Science Frontiers
2022; 3(2): 66-73
http://www.sciencepublishinggroup.com/j/sf
doi: 10.11648/j.sf.20220302.12

Applying Deep Learning Technology on Prediction of Gini Coefficient

Ruijian Zhang

Department of Computer Science, Purdue University, Hammond, USA

Email address:
zhang45@purdue.edu

To cite this article:
Ruijian Zhang. Applying Deep Learning Technology on Prediction of Gini Coefficient. *Science Frontiers*. Vol. 3, No. 2, 2022, pp. 66-73.
doi: 10.11648/j.sf.20220302.12

Received: May 2, 2022; **Accepted**: May 18, 2022; **Published**: May 31, 2022

Abstract: Over the last few decades, income inequality has grown dramatically in the United States, with the top twenty percent earning more than the bottom eighty percent combined. The studies about the income inequality is becoming more and more important in the economics and politics area. The research of this paper combined the computer technology and the economic problem solving together. The project applied Deep Learning technology for creating a model to predict the most important indicator of income inequality, Gini coefficient in the future based on the historical relevant data, so that observe the ever widening gap between the rich and the poor. It also found the major elements that governed this widening behavior through analyzing the impacts of the related attributes. This study obtained and analyzed substantial amount of data, which contain information on the income, expenses and the financial footprints of families in the United States, to draw empirical conclusions. The results may help the public and the economic research society for their decision making. Using Deep Learning algorithms, the data analysis was far more efficient, and by generating the Deep Learning multi-layer Neural Networks, the prediction was quite accurate. This study has obtained some promising results. It showed an encouraging direction on the prediction of Gini coefficient, through applying Deep Learning models.

Keywords: Deep-Learning, Neural Networks, Feature Correlation, Prediction, Income Inequality, Gini Coefficient

1. Introduction

The levels of income inequality, once a problem for developing nations, is now a shaping social and political issues in the United States. Over the last few decades, income inequality has grown dramatically in the United States, with the top twenty percent earning more than the bottom eighty percent combined. The research about the income inequality is becoming more and more important in the economics and politics area [1].

Income is defined as household disposable income in a particular year. It consists of earnings, self-employment, capital income and public cash transfers; income taxes and social security contributions paid by households are deducted. The income of the household is attributed to each of its members, with an adjustment to reflect differences in needs for households of different sizes. Income inequality among individuals is measured by some indicators.

Gini coefficient is a commonly-used indicator and method in economics to measure the income distribution within a society of a nation or region [2, 3]. It was proposed by an Italian economist Corrado Gini in 1922 and derived from the Lorenz curve. Lorenz curve is a graphical representation of how equal the income or wealth distribution is in a society. It was developed by an American statistician M. O. Lorenz in 1905, who constructed 10 income groups by ranking people's income, with each group accounting for 10% of the total population. The proportion of the overall income assumed by each group is then calculated. In Figure 1, the percentage of the households is plotted on the x-axis, the percentage of income is plotted on the y-axis. The actual income distribution curve, which is named the Lorenz Curve, can be drawn in the diagram.

In Figure 1, the line OL at 45 degree indicates perfect equality of income distribution, under which scenario, each 10% of the population gets 10% of the total income. Hence it is named Line of Perfect Equality. The line OXL shows perfect inequality of income distribution, under which scenario, one person has all the income. Therefore it is named Line of Total Inequality. The curve ODL, falling between OL and OXL and representing actual income distribution, is the Lorenz curve. It indicates a more equal income distribution when closer to the Line of Perfect Equality, and a less equal income distribution when closer to the Line of Total Inequality.

Gini found the criteria to represent the equality of income distribution based on the Lorenz curve. He used A to express the area between the Line of Perfect Equality and the curve of actual income distribution, and B to express the area under the Lorenz curve. The inequality level is represented by the ratio of A over A+B. This ratio is named Gini coefficient or Lorenz coefficient (see Figure 1). If A is zero, the Gini coefficient is zero, representing perfect equality; if B is zero, the coefficient is one, meaning total inequality. The Gini coefficient could be any value between zero and one. The more balanced the income distribution, the flatter the Lorenz curve, and the smaller the Gini coefficient; on the contrary, the less balanced the income distribution, the more arched the Lorenz curve, and the greater the Gini coefficient.

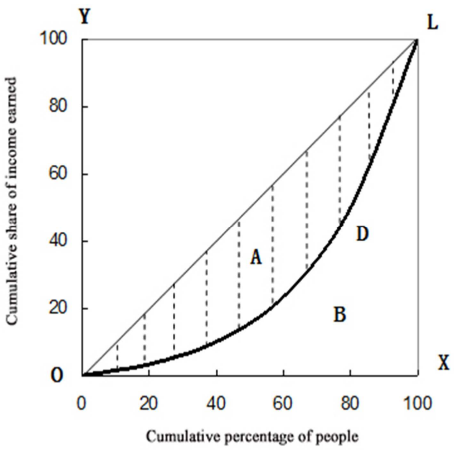

Figure 1. *Lorenz Curve.*

According to the United Nations' standard, a Gini coefficient less than 0.2 represents perfect equal income distribution; one between 0.2 and 0.3 represents relatively equal distribution; one between 0.3 and 0.4 represents relatively reasonable distribution; one between 0.4 and 0.5 represents relatively unequal distribution; and one above 0.6 represents unequal distribution.

This study applied Deep Learning technology for prediction of Gini coefficient in the future based on the historical relevant data. It may help the public and the economic research society for their decision making.

The project applying Deep Learning technologies is a process of extracting patterns from data. The process consists of the following steps: data collecting, data preparation, and Gini coefficient prediction. The project designed and implemented a Deep Learning system for income inequality modeling.

Generally speaking, the growing income inequality has been caused by the stagnation of real wages in the middle and lower class, while the income of the top one percent has nearly tripled [1]. Modern economics provides more causes for increased income inequality in developed countries, such as employment, minimum wages, wealth, health, childbearing and development, education, gender discrimination, political reform, and even the rise of technology and automation in the workforce. For example, one potential cause for the growth of income inequality is inflation. While the rich have resources to grow their wealth, such as the stock exchange or private investments, the poor have limited or no access to these resources. At the same time, all wealth that is not invested and growing will not only stagnate but steadily lose value as inflation decreases the value of currency [4, 5].

The project identified data resources that were related with the causes of the income inequality. The data described income changes, cost of living, and inflation, etc., in the United States over time. The relevant data sources were found from previous studies, government publications, and digital libraries [6-12]. The project uniformly formatted the data, relative to time, to provide a clean, solid base for data mining-deep learning system. The data cleaning, data format and normalization, missing or invalid data treatment and other data preprocessing and preparation work were done before being used.

The project applied R, a statistical analysis tool, to analyze the gathered data for underlying correlation between Gini coefficient and other related attributes. The created intelligent system verified the existence of links between Gini coefficient and the other related attributes through the data set. It only kept a group of the most correlated attributes in the data set. Then the project built models of the data to show that Gini coefficient and the group of correlated attributes are intrinsically linked.

The project established a Deep Learning system, which applied multi-layer Neural Network for multiple regression. It built an intelligent Gini coefficient prediction system. The project quantified the gap between the rich and the poor, built a basic model highlighting the trends of income inequality. A data model was designed, that predicted the Gini coefficient in the future. We applied Neural Networks technologies with Python programming language to perform Gini coefficient prediction. We also employed a Deep Learning system with multi-layer Neural Networks and TensorFlow to improve the prediction accuracy.

Furthermore, the visualized results obtained from this project were published in an interactive website so that the public and the economic research society can use the information for their decision making.

This paper is organized in seven sections. Section two to five elaborate the design and implementation of the intelligent Gini coefficient prediction system. Section Two is about data collection and data preprocessing. Section Three is about feature selection and dimensionality reduction. Section four is about creating the datasets for Gini coefficient prediction. Section Five is about the deep-learning system for Gini coefficient prediction. Section Six discusses and analyzes the experimental results. Finally, Section Seven provides concluding remarks.

2. Data Collection and Data Preprocessing

We used data to “train” the Gini coefficient prediction model. The data sources were collected from government publications, previous studies, and digital libraries. One of them is from the Federal Reserve Economic Data (FRED), which is an online database consisting of hundreds of thousands of economic data time series from scores of national, international, public, and private sources. FRED, created and maintained by the Research Department at the Federal Reserve Bank of St. Louis, has contained many of the databases reported by the Board of Governors, Bureau of Economic Analysis, Bureau of Labour Statistics, and Census, etc. Through time, FRED has expanded its collection to include many more international, national, and regional data series. Naturally, care will be taken to add data in a thorough and prudent manner. Furthermore, FRED goes far beyond simply providing data, it combines data with a powerful mix of tools that help the user understand, interact with, display, and disseminate the data. The data are accessible from a variety of different hardware and software, with the primary point of access being the FRED website [6]. From the homepage, users can choose to search for the data by typing in their search term or alternatively can browse the data through other organized points of access.

Another data source is from the Organization for Economic Cooperation and Development (OECD), which is helping governments to understand how the economy is changing, so they can improve their policies to make the economies stronger and fairer. The data is from the OECD Income Distribution Database [7], which provides comparable data on income, income inequality and poverty across many countries including United States.

The third data source is from the University of Michigan’s PSID (Panel Study of Income Dynamics). The University of Michigan has been conducting study on “Rich getting richer” with families in the United States and has been successful in collecting data from 1968. This data source is categorized into five sections, each of that is used based on what is needed for the specific case study to train the proposed model. The most important section of the PSID is the Family Identification Mapping System, which gathers inter- and intra-generational data regarding all familial ties of an individual, which can be used to observe trends in the income dynamics of individual households over time [8, 9]. The researchers can replicate results of their previous studies by access these datasets.

Other data sources include the United States Census’ Current Population Survey, which examines the whole U.S. population’s income and economic data [10]; the World Bank Open Data [11], which has free and open access to global development data and can be browsed by Country or Indicator, and the data of Trends in the Distribution of Household Income, from the US *Congressional Budget Office* [12]. By using data from various sources, any objective bias should be eliminated.

The data preprocessing and preparation were done before the data sets being used. It included data cleaning, data format and normalization, missing or invalid data treatment and so on. For example, the data normalization standardized the data of all the attributes into a range of [0, 1], with the formula below:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Here, x is an input data value of any attribute, x’ is the normalized value for the deep learning system.

The data used in our project was from FRED and OECD. After preprocessing, it included 740 instances, with 49 attribute of either floating-point or integer data type. The time range was monthly from January 1959 to August 2020. We split the data into a training dataset with 493 instances and a testing dataset with 247 instances (with that record number is multiple of 3). The data splitting is shown as the Figure 2.

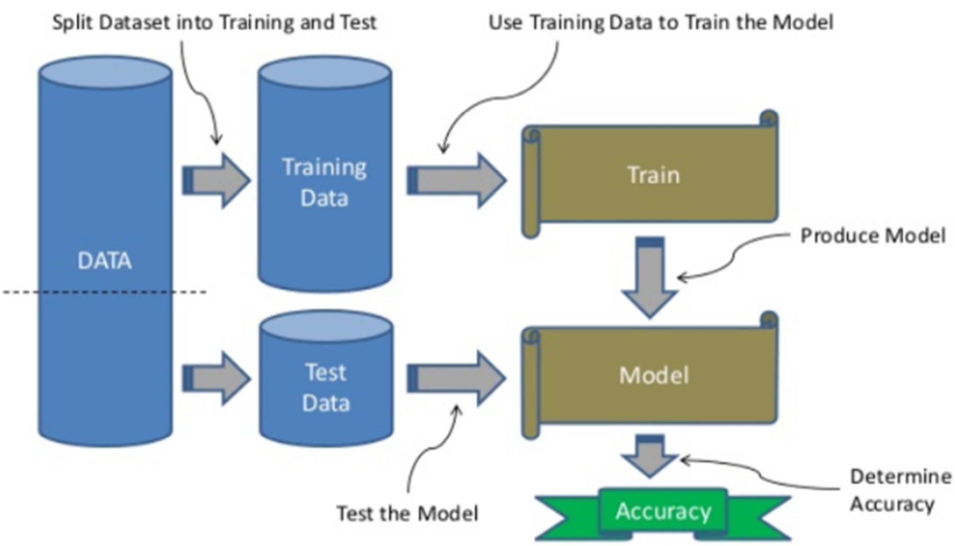

Figure 2. *Data Splitting into Training Set and Testing Set.*

3. Feature Selection and Dimensionality Reduction

Feature selection and dimensionality reduction are techniques to help reduce the burden on the computation of the model. These techniques help us create a shortlist of the most impactful attributes.

First of all, we eliminated the attributes with very low variances. Extremely low variance is an undesirable quality as it tends to not serve the deep learning model well. The function of calculating the variances generated the Figure 3 below containing the variance value for each attribute.

```
Variance Unnamed: 0            2.961000e+03
inflation                 1.552813e+01
GDP_growth                6.418480e-01
short_term_interest       1.474341e+01
long_term_interest        9.217807e+00
DisposableIncome          9.593112e+06
FedralTax                 3.267634e+05
HoursWorkedTotal          1.288689e+03
HouseholdTax              3.117463e+11
Housing_Price_Indexed     8.993326e+02
National_Debt             3.347915e+13
PPI                       5.859160e+02
GINI_Coefficient          6.294570e+00
Per_Capita_GDP            9.313907e+07
Share_Prices              1.094151e+03
```

Figure 3. *Variance Value for Attributes.*

The variance ranges from 0.64 to 3.34e+13, with GDP_growth having the lowest variance and National_Debt having the highest variance. These variance values will need to be taken into consideration when building our model. Typically attributes with higher variance tend to be better suited for training models.

Correlation was also calculated between the attributes. The project set up a framework to provide validation of the correlation analysis. It analyzed data in the brute force way (exhaustive search), run traditional correlation analysis algorithms (Pearson's r), which is slow but yields accurate results. For reducing calculation time, the system also provided an alternative of correlation analysis, which analyzed data in the predictive way, using machine learning to predict correlation metrics by p-value, which is fast but yields less accurate results. Overall there were high correlation between many of our attributes. This could possibly negatively affect the generation of some simpler models such as linear and logistic regression, but may benefit other models such as multivariate multiple regression, that was in this deep learning system for the Gini coefficient prediction. These correlation values would also need to be taken into account when evaluating the relative importance for building our model.

We applied the Random Forest Classifier to evaluate the relative importance of each attribute for Gini coefficient prediction. Random Forests use random decision trees with bootstrapping to optimize feature selection. In this study we run Random Forests with major attributes for the importance by the correlation. The results are shown in Figure 4 below:

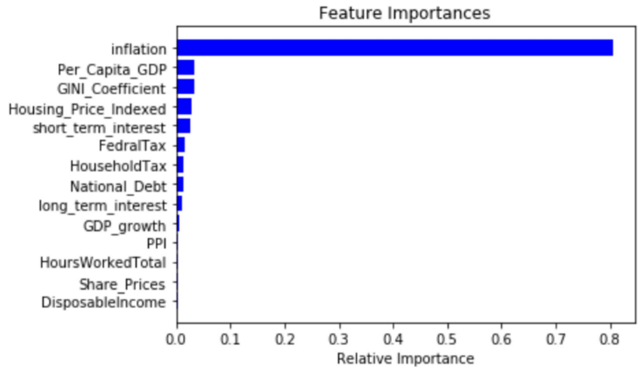

Figure 4. *Random Forest's Features Rank of Importance.*

Nineteen attributes were selected from forth-nine candidates by the function of variance and correlation analysis and the Random Forest importance evaluation. That include Consumer Price Index, Consumer Price Index Core, Producer Price Indices, Unemployment Rate, Minimum Wage, Gross Domestic Product, Disposable Income, Palma Ratio, Average Wage Index, Federal Tax Rate, Corporate Tax Rate, Cost of Living, and Gini Coefficient, etc.

Among them, Inflation and Consumer Price Index (CPI) are crucial elements affecting the Gini coefficient prediction. Inflation is measured in terms of the annual growth rate with a breakdown for food, energy and total of others excluding food and energy. It measures the erosion of living standards. Inflation and income inequality have been shown having a close relationship in this study by correlation analysis. Inflation decreases the value of currency so that all wealth that is not invested and growing will not only stagnate but steadily lose value. While the rich have resources to grow their wealth, such as the stock exchange or private investments, the poor have limited or no access to these resources. CPI is defined as the change in the prices of a basket of goods and services that are typically purchased by specific groups of households. This attribute characterizes inflation and the growth of the income inequality in terms of the US's economy. Our Deep Learning model makes prediction of the Gini coefficient based on the historical data of these attributes.

The future Gini coefficient is what the deep learning system of this project predicts so the historical data of the Gini coefficient attribute was selected even though it naturally has very low variance.

4. Datasets for Gini Coefficient Prediction

In the training datasets for Gini Coefficient prediction, the attributes should include the previous months' data as the inputs of the prediction system and the current month's Gini Coefficient as the output of the prediction system. For example, if the system is for prediction of a month's Gini coefficient (say, Month 2) by the previous one month's data (say, Month 1), the training case should include all the attributes of Month 1 as input and the Gini coefficient of the Month 2 as output, and so on. If the system is for prediction of a month's Gini coefficient (say, Month 3) by the previous two month's data (say, Month 1 and Month 2), the training case should include all the attributes of the previous two month (Month 1 and Month 2) as input and the Gini coefficient of the third month (Month 3) as output, and so on. For better prediction accuracy, the system can use longer period's data (for example, three or four months) to predict following month's Gini coefficient.

Figure 5 and Figure 6 below show the samples of the datasets using one previous month's or two previous months' economic information to predict the following month's Gini coefficient respectively. In these two figures, the index of an attribute indicates the first month or the second month of the previous months. For example, "CPI. 1" represents the Consumer Price Index of the first month and "CPI. 2" represents the Consumer Price Index of the second month of previous months.

First Month					**Prediction**
CPI.1	**PPI.1**	**Minimum Wage.1**	**……**	**Unemployment Rate.1**	**Second Month's Gini Coefficient**
221.6	17.2	7.25		4.5	0.36
220.6	17.6	7.35		4.8	0.37

Figure 5. *Apply Previous One Months' Economic Information to Predict the Next Month's Gini Coefficient.*

First Month					**Second Month**					**Prediction**
CPI.1	**PPI.1**	**Minimum Wage.1**	**……**	**Unemploy-ment Rate.1**	**CPI.2**	**PPI.2**	**Minimum Wage.2**	**……**	**Unemploy-ment Rate.2**	**Third Month's Gini Coefficient**
221.6	17.2	7.25		4.5	220.6	17.6	7.35		4.8	0.36
220.6	17.6	7.35		4.8	220.8	17.8	7.35		4.3	0.37

Figure 6. *Apply Previous Two Months' Economic Information to Predict the Third Month's Gini Coefficient.*

5. Deep Learning System for Gini Coefficient Prediction

The rebirth of AI started in the 2010's as the semiconductor industry exponentially increased computing power and lowered hardware costs. At the same time, substantial computing power was becoming more accessible to the public via cloud computing technologies. This immense power enabled the practical use of statistical and machine learning methods, as represented by the Artificial Neural Network (ANN) and later Deep Learning (DL). By providing manually-validated training datasets, practitioners had the machine learning algorithms look at the datasets and automatically "learn" their internal patterns and hidden knowledge by induction. The algorithms then used the learned knowledge (called models) to deduce results from unseen data; thereby providing a generalized learning capability. In this Gini coefficient prediction project, the approach had shown excellent performance for solving data analysis related problems [13, 14].

The deep learning system applied Multi-layer Neural Network for multiple regression. The structure of the system is shown as the Figure 7.

Here, each node represent a neuron. In this study, the hidden layer number N equal to 6, with 200 neurons per layer. For each layer, an activation function was emplyed to introduce non-linear properties to the neural network. The activation function is differentiable to perform backward propagation in the model to improve the accuracy. With the activation function, a summation of the weighted inputs will produce the output. The Figure 8 below shows how the activation function f works with the weights w:

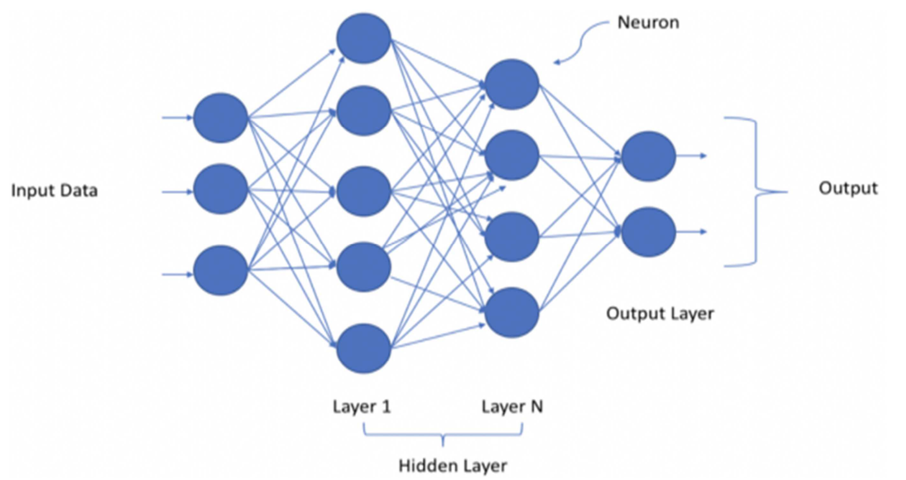

Figure 7. *The Structure of the Multi-layer Neural Network.*

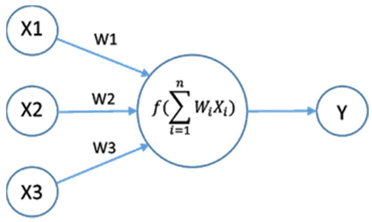

Figure 8. *The Weights W_i and the Activation Function f.*

Here, W_i is the weight of the input X_i, Y is the output. The more detailed formular below shows that with the activation function, a summation of the weighted inputs will produce the output:

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \quad (2)$$

Here, σ represent the activation function; *w* is the weight; *a* is the value of each neuron; *b* is the bias adjuster, *k* is the index number of the input neuron; *j* is the total number of the hidden layers; and *l* is the index number of the current hidden layer.

In this Gini coefficient prediction system, the Sigmoid Activation Function was emplyed to introduce non-linear properties to the DL neural network. The main reason why sigmoid function was used is because it is a non-linear function and the outputs exist between 0 and 1**.** Therefore, it is especially suitable for models where we predict the Gini coefficient as an output. Since Gini coefficient exists only in the range of 0 to 1, sigmoid was the right choice for the Gini coefficient prediction system [15, 16]. The sigmoid function is represented as:

$$F(\mathrm{x}) = \frac{1}{1+e^{-x}} \quad (3)$$

The function is continuously differentiable so that users can find the slope of the sigmoid curve at any two points. The derivative of the function is: f'(x) = 1 - sigmoid(x). Also, sigmoid function is not symmetric about zero so that the signs of all output values of neurons will be the same. The function provides a smooth gradient, so that it prevents jumps in output values. The Figure 9 below represents the S-shape of the sigmoid activation function:

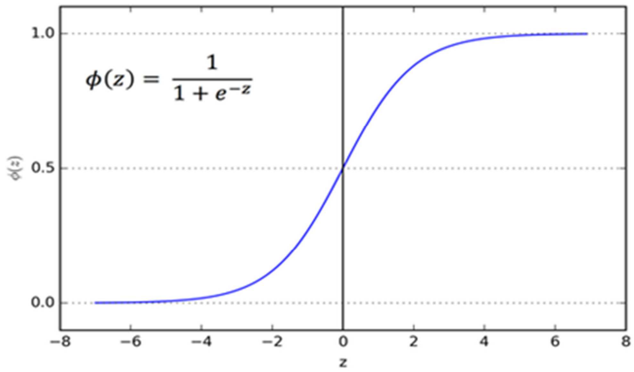

Figure 9. *The S-shape of the Sigmoid Activation Function.*

In this DL Multi-layer Neural Network system, user can adjust the number of hidden layer and the number of neurons of each layer based on computational resources, time available and the request of the prediction accuracy. Generally speaking, the more layers and more neurons, the better accuracy of the prediction, but with increased costs. In this study, according to the experiment results, we use 6 layers with 200 neurons per layer, which achieved an acceptable prediction accuracy. Further increasing the layers and neurons would not produce notable improment of the prediction accuracy in the experiments.

The DL Neural Network Regression system applied Tf. estimator. DNN Regressor [17], a TensorFlow model developing tool, for multiple regression and prediction of Gini coefficient, which are numeric values in a continuous range.

The project employed the Adam Optimization Algorithm, instead of the classical stochastic gradient descent procedure, to update the neural network weights iteratively based on the training data [18]. Applying Adam Optimization Algorithm, a learning rate is mentioned for each network. Weight is separately adapted as learning unfolds.

6. Results and Analysis

This research performed multiple regression prediction of Gini coefficient using DL Multi-layer Neural Network system. The experiment results are shown in Figure 10 below.

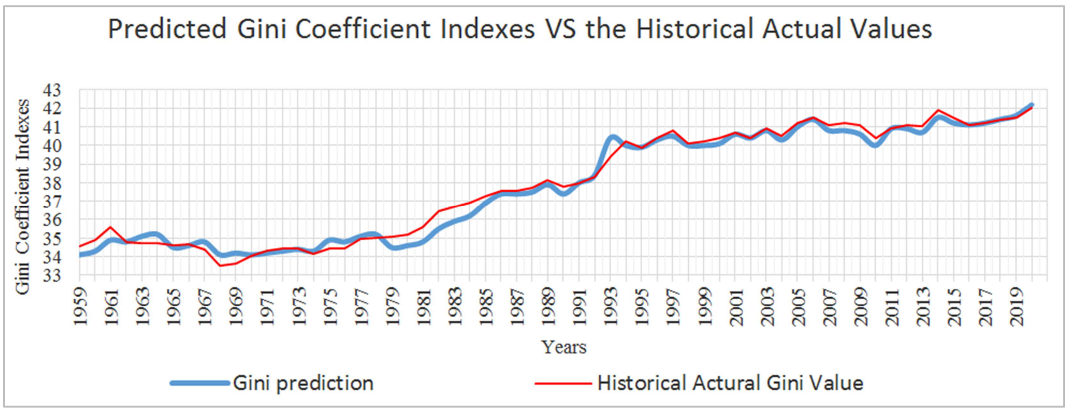

Figure 10. *Predicted Gini Coefficient Indexes VS the Historical Actual Values.*

Here, the abscissa axis is the time line. The time range was from January 1959 to August 2020, more than 61 years. For the purpose of showing clearly, the axis only shows the year numbers of every three years. The ordinate axis is the Gini coefficient index. The Gini coefficient index is the Gini coefficient expressed as a percentage, and is equal to the Gini coefficient multiplied by 100. The calculation formula is shown below:

$$\text{Gini coefficient index} = \text{Gini coefficient} \times 100 \quad (4)$$

The range of the Gini coefficient indexes shown in Figure 9 is from 33 to 43.

In Figure 9, the blue line shows the results of the predicted Gini coefficient indexes; the red line shows the historical actual values of the Gini coefficient indexes. Comparing each pair of the two values of the same time, we get Absolute Error and Relative Error of each case. Figure 9 shows the experiment results of apply previous two months' economic information to predict the third month's Gini coefficient index. The obtained average Absolute Error of the Gini coefficient index prediction was 0.29. The obtained average Relative Error of the Gini coefficient index prediction was 0.79%. The Absolute Error and the Relative Error of the prediction are defined as the formula below:

$$\text{Absolute Error} = \text{ABS (Prediction Result - Actual Value)} \quad (5)$$

$$\text{Relative Error} = \frac{\text{ABS (Prediction Result} - \text{Actual Value)}}{\text{Actual Value}} \times 100 \quad (6)$$

Figure 11 below shows the comparison of the experiment results of the Relative Errors of applying previous one month's economic information to predict the next month's Gini coefficient index, applying previous two months' economic information to predict the third month's Gini coefficient index, applying previous three months' economic information to predict the fourth month's Gini coefficient index, and applying previous four months' economic information to predict the fifth month's Gini coefficient index.

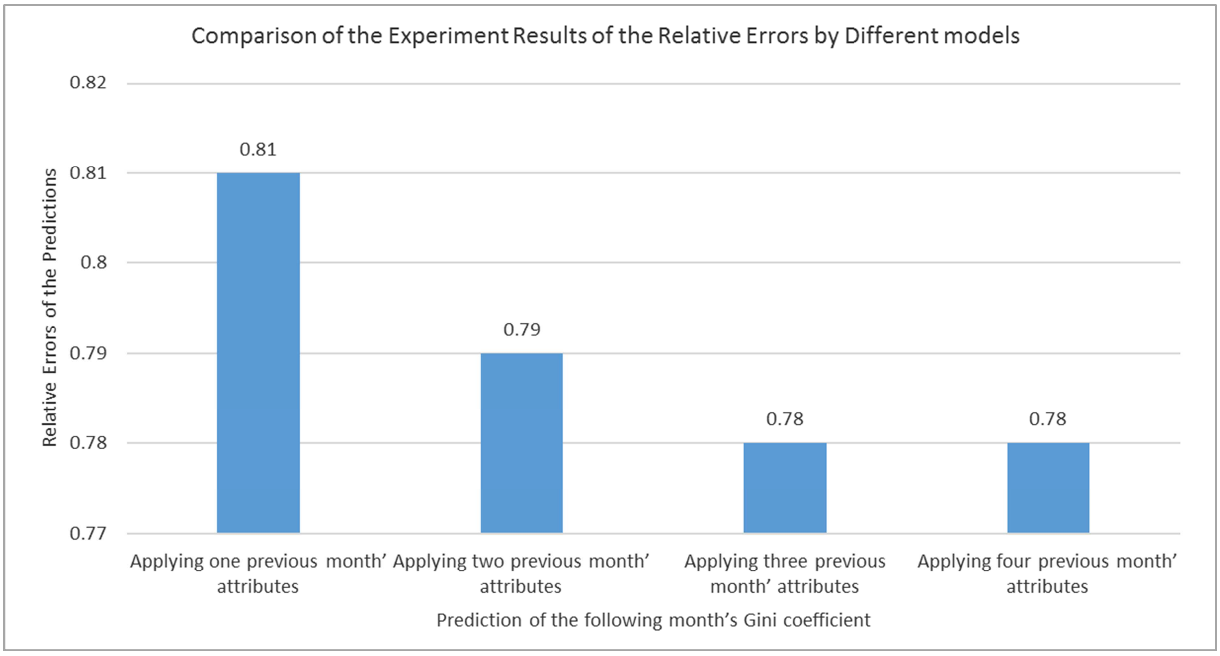

Figure 11. *Comparison of the Experiment Results of the Relative Errors by Different models.*

Here, the abscissa axis shows four different models of the Gini coefficient prediction. The ordinate axis is the Relative Error of the predictions. The experiment results showed that the Relative Error of the model of applying previous two months' economic information is lower than that of the model of applying previous one month's economic information by 0.02% (0.79% vs 0.81%); the Relative Error of the models of applying previous three months' and four months' economic information are both lower than that of the model of applying previous two month's economic information by 0.01% (0.78% vs 0.79%); the Relative Error of the model of applying previous three months' economic information is almost the same with that of the model of applying previous four month's economic information (both are 0.78%).

The experiment results indicated that using longer (more than three months) period's economic information to predict the Gini coefficient did not create meaningful improvements of the prediction accuracy, but needed more computational resources and longer time to run. Here, user of the prediction models should select a balance of the tradeoff between the prediction accuracy and the computational costs [19].

Applying Deep-learning Multi-layer Neural Network system, the Gini coefficient prediction accuracy is quite encourage [20]. In the future research work, if applying more cutting-edge technologies, such as Reinforcement Machine Learning, it is expected that the system could obtain better results.

7. Conclusion

This study combined the computer technology and the economic problem solving together, examined the gap between the rich and the poor, and observed the historical trends of income change in various economic brackets. The project applied Deep Learning technology for creating a model on income inequality and Gini coefficient in the United States. It built an intelligent system that could predict Gini coefficient in the future based on the historical relevant data, so that observe the ever widening gap between the rich and the poor. It also found the major elements that governed this widening behavior through analyzing the impacts of the related attributes. The results may help the public and the economic research society for their decision making.

Using DL algorithms, the data analysis was far more efficient, and by generating the DL multi-layer neural networks, the prediction was quite accurate. This study has obtained some promising results. It showed an encouraging direction on the prediction of Gini coefficient, through applying DL models.

References

[1] Edgar A. Ghossoub and Robert R. Reed, "Financial development, income inequality, and the redistributive effects of monetary policy", Journal of Development Economics, vol. 126, issue C, 167-189, 2017.

[2] Lisa Smith, 'The Gini Index: Measuring Income Distribution'. August, 2021.

[3] Gini, Corrado (1936). "On the Measure of Concentration with Special Reference to Income and Statistics", Colorado College Publication, General Series No. 208, 73–79.

[4] R. P. Cysne, W. L. Maldonado, and P. K. Monteiro, "Inflation and income inequality: a shopping-time approach", Journal of Development Economics, 78, pp. 516-528, 2005.

[5] Aleš Bulíř, "Income Inequality: Does Inflation Matter?" *IMF Staff Papers*, vol. 48, no. 1, pp. 139–159. 2001. JSTOR.

[6] Federal Reserve Economic Data (FRED), from Research Department at the Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/

[7] Data from Organization for Economic Cooperation and Development. https://data.oecd.org/

[8] Institute for Social Research, Survey Research Center, University of Michigan: "Panel Study for income Dynamics." https://psidonline.isr.umich.edu/default.aspx

[9] Katherine A McGonagle, Robert F Schoeni, Narayan Sastry, Vicki A Freedman, (Institute for Social Research, University of Michigan): "The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research". https://psidonline.isr.umich.edu/llcs2012.pdf

[10] U.S. population's income and economic data, from United States Census' Current Population Survey. https://www.census.gov/data.html

[11] World Bank Open Data. https://data.worldbank.org/

[12] "Trends in the Distribution of Household Income", from the US *Congressional Budget Office.* https://www.cbo.gov/publication

[13] Hubert Ooghe, Heidi Claus, Nathalie Sierens and Jan Camerlynck "International Comparison of Failure Prediction Models From Different Countries: An Empirical Analysis", September 1999.

[14] Hans Risselada, Peter C. Verhoef, Tammo H. A. Bijmolt, "Staying Power of Churn Prediction Models", Journal of Interactive Marketing, Volume 24, Issue 3, August 2010, Pages 198-208.

[15] Siddharth Sharma, Simone Sharma, Anidhya Athaiya, "ACTIVATION FUNCTIONS IN NEURAL NETWORKS", International Journal of Engineering Applied Sciences and Technology, 2020 Vol. 4, Issue 12, ISSN No. 2455-2143, Pages 310-316.

[16] Forest Agostinelli, Matthew Hoffman, Peter Sadowski, Pierre Baldi, "Learning Activation Functions to Improve Deep Neural Networks", Proceeding of the International Conference on Learning Representations (ICLR) 2015.

[17] Valerie Sarge, Shashank Verma, Ben Barsdell, James Sohn, Hao Wu, and Vartika Singh, "Improved TensorFlow 2.7 Operations for Faster Recommenders with NVIDIA", *from* NVIDIA.

[18] Imran Khan Mohd Jais, Amelia Ritahani Ismail, Syed Qamrun Nisa, "Adam Optimization Algorithm for Wide and Deep Neural Network", Journal of Knowledge Engineering and Data Science, Vol 2, No 1, 2019.

[19] Ruijian Zhang, "Applying Parallel Programming and High Performance Computing to Speed up Data Mining Processing", Proceedings of the 16th IEEE International Conference on Computer and Information Science, May 2017.

[20] Ruijian Zhang, "Applying Data Mining Technology on Inflation Prediction in the United States", IEEE Proceedings of the Eighth International Conference on Computational Science, Intelligence & Applied Informatics (CSII 2021), September 2021.