Team Number:	APMCM2213288
Problem Chosen:	E

2023 APMCM summary sheet

With the development of science and technology, the economic strength and military strength of all countries in the world have been fully increased. Nuclear weapons play an important role in safeguarding national security and regional stability. Nuclear weapons are extremely destructive and do great harm to human beings, the biosphere and the environment. Therefore, today's international organizations strictly control nuclear weapons. Because of the huge power of nuclear weapons, it is necessary to investigate and analyze which countries may possess nuclear weapons, how many nuclear weapons each country may have in the future, and the explosive power of nuclear weapons.

Therefore, this paper will analyze the above problems according to the given data, apply a variety of mathematical models and in-depth learning, and establish a corresponding model for prediction and analysis using a new statistical calculation method of neural networks, and give a variety of model evaluation parameters for model reliability testing.

For question 1, we made statistics based on the data given, compiled a Python code for data analysis, and drew the corresponding conclusions.

For question 2, we have established a variety of deep learning models, such as LSTM, MLP and basic machine learning models, random forest, to do regression analysis on the given data. Based on the fitting analysis results, we have given a variety of evaluation parameters, and made predictions based on the trained regression model. We have obtained relatively reasonable answers, and can make reasonable explanations using the international situation and the corresponding international treaties.

For question 3,We use the dimensional analysis method and Buckingham's theorem to establish the estimation model of nuclear bomb explosion power.

For question 4,we give suggestions to all countries in the world according to the analyzed questions 1, 2 and 3.

Keywords: LSTM MLP Random Forest Dimensional Analysis

Contents

I. Overview

1.1 Background

With the continuous deepening of human research on matter, the research on macro properties has gradually transformed into the research on micro composition. When human beings continue to study the micro world, they discovered nuclear energy. Nuclear energy can be traced back to the discovery of electrons by British physicist Thomson at the end of the 19th century. Since then, mankind has gradually uncovered the mystery of atomic nuclei.

With the continuous deepening of human research on nuclear energy, people found that a small atomic nucleus contains a force that can not be ignored. Einstein first proposed the mass energy equation:$E = mc^2$.Let people realize that nuclear energy has great potential in all aspects. Nuclear energy is not easy to control, so people began to use it to make destructive weapons - nuclear bombs.

In World War II, in order to speed up Japan's surrender, the United States dropped two atomic bombs on Hiroshima and Nagasaki, destroying all the buildings in the city and killing hundreds of thousands of people living in the city. This is also the only time in human history that a nuclear bomb has been put into practice. It was this war that made people realize the horror of nuclear bombs.

In order to maintain international peace and avoid the Third World War, it is extremely necessary to restrict nuclear tests. In addition, using the model to predict and analyze the number of nuclear bombs in the future can also help us better maintain regional stability and national security.

Figure 1 The ruins after the nuclear explosion

1 . 2 Restatement of the Problem

Problem 1: We need to perform some basic data analysis based on the data set given in the question, which countries have ever possessed nuclear weapons, which countries have seen their nuclear weapons stockpiles decrease or increase the most in the last 20 years, in which five years nuclear weapons testing has occurred the most, in which countries nuclear weapons research has been most active in the last 10 years and which countries have moved from ”not thinking about nuclear weapons” to ”having nuclear weapons” the fastest.

Problem 2: We need to model according to a given data set or an open data set collected by ourselves, and predict which countries may have nuclear weapons within 100 years, the change trend of the total international nuclear weapons within 100 years, the total nuclear weapons after 100 years, and the number of nuclear weapons of each country according to the given model.

Problem 3: We need to derive the formula for calculating the lethality of nuclear bombs, estimate the power of existing nuclear bombs, learn the energy or other data required by various methods of destroying the earth through consulting information, calculate the number of nuclear bombs required to destroy the earth, and give suggestions on the future storage of nuclear weapons in various countries.

Problem 4: Write a non-technical article for the United Nations explaining the team's findings and making a few recommendations for all countries.

II. Symbol Descriptions

In our paper,we may use some symbols.Some of the main symbols and their meanings are included in the table below.

Table 1 Symbol Descriptions

Symbols	Meaning	Unit
R	Killing radius of nuclear bomb	m
t	Time after explosion	s
E	Energy generated by explosion	J
P	Atmospheric pressure	pa
ρ	Air density	kg/m^3
T	Tnt equivalent	t
Q	Quantity of heat	J

III. Model building and solution of question 1

In recent years, the degree of development of big data has become more and more obvious, and many fields have developed in a better direction due to the use of big data analysis. python appears to be more active in data analysis and interaction, exploratory computing and data visualization, especially the Numpy library included in python, which can be said to have unparalleled advantages in handling medium-sized data and has become The use of Python can greatly improve the efficiency of data analysis, which is the reason why we choose Python to analyze the given data. Also for some of the data we used ECharts for data visualization.

3.1 Countries that have possessed nuclear weapons

In the dataset given in the title, our method is to count the countries with non-zero stockpiles of nuclear weapons in any year from 1945 to 2022 directly in the stockpiles table, representing those countries that have nuclear weapons. The countries we counted are: China, France, India, Israel, North Korea, Pakistan, Russia, South Africa, United Kingdom, United States.

3.2 Analysis of changes in nuclear weapons stockpiles

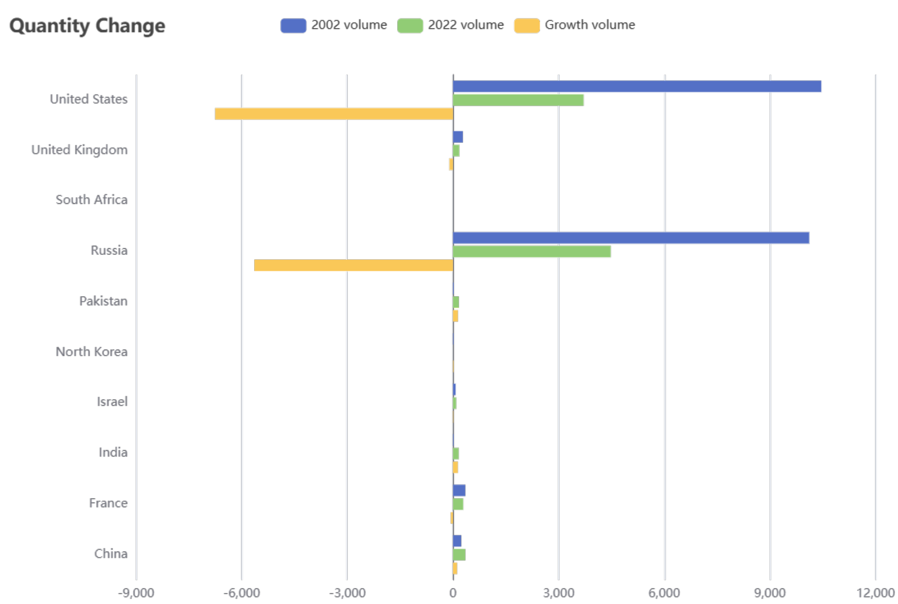

Figure 2 Changes in the number of nuclear weapons by country from 2002 to 2022

Table 2 Changes in the number of nuclear weapons by country from 2002 to 2022

Country	Inventory in 2022	Inventory in 2002	Amount of change
China	350	235	+115
France	290	350	-60
India	160	23	+137
Israel	90	76	+14
North Korea	20	0	+20
Pakistan	165	26	+139
Russia	4477	10114	-5637
South Africa	0	0	0
United Kingdom	180	280	-100
United States	3708	10457	-6749

In this question we answer by making a difference, i.e., subtracting the nuclear weapons stockpile in 2002 from the nuclear weapons stockpile in 2022 among all nuclear-armed states, and then tallying the increase or decrease in each country's nuclear weapons over a 20-year period, leading to the table above.

From the above table 2 and figure 2 we conclude that Pakistan has the largest increase in its nuclear weapons stockpile and the United States has the largest decrease.

3.3 Statistics on the number of nuclear tests

To calculate the number of nuclear weapons tests in the world within 5 years, our idea is to first calculate the sum of world nuclear tests for each year from 1945 to 2019. Then we count the sum of the world nuclear tests within 5 years in 5-year rounds starting from 1945, and finally find the maximum value of the statistical results.

The results are plotted in a line figure 3 and a bar figure 4 for a better visualization of the results. As can be seen from the graph, the most nuclear weapons tests occurred in 1962-1966.

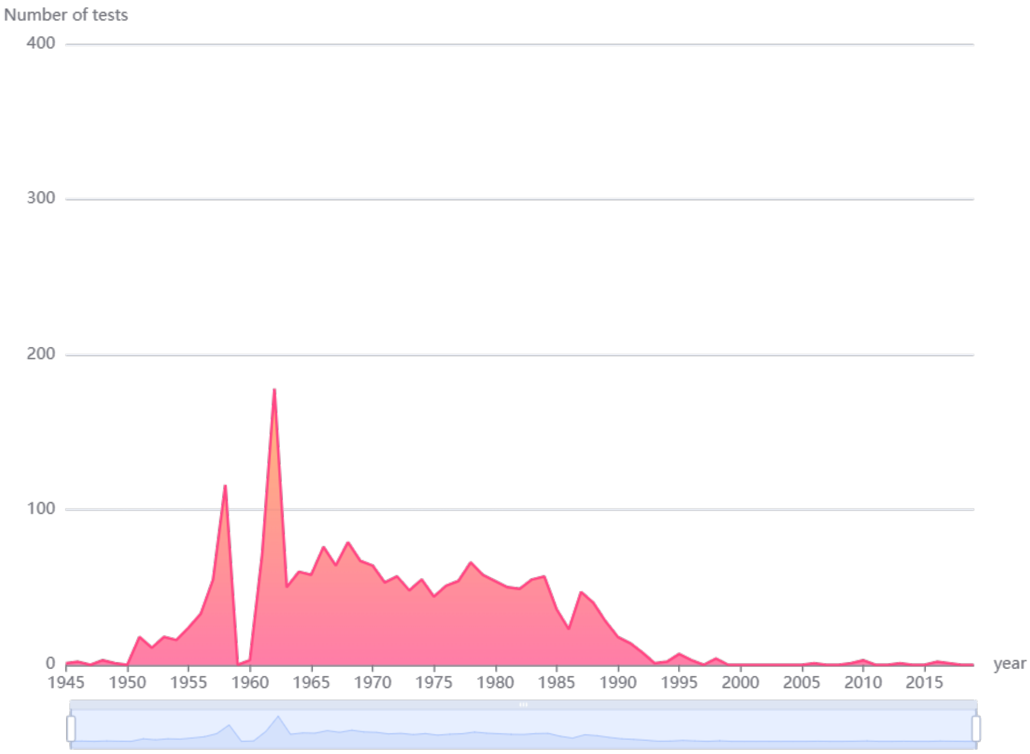

Figure 3 Total world nuclear tests from 1945 to 2019

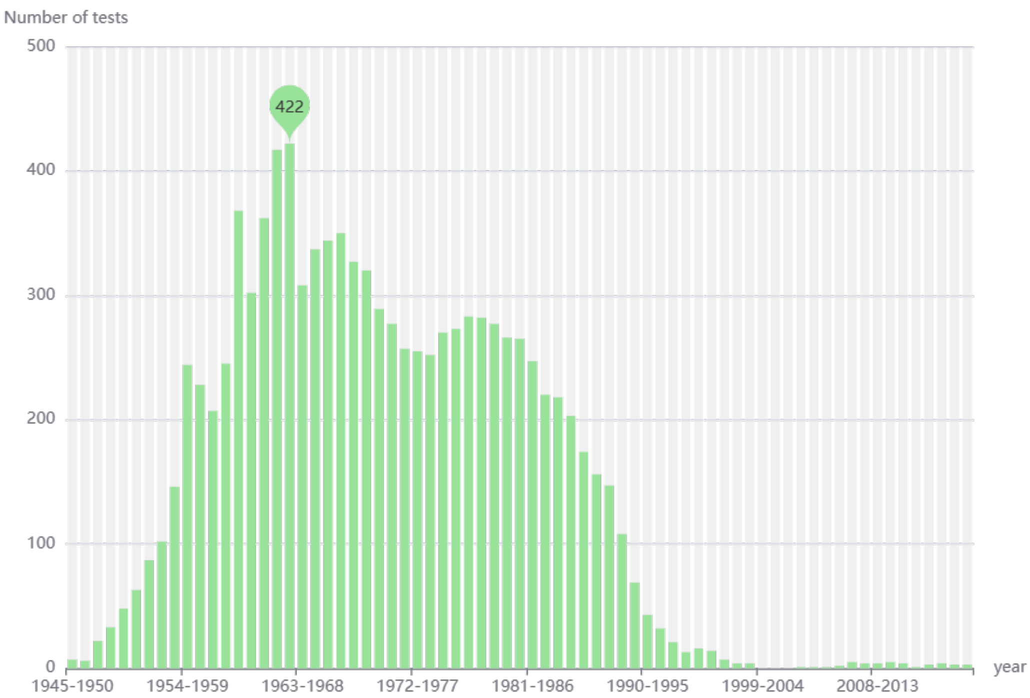

Figure 4 Total world nuclear tests every five years

3.4 Nuclear Weapons Research Activity Analysis

To find how active each country's nuclear weapons research has been in the past 10 years, we mainly analyzed the tests dataset, in which we summed up the number of nuclear tests of each country within 10 years (2009-2019) and used the sum of each country's 10-year nuclear tests as an evaluation criterion for nuclear weapons research activity. We calculated that all other countries had 0 tests in the 10-year period, and North Korea had 8 tests. Therefore, we conclude that North Korea has been the most active in nuclear weapons research in the past 10 years.

3.5 Changing attitudes toward nuclear weapons

Our solution is to find all countries that had nuclear weapons in the Positions dataset, and among them, find the last year they "did not consider nuclear weapons" and the first year they "had nuclear weapons", and subtract the latter year from the former. The country with the smallest difference is the country with the fastest transition. Again, we visualized the data for this question for a more intuitive understanding in figure 5 and table 3.We concluded that the country that made the fastest transition from "not considering nuclear weapons" to "having nuclear weapons" was the United States.

Table 3 The Shift from Don't Think About Nuclear Weapons to Have Nuclear Weapons

Country	Disregard	Possession	Time spent(years)
China	1951	1964	13
France	1944	1960	16
India	1947	1987	40
Israel	1948	1967	19
North Korea	1961	2006	45
Pakistan	1971	1987	16
Russia	1941	1949	8
South Africa	1968	1979	11
United Kingdom	1939	1952	13
United States	1938	1945	7

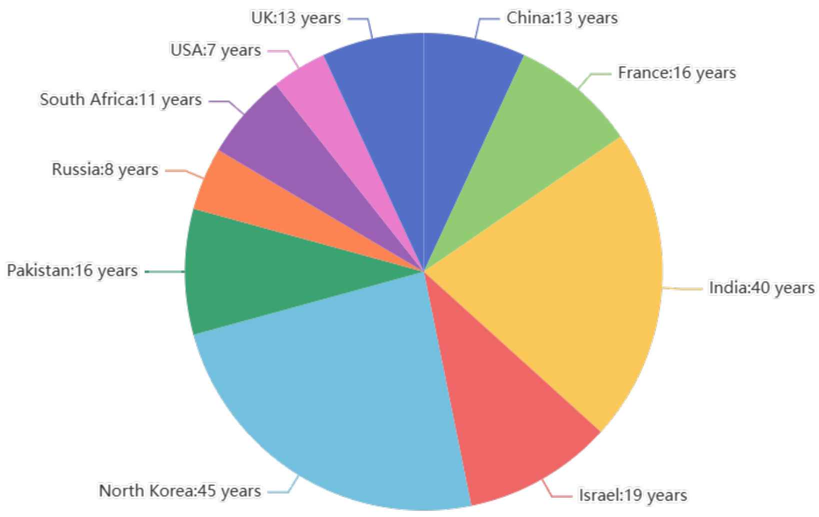

Figure 5 The Shift from Don't Think About Nuclear Weapons to Have Nuclear Weapons

IV. Model building and solution of question 2

We know that with the continuous development of statistical methods and the continuous improvement of modern computer computing power, the ability of computer to process complex data has also been significantly improved. Under the principle of statistics, two statistical methods, machine learning and deep learning, have been developed. These two methods have been widely used in various fields of computer. Question 2 will be solved by establishing regression prediction model based on deep learning and machine learning.

4.1 Basic assumptions and relevant definitions

The number of nuclear weapons is related to many factors, such as the international situation, religious beliefs, economic development, etc. Since many factors cannot be analyzed quantitatively, this model assumes that the international situation, religious belief and other factors have little impact on the number of nuclear weapons, and that the number of nuclear weapons is related to the time series.

Based on the above assumptions, we use deep learning and machine learning methods to predict in the model. Neural network is a new mathematical model proposed in recent years.It is a model established by computer scientists by simulating the information transmission process between human neurons. Its main idea is shown in the figure

6 below. The current neuron receives the information given by the previous neuron according to its weight, and then transfers it to the next neuron through the activation function after processing. The activation function is similar to the impulse threshold in the human nervous system. Only when the impulse reaches a certain value can the neuron be excited, which is still similar in computers. The activation function can select nonlinear function, which enables our neural network to fit nonlinear function. Multi-layer perceptron (MLP) can approach any function with any precision[1], which makes it possible to use MLP to do regression tasks.

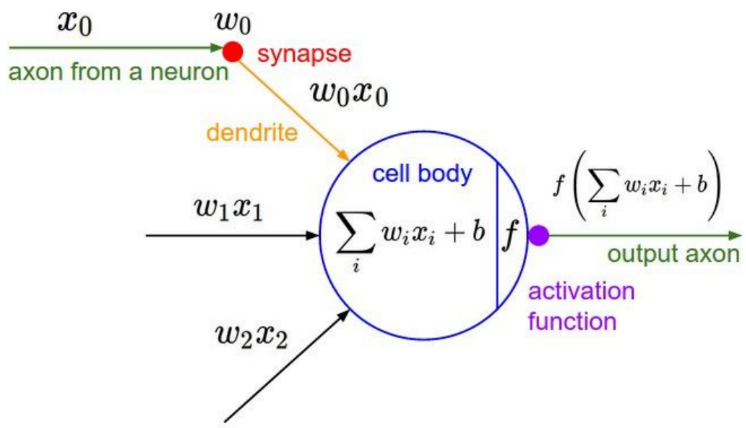

Figure 6 The Working principle diagram of neural network

After establishing the neural network, we can optimize it through the Backpropagation (BP) algorithm. BP algorithm[5] can be used to calculate the gradient of complex functions, which mainly uses the chain rule in calculus of universities.

$$\frac{\delta f}{\delta x} = \frac{\delta f}{\delta u}\frac{\delta u}{\delta x}$$

However, any expression of an extremely complex multivariate function can be decomposed into a composite form of several elementary functions and obtained through addition, subtraction, multiplication and division. So in the neural network, even if the expressed function is very complex, we can calculate the corresponding gradient through the Backpropagation algorithm, and then solve the minimum value of the function according to the gradient descent method. The gradient of the final output function relative to a neuron is equivalent to the upstream gradient value multiplied by the current gradient value. It is worth noting that the object of our final optimization is the loss function selected by the entire neural network. We optimize the loss function to the minimum by gradient descent method, that is, we get our trained fitting model.

In addition, the traditional machine learning method random forest can also be used in regression tasks[2], and random forest is an extended variant of Bagging. On the basis of building bagging ensemble with decision tree based learners, random forest further introduces the selection of random attributes in the training process of decision tree, which has superior performance on data with large fluctuations.

In the development, long short term memory (LSTM) has also performed amazingly in the task of predicting time series[3]. LSTM is essentially a recurrent neural network (RNN), which can handle long-term dependencies. Its working principle is shown in the following figure 7.

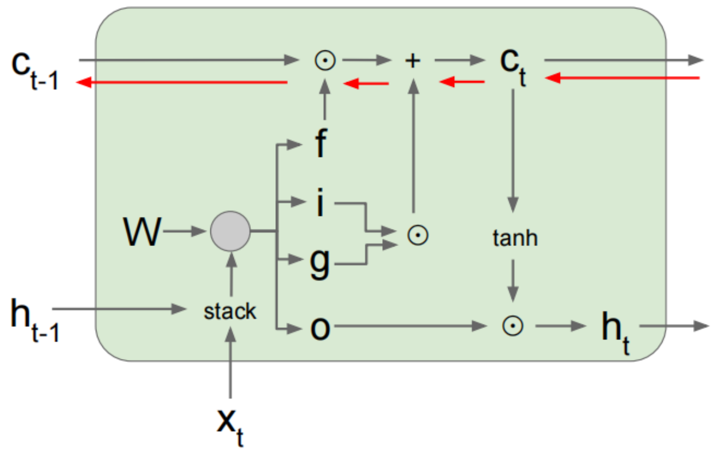

Figure 7 The Working principle diagram of LSTM

It consists of several layers and point by point operations, which act as the gate for data input and output and provide information for LSTM unit status. This unit state maintains long-term memory and context through network and input. Therefore, it can perform well in natural language processing (NLP) or time series prediction.

This paper will solve problem 2 by establishing LSTM, MLP and random forest prediction models, and apply the mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), R^2 score and other index evaluation models, and give the prediction results of the corresponding problems.

4.2 Establish different prediction models

If we directly use the given data to model, it is easy to cause strong noise in the image fitted by the model, and the fitting error is large, so we first consider processing the data. Since the given data fluctuates widely, and there is a large gap between the possession of nuclear weapons among different countries, we first forecast the total amount. In addition, the total amount of data is relatively large. First, we process the total amount of nuclear weapons, take the logarithm of 2 as the base, compress the total amount to [0,16], and process the year dimension to uniformly map the given year to the interval range of [- 40,40]. In this way, the relative fluctuation of data is not large, which is more conducive to our model prediction.The following figure8 shows the change curve of the processed data.

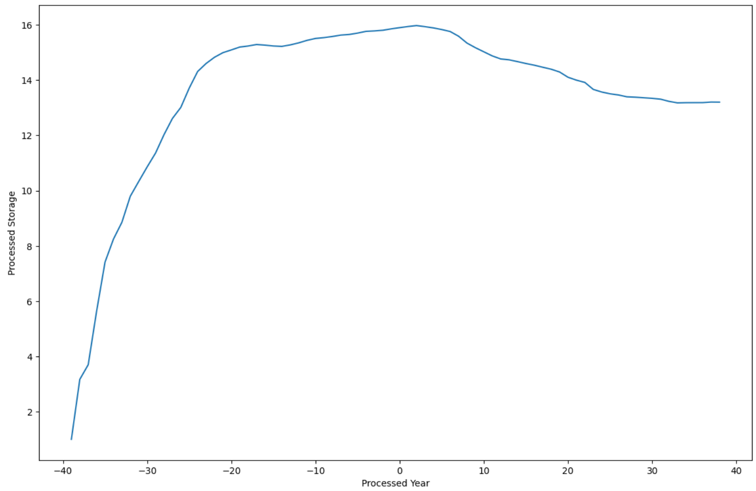

Figure 8 Processed data curve

We can see that the processed data is relatively gentle, which is conducive to our model fitting. Next, we use the MLPRegressor built in Python's Keras library to build a fitting model. We choose ReLU activation function as the connection between neurons. Since the whole MLP has strong continuity, it can be optimized by gradient descent method. We choose Adam to optimize the loss. Because we need to fit the curve, MSE is preferred as the loss function.After the fitting is completed, we will use the trained model to predict. After the prediction, we will reprocess the data into the same form as the input data, that is, restore the year, and do the power operation for the nuclear weapons reserve.The fitting and prediction results are shown in the figure 9 below.

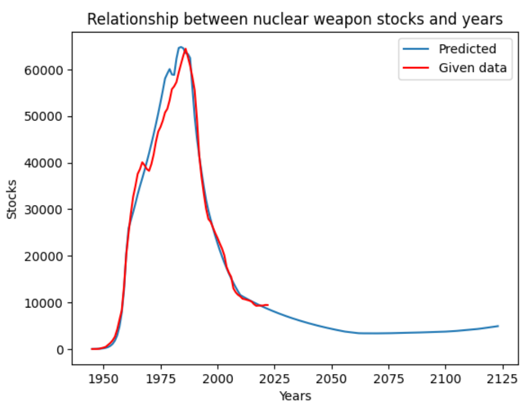

Figure 9 MLP

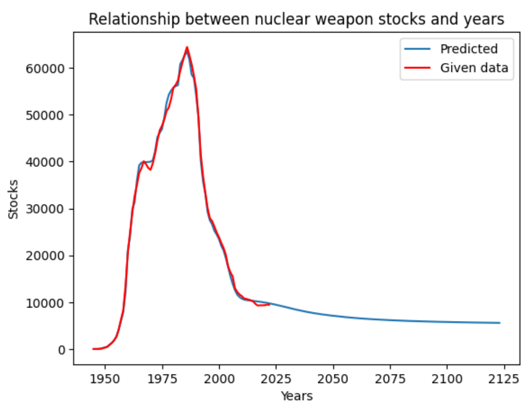

Figure 10 LSTM

Next, we use LSTM model to predict the time series. Because the data is still large, we select Tanh activation function from the activation functions of some full connection layers to prevent the loss from being too large, resulting in the gradient decline optimization failing to converge in time. Tanh activation function can only be used in some neurons, and ReLU activation function should be used in other activation functions to prevent gradient loss. Adam is still used in our optimizer, but SGD random gradient descent method is not selected. If the SGD random gradient descent method is adopted, the convergence speed may be too low. The LSTM fitting effect is shown in the above figure 10.

Finally, we applied the random forest method with the best fitting effect. We applied the random forest regression model in the Python library, repeatedly trained for 1000 rounds, and obtained the following fitting curve figure 11.

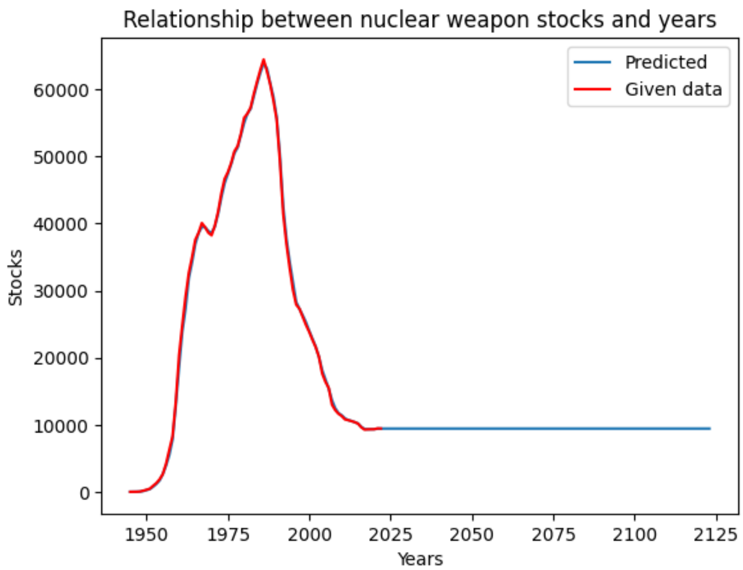

Figure 11 Random Forest fitting effect figure

However, we found that although our model fitted well, it was trained with all data sets, and the accuracy of the final prediction effect of the model could not be guaranteed. Therefore, we considered to divide the data set, taking 1/8 of the data as the test set, and the remaining 7/8 as the training set training data. We will judge the generalization ability of the model according to the difference between training data and test data.We used LSTM and random forest methods to get a new regression model, and the effect is shown in the following figure 12 and figure 13.

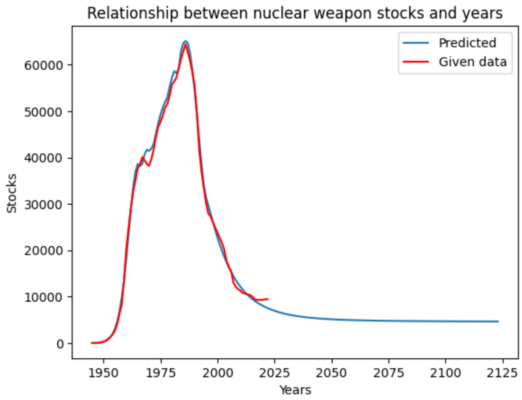

Figure 12 LSTM

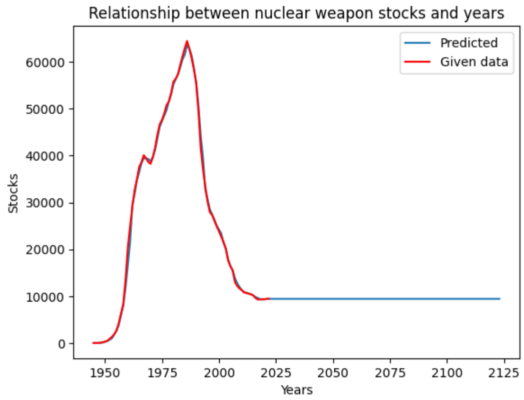

Figure 13 Random Forest

Similarly, we use the same method to model and analyze the data of each country, get the model, and use the model to predict how many nuclear bombs each country may have in 2123. Due to space constraints, the trend chart of each country's possession of nuclear bombs is given in the appendix.The specific data is shown in the following figure 14 and figure 15.

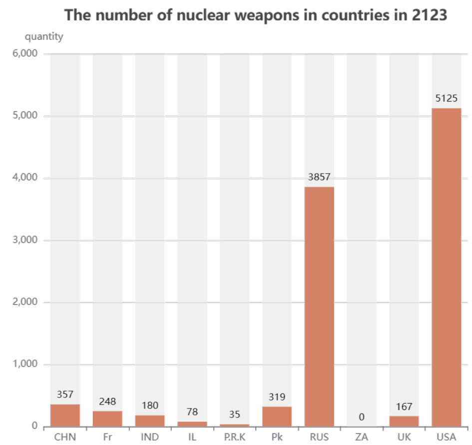

Figure 14 The number in 2123

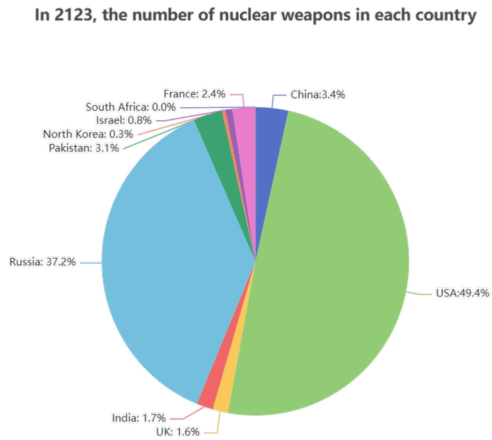

Figure 15 Pie Chart

4.3 The evaluation of models

In this part, we will use multiple parameters to evaluate the fitting effect of the model. The specific meanings are as follows:

$$MSE = \frac{1}{n}\sum(y_i - \widehat{y_i})^2$$

RMSE: Root mean square error (RMSE) is a typical indicator of regression model, which is used to indicate how much error the model will produce in prediction. For larger errors, the weight is higher.

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \widehat{y_i})^2}$$

MAE: The average absolute error (MAE) is used to measure the average absolute error between the predicted value and the real value. The smaller the MAE, the better the model. Its definition is as follows:

$$MAE = \frac{1}{n}\sum|y_i - \widehat{y_i}|$$

R^2 Score: The denominator is the variance of the real value. The larger the variance, the more information it carries. The closer R2 is to 1, the better, and it will be less than 0 in case of extremely poor model.

$$R^2 = 1 - \frac{\sum(y_i - \widehat{y_i})^2}{\sum(y_i - \overline{y_i})^2}$$

Relative error on the test set: describes the gap between the predicted data and the real data. Since we cannot know the number of nuclear weapons in the next 100 years, we can only evaluate the prediction reliability of the model based on the known data.

$$Error = \frac{1}{n}\sum\frac{|y_i - \widehat{y_i}|}{y_i}$$

In the expressions,y_i is given data and y_i is predicted data.The evaluation indicators of different models are shown in the following table 4.

Table 4 Evaluation parameter table

Method	**MSE**	**RMSE**	**MAE**	$R^2 scores$
LSTM(Full data)	725434	851	640	0.998
MLPRegression	14454942.4	3801	2226	0.963
Random Forest Regression(Full data)	259676	509	341	0.999
Random Forest Regression(part of the data)	716390	846	477	0.998
LSTM(part of data)	1551837	1245	987	0.996

The LSTM model that uses part data for prediction, the comparison between the predicted value and the actual value is shown in the following table 5:

It can be seen from the above tables:table 4 and table 5,that the fitting effect of LSTM is good, and the prediction effect of the model obtained by using some data

Table 5 Difference between predicted data and given data.

Given data	Predicted data	Relative error
1290	1156	0.10
24876	19449	0.22
38621	42424	0.10
50757	46204	0.08
62927	59176	0.06
36983	40406	0.09
22656	18447	0.18
11635	11507	0.01
9272	8861	0.04

training is also good. The maximum error is not more than 22 percent, and the average error is about 10 percent. Relatively speaking, the reliability of the model is high. Although MSE and MAE are relatively high, this is due to the large range of data itself. When analyzing the model from multiple dimensions, the model fitting effect is good.

The effect of random forest regression model is also very good, but according to the curve, over fitting may occur, but on the whole, the regression effect is also very good.

4.4 Solution of the problems

With regard to question 1, we have noticed that some countries have changed their attitudes towards nuclear weapons in recent years. For this reason, we have counted the countries that are more concerned about nuclear weapons in the past decade. Among them,**Brazil, Iran and Syria** are more and more concerned about their attitudes towards nuclear weapons in recent years.We believe that these three countries are the most likely countries to develop new nuclear weapons in the next 100 years.

As for question 2, the three models we have built believe that the trend of nuclear weapons development in the future is generally decreasing. According to the best prediction model, the number of nuclear weapons that each country may have in 2123 is shown in the figure 14 and figure 15. In 2123, the total number is 10366, which is similar to the result predicted by our random forest model which has shown in figure 11 and figure 13.

4.5 The interpretation of predicted data

On the whole, the number of nuclear weapons in the future should be stable. With the constraints of the Treaty on the Non Proliferation of Nuclear Weapons and the strict control of the United Nations on nuclear tests, and the growing trend of economic globalization, the trend of all countries in the future is to retain only some nuclear weapons to protect countries, and the number should tend to remain unchanged.

V. Model building and solution of question 3

We found it difficult to directly derive the formula of atomic bomb explosion power through existing knowledge, so we then used dimensional analysis to indirectly derive the mathematical model of nuclear bomb explosion power.

5.1 Introduction to usage and theorem

There is a relationship between various physical quantities, which indicates that their structures must be composed of a number of unified basic components, and they form a myriad of differences between quantities according to the number of components, just as everything in the world is composed of more than a hundred chemical elements. This basic component of physical quantity is called dimension. Because physics studies the evolution and movement of matter in time and space, all quantitative problems ultimately depend on the three basic quantities of mass, time and length. Therefore, it is most suitable to select M, T and L as the dimensions of these three basic quantities. The dimensions of all other derived quantities can be combined into the dimensions of these three basic quantities according to definitions or objective laws. There are many ways to take basic quantities. In mechanics, mass, length and time are usually taken as basic quantities. Other quantities (such as speed, force, etc.) can be derived from basic quantities according to certain rules. Any other derived quantity with three independent dimensions can also be used as the basic quantity. Two physical quantities that are completely different in nature can have the same dimension, such as work and torque. Any equation that correctly reflects the laws of physical phenomena must have the same dimensions at both ends.

The size of a physical quantity is usually expressed by one or several real numbers together with the units used, except for those measured by numbers. This kind of number is generally called ”nominal number”, which means no meaningful number without indicating the name of the unit. The real value of the nominal number can vary with different objects in different time or space. This is a substantial change due to different objects or changes in the object itself. However, the name and value will also change with the size of the unit used, and it is a continuous function of the size of the unit. Since the size of the unit can be selected at will, the above changes in name and value are not objective and substantial changes. The law of substantial change is the research object of the discipline itself. The various physical laws obtained from the research are expressed in the form of mathematical equations, which control the growth and decline of the relevant quantities themselves. The non substantial change does not involve the substantial objective process, but reflects the subjective choice of the unit. Of course, objective laws do not depend on subjectivity, which requires that the non substantial change of values must ensure the absoluteness of the objective size of things. Specifically, no matter what the unit of measurement is, the relative size of any two similar quantities of a certain size will never change, that is, their ratio must be a fixed value for any unit. The invariance of the relative size of the same kind of quantity to the unit is the fundamental principle of measurement. In violation of this principle, measurement will be meaningless. According to this principle, the following important conclusions can be drawn: in the determined unit system, all physical quantities have the power product form of the basic quantity dimension (the proof is omitted), that is, their form can be written as $x^a y^b z^c$. Including x,y,z Is the dimension of the basic quantity; Power a,b,c are constants, but not necessarily integers.

The theoretical basis of dimensional analysis is π Theorem, which was proposed by E. Buckingham in 1914:

Suppose a physical problem involves n physical quantities (including physical constants)$P_1, P_2......P_n$, and there are m basic quantities in the unit system we choose, then n-m dimensionless quantities can be formed. There is a functional relationship between physical quantities$P_1, P_2......P_n$

$$f(P_1, P_2......P_n) = 0$$

It can be expressed as dimensionless form:

$$F(\pi_1, \pi_2......\pi_n) = 0$$

(In the case of n=m, there are two possibilities: if the dimensions of$P_1, P_2......P_n$are independent of each other, they cannot form a dimensionless quantity; if they are not independent, they can also form a dimensionless quantity)[4].

5 . 2 Establishment and solution of nuclear explosion model

Assume that the radius R of shock wave formed by explosion only depends on the time t after explosion, the energy E released at the moment of explosion, the atmospheric pressure P, and the air density $\rho, thatis$

$$R = \varphi(t, E, P, \rho)$$

that is

$$f(R, t, E, P, \rho) = 0$$

This problem is a dynamic problem. The only dimensions involved are length dimension L, mass dimension M, and time dimension T

The dimensions of physical quantities involved in this question are:

$$[R] = L; [t] = T; [E] = L^2MT^{-2}; [P] = L^{-1}MT{-}2; [\rho] = L^{-3}M$$

The dimension matrix is:

$$A_{3*5} = \begin{bmatrix} 1 & 0 & 2 & -3 & -1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & -2 & 0 & -2 \end{bmatrix}$$

The number of rows is 3, the number of columns is 5, and the rank of the matrix is 3,The derived equation has two basic solutions

$$y_1 = (1, -\frac{2}{5}, -\frac{1}{5}, \frac{1}{5}, 0)$$

$$y_2 = (0, \frac{6}{5}, -\frac{2}{5}, -\frac{3}{5}, 1)^T$$

That is, two dimensionless quantities are solved:

$$\pi_1 = Rt^{-\frac{2}{5}}E^{-\frac{1}{5}}\rho^{\frac{1}{5}} = R(\frac{\rho}{t^2E})^{\frac{1}{5}}$$

$$\pi_2 = t^{\frac{6}{5}}E^{-\frac{2}{5}}\rho^{\frac{3}{5}}P = (\frac{t^6P^5}{\rho^3E^2})^{\frac{1}{5}}$$

Because

$$F(\pi_1, \pi_2) = 0$$

we can deduce

$$R(\frac{\rho}{t^2E})^{\frac{1}{5}} = \varphi((\frac{t^6P^5}{\rho^3E^2})^{\frac{1}{5}})$$

so

$$R = (\frac{t^2E}{\rho})^{\frac{1}{5}}\varphi((\frac{t^6P^5}{\rho^3E^2})^{\frac{1}{5}})$$

Because the explosion time is very short and the energy is very large, we approximately believe that

$$(\frac{t^6P^5}{\rho^3E^2})^{\frac{1}{5}} \approx 0$$

so $\varphi(0)$ is a constant.

Therefore,We get the formula of nuclear bomb explosion power:

$$R = C(\frac{t^2E}{\rho})^{\frac{1}{5}}$$

According to the known small nuclear explosion experiment:

$$C = 1$$

so

$$R = (\frac{t^2E}{\rho})^{\frac{1}{5}}$$

This calculation method is somewhat complicated, and we can also estimate the nuclear bomb power through simple methods. A simple formula for estimating the nuclear bomb is as follows:

$$R = CT^{\frac{1}{3}}$$

(Where, R is the killing range radius of the nuclear bomb, T is the explosion yield of the nuclear bomb, and C is the proportional constant,$C \approx 1.493885$)

According to the formula, the killing radius of the most powerful nuclear bomb ”czar bombs” detonated by humans is 25.54km

5.3 Several Methods and Calculations of Bombing the Earth

5.3.1 *Simple calculation by area*

According to the formula we obtained above, we can get the radius of the killing range of the nuclear bomb. Then we can estimate the killing area through the area formula of the circle. Then we can divide the surface area by the killing area to get the number of nuclear bombs needed.

Through the above process, we can calculate that the killing area of the ”czar bombs”is about 2050 square kilometers,as the following figure16,while the surface area of the earth is about 510072000 square kilometers.It can be calculated that it takes about 250000 ”czar bombs” to destroy the earth.

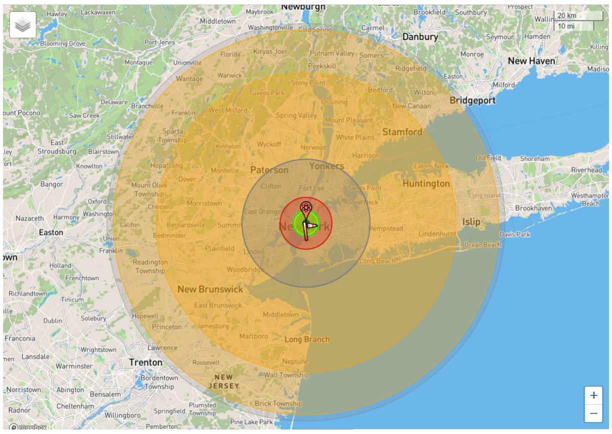

Figure 16 Killing range of nuclear bomb

5.3.2 *Analysis from the perspective of energy*

According to the calculation, the energy released by the explosion of the ”czar bombs” is about 2.1 x $10^{17}J$.

According to the data, the energy required to blow up the earth’s surface is 1.5 x $10^{30}J$.

It can be calculated that it will take about 7.1 x 10 ^{12}”$czarbombs$”.

However, it is not necessary to blow up the earth’s surface to destroy the earth’s ecological environment. We can also try to estimate it from the thermodynamic formula.According to thermodynamic formula: $Q = cm\Delta t$. The energy required to melt the earth’s surface is 2.9 x $10^{28}J$.

It can be calculated that it will take about 1.4 x 10 ^{11}”$czarbombs$”.

But considering that it only needs to evaporate all the water on the earth to kill all the life on the earth, according to the above formula, we calculate that the energy required to evaporate all the water on the earth is 4.5 x $10^{27}J$.

It can be calculated that it will take about 2.1 x 10 ^{10}”$czarbombs$”.

Taking into account the protective effect of the earth’s atmosphere, we believe that without the protection of the atmosphere, the earth’s ecology will be destroyed by radiation. Therefore, only the earth’s atmosphere needs to be blown up to destroy the earth. According to the data, the energy needed to blow up the earth’s atmosphere is 3.2 x 10 ^{26}J.

It can be calculated that it will take about 1.5 x 10 ^{9}”$czarbombs$”.

5.3.3 *Analysis of other ecological aspects*

In fact, the destruction of the earth does not require so many nuclear bombs. From an ecological perspective, few nuclear bombs can cause great damage to the ecology.

For example, the nuclear winter effect produced by nuclear bombs is worth considering.

When calculating the "nuclear winter" phenomenon that may be caused by nuclear war, it is necessary to consider the impact of radiation cloud generated by nuclear weapon explosion on the earth surface temperature. Specifically, the shielding degree of radiation cloud to the earth surface temperature can be calculated, and then the shielding degree can be converted into the variation of the earth surface temperature. In order to determine the shielding degree of radiation cloud to the earth surface temperature, the atmospheric propagation model can be used to simulate the propagation process of radiation cloud. Specifically, the following equation can be used:

$$I = I_0 e^{-x}$$

Where, I is the intensity of the radiation cloud at the depth of x, and I_0 is the initial intensity of the radiation cloud, It is the absorption rate of radiation cloud in the earth's atmosphere.

By calculating the intensity of the radiation cloud, we can calculate the shielding degree of the radiation cloud to the earth surface. The degree of obscuration can be expressed by the following equation:

$$T = (1 - I/I_0) * T_0$$

Among them,ΔT is the degree to which the radiation cloud obscures the earth's surface temperature, ΔT_0 is the change of the earth's surface temperature when the radiation cloud does not exist. Finally, the mathematical model of "nuclear winter" phenomenon that may be caused by nuclear war can be obtained by substituting the shielding degree of radiation cloud to the earth surface temperature into the greenhouse effect equation.

In this way, only a few thousand Tsar bombs will bring great impact on the earth's ecology.

5.4 Answers of Question 3

For 3.a, the true meaning of "destroying the earth" requires $7.1x10^{12}$ "czar bombs".But the destruction of human civilization does not require so many nuclear bombs.The earth we live on is very powerful. Nuclear war may destroy us, but for the earth, it may just be that the earth has sent away a group of passers-by. Without human beings, the earth will continue to develop hundreds of millions of years later. It is always human civilization, not the earth, that is easy to destroy.

For 3.b, the yield of the most powerful nuclear bomb ”czar bombs” is $2.1x10^{17}J$, and the killing area is $2050km^2$. Even if the existing bombs on the earth are ”czar bombs”, they are far from the nuclear bombs needed to destroy the earth.

Table 6 The upper limit of nuclear weapons possession of each country

Country	**Upper limit**
China	1000
France	900
India	200
Israel	100
North Korea	20
Pakistan	200
Russia	8500
South Africa	0
United Kingdom	500
United States	8500

For 3.c, we believe that the total number of nuclear bombs on the earth should be less than one tenth of the number of nuclear bombs that can cause mass extinction. With reference to the ”Hiksurub meteorite” (the culprit of the extinction of dinosaurs), the energy of the meteorite is about $5.0x10^{23}J$, which is equal to about 238000 ”czar bombs”, so the total number of nuclear bombs on the earth should be controlled below 20000.The reason why more nuclear bombs are allowed to remain is that mankind still needs weapons to face emergencies.The specific situation of each country is shown in the table 6 above.

VI. The solution of question 4

Nuclear weapons are among the most destructive weapons in human history, and in the past few decades, almost all major powers have possessed them. However, as competition between countries around the world has increased, the threat of nuclear weapons has grown. Nine countries worldwide currently possess nuclear weapons: the United States, Russia, China, the United Kingdom, France, India, Pakistan, North Korea, and Israel. Of these, the United States and Russia have the largest number of nuclear weapons. Other countries also possess some number of nuclear weapons, but fewer overall.This study finds that the total number of nuclear weapons in the world will show a clear downward trend in the coming period, and the number of nuclear weapons of each nuclear power will also basically show a decreasing or unchanging trend. If nuclear weapons were used for military purposes, they would result in enormous loss of life and property. In addition, the use of nuclear weapons could also lead to serious environmental pollution and possibly even global climate change. In view of the current situation, on the one hand, the UN should take some actions to reduce the threat of nuclear weapons. First, the UN should promote the accession of all nuclear-armed states to the NPT and ensure that countries that already have nuclear weapons do not transfer them to other countries. Second, the UN should also promote dialogue and cooperation among nuclear-armed states to reduce tensions and lower the risk of nuclear weapons use. Finally, the UN should also support anti-nuclear campaigns to help raise public awareness of nuclear weapons and promote public opposition to their use. On the other hand, we make several recommendations for all states: First, nuclear-armed states should strengthen the development and application of nuclear non-proliferation technologies to further enhance the security of nuclear weapons and to effectively prevent accidental leaks and misuse. Second, nuclear security cooperation among countries should be strengthened to control the use and proliferation of nuclear weapons and to ensure global nuclear security. Third, nuclear-armed states should take the initiative to promote the nuclear disarmament process, reduce the risk of a nuclear arms race and reduce the risk of nuclear proliferation, and lower the cost of military deterrence. In general, to truly reduce the threat of nuclear weapons, this goal must be achieved by joint efforts of all countries and through international cooperation. Only through joint efforts can we build a truly peaceful world in which humanity is no longer threatened by nuclear weapons.

VII. Conclusions

7.1 Our contribution

- We use a variety of statistical methods to fit the given data, and give a variety of evaluation indicators to evaluate the model.
- Because we do not know the data for the next 100 years, we will divide the given data into training set and test set, so that the trained model can predict on the test set, and give the relative error to judge the prediction ability of the model.
- We used the dimensional analysis method to establish an explosion lethality model, and calculated the lethality range of the nuclear bomb according to this model. Based on this model, we solved the third question.

7.2 Limitation of our work

- Because the given data is relatively few, we can not get an accurate model to predict.
- Both the prediction of the power of nuclear explosion and the number of nuclear weapons are related to many factors, and we assume that they are decoupled from other factors in this paper. This simplifies the model to some extent, but cannot completely simulate the actual situation.
- We did not consider the impact of nuclear bomb detonation on climate, biosphere and other factors, as well as the impact of nuclear radiation on human body, which makes us still have errors in estimating the power of nuclear explosion.

7.3 Future work

- If military strength, economic factors, international situation and other factors can be analyzed quantitatively, the prediction model in question 2 will be closer to reality and the prediction results will be more reliable.
- In addition, if we can establish a more complex nuclear bomb detonation model and take into account the harm to the environment, then the prediction of the real power of nuclear bombs will be more accurate.

VIII. References

[1] Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. Acta Numerica, 8, 143-195. doi:10.1017/S0962492900002919

[2] Segal M R. Machine learning benchmarks and random forest regression[J]. 2004.

[3] Kaselimi M, Doulamis N, Doulamis A, et al. Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring[C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 2747-2751.

[4] Buckingham,E,Phys.Rev.,4(4),345(1914);J.Wash.Acad.,Sci,3,347(1914).

[5] Werbos P J. Backpropagation through time: what it does and how to do it[J]. Proceedings of the IEEE, 1990, 78(10): 1550-1560.

IX. Appendix

Listing 1: The python code for question 1-a

```
import pandas as pd
data = pd.read_csv(r"C:\Users\dell\Desktop\2.csv")
data1=data[data["Stockpile"]>0]
a=data["Country"]
print(set(a))
```

Listing 2: The python code for question 1-b

```
import pandas as pd
data = pd.read_csv(r"C:\Users\dell\Desktop\2.csv")
data21=data[data["Year"]==2022]
data22=data[data["Year"]==2002]
a1=data21["Stockpile"]
a2=data22["Stockpile"]
a3 = list(map(lambda x: x[0]-x[1], zip(a1, a2)))
data21["Difference"]=a3
up=data21[data21["Difference"]==max(data21["Difference"])]
down=data21[data21["Difference"]==min(data21["Difference"])]
print(up)
print(down)
```

Listing 3: The python code for question 1-c

```
import numpy as np
data=np.loadtxt(r"C:\Users\dell\Desktop\3.csv",
              delimiter=',',skiprows=1,usecols=(2,3))
np.set_printoptions(suppress=True)
a=[]
for i in range (1945,2020):
   data1=data[data[:,0]==i]
   data11=data1[:,1]
   sum=np.sum(data11)
   a.append(sum)
sum1=0
```

```
index=0
for i in range(0,71):
    sum2=a[i]+a[i+1]+a[i+2]+a[i+3]+a[i+4]
    if sum2>sum1:
        sum1=sum2
        index=i
print(index+1945)
```

Listing 4: The python code for question 1-d

```
import pandas as pd
data = pd.read_csv(r"C:\Users\dell\Desktop\3.csv")
data31=data[data["Year"]>2008]
data32=data31[data["Tests"]>0]
print(data32)
```

Listing 5: The python code for question 1-e

```
import pandas as pd
import numpy as np
data = pd.read_csv(r"C:\Users\dell\Desktop\4.csv")
data5 = pd.DataFrame(columns=["Country", "Difference"])
data5["Country"]=["China","France","India","Israel","North
    Korea","Pakistan","Russia","South Africa","United Kingdom","United
    States"]
a=[]
for i in data5["Country"]:
    data51=data[data["Country"]==i]
    data512=data51[data["Status"]==3]
    data512=data512[data512["Year"]==min(data512["Year"])]
    data511=data51[data["Status"]==0]
    data511=data511[data511["Year"]<min(data512["Year"])]
    data511=data511[data511["Year"]==max(data511["Year"])]
    a.append(int(data512["Year"])-int(data511["Year"]))
data5["Difference"]=a
print(data5)
```

Listing 6: The Python code for data processing of question 2

```
data_position=np.loadtxt("./China_storage.csv",skiprows=1,
                     delimiter=",",usecols=(2,3))
data_storage=np.loadtxt("./storage.csv",skiprows=1,delimiter=",",
                     usecols=(2,3))
a=[]
X_label=[]
X_train=[]
Y_train=[]
for i in range(1945,2023):
    data1=data_storage[data_storage[:,0]==i]
    storage=data1[:,1]
    sum=np.sum(storage)
    Y_train=np.append(Y_train,sum)
    a.append(sum)
    X_label.append(i)
    X_train=np.append(X_train,i-1984)
from sklearn.neural_network import MLPRegressor
import matplotlib.pyplot as plt
import numpy as np
import math
plt.figure(figsize=(14,9))
iter = 1000
X_train=X_train.reshape(-1, 1)
maxx=np.max(Y_train)
minn=np.min(Y_train)
YY_train=[]
for y in Y_train:
    if y>0:
        y=math.log2(y)
    YY_train=np.append(YY_train,y)
YY_train=YY_train.reshape(-1, 1)
origin_x=X_train
origin_y=YY_train
new_data_train=[]
new_data_trainlabel=[]
new_data_test=[]
```

```
new_data_testlabel=[]
for i in range (1,78):
    if(i%8==0):
        new_data_test=np.append(new_data_test,X_train[i])
        new_data_testlabel=np.append(new_data_testlabel,YY_train[i])
    else:
        new_data_train=np.append(new_data_train,X_train[i])
        new_data_trainlabel=np.append(new_data_trainlabel,YY_train[i])
```

Listing 7: Random forest model

```
# Fitting Random Forest Regression to the dataset
# import the regressor
from sklearn.ensemble import RandomForestRegressor
for i in range(1,10):
    plt.subplot(330+i)
    plt.scatter(X_train,YY_train,s=5)
    rg = RandomForestRegressor(n_estimators = 100, random_state = 0)
    rg.fit(X_train, YY_train)
    y_pre = rg.predict(X_train)
    plt.plot(X_train,y_pre,c='red')
    plt.title('iterations:'+str(iter))
    iter += 1000
origin_x=X_train
origin_y=YY_train
```

Listing 8: LSTM model

```
new_data_train=new_data_train.reshape(-1,1)
new_data_trainlabel=new_data_trainlabel.reshape(-1,1)
new_data_test=new_data_test.reshape(-1,1)
new_data_testlabel=new_data_testlabel.reshape(-1,1)
model = tf.keras.Sequential()
model.add(LSTM(100, return_sequences=False))
model.add(tf.keras.layers.Dense(1024, activation='tanh'))
model.add(tf.keras.layers.Dense(256, activation='relu'))
model.add(tf.keras.layers.Dense(64, activation='relu'))
model.add(tf.keras.layers.Dense(1))
```

```
model.build((None,90,1))
model.summary()
model.compile(loss=tf.keras.metrics.mean_squared_error,
    optimizer='adam',
                metrics=['accuracy'])
history = model.fit(new_data_train, new_data_trainlabel,
                     epochs=10000,
                     validation_split=0.1,
                     batch_size=30,
                     shuffle=True,
                     )
LSTM_pred=model.predict(X_train)
plt.plot(X_train,LSTM_pred,c='red')
plt.plot(origin_x,YY_train)
```

Listing 9: MLP model

```
for i in range(1,10):
    plt.subplot(330+i)
    plt.scatter(X_train,YY_train,s=5)
    rg = MLPRegressor(activation='relu',max_iter=iter)
    rg = rg.fit(X_train,YY_train)
    y_pre = rg.predict(X_train)
    plt.plot(X_train,y_pre,c='red')
    plt.title('iterations:'+str(iter))
    iter += 10000
origin_x=X_train
origin_y=YY_train
```

Listing 10: Draw figure

```
plt.plot(X_train,yy_pre)
plt.plot(origin_x,origin_y,color='red')
plt.xlabel("Years")
plt.ylabel("Stocks")
plt.legend(labels=["Predicted","Given data"],loc="upper
    right",fontsize=10)
plt.title("Relationship between nuclear weapon stocks and years")
```

```
plt.show()
plt.show()
```

Listing 11: Calculate the parameters of the evaluation model

```
mse=0
total=0
c=0
mae=0
for i in range(1,78):
    mse+=(yy_pre[i]-origin_y[i])**2
    total+=origin_y[i]
    mae+=abs(yy_pre[i]-origin_y[i])
average=total/78
for i in range(1,78):
    c+=(origin_y[i]-average)**2
R=0.0
R=1-mse/c
mse=mse/78
print("The Mean Absolute Error is:")
print(mse)
print("The Root Mean Square Error is:")
Rmse=math.sqrt(mse)
print(Rmse)
print("The value of coefficient of determination is")
print(R)
mae=mae/78
print("Mean Absolute Percentage Error is")
print(mae)
```

The figures for predicted the number of nuclear weapons possessed by each country in the next decade are below.

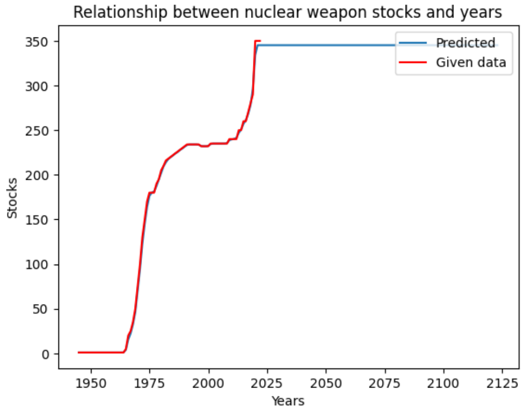

Figure 17 The storage of China

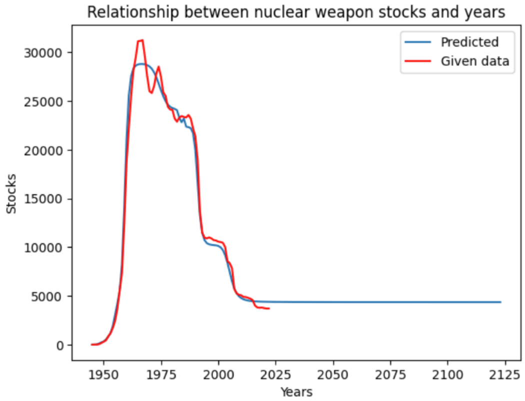

Figure 18 The storage of USA

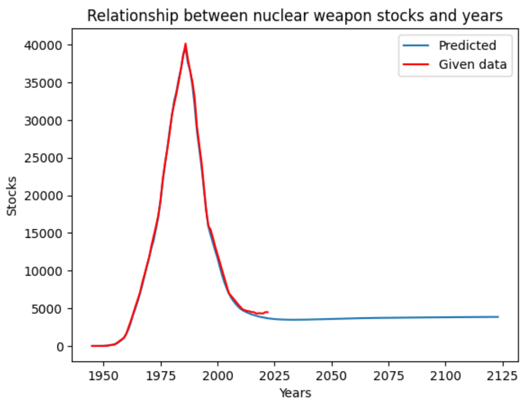

Figure 19 The storage of Russia

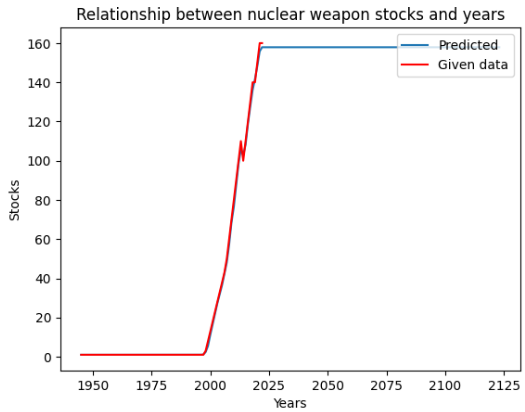

Figure 20 The storage of India

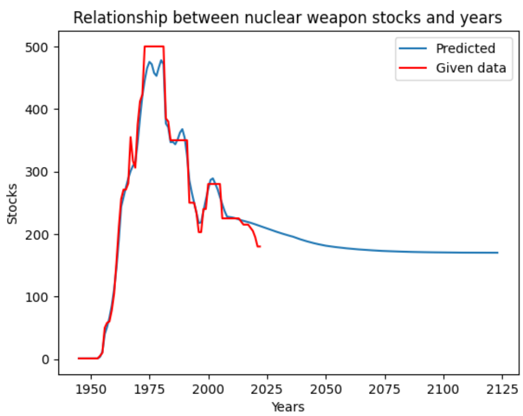

Figure 21 The storage of UK

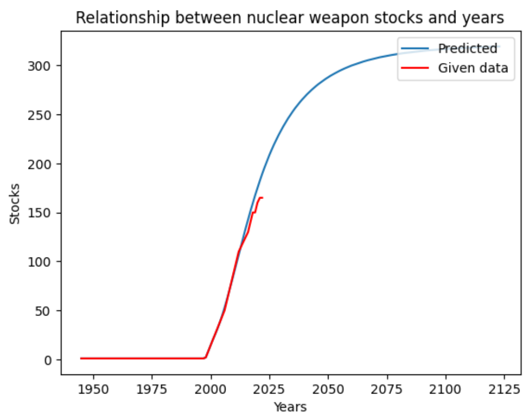

Figure 22 The storage of Pakistan

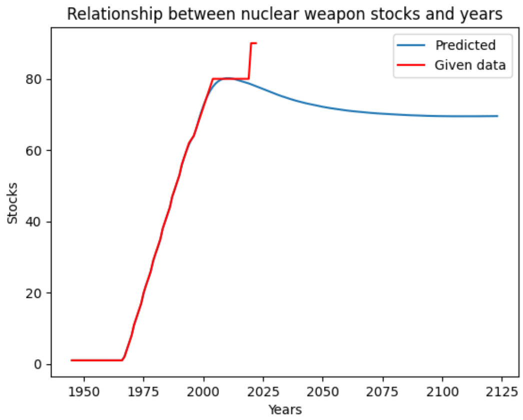

Figure 23 The storage of Israel

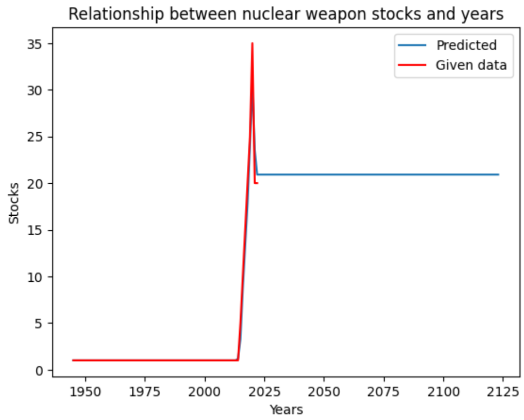

Figure 24 The storage of North Korea

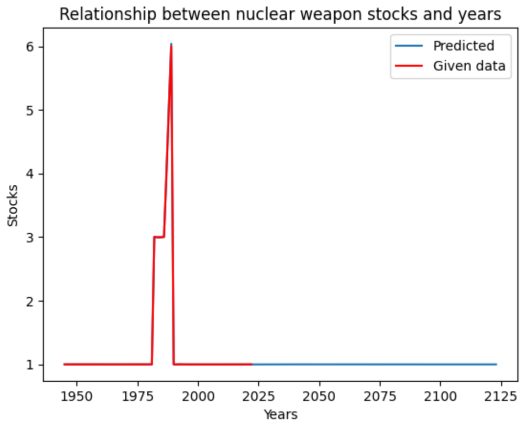

Figure 25 The storage of South Africa

Figure 26 The storage of France