverageTemperature']/len(df['AverageTemperature'])*100

Out[7]: 4.599940629742826
```

Figure 1：Number and percentage of null values

According to the first assumption of the model, our team retrieved all the data with uncertainty greater than 0.5 and removed the whole rows.

5. 2 Data Clustering & Dimensionality Reduction

The labels allow clustering the data and never achieve dimensionality reduction.

For problem 1's a, by clustering all the temperature data with the time label of March and finding the average temperature, and then taking the data every ten years, we can obtain the data set of global average temperature in March with a time interval of ten years.

For problem 1's c, when a time series prediction model in years is needed, our team clusters the attached dataset in years and averages the temperatures of all months in a year to create a new dataset.

For problem 2, b, data on temperature changes before and after a natural disaster are needed. We perform clustering by city search and reduce the dimensionality of the dataset.

5. 3 Building of ARIMA model

ARIMA model is a linear model proposed by Box and Jenkins for time linear model for time series analysis and forecasting. In the time series analysis, three parameters are required In the time series analysis, three parameters are required, which are autoregressive order (p), difference order (d) and moving average order (q). (The general form of ARIMA(p, d, q) is as follows is as follows:[2]

$$\begin{cases} \phi(B)\nabla^d y_t = \theta(B)\varepsilon_t \\ E\left(\varepsilon_t\right)=0 \quad \operatorname{var}\left(\varepsilon_t\right)=\sigma_\varepsilon^2, E\left(\varepsilon_t\varepsilon_s\right)=0, s\neq t \\ E\left(x_t\varepsilon_s\right)=0 \quad s<t \\ \phi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p \\ \theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q \end{cases}$$

where: B is the back-shift operator; d = (1 - B) d is the higher-order difference; φi (i = 1, 2, ..., p) and θj (j = 1, 2, ..., q) are the autoregressive and shift-average parameters, respectively; εt is a εt is the error term that conforms to the N(0, σ2) normal distribution of the error term.

The time series yt should first be selected using the ADF test to select the difference order to time series into a smooth series, and then select the best model according to the Akaike Information Criterion (AIC) to select the best model.

The ARIMA model requires the series to satisfy smoothness, and the results of the ADF test are viewed to analyze whether it can significantly reject the hypothesis that the series is not smooth ($p<0.05$) based on the analyzed t-values.[2] ADF Inspection Form is shown in Form 1:

Form 1: ADF Inspection Form

ADF Inspection Form

Variables	Difference order	t	P	AIC	Threshold value		
					1%	5%	10%
AT	0	-3.726	0.004***	3178.267	-3.434	-2.863	-2.568
	1	-15.86	0.000***	3189.304	-3.434	-2.863	-2.568
	2	-19.313	0.000***	3360.945	-3.434	-2.863	-2.568

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively.

The model requires that the series must be smooth time series data. By analyzing the t-values, it is analyzed whether it can significantly reject the original hypothesis that the series is not smooth.

The results of this series test show that the series is a smooth time series based on the variable AT:0 with a significance p-value at the difference of order 1, which presents significance at the level of 0. The original hypothesis is rejected.

According to the above information, an ARIMA model was developed with the following model parameters.

Form 2：ARIMA model (2,1,2) test table

ARIMA model (2,1,2) test table

item	Symbols	Value
Sample size	Df Residuals	1810
	N	1816
Q-statistic	Q6(P)	3.531(0.060*)
	Q12(P)	28.613(0.000***)
	Q18(P)	139.989(0.000***)
	Q24(P)	151.122(0.000***)
	Q30(P)	169.654(0.000***)
Information Guidelines	AIC	3778.621

ARIMA model (2,1,2) test table

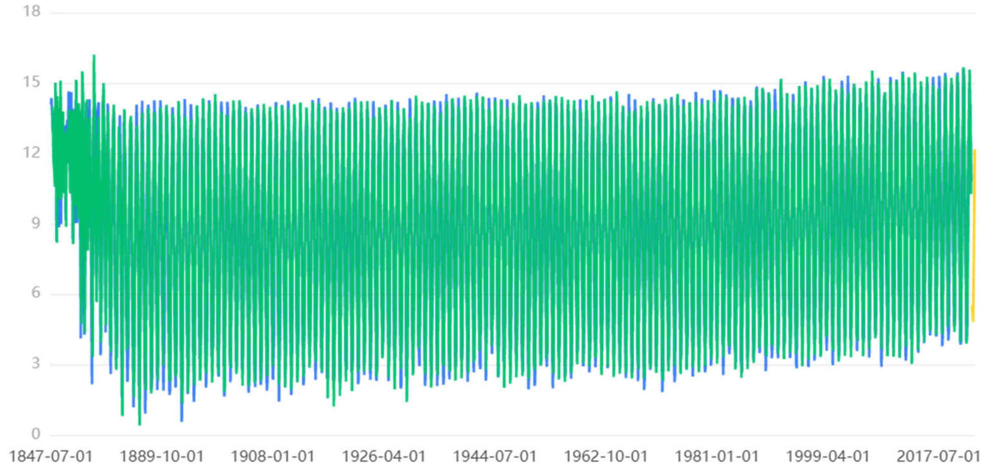
item	Symbols	Value
	BIC	3811.644
Goodness of fit	R²	0.973

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively

The equation generated by the ARIMA model is:

$$y(t)=-0.002+1.721*y(t-1)-0.99*y(t-2)-1.759*\varepsilon(t-1)+0.809*\varepsilon(t-2)$$

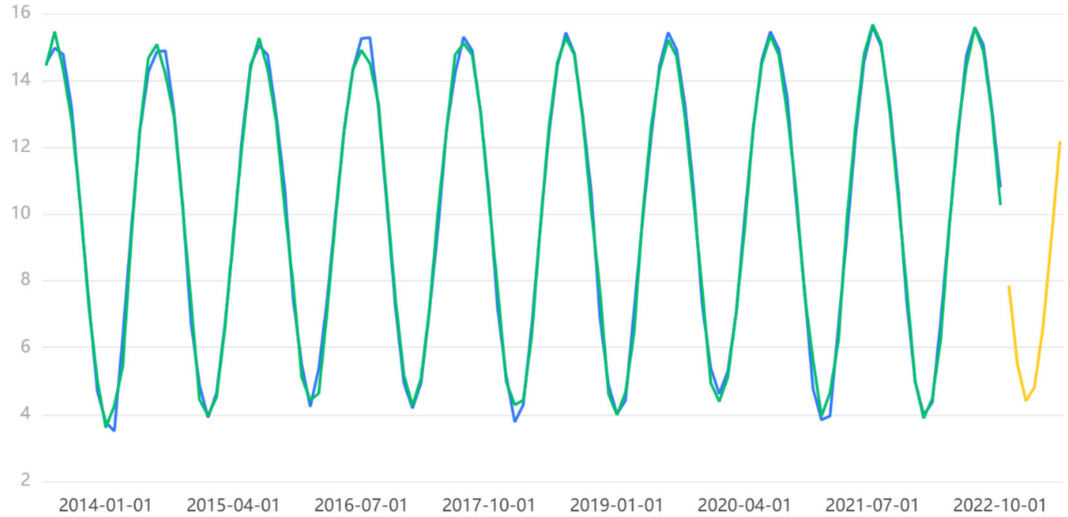
Using the ARIMA model, the global temperature for each month is fitted to obtain Figure 2



Note: The blue line is the original data line, the green line is the fitted line, and the orange line is the predicted line

Figure 2：Time Series Chart

Zoom in on the last part to see the predicted results more clearly, as shown inFigure 3.

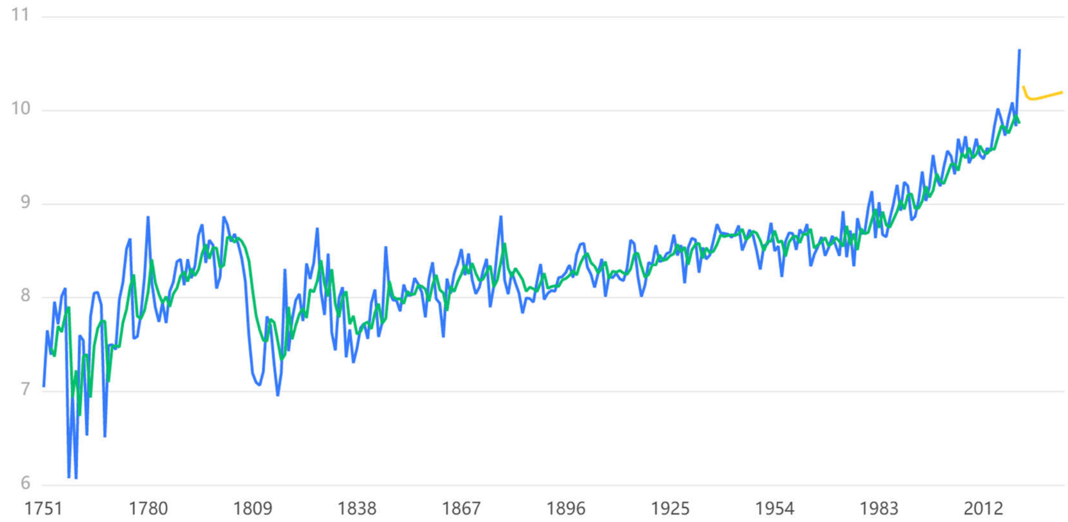


Note: The blue line is the original data line, the green line is the fitted line, and the orange line is the predicted line

Figure 3: Partial Enlargement of Figure 2

Meanwhile, using a similar approach, our team also developed an ARIMA model of the global average temperature on an annual basis and made future projections. Based on the variable temp, we automatically find the optimal parameters based on the AIC information criterion, and the model results in an ARIMA model (1,1,1) test table and based on 1-difference data with the following model equation.

$$y(t) = 0.01 + 0.305 * y(t-1) - 0.819 * \varepsilon(t-1)$$



Note: The blue line is the original data line, the green line is the fitted line, and the orange line is the predicted line

Figure 4: temperature prediction on an annual basis

5. 4 Building of LSTM model

A recurrent neural network (RNN) (shown in Figure 3) is a type of neural network that can to process time series data, and its current state will The current state contains the previous information of the whole time series.[2]

RNN can reflect the serial correlation characteristics of financial time series data, but there is the problem of gradient disappearance or gradient explosion, and its mining of historical information of financial time series data is very limited. In contrast, long short-term memory (LSTM) neural network is a kind of neural network that can well handle.[3]

special RNN with long-term dependence of time series data. the LSTM neural network structure (shown in Figure 3) contains a series of recurrently connected subnetworks (i.e., memory modules), each of which contains one or more self each memory module contains one or more self-connected cells, and three gates controlling the flow of information: input gate, output gate, and forget gate. Each memory module contains one or more self-connected cells (cells) and a system of three threshold cells: input gates, output gates, and forgetting gates that control the flow of information. In the LSTM network

the execution steps can be summarized as follows:[4]

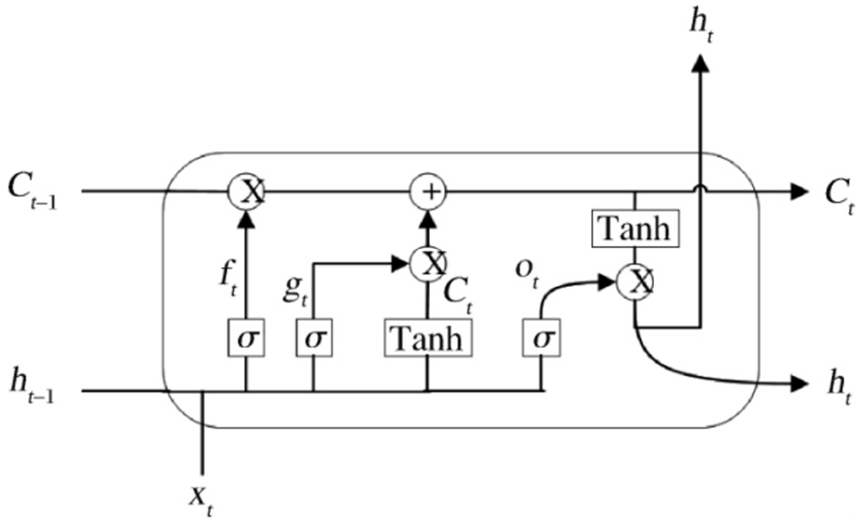


Figure 5: LSTM architecture [4]

First, the forgetgate f_t determines the information to be removed from the cell,

$$f_t = \sigma(b_f + W_f x_t + U_f h_{t-1})$$

where σ is the sigmoid activation function that sets the information flow weights to values between 0 and 1, where 0 means the information is completely removed and 1 means all information is retained. x is the current input vector, h_t is the current hidden layer vector, and b_f, W_f, U_f are the bias, input weights, and loop weights of the forgetting gate, respectively.[5]

Secondly, the information state in the cell is updated. Let g_t be controlled by the sigmoid activation function between 0 and 1 of the external input gate.

$$g_t = \sigma(b_g + W_g x_t + U_g h_{t-1})$$

Cell state C updated on the basis of C.

$$C_t = f *_t C_{t-1} + g_t * \tanh\left(b_c + W_c x_t + U_c h_{t-1}\right)$$

Finally, the information output is performed by the output gate o_t controlled.

$$h_t = o_t * \tanh(C_t)$$
$$o_t = \sigma(b_o + W_o x_t + U_o h_{t-1})$$

Therefore, the LSTM neural network contains not only the external circulation between cells in the hidden layer involved in RNN, but also the self-loop within cells, which is more complete for mining the historical information of global average temperature time series data and considering the sequence dependence, and it is feasible to build the LSTM deep neural network model for global average temperature time series data.[6]

The model parameters are shown in the following table:

Form 3: Parameters of LSTM Neural Network with year as time interval

Parameters	Values
Percentage of training set	0.95
Validation set percentage	0.05
Optimizer	adam
Loss	mean_squared_error
Batch_size	50
Epochs	5

```
model.summary()

Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_4 (LSTM)               (None, 60, 128)           66560

 lstm_5 (LSTM)               (None, 64)                49408

 dense_4 (Dense)             (None, 25)                1625

 dense_5 (Dense)             (None, 1)                 26

=================================================================
Total params: 117,619
Trainable params: 117,619
Non-trainable params: 0
_________________________________________________________________
```

Figure 6: LSTM network structure

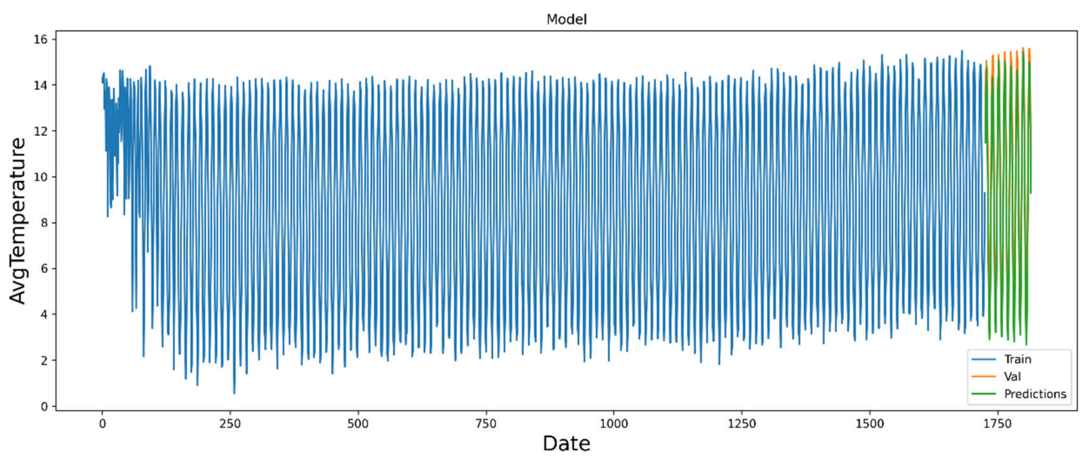


Figure 7: Results of fitting LSTM neural network with month as time interval

The model parameters are shown in the following table:

Form 4: Parameters of LSTM Neural Network with month as time interval

Parameters	Values
Percentage of training set	0.95
Validation set percentage	0.05
Optimizer	adam
Loss	mean_squared_error
Batch_size	15
Epochs	40

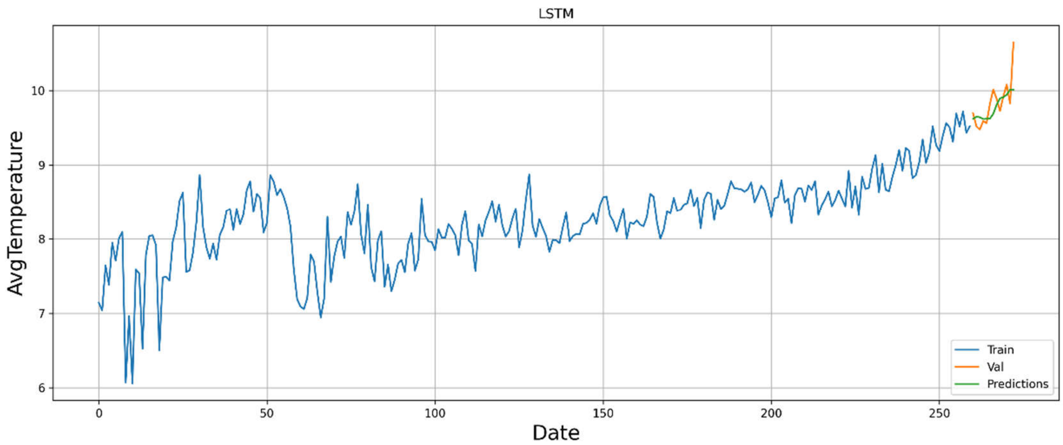


Figure 8: Results of fitting LSTM neural network with year as time interval

5. 5 Answers to questions

Qusetion 1,a:

Our team's answer is:

Partially agree with the title. the warming in March 2022 is indeed very significant, with March 2020 ranking second in terms of warming over the one hundred and forty years covered by the data, and about the same increase as the first. Considering that 2022 itself is much warmer than 1880, our team believes that the view of the title is correct.

Also, our team believes that such a high warming in 1880 is inextricably linked to the second industrial revolution.

The following is the process of our team's modeling analysis.According to the following equation:

$$ATI = T_{(i)y} - T_{(i-10)y}$$

Substitute the data into the formula to calculate and get the ATI, and turn the ATI into a line graph with time,shown in Figure 9.

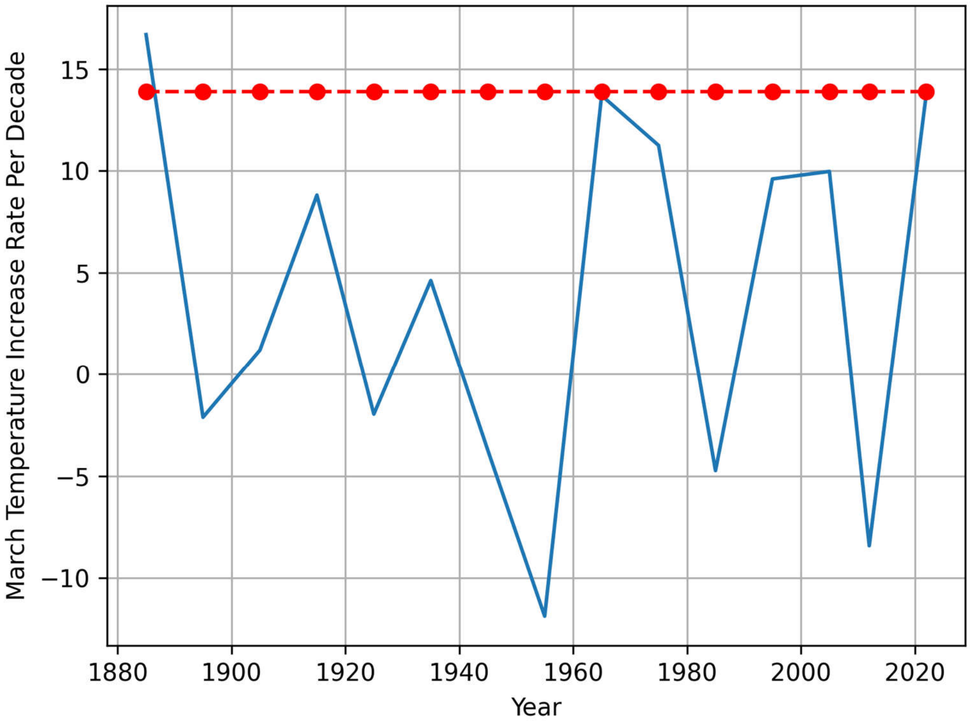


Figure 9：Plot of temperature increase in March for a time interval of 10 years

Our team drew two heat maps of the world map, based on the average temperature of each country counted in 1910 and 2010.

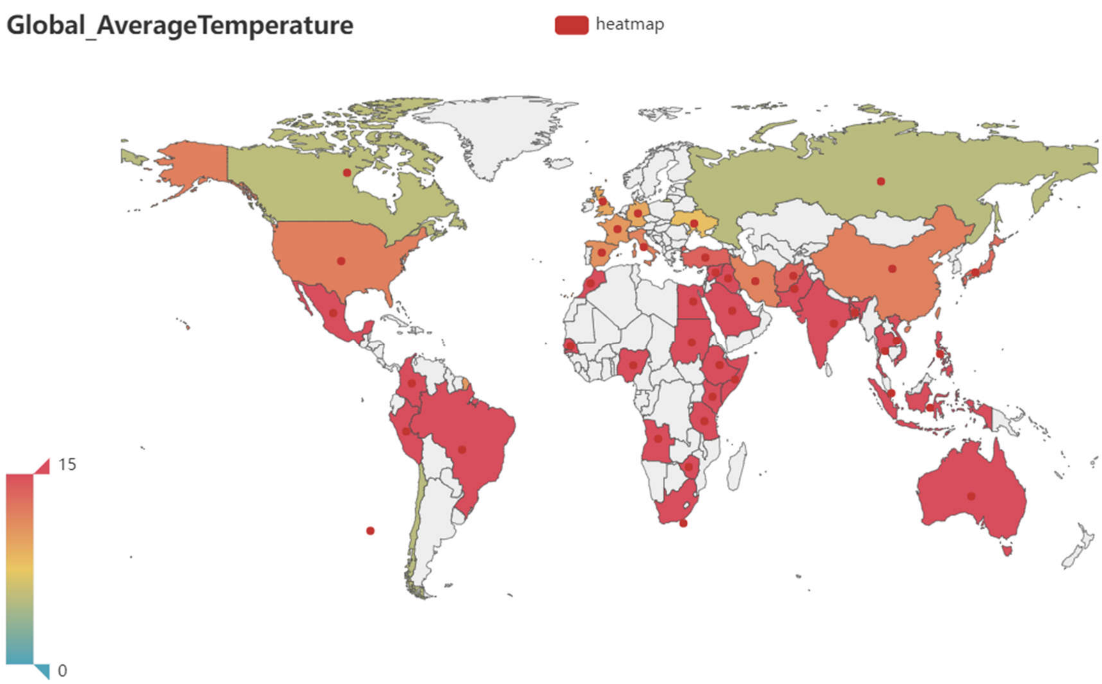


Figure 10:1910

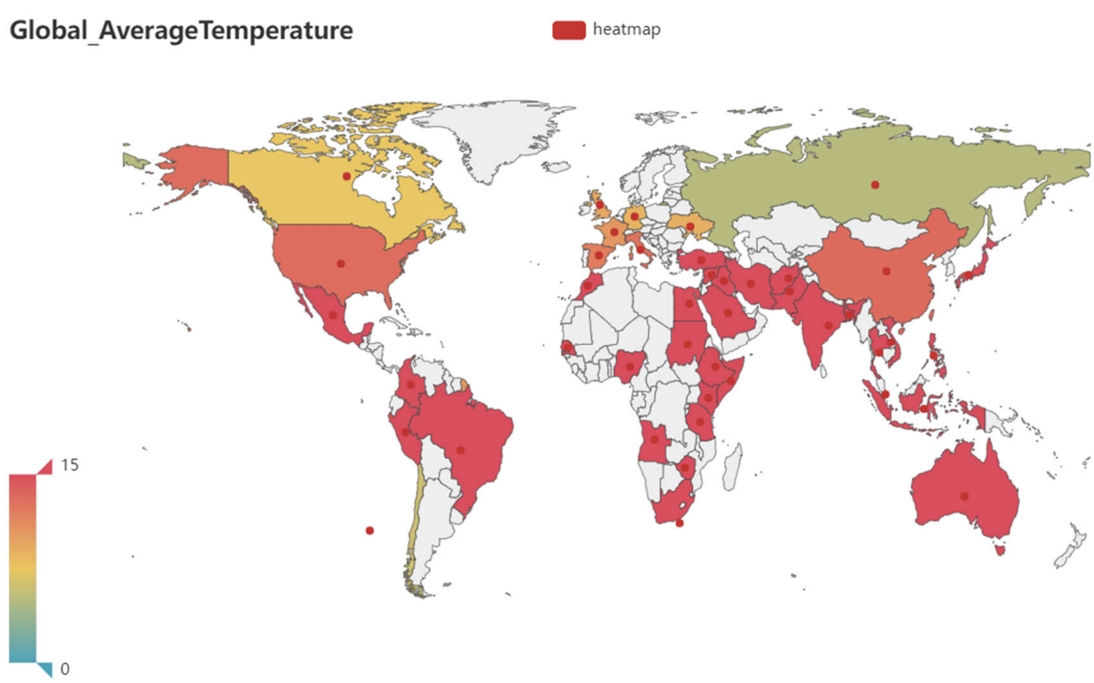


Figure 11:2010

We can clearly see that after 100 years, the red hot countries have started to spread significantly from the equator to the two levels, and the global warming trend is very significant.

Question 1,c:

Our model gives an expected global average annual temperature of 20℃ in 2200 years.

Question 1,d:

The ARIMA model predicts more linear results and does not fit well for the nonlinear part. The LSTM model, on the other hand, fits the linear and nonlinear parts well, with an RMSE value of about 0.3.

RMSE can be obtained as following:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{n}(T_t - f(t))^2}$$

```
# Get the root mean squared error (RMSE)
rmse = np.sqrt(np.mean(((predictions - y_test) ** 2)))
rmse

1/1 [==============================] - 0s 17ms/step

Out[10]: 0.2985776256469614
```

Figure 12：RMSE

Therefore, our team believes that the LSTM model is more accurate in its predictions.

6. Correlation analysis of factors affecting temperature change

6. 1 Global temperature in relation to time, longitude and dimension

The Pearson correlation coefficient, also known as the product-difference correlation coefficient, is a statistical indicator that expresses the degree and direction of linear correlation between two variables. The correlation coefficient of a sample is denoted by the symbol r.[7] The calculation formula is:

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}} = \frac{l_{XY}}{\sqrt{l_{XX} l_{YY}}}$$

Among it:

$$l_{XX} = (X-\bar{X})^2 = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$$

$$l_{YY} = (Y-\bar{Y})^2 = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$$

$$l_{XY} = \sum (X-\bar{X})(Y-\bar{Y}) = \sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}$$

Using the above mathematical model the correlation between the variables can be evaluated and the correlation coefficients can be obtained.

In the above analysis, we have been able to see the significant effect of time on the global average temperature, while it is clear that longitude and dimensionality do not have any relationship with time. Therefore, our team transformed the original problem into a mathematical model to find the relationship between global average temperature and longitude and dimensionality.

The global temperatures of 1992, 2002 and 2012 were selected for three separate analyses by searching the databases given in the Appendix.

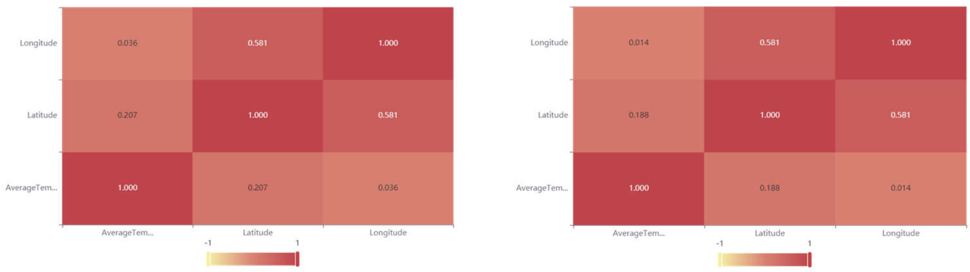


Figure 13:Heat Map of 1992(left) and 2002(right)

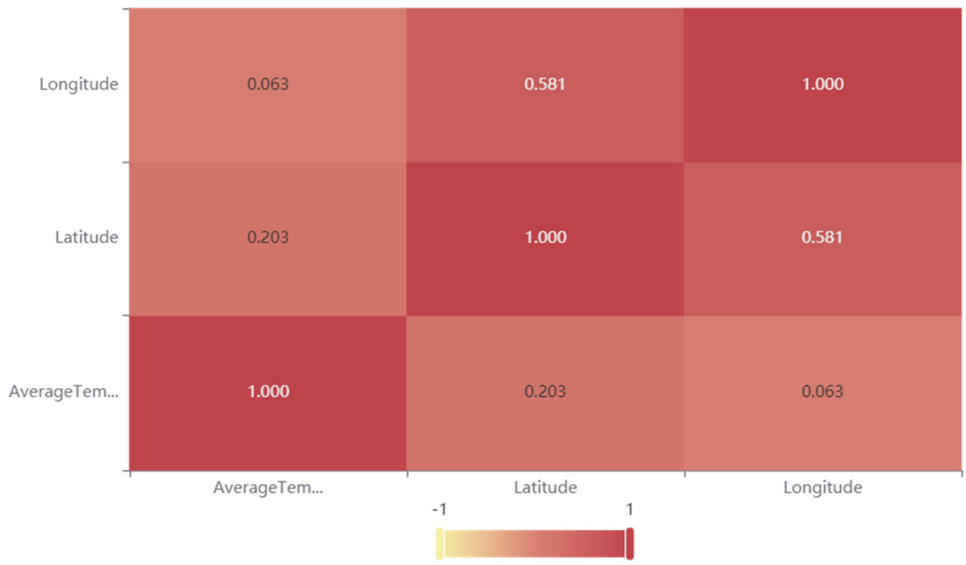


Figure 14: Heat Map of 2012

It is easy to see from these three graphs that the average temperature forms a significant positive correlation with latitude, but the average temperature does not correlate with longitude.

6. 2 Impact of natural disasters on global temperature

For problem 2, b, data on temperature changes before and after a natural disaster are needed. We perform clustering by city search and reduce the dimensionality of the dataset.

In the analysis of this question, we take the Chinese city of Changchun as an example, because the Tianchi volcano near Changchun erupted in 1903.

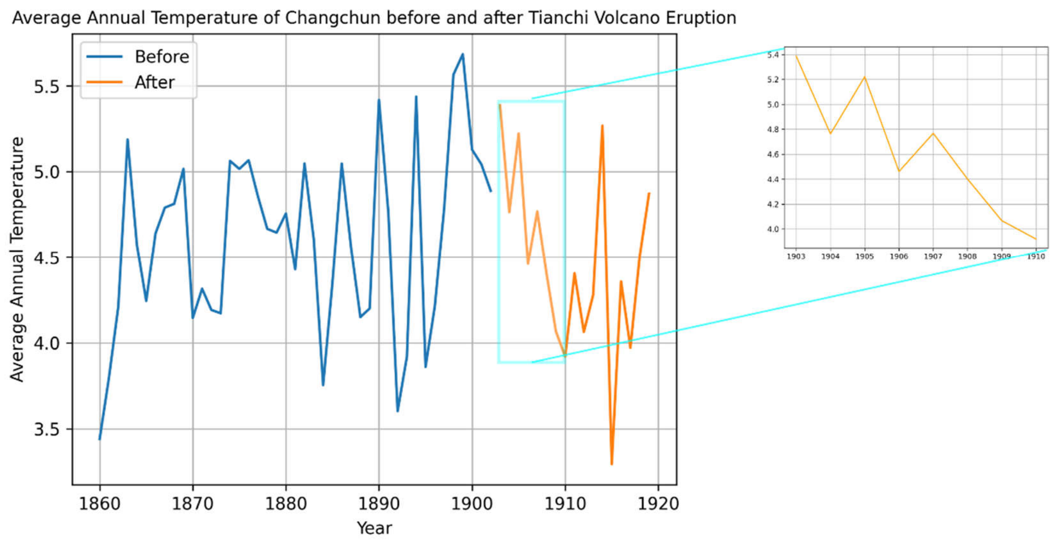


Figure 15：Average temperature in Changchun before and after the volcanic eruption

We can see that after the eruption of the volcano in 1903, we can see in the local magnification that the temperature dropped rapidly within two years and swung back down in the following years.

Volcanic eruptions make the climate in some regions colder due to the large amount of volcanic dust constantly sent into the atmosphere, which can be up to 0.5 - 3 km thick and can wander in the troposphere and stratosphere for 1 or 2 years, thus reducing the solar radiation heat in some regions by 10% - -30%. Coupled with the fact that volcanic dust is the nucleus for the formation of clouds and water droplets in the atmosphere, the sky is full of volcanic dust, and it is very easy to form clouds and rain. More clouds and water droplets in the sky obviously weaken the solar radiation heat. Therefore, during the 1 or 2 years of volcanic eruption, the climate in some regions of the Earth will appear colder, especially in the summer.

Similar facts are, in 1980, May 18 and 25, June 12, July 22, located in Washington State, the United States of America, the Mount St. Helens volcano four eruptions, each eruption, consisting of soot from the huge mushroom cloud rapidly rising, like a hydrogen bomb explosion, all of a sudden, the sky is dark, the day suddenly turned into night, can not see five fingers. As the volcanic dust with the airflow moving in the sky, this year China's Yangtze River basin there is a significant cold summer, August is particularly abnormal, Shanghai, for example, the average monthly temperature of only 24.8 ℃, 3.0 ℃ lower than the normal annual scenario. True enough, the volcanic eruption did make the climate colder in some areas.

6. 3 The main causes of global temperature change

Our team has collected a dataset of carbon emissions from various countries in recent years, which is attached to the supporting material. Unfortunately, for some reasons, our team was not able to complete the detailed data analysis in this section, and here we can only give the main causes of global warming inferred from the conclusions on the effect of volcanic eruptions on temperature.[8], [9]

Our team believes that global warming is mainly caused by the greenhouse effect of excessive CO_2 emissions, and we believe that this ecological damage is likely to be irreversible, the global warming trend has been established, and all we can do is to do our best to mitigate the global warming rate.

6. 4 Measures to mitigate global warming

1) Shut down coal-fired power plants and other high-emissions equipment.
2) Support new energy enterprises, such as wind, solar, and electric vehicles, to create more green economy jobs.
3) It is necessary to reduce the destruction of forests, grasslands and wetlands, and reduce marine pollution

Reference

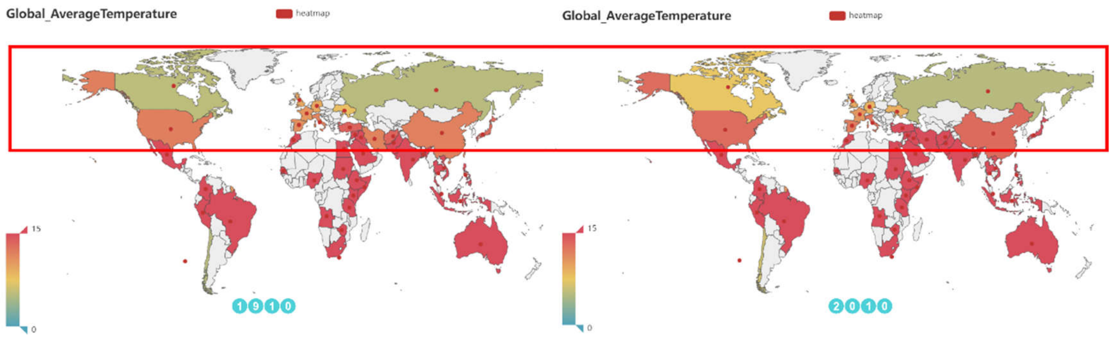
[1] Zhao Zongci, Luo Yong, and Huang Jianbin, "Evidence of human influence on global warming," Science and Technology Herald, vol. 39, no. 19, pp. 18–23, 2021.

[2] Wang, Y.W. and Ma, Shucai, "Time series forecasting based on hybrid models of ARIMA and LSTM," Computer Applications and Software, vol. 38, no. 02, pp. 291–298, 2021.

[3] Ouyang Hongbing, Huang Qiang, and Yan Hongju, "Financial time series forecasting based on LSTM neural network," China Management Science, vol. 28,

no. 04, pp. 27–35, 2020, doi: 10.16381/j.cnki.issn1003-207x.2020.04.003.
[4] A. Graves, “Long short-term memory,” *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.
[5] K. Smagulova and A. P. James, “A survey on LSTM memristive neural network architectures and applications,” *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019.
[6] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: LSTM cells and network architectures,” *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
[7] Xu Weichao, "A review of correlation coefficient studies," Journal of Guangdong University of Technology, vol. 29, no. 03, pp. 12–17, 2012.
[8] J. Houghton, “Global warming,” *Reports on progress in physics*, vol. 68, no. 6, p. 1343, 2005.
[9] R. A. Kerr, “Global warming is changing the world,” *Science*, vol. 316, no. 5822, pp. 188–190, 2007.

Report

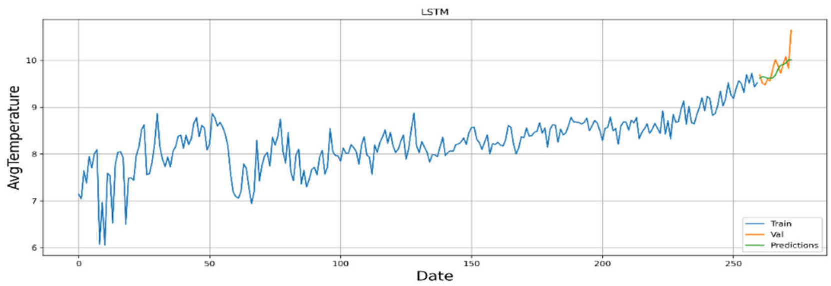


We can clearly see that after 100 years, the red hot countries have started to spread significantly from the equator to the two levels, and the global warming trend is very significant.

If we pay attention to the areas selected by the red border, we can clearly see that the red color of the countries in the temperate zone is more obvious, while the countries near the cold zone are no longer the green color they used to be, but mostly brown and tan. The trend of temperature increase has started to spread from the equator to the temperate and cold zones, and the global temperature is increasing with the naked eye.

The graph shows only the average of the year's temperature, in fact, in the summer the highest global temperature has been very scary, and caused a number of casualties.


Showing global temperature rise in two world heat maps may not be enough to make you realize at what rate our planet is warming up, so our team has plotted the global temperature trend and made a short time forecast.



According to our team's analysis, the main cause of global temperature increase is the large amount of carbon dioxide emissions, according to which we propose the following sexually effective measures.

- ➢ Shut down coal-fired power plants and other high-emissions equipment.
- ➢ Support new energy enterprises, such as wind, solar, and electric vehicles, to create more green economy jobs.
- ➢ It is necessary to reduce the destruction of forests, grasslands and wetlands.

Appendix

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import time
pd.set_option('mode.chained_assignment', None)
data=pd.read_csv('C:/Users/Tony/Desktop/Changchun.csv',parse_dates = ['year'])
data['year']=data['year'].dt.year
dif=83
end=100
plt.plot(data['year'][40:dif],data['AverageTemperature'][40:dif])
plt.plot(data['year'][dif:end],data['AverageTemperature'][dif:end])
plt.grid(True)
plt.title('Average Annual Temperature of Changchun before and after Tianchi Volcano Eruption',fontsize=10)
plt.xlabel('Year')
plt.ylabel('Average Annual Temperature')
plt.legend(['Before','After'],loc='upper left')
plt.savefig('Changchun.png', dpi=300) #指定分辨率保存
plt.figure()
plt.grid(True)
plt.plot(data['year'][dif:dif+8],data['AverageTemperature'][dif:dif+8],color='orange')
plt.savefig('Changchun_1.png', dpi=300) #指定分辨率保存
```

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data=pd.read_csv('C:/Users/Tony/Desktop/orginal.csv',parse_dates = ['T'])[['T','AT']]
training_data_len=int(np.ceil( len(data['AT']) * .95 )) #get the 95% of rows

training_data_len
AT=np.array(data['AT'])
temp_values=AT.reshape(AT.shape[0],1)
from sklearn.preprocessing import MinMaxScaler #preprocessing
scaler = MinMaxScaler(feature_range=(0,1))
scaled_data = scaler.fit_transform(temp_values)
scaled_data
# Create the training data set
# Create the scaled training data set
train_data = scaled_data[0:int(training_data_len), :]
```

```
# Split the data into x_train and y_train data sets
x_train = []
y_train = []

for i in range(60, len(train_data)):
    x_train.append(train_data[i-60:i, 0])
    y_train.append(train_data[i, 0])
    if i<= 61:
        print(x_train)
        print(y_train)
        print()

# Convert the x_train and y_train to numpy arrays
x_train, y_train = np.array(x_train), np.array(y_train)

# Reshape the data
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
# x_train.shape
from keras.models import Sequential
from keras.layers import Dense, LSTM

# Build the LSTM model
model = Sequential()
model.add(LSTM(128, return_sequences=True, input_shape= (x_train.shape[1], 1)))
model.add(LSTM(64, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))

# Compile the model
model.compile(optimizer='adam', loss='mean_squared_error')

# Train the model
model.fit(x_train, y_train, batch_size=50, epochs=5)
# Create the testing data set
# Create a new array containing scaled values from index 1543 to 2002
test_data = scaled_data[training_data_len - 60: , :]
# Create the data sets x_test and y_test
x_test = []
y_test = temp_values[training_data_len:, :]
for i in range(60, len(test_data)):
    x_test.append(test_data[i-60:i, 0])

# Convert the data to a numpy array
```

```
x_test = np.array(x_test)

# Reshape the data
x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1 ))

# Get the models predicted price values
predictions = model.predict(x_test)
predictions = scaler.inverse_transform(predictions)

# Get the root mean squared error (RMSE)
rmse = np.sqrt(np.mean(((predictions - y_test) ** 2)))
rmse
temperature=data['AT']
train = temperature[:training_data_len]
valid=pd.DataFrame(columns=['AvgTemperature', 'Predictions'])
valid['AvgTemperature'] = temperature[training_data_len:]
valid['Predictions'] = predictions
# Visualize the data
plt.figure(figsize=(16,6))
plt.title('Model')
plt.xlabel('Date', fontsize=18)
plt.ylabel('AvgTemperature', fontsize=18)
plt.plot(train)
plt.plot(valid[['AvgTemperature', 'Predictions']])
plt.legend(['Train', 'Val', 'Predictions'], loc='lower right')
# plt.show()
plt.savefig('plot1_2_2_month.png', dpi=300) #指定分辨率保存
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import time
pd.set_option('mode.chained_assignment', None)
dif='2.59  3.20  5.29  8.29  11.28  13.43  14.31  13.84  12.04  9.20  6.07  3.63'.split('  ')
dif=[float(item)for item in dif]
dif
def draw(x):
    plt.plot(x[0],x[1])
    plt.xlabel('Year')  # x 轴标题
    plt.ylabel('March Temperature Increase Rate Per Decade')   # y 轴标题
    plt.grid(axis="x")  # 坐标网格
    plt.grid(axis="y")  # 坐标网格
    plt.savefig('plot1.png', dpi=300) #指定分辨率保存
```

```
data = pd.read_table('C:/Users/Tony/Desktop/11.txt',sep=',',header=None)
data=data[[1,2,3,4]]
data[3]=[float(item)for item in data[3]]
data[4]=[float(item)for item in data[4]]
for i in range(int(len(data[3])/12)):
    for j in range(12):
        data[3][12*i+j]+=dif[j]

for j in range(10):
    data[3][3264+j]+=dif[j]

data_1=[]
for i in range(int(len(data[4]))):
    if data[4][i]<0.5:
        data_1.append([data[1][i],data[2][i],data[3][i],data[4][i]])
data_2=pd.DataFrame(data_1)
data=data_2
data
March=[]
for i in range(len(data[2])):
    if data[1][i]==3:
        March.append([data[0][i],data[2][i]])
March=pd.DataFrame(March,columns=['year','AT'])
March
ten_m=[]
for i in range(int(len(March['AT'])/10)):
    ten_m.append([March['year'][10*i],March['AT'][10*i]])
ten_m.append([2012,March['AT'][137]])
ten_m.append([2022,March['AT'][147]])
ten_m=pd.DataFrame(ten_m)
# plt.plot(ten_m[0],ten_m[1])
# plt.xlabel('Year')  # x 轴标题
# plt.ylabel('Global mean temperature for March')   # y 轴标题
# plt.grid(axis="x")  # 坐标网格
# plt.grid(axis="y")  # 坐标网格
# plt.savefig('plot1.png', dpi=300) #指定分辨率保存
diff=[]
for i in range(1,len(ten_m[1])):
#    diff.append([ten_m[0][i],(ten_m[1][i]-ten_m[1][i-1])/ten_m[1][i-1]*100])#增长率
    diff.append([ten_m[0][i],(ten_m[1][i]-ten_m[1][i-1])/ten_m[1][i-1]*100])#增长绝对值
diff=pd.DataFrame(diff)
draw(diff)
```

```
plt.plot(diff[0],[13.9]*15,color='r',marker='o',linestyle='dashed')
plt.savefig('plot1_1.png', dpi=300) #指定分辨率保存
diff
#第三次工业革命
# 导出 csv
# data['T']=pd.to_datetime(data[0].astype('str')+'-'+data[1].astype('str'))
# data['AT']=data[2]
# data.to_csv('orginal_month.csv',columns=['T','AT'])
# 导出 csv
data_1=pd.DataFrame(data.groupby(0)[2].mean())
data_1
data_1.to_csv('orginal_year.csv')
```