Data Analysis of Major Industries in the Country Based on Economic Indicators and Machine Learning Technology

Yinzhou Xiao[1], Meili Liu[2], Chun-Te Lee[3](✉), and Jeng-Eng Lin[4]

[1] College of Business and Public Management, Wenzhou-Kean University, Wenzhou, China
yinzhoux@kean.edu
[2] Institute of Artificial Intelligence, Media Group, Shenzhen, China
[3] Department of Mathematics, Wenzhou-Kean University, Wenzhou, China
chulee@kean.edu
[4] Department of Mathematical Sciences, George Mason University, Washington, DC, USA
jelin@gmu.edu

Abstract. In this article, we tend to prove whether some economic indicators related to the three industries are characteristics of the development status of each country. Therefore, we first classify each country into developed, moderately developed, and developing based on per capita GDP. After that, we conducted surveys on 42 countries based on three industrial indicators: the employment rate of the primary industry, the employment rate of the secondary industry, the employment rate of the tertiary industry, the proportion of agricultural added value in GDP, the proportion of industrial added value in GDP, and the proportion of industrial added value. *Classification.* The added value of the service industry accounts for GDP and agricultural production index. Based on these data, we standardize to avoid bias due to different measurements of these variables. Then, apply correlation analysis to eliminate some variables. Next, hierarchical clustering and decision trees help us find the criteria for classifying these countries into three categories. After obtaining the category, we matched the classification result with the category derived from GDP per capita, and successfully verified our hypothesis through the chi-square test. Finally, we put forward some suggestions for the development of moderately developed countries and developing countries based on our research results.

Keywords: Main Industries · Developing Country · Developed Country · Normalization · Correlation · Decision-Tree · Chi-Square Test · Policies

1 Introduction

The three industries have different behaviors in countries with different development situations. Zhang, Lin & Gong [14] point out that developing countries prefer a bank-based financial system, while the financial structure of developed countries is more dependent on the market [18]. The difference lies in the structure of the tertiary industry.

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 679–693, 2023.
https://doi.org/10.2991/978-94-6463-010-7_70

Bogoviz, Osipov, Chistyakova, and Borisov (2018) remark that developed countries have matured in Industry 4.0 and the secondary industry involves artificial intelligence, but developing countries have difficulties in policy and funding issues [1]. Notice that the situation in the agricultural industry is similar to other situations of main industry. From the article of Keskin (2018), it can be seen that compared with developing countries, precision agriculture through high-tech management of agriculture is more widely used in developed countries [11]. These differences mean that the conditions of the three industries are characteristics of different development national conditions. According to Gryshova et al. (2020), research results pointed out that the situation of the three major industries will affect the economic development and people's well-being in all countries [6]. Therefore, it is believed that economic indicators can reflect the situation of the three industries. Our goal in this article is to examine the characteristics of countries with different development status. We aim to use machine learning technology to classify countries through economic indicators related to the three major industries to show the development status and various development status of a country.

2 Sorting and Analyzing Data

2.1 Data Collating

We first selected data related to three industries in 2019 from the "China Statistical Yearbook", for a total of 42 countries. In addition, there are 7 variables, for convenience, we use symbols to mark them. The variables and their symbols are shown in Fig. 1.

According to Boyle (2021), those countries whose GDP per capita are over \$25000 are classified as developed countries by economists, but mainly developed countries have GDP per capita more than \$40000. Based on Boyle, in our paper, we classify 42 countries into 3 categories, which are developed (GDP per capita > \$40000), moderate developed (\$25000 < GDP per capita < \$40000) and developing countries (GDP per capita < \$25000). It is the category that we are going to match after we analyze on our 7 variables.

Var 1	Employment rate of primary industry
Var 2	Employment rate of secondary industry
Var 3	Employment rate of tertiary industry
Var 4	Proportion of agricultural added value in GDP
Var 5	Proportion of industrial added value in GDP
Var 6	Proportion of service industry added value in GDP
Var 7	Agricultural production index

Fig. 1. Notations of variables

Country	Var1	Var2	Var3	Var4	Var5	Var6	Var7
China	25.4	28.2	46.4	7.10	39.0	53.9	147.449550
Bangladesh	38.6	21.3	40.2	12.70	29.6	52.8	152.823379
Brunei	1.4	15.9	82.8	1.00	62.5	38.2	181.058654
Cambodia	32.3	29.0	38.7	20.70	34.2	38.8	189.127440
India	42.4	25.6	32.0	16.00	24.9	49.9	152.228645
Indonesia	28.6	22.5	48.9	12.70	38.9	44.2	150.155362
Iran	17.9	30.6	51.5	9.50	34.9	54.4	106.227031
Japan	3.4	24.3	72.3	1.20	29.1	69.3	89.081210
Kazakhstan	15.8	20.5	63.7	4.40	33.1	55.5	160.697025
Korea	4.9	25.1	70.0	1.70	33.0	56.8	105.161032
Laos	62.4	11.9	25.7	15.30	30.9	42.7	301.424495
Malaysia	10.4	27.0	62.6	7.30	37.4	54.2	125.161677
Mongolia	27.4	19.4	53.2	11.00	39.1	39.0	183.619482
Myanmar	48.9	16.1	35.0	21.40	38.0	40.7	142.953683
Pakistan	36.7	25.3	38.0	22.00	18.3	53.9	152.135794
Philippines	23.4	19.4	57.2	8.80	30.2	61.0	110.932108
Singapore	0.7	15.5	83.8	0.04	24.5	70.4	119.294839

Fig. 2. Snapshot of part of the original data

It is known that countries with a per capita GDP of more than US$25,000 belong to developed countries, but Boyle (2021) infers that the per capita GDP of developed countries should exceed US$40,000 [2]. Based on the analysis of [2], we analyze 42 countries and divide these countries into 3 categories, namely the developed countries (per capita GDP $>$ 40,000 US dollars), moderately developed countries (25,000 US dollars $<$ per capita GDP $<$ 40,000 US dollars) and Developing countries (per capita GDP $<$ 25,000 USD). In these categories, we selected and analyze seven important feature variables regarding the development of a country. Notice that However, there are missing data in the agricultural production index. From the 2020 China Statistical Yearbook, we only obtained data from year 2014 to 2016. In order to fill in the missing data in 2019, we did research in the 2015–2020 yearbook and obtained the 2011–2016 agricultural production index. Then the gray model is used to predict the index in 2019. After integrating the data for 2019, we import the excel file containing the original data into the Jupyter notebook, as shown in Fig. 2.

2.2 Data Processing and Mapping

By importing the data, we found that these 7 variables have different dimensions. According to Daisy (2017), it is seen that different dimensions will influence the final result of data analysis, so it requires normalization [9]. Therefore, we started to choose the method for normalization. The first method we tried was Z-score normalization. According to Linda (2018), Z-score normalization requires that the original data distribute as normal distribution. Therefore, we were supposed to test whether the original data meet

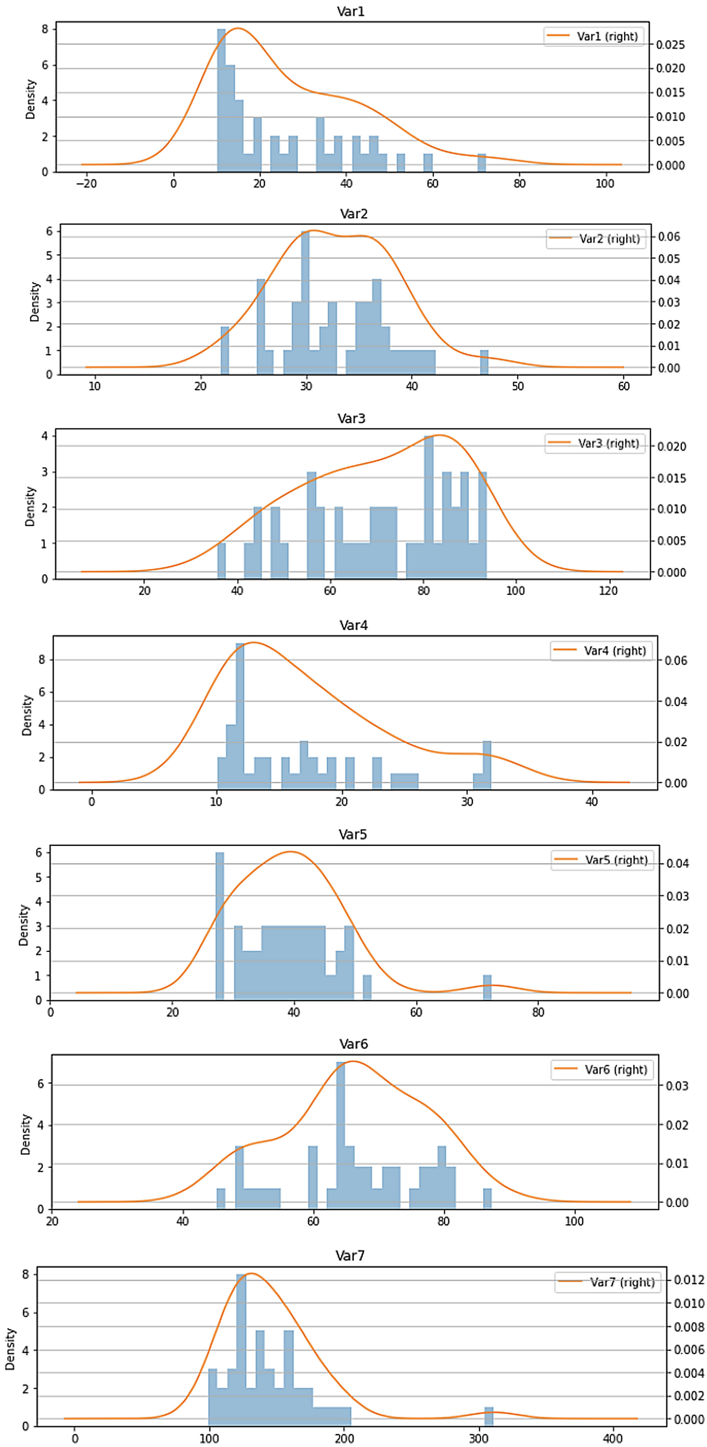

Fig. 3. Check for normal distribution.

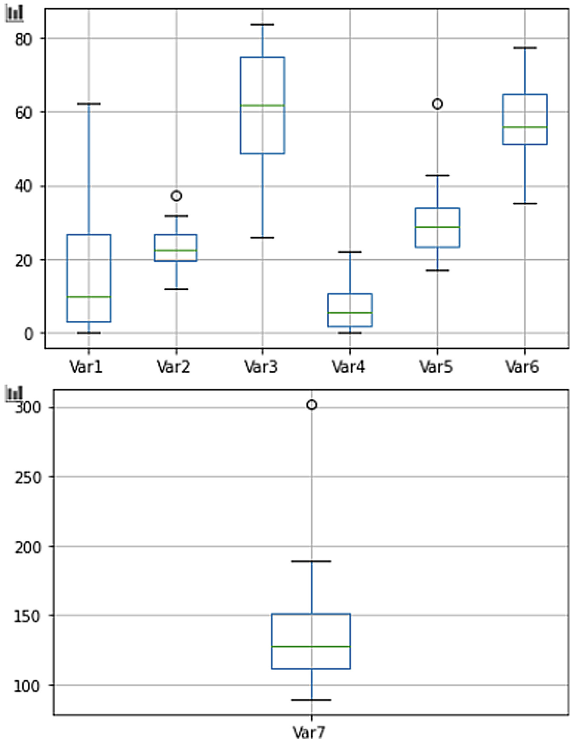

Fig. 4. Check for outliers.

the requirement. The results that we did normalization for every variable are shown as follows:

From Fig. 3, we observe that the original data of all variables do not follow a normal distribution. Therefore, it is impossible to apply Z-score normalization to the data. It is suggested in [9] that min-max normalization is applied, but min-max normalization can be performed efficiently under the requirement of no outliers in the data. Therefore, we need to check whether our data has extreme values. The result is shown in Fig. 4.

Seen from the Fig. 4 that there are extreme values for every variable, therefore, min-max normalization is not appropriate here; [7]. According to Santhakumaran (2011), when there are extreme data, we could apply non-linear normalization, such as sigmoid normalization which processes as Eq. (1). However, the results are not satisfied because most data after normalization are close to 1, which means there are no big difference between data. Therefore, another non-linear normalization is applied, which is log normalization which processes as Eq. (2).

$$\mathrm{x}' = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{1}$$

$$\mathrm{x}' = \frac{\log_{10} x}{\log_{10} max} \tag{2}$$

Country	Var1	Var2	Var3	Var4	Var5	Var6	Var7
China	0.78256	0.92272	0.86651	0.63412	0.88595	0.9168	0.87474
Bangladesh	0.8838	0.84518	0.83413	0.82225	0.81926	0.91205	0.88101
Brunei	0.0814	0.76439	0.99729	0	1	0.83763	0.91071
Cambodia	0.84069	0.93045	0.82554	0.9803	0.85419	0.84121	0.91835
India	0.90652	0.89599	0.78261	0.89698	0.77745	0.89906	0.88033
Indonesia	0.81126	0.86033	0.87836	0.82225	0.88533	0.87117	0.87793
Iran	0.6979	0.94529	0.89006	0.72833	0.85909	0.91892	0.8173
Japan	0.29606	0.88159	0.96667	0.05898	0.81514	0.97458	0.78646
Kazakhstan	0.66771	0.8346	0.93807	0.47932	0.84629	0.92352	0.88981
Korea	0.38447	0.89054	0.95937	0.17167	0.84555	0.92885	0.81553
Laos	1	0.68432	0.7331	0.8825	0.82965	0.86324	1
Malaysia	0.56653	0.91071	0.93414	0.64311	0.87582	0.91807	0.84603
Mongolia	0.80089	0.81936	0.8974	0.77576	0.88657	0.84239	0.91317
Myanmar	0.94102	0.76784	0.80285	0.99105	0.87967	0.8522	0.86932
Pakistan	0.87159	0.89274	0.82142	1	0.70297	0.9168	0.88022
Philippines	0.76272	0.81936	0.91377	0.70357	0.82411	0.94525	0.82489
Singapore	-0.08629	0.75735	1	-1.04136	0.77353	0.9782	0.83762

Fig. 5. Part of normalized data

	Var1	Var2	Var3	Var4	Var5	Var6	Var7
Var1	1.000000	0.102036	-0.830280	0.764683	0.361917	-0.519312	0.384108
Var2	0.102036	1.000000	-0.117788	0.106600	0.103882	0.072356	-0.244377
Var3	-0.830280	-0.117788	1.000000	-0.744533	-0.307699	0.566869	-0.554189
Var4	0.764683	0.106600	-0.744533	1.000000	0.269565	-0.599834	0.453529
Var5	0.361917	0.103882	-0.307699	0.269565	1.000000	-0.750921	0.256285
Var6	-0.519312	0.072356	0.566869	-0.599834	-0.750921	1.000000	-0.570165
Var7	0.384108	-0.244377	-0.554189	0.453529	0.256285	-0.570165	1.000000

Fig. 6. Correlation between variables

Finally, the normalized results are satisfied, and the part of data imported in excel file is as shown in Fig. 5.

After normalization, we can start the analysis. The first step we need to do is to test for correlation, because there may be correlations between variables. In addition, according to Sympa, a major in statistics, when the correlation between two variables is close to -1 or 1, one of them can be omitted. The relevant results are shown in the Fig. 6.

From the results, it clearly displays that the pairs of variables have high correlation with each other are Var1 and Var3, Var3 and Var4, Var5 and Var6. Finally, we determined to choose Var2, Var3, Var5 and Var7 as variables of our project.

3 Data Processng by Machine Learning

3.1 Data Visualization and Hierarchical Clustering Method

Having selected variables that we tend to analyze. We initially add one column into excel file which contains the normalized data. The column shows the classification of countries based on the GDP per capita which classifies the countries into developed, moderately developed and developing as we mentioned previously. In Fig. 7, we show the definition of the developed, moderately developed and developing countries.

The purpose of adding such information in the excel file is to match our decision tree classification with the per capita GDP classification as much as possible. This is similar to the situation where we want to figure out the rules of the game. What we have to do is to make the rules we predict reach the real rules, and the real rules are the benchmark in the process of exploration. The respective numbers of each category of country are as shown in Fig. 8.

The classification method we chose is hierarchical clustering, because compared with other machine learning methods such as KNN and K-Means, hierarchical clustering can clearly show which countries belong to a category. KNN and K-Means can visually display classification when processing two-dimensional or three-dimensional data. Once we have four variables, we cannot clearly determine the category of the country. In addition, the hierarchical clustering also generates a heat map that describes the correlation between the country and our 4 variables. The clustering results and related graphs are shown in Figs. 9 and 10.

According to the results of hierarchical clustering, the 42 countries could be classified into 3 categories as well, and we define these categories as A, B and C, following the appearing order in Fig. 9 as yellow, green and red respectively. To sum the result up, the chart is provided in Fig. 11.

In Fig. 12, we show the classification results of per capita GDP prosperity.

From Fig. 12, it can be inferred from Fig. 12 that C represents category 3, namely developed countries. Because of their equal numbers, only two countries classified as C belong to the first and second categories in the hierarchical cluster. In category A, one belongs to category 2 and one belongs to category 1. In category B, 11 belong to category 1. Therefore, we can suspect that C represents category 3 (developed countries), but A and B contain more category 1 than other categories. Therefore, we cannot even come up with hypotheses to test whether A, B, C and 1, 2, and 3 have a one-to-one correspondence.

1	Developing countries
2	Moderately developed countries
3	Developed countries

Fig. 7. Notation of different countries

1	Developing countries	28
2	Moderately developed countries	4
3	Developed countries	10

Fig. 8. The amount of different countries

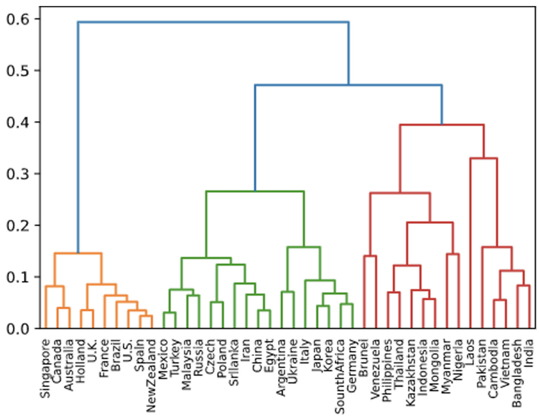

Fig. 9. Diagram of hierarchical clustering

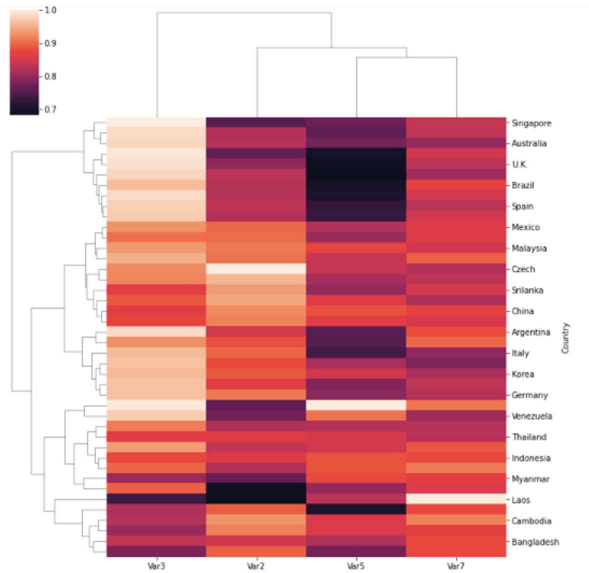

Fig. 10. Diagram of correlation between countries and variables

In addition, despite the heat map, we cannot know the exact classification criteria. We can observe that Var 7 and Var 5 have a relatively large impact on country classification. Countries perform differently on each variable, but we don't know the exact criteria for classification, which means we don't know how different behaviors of variables affect The category of the country.

A	15
B	15
C	10

Fig. 11. The results of classification via hierarchical clustering

1	28	A	15
2	4	B	15
3	10	C	10

Fig. 12. The comparison of results via two classifications

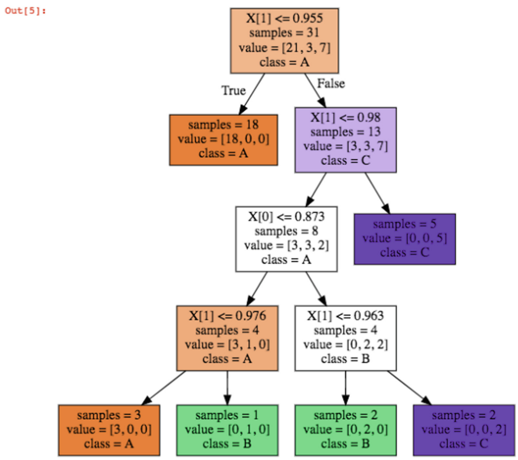

Fig. 13. The classification results of decision tree

3.2 Applying Decision Tree and Analyzing the Results

In order to obtain more accurate classification and its standards. We choose another machine learning method, the decision tree. The decision tree can visually display the specific criteria of the classification in numerical form. When applying the decision tree, we only imported 4 of the 7 original variables for analysis. The results of the decision tree classification are shown in Fig. 13.

In decision tree, there are X[0], X[1], X[2] and X[3]. When python deal with the variables, the first one will be marked as index 0, therefore, X[0], X[1], X[2] and X[3] represent for Var 2, Var 3, Var 5 and Var 7 respectively. In addition, we could figure out the most important factor from the diagram of decision tree. According to Learn by

1	28	A	28
2	4	B	3
3	10	C	11

Fig. 14. The comparison of results via two classifications

Observed	10	4	28
Expected	11	3	28

Fig. 15. The comparison chart for Chi-square test

Marketing, the most crucial factor appears at the top of decision tree. Obviously, the most important one in our classification is X[1], which is the Var 3 (employment rate of tertiary industry).

Next, to make sure the accuracy of our classification, we filter obtain the categories of 42 countries in excel based on the criteria of the decision tree. Also, we define these categories as A, B and C The results of matching are as shown in Fig. 14:

To make sure A represents 1, we have to compare whether countries that A include are countries included in 1, so as B and C. Fortunately, most countries belong to A are included in 1, only one country comes from category 3; and countries in B are all in category 2; and most countries classified as C are in category 1, only one country belongs to category 2. Therefore, the classification of decision tree is perfect, we could infer that classification from decision tree could be matched with the result from GDP per capita. However, for analysis purposes, we choose some statistics to prove accuracy. In [12], the recommended chi-square is a non-parametric test because it can effectively test the relationship between two variables. Therefore, in our project, we want to test whether the conditions of these three industries are related to the development of countries around the world. Chi-square is appropriate here, so the reason for applying the chi-square test is. Initially, we set the following assumptions:

H_0: we could classify developing, moderately developed and developed countries based on data related to three industries.
H_1: data related to three industries cannot be regarded as features to classify countries as developing, moderately developed and developed.

When processing the Chi-square test, we have concluded the results in Fig. 15.

In addition, since the total number of countries is 42 and the number of categories is 3, when the number of 2 categories is determined, the number of remaining categories is fixed. Therefore, we only have 2 degrees of freedom, and we have chosen a 95% confidence interval. Finally, the value of Chi-square is only 0.42, and the critical value is 5.99. When the critical value is greater than the score of chi-test, we do not reject

H0. Therefore, H0 is accepted, which means that the situation of the tertiary industry (employment rate of the secondary industry, employment rate of the tertiary industry, the proportion of industrial added value in GDP, agricultural production index) is related to the development of various aspects. Country.

4 Policies Based on the Standard of Classification

4.1 Analyze the Standard of Classification via Decision Tree

First, we can conclude from the decision tree that Var 3 (employment rate in the tertiary industry) is the most important variable. In (Learn by Market), the factors at the top of the decision tree are the most important. In the results of our project, the variable at the top of the decision tree is Var 3 (employment rate in the tertiary industry), so it is the most important. In addition, Var2 (employment rate in the secondary industry) is also important because the decision tree only classifies countries based on Var2 (employment rate in the secondary industry) and Var3 (employment rate in the tertiary industry).

From the results of decision tree, we could transform the normalized date to original ones and conclude the features of developed countries, moderately developed countries and developing countries on three industries as follows:

$$\begin{cases} \textbf{Developed Countries}: \\ \text{Var}\,3 > 76.7\ \% \\ \text{or} \\ 71.14\% < \text{Var}\,3 < 76.7\%\ \text{and}\ \text{Var}\,2 > 23.56\% \end{cases}$$

$$\begin{cases} \textbf{Moderately Developed Countries}: \\ 68.66\% < \text{Var}\,3 \le 71.14\%\ \text{and}\ \text{Var}\,2 > 23.56\% \\ \text{or} \\ 75.35\% < \text{Var}\,3 \le 76.7\%\ \text{and}\ \text{Var}\,2 \le 23.56\% \end{cases}$$

$$\begin{cases} \textbf{Developed Countries}: \\ \text{Var}\,3 > 68.66\ \% \\ \text{or} \\ 68.66\% < \text{Var}\,3 \le 75.35\%\ \text{and}\ \text{Var}\,2 \le 23.56\% \end{cases}$$

4.2 Policies for Developing Countries and Moderately Developed Countries

On the basis of the classification standard, we have also proposed some more developed policies for moderately developed countries and developing countries. For developing countries, our policy is first to help them achieve appropriate development; we tend to help moderate countries become developed countries.

It can be seen that there are two types of developing countries. One is that the employment rate of the tertiary industry is lower than 68.66%, the second is that the employment rate of the tertiary industry is between 68.66% and 75.35%, and the employment rate of the secondary industry is less than or equal to 23.56%. The comparison between the two types of developing countries and the two types of moderately developed countries is as follows:

	Developing Countries		Moderately Developed Countries	
	1	2	1	2
Var2		<=23.56%	>23.56%	<=23.56%
Var3	<=68.66%	(68.66%, 75.35%]	(68.66%, 71.14%]	(75.35%, 76.7%]

From the table shown above, we could see clearly that the first category of developing countries has two situations: employment rate of the secondary industry is more than 23.6% and less than 23.6%. In [5], a strong secondary industry is considered to be the key to the development of the tertiary industry. Therefore, in the first category of developing countries (the employment rate of the tertiary industry is less than 68.66%), the developing countries with the employment rate of the secondary industry below 23.6% can initially increase the employment rate of the secondary industry by 23.56%, and then develop the third industry. Industrial employment. Finally, their status will be the same as that of the first category of moderately developed countries. Literature [3] pointed out that technological innovation is an important factor to promote the development of the secondary industry. The research results of [10] show that the employment elasticity coefficient of the secondary industry can reflect the development of the secondary industry and increase the employment of the secondary industry. Therefore, considering countries with a weak secondary industry base, we believe that developing first and innovating first is not a good choice. It can be seen in [13] that the educational factor plays a very important role in the development of the secondary industry. For such developing countries, the policy we suggested is that national governments should consider to invest more budgets on education, especially the science and engineering education that are related to innovation.

Note that in our research results, developing countries (the employment rate in the tertiary industry is less than 68.66%) belong to the first category, and the employment rate in the secondary industry exceeds 23.6%, indicating a strong industrial foundation, because the secondary industry is basically the same as some moderately developed countries and some developed countries. The focus is to increase the employment rate of the tertiary industry to more than 68.66%. The research results in [4] pointed out that technological development is a key factor in the development of the tertiary industry. In [10], the author believes that the huge employment elasticity coefficient of the tertiary industry indicates that the development of the tertiary industry will greatly increase employment in this industry. Therefore, our research results indicate that countries with strong industrial foundations should continue to import high-innovation technologies from developed countries and implement these technologies on their other parts of strong industrial foundations. In addition, these countries may consider setting up specialized departments for technological exploration to meet their own needs as much as possible.

The strategies of the second category of developing countries are the same as those of the former category. For middle developed countries, we also compare these countries with developed countries, and the results are as follows:

	Developing Countries		Moderately Developed Countries	
	1	2	1	2
Var2	>23.56%	<=23.56%		>23.56%
Var3	68.66%, 71.14%	(75.35%, 76.7%]	>76.7%	(71.14%, 76.7%]

From a comparative point of view, the employment rate of the tertiary industry of the first category in the middle developed countries (the employment rate of the secondary industry is more than 23.56%) should be increased to more than 71.14%. It can be seen in [4] that the factors for the development of the tertiary industry are globalization, technology and deregulation. The employment rate in the secondary industry is in the same range as in developed countries. Countries belonging to the first category of moderately developed countries have a solid foundation. Belongs to the secondary industry. In [13], the author shows that with the influence of education on the secondary industry, strong industries reflect high-quality education. Therefore, with a relatively strong economic, educational, and industrial foundation, the development direction of technology will be different from the strategy of developing countries. The policies we suggested are (1) the government should invest more on construction of research environment, such as laboratory and research budgets for innovation; (2) the government may consider setting less strict regulations on developing technology and innovation, such as permission to some materials.

For the second category of moderately developed countries (the employment rate in the secondary industry is less than 23.56%), it can be more developed by increasing the employment rate in the secondary industry to more than 23.56%. The policies of these countries are similar to those of the second type of developing countries, but there are differences due to their economic strength. The policy we suggested are as follows: (1) the government should import high innovation technology from developed countries, and realize these technologies domestically relying on the strong industrial basis; (2) The government may consider to invest more on construction of research environment, such as laboratory and research budgets for probing the more advanced industrial technology from abroad.

5 Conclusion

From our machine learning modeling, we can draw conclusions from the decision tree and chi-square that the conditions of these three industries can be characteristics of developed, moderately developed, and developing countries. The result of the decision tree also tells us that among the seven indicators of our project, the employment rate in the tertiary industry and the employment rate in the secondary industry are the most significant characteristics of countries with different development status, because these two variables are unique. The decision tree relies on two criteria when dealing with the classification of 42 countries. In addition, we compared these most important factors between developing countries and moderately developed countries, and the relationships between moderately developed countries and developed countries. Based on the status quo of developing countries and the results of our projects, we proposed an overall strategies for developing countries to develop into and help moderately developed countries.

References

1. Bogoviz AV, Osipov VS, Chistyakova MK, Borisov MY (2019) Comparative analysis of formation of industry 4.0 in developed and developing countries. In: Popkova E, Ragulina Y, Bogoviz A (eds) Industry 4.0: Industrial Revolution of the 21st Century. Studies in Systems, Decision and Control, vol 169. Springer, Cham. Accessed 06 May 2021. https://doi.org/10.1007/978-3-319-94310-7_15
2. Boyle (2021) Top 25 Developed and Developing Countries, 25 April 2021. https://www.investopedia.com/updates/top-developing-countries/. Accessed 10 May 2021
3. BusinessEx Staff (2019) Here're the Technological Innovations that Manufacturing Industry Should Adopt, 24 April 2019. https://www.entrepreneur.com/article/332720. Accessed 6 June
4. Dzhain (2012) Development of the Tertiary Sector A Comparative Analysis Between BRIC Countries and Select Developed Economies, 01 May 2012. https://www.theseus.fi/bitstream/handle/10024/46284/Dzhain.Nikita.pdf;jsessionid=8F7030F6160A62D9D110FCE3817EEE5F?sequence=1. Accessed 6 June
5. Ding (2017) Industrial prosperity is still the cornerstone of service industry development, 04 May 2017. http://www.360doc.com/content/17/0504/17/363711_651076850.shtml. Accessed 6 June 6
6. Gryshova (2020) Assessment of the Industrial Structure and its Influence on Sustainable Economic Development and Quality of Life of the Population of Different World Countries, 15 January 2020. Accessed 06 May 2021
7. Jayalakshmi, Santhakumaran (2011) Statistical Normalization and Back Propagation for Classification, February 2011. http://www.ijcte.org/papers/288-L052.pdf. Accessed 12 May
8. Learn by Market (n.d.) Tree Based Variable Importance. http://www.learnbymarketing.com/603/variable-importance/. Accessed 10 May
9. Linda (2018) The method and function of normalization, 22 November 2018. https://blog.csdn.net/qq_23100417/article/details/84347475. Accessed 12 May
10. Liu (2020) Analysis of the influence of the development of the three industries on the employment rate, 09 October 2020. https://wenku.baidu.com/view/277c61f303020740be1e650e52ea551811a6c901.html. Accessed 06 June
11. Say K, Sehri (2018) Adoption of precision agriculture technologies in developed and developing countries, January 2018. http://tojsat.net/journals/tojsat/volumes/tojsat-volume08-i01.pdf#page=16. Accessed 06 May 2021
12. Xia (2018) Chi- Square Test, 16 August 2018. https://blog.csdn.net/ludan_xia/article/details/81737669. Accessed 17 May
13. Yao (2017) The influence of technological progress and educational development on industrial structure. https://xueshu.baidu.com/usercenter/paper/show?paperid=135e0gu0r75d0ev0857f0xg0xq252299&site=xueshu_se. Accessed 06 June
14. Zhang L, Gong (2018) Financial structure, industrial structure, and economic development: a new structural economics perspective, 05 September 2018. https://onlinelibrary.wiley.com. Accessed 06 May 2021. https://doi.org/10.1111/manc.12240