Research on Educational Donation Mechanism Based on Data Mining and Economics

Yu Zhou[1], Xuesong Zhang[2], and Wei Zhou[3]

[1]College of Mathematics and Systems Science, Shandong University of science and technology, Qingdao, Shandong, 266590, China

[2]Mathematics School, Jilin University, Changchun, Jilin, 130012, China

[3]Liren College of Yanshan University, Department of Mechanical Engineering, Qinhuangdao, Hebei, 066004, China

Keywords: pac, re, rgr, economics, dea cost-effectiveness analysis model

Abstract: First, preprocess the data. Perform data screening based on the completeness and redundancy of the information, delete data with less than the threshold, and use pca to combine indicators of different attributes. Normalize the retained school indicators. Determine the weight between the indicators. Calculate the school's score and allocate funds according to the school's score. Second, define the index for evaluating investment benefits, and reflect the changes in growth efficiency re and growth efficiency rate rgr to reflect the cost-related benefits. rgr determines the investment period as 7 years. Starting from the cost-effectiveness of the investment, evaluate and analyze the effectiveness of the investment, establish a dea cost-effectiveness analysis model, and determine the school with the best unit return on investment. Consider the subjective processing in data preprocessing. The corresponding results are compared with the original results to illustrate the rationality of the subjective treatment.

1. Background

The significance of existence in higher education is no doubt that with the development of society, especially skilled employees can complete complex tasks, and more talents are required to be competent for scientific research. The proportion of the United States in global higher education is only for the eleventh, this makes financial support for large charities necessary, and investment in education is also responsible for the country's future.

In this context, in order to improve the educational performance of U.S. undergraduates studying in colleges and universities, the goodgrant foundation has decided to invest in eligible schools. Therefore, the foundation needs to determine the schools to invest in and the investment amount of each school. Understand the return of this investment and the time required for continuous investment. In order to accurately obtain these results, you need to build a reasonable model based on the relevant data given, extract the valid part of the data and evaluate the sample to determine the investment school and amount we considered the return of the school in academic and economic perspectives, and then determined the investment time needed to invest. Based on our model, we made reasonable suggestions for the foundation in the letter.

In recent years, big data has become increasingly popular, and many economic fields, including philanthropy, need big data to guide it. In this article, we use data mining methods to process data, and build an evaluation system based on principal component analysis and RE, the RGR investment term model has successfully determined the optimal investment strategy for the Goodgrant Foundation. Based on this, the investment strategy is analyzed based on the DEA cost-performance analysis model combined with economics.

William Bernbach proposed a new indicator called "return on investment". All the economic returns from an investment activity reflect the value of the investment. On this basis, add academic returns to fully reflect the foundation the value of investment. The American consulting group proposed the concept of pac, which can be programmed as an automation controller. In the early

stage, people did not know much about the new technology of pac. In order to quickly increase the popularity of pac technology, manufacturers gave it a slogan. This combination the slogan of known technology is continuously spreading to more and more researchers, and has attracted the attention of many related technical personnel in the field of industrial control. However, this misleading interpretation of pac is only a gimmick that attracts people's attention. It can't fully explain the true role of pac, there is still a gap with the exact meaning of pac, and it will also cause those skilled in the field of ipc and plc to have unnecessary doubts about pac.

To be precise, pac is a comprehensive product that combines the advantages of both IPC and PLC. Involving the fields spans the two. In the remaining indicators, in order to further select several indicators that are important for the decision of the investment amount, the principal component analysis method is used to extract compliance. The index required for the contribution rate, which is a key indicator for evaluating potential schools. If you want to understand the relationship between these three, it is very easy, just like the relationship between suv, off-road vehicle and car in the automotive industry. plc is similar to suv that combines the outstanding road driving ability of off-road vehicles and the comfort and fastness of car driving. PAC not only inherits the advantages of strong processing power and stable control of programmable logic controller plc, but also has complex IPC processing. The advantages of calculation.pac as a balance product of the performance of the two is by no means a simple "1 + 1", which incorporates many innovative ideas. Pac naturally cannot match the response speed of plc's interface, nor can it beat the computing power of ipc. Neutralization among the participants, pac provides a good solution to the problem for modern control methods.

Pac can be called a leading-edge integrated product in the field of automation. In the process of technological improvement, many similar products have emerged, such as integrated inverters and other products. Although their appearance cannot shake the industrial automation industry, which has been formed for many years, Pattern, but it can provide corresponding help for many users with related needs, and also makes people think about the future improvement of automation products. Nowadays, the development of networks and information is remarkable, and industrial automation will definitely become the future development direction. Such a background trend is bound to prompt automation products to make relevant improvements, integrate computer technology as much as possible, and effectively combine the IT industry with the automation industry. This will inevitably accelerate the integration of automatic control and information management in practical applications. In order to achieve such future development goals, the design of automation products must gradually develop in the direction of intelligent networking.

The combination of cultivating self-sustaining behavior and pac personality structure theory can become a powerful tool to enhance the effect of interpersonal communication. Their common goal is to make employees feel "good" to themselves and others. At the same time, they also ensure the use of more adult-based Mindset problem solving methods. As a result, they improve communication and interpersonal cooperation. Although they can be implemented by individuals, these tools will be most effective when they are widely used throughout the organization and supported by senior management. At the same time, they create important foundations for solving the more complex challenges faced by small groups and people working on committees.

The emergence of pac products is very timely. Its internal operating platform has strong compatibility, good versatility, and a design system with free programming by the operator, which can fully meet the different needs of users, which in turn will inspire far more satisfaction than ever. Improved decision-making required by modern automation. At present, domestic research on pac technology is still in its infancy, compared with fewer users using pac abroad, but with the continuous development of the domestic industry, pac technology will also be rapidly obtained in China development of.

2. Establishment of Optimal Investment Strategy

According to the literature [2], we classify all attribute values and divide them into eight categories of first-level indicators, which are college characteristics, degree granting rate, student

retention rate, student registration, student funding, family income, and student graduation Income, student debt repayment price, there are secondary indicators under the primary indicator.Taking the characteristics of the institution as an example, the secondary indicator is the Control of institution.

First, we preprocess the data. We perform data screening based on the completeness and redundancy of the information, delete the data with less than the threshold, and use pca to merge the indicators of different attributes. For normalization of the retained school indicators, we the weights between the indicators are determined according to the pca. Then, we calculate the school's score, select the school with the priority investment, and allocate funds based on the school's score.

Secondly, we define indicators for evaluating investment benefits, and determine the value of the indicators by collecting data to build an investment benefit evaluation model. Based on this, we can obtain the change trend of growth efficiency re and growth efficiency rate rgr. These two indicators reflect the cost-related Income, which is an important factor that charitable organizations should consider. Therefore, the investment period is determined to be 7 years based on re and rgr.

The third is to start from the aspect of investment cost-effectiveness, evaluate and analyze the effectiveness of the investment. To this end, we have established a DEA cost-effectiveness analysis model to determine the school with the best degree of unit investment return. Determine the 10 schools for investment as an example. It can be determined that ArizonaStateUniversity-Tempe can make the most effective use of limited financial investments and has the highest cost performance, so its unit investment return is better than other schools.

The foundation needs to determine the schools to invest in and the amount of investment for each school, and it needs to understand the return on this investment and the time required to continue investing. In order to accurately obtain these results, it is necessary to establish a reasonable model based on the relevant data given, Extract the valid part of the data and evaluate the sample to determine the investment school and amount. We consider the return of the school in various aspects of academic and economic, and then determine the investment time that needs to be invested. Based on our model, we are the foundation Proposed reasonable suggestions.

2.1 Data normalization

(1) For indicators with larger data values reflecting their worse attributes, normalize the data using the following method:

$$b_{ij} = 1 - a_{ij} / a_{jmax} \tag{1}$$

(2) For such indicators that the larger the data value reflects the better attributes and the existence of uncertain factors such as customs and habits, the data is normalized by the following methods:

$$b_{ij} = \frac{a_{ij} - a_{ij}^{min}}{a_{ij}^{max} - a_{ij}^{min}} \tag{2}$$

2.2 Get indicator weight

• Calculate the correlation coefficient matrix r:

$$R_{ij} = (r_{ij})_{m\times m},\ r_{ij} = \frac{\sum_{k=1}^{n} a_{ki}a_{kj}}{n-1},\ i,j = 1,2,...,m \tag{3}$$

• Calculate eigenvalues and eigenvectors.

Calculate the eigenvalue of the correlation coefficient r and the corresponding eigenvector and each eigenvalue λ_u corresponds to a unique eigenvector $\gamma_u = \left[\gamma_{1u}, \gamma_{2u}, \gamma_{3u}, ..., \gamma_{mu}\right]^T$. A new main component is composed of the eigenvectors:

$$y_x = \gamma_{1x}a_1 + \gamma_{2x}a_2 + ... + \gamma_{mm}a_m \tag{4}$$

• Extract the main ingredients.

Calculate the values of the main component $y_1, y_2, y_3, ..., y_p$ cumulative contribution rate α_p. When α_p is close to 1, select the first p the indicator variable replaces all the original indicator

variables for analysis. The formula for calculating the contribution rate is

$$\alpha_p = \frac{\sum_{k=1}^{p} \gamma_k}{\sum_{k=1}^{m} \gamma_k} \tag{5}$$

• Calculate the score of each sample and comprehensively evaluate the sample.

Define z as the comprehensive score $b_j = \frac{\gamma_j}{\sum_{k=1}^{m} \gamma_k}$ as the contribution rate of the jth principal component, then the score of the comprehensive score is

$$Z = \sum_{j=1}^{p} b_j y_j \tag{6}$$

3. Priority investment selection and funding allocation of potential schools

In order to get the indicators for evaluating the recommended schools, we provide the following steps:

Step1: For the secondary indicators, we use the principal component analysis method to select all the major components whose contribution rate exceeds μ_1 and calculate their corresponding weights.

Step2: Using the principal component analysis method again, select the primary component whose cumulative contribution rate exceeds μ_2 in the primary index, and according to its corresponding weight and the new principal component (y_1, y_2... y_x), calculate the comprehensive score.

After the above two steps of processing, a total of 8 first-level indicators with a cumulative contribution rate of 100.00% after the index is extracted through the principal components are used as evaluation indicators for potential schools.Because the contributions of several indicators with lower contribution rates are similar, the comparison is not removed. Index.

Through analysis, it can be found that the first-level indicators of college characteristics, degree award rate, and student retention rate have a greater weight in potential schools, while indicators such as student family income have a smaller weight in potential schools.

Substitute the sample data according to formula (6) to get the scores of each potential school, and get the priority investment school based on the score. The score of each potential school is used as a weight to determine the investment amount of each potential school. The details are as follows:

Table.1. Priority investment schools and corresponding investment amounts

Priority Investment School	School score	Investment amount ($)
University of Central Florida	1.29	12883722.16
Pennsylvania State University-Main Campus	1.24	12354083.72
Ohio State University-Main Campus	1.19	11900117.30
Liberty University	1.14	11449781.50
Houston Community College	1.02	10162421.79
Texas A & M University-College Station	0.94	9443631.589
Michigan State University	0.93	9297720.369
The University of Texas at Austin	0.79	7870040.041
Lone Star College System	0.76	7636409.946
Arizona State University-Tempe	0.70	7002078.836

4. Choice of non-potential school priority investment

Because we take potential schools as the main analysis object, we first established an evaluation system for potential schools, and obtained scores for various indicators, and brought the determined scores to non-potential schools. The score ranking of non-potential schools is shown in Table 2

below:

Table.2. Non-potential priority investment schools

Priority Investment School	School score
Aerosim Flight Academy	0.59
The International Culinary Center	0.47
Fortis Institute-Birmingham	0.42

As can be seen from Table 2, the highest score of non-potential schools is 0.59, which is still lower than the score of the last-priority investment school among potential schools. Therefore, the subsequent analysis is only targeted at those schools with potential investment.

5. Investment benefit indicators

In order to solve the investment benefits of the goodgrant foundation in various schools, we have defined the following indicators:

5.1 Definition of investment indicators

• Economic Indicators

Secondary indicator: return on investment roi

Indicator meaning: refers to the economic return obtained in an investment activity, and refers to the value that should be returned through investment. According to reference [3, 4], this indicator is an important indicator for measuring the value of investment, and it is also effective for improving investment analysis. It is an important complement to reliability and reliability. The data comes from [5].

Secondary indicator: Input-to-investment ratio Pro

Indicator meaning: The ratio of net income to total investment during the investment time is used to measure the degree to which a school creates revenue. The larger the value, the higher the degree of benefit created.

Calculation method: Net return funds accumulated over the investment period divided by total capital investment.

• Academic indicators

• Institutional characteristics

Secondary indicators: graduation rate, degree awarded, degree awarded, student retention, teaching environment, teacher-student ratio, international teacher ratio, international student ratio.

Indicator meaning: It reflects the quality of a school's learning atmosphere, academic strength, popularity, etc. The data comes from [6].

• Academic reputation

Secondary indicators: the degree of influence in the academia, such as humanities and arts, engineering technology, life sciences and medicine, natural sciences, social sciences, etc.

Indicator meaning: It reflects the academic research ability of a school and its influence on the academic community. The data comes from [7].

• Talent development

Secondary indicators: the income of graduates, the popularity of alumni, and the achievements of graduates.

Indicator meaning: It is an indicator to measure whether the school can cultivate students with excellent academic performance and social value. The data comes from [8].

5.2 Obtaining investment benefits

Define the recovery benefit function here, use this function to express the functional relationship between input and investment, and get the relationship between input and investment according to the determined functional relationship. Let t be time, and kt + b be the status of the school. Functional relationship that varies linearly over time.

With the investment of funds, the return on benefits shows a long-term, continuous, and gradual decrease in growth rate. After the funds are invested in the school, the return on benefits shows a long-term, continuous, and gradual decrease in growth rate. Based on this, we construct Functions about re and rgr.

5.3 Conclusion and analysis

According to the collected first-level indicator data, in order to eliminate the impact of different dimensions between these indicators. First, the data is normalized, and from the reference [6], we get the weights between academic indicators; second, using the obtained weights and normalized data to calculate re, the change trend of re and rgr over time is obtained, as shown in Figure 1.

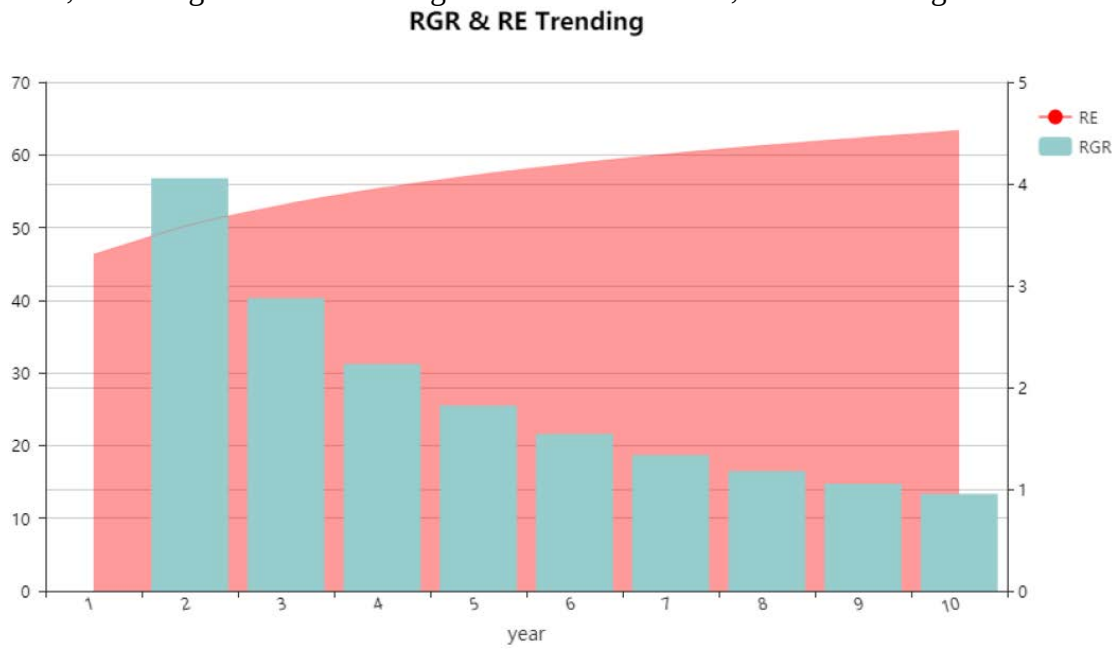

Figure 1. Re and rgr change trends

From the above analysis, we can see that as time changes, re shows a slowly increasing trend, but rgr is gradually decreasing, which is in line with the basic situation of foundation investment. Therefore, according to the changing trend of rgr, we can know that when the growth trend of rgr gradually slows down Best time for investment.

It can be determined that the investment period is 7 years, but a deeper analysis of the recovery efficiency function, combined with the basic situation of the foundation's investment, can be concluded that, whether for the school or the foundation, the continuous and long-term Investment and support in education have long-term benefits and higher value for both parties.

6. Establish dea cost-effectiveness analysis model

6.1 Model establishment

At present, domestic scholars use the DEA method to evaluate the efficiency of public goods supply. Tu Jun and Wu Guisheng use the two-step method of DEA-Tohit to measure the efficiency of agricultural innovation systems in 30 provinces, cities, and autonomous regions. The results show that the influencing factors including rural basic education levels, natural disaster areas, and government funding for science and technology is not significant. Zhu Yuchun and other researchers used the DEA method to study the efficiency of rural public services in 28 provinces, and calculated technical efficiency, pure technical efficiency, scale efficiency, and their coefficients of variation. The results show that the technological efficiency of eight provinces and autonomous regions in Liaoning, Shanghai, Fujian, Shandong, Guangdong, Hunan, Inner Mongolia, and Sichuan is at the forefront of production, that is, the technology is effective; pure technical efficiency and scale efficiency show a gradient change characteristic, that is, eastern> central> West China. Li Yanling,

Tang Juanli, and Wang Fang used the DEA method to study the efficiency of rural public service supply in Hunan, Shaanxi, and Chongqing. Sun Lu, Wu Ruiming, and Li Yun applied factor analysis and DEA to the 16 cities in the Yangtze River Delta. Public service performance was compared and analyzed. Cui Yuanfeng and Yan Lidong used the DEA model to improve the effect of financial support for agriculture.Li Yanling and Zeng Fusheng used the Tobit model to study the influencing factors of rural public goods supply efficiency. Institutional factors, management factors, and financial factors were the main factors, and the efficiency of local public goods supply in China showed a trend of high in the east and low in the west. The literature uses the DEA method to provide deep insights on the efficiency of public goods supply from different perspectives. However, there are few literatures on the public goods supply efficiency in Qinghai Tibetan area. This is what needs attention. The multi-index evaluation of the DEA method measures the similar decision units. The efficiency problem has proven to be quite effective, and it is widely used in the evaluation of productivity and efficiency in industry, cities, regions, and the world. The classic DEA model is called the CCR (C2R) model.

In the investment industry, it is usually faced with the situation that the investment is more profitable, the investment is low, and the harvest is low, which is difficult to analyze. Therefore, we have established a new model to evaluate the effectiveness of investment and investment.

In order to evaluate and analyze the effect of investment strategies, analyze from the aspect of investment cost-effectiveness. Let the annual input and surrender (Input and surrender) be an IAS unit, a total of ... units (such as results). X is the input variable, Y is the investment Output variables, each unit has m kinds of inputs and s kinds of inputs, v_i is the weight of the i-th input, and u_r is the weight of the r-th input, so the efficiency evaluation score of the decision unit IAS is h = (uY) / (vX).

6.2 Conclusion and analysis

After making a certain amount of investment in each school, determine the ten schools to invest in as an example. The Arizona State University-Tempe, The University of Texas at Austin, and Lone Star College System have higher comprehensive returns and better cost-effectiveness, and the school Liberty The comprehensive cost-effectiveness of University, University of Central Florida, and Pennsylvania State University-Main Campus is relatively low. It can be seen that Arizona State University-Tempe can make the most effective use of limited financial investment and has the highest cost-effectiveness, so its unit investment return is better than other schools.

Compared with the contribution rate of the first-level indicators, the order of the size of the first-level indicators has not changed. The degree of contribution rates and the degree of contribution of the student funding indicators have fallen significantly. The contribution rates of the students' graduation income and student repayment price indicators are obvious. Increase, while other indicators have only small-scale fluctuations.

Calculate school scores based on the weights obtained after retaining all household income information.The top ten schools are mainly arts colleges. Considering that arts colleges have higher expenses, family income is better, which will affect the foundation. This investment decision was avoided after averaging household income, so we verified the correctness of our averaging of household income.

7. Analysis of the advantages and disadvantages of the model

7.1 Analysis of the advantages of the model

• Effective extraction of data information

In the context of big data, valuable information is overwhelmed by useless information. For this situation, it is accurate

To extract valuable information, we use the pca method to extract valuable information from the original data, and propose different normalization methods for different attribute data to eliminate the dimensional impact, making our pca more flexible. Finally, we give weights from the perspective of the data from the perspective of data, objectively analyze the samples with the best scores among

many samples.

• Consider both academic and economic directions

Borrowing the idea of general regression, fully considering the return benefits of schools in the academic and economic directions, not just the test

Consider the economic return. Use the log function to fit the return benefit and determine the investment period, define the return benefit rate function RGR, and intuitively determine the investment period from the perspective of return benefit data.

• Consider cost performance

Establish a dea cost-performance analysis model to evaluate the ability of invested schools to use funds and the ability to create social value,

Evaluate the cost-effectiveness of school value creation from an academic and economic perspective.

7.2 Analysis of the shortcomings of the model

• In the dea analysis, we only analyzed the cost-effectiveness of the comprehensive return. If we know the specific allocation ratio of a school in academic direction, school construction direction, student investment, and education investment, we can rebuild the dea model for a school. The cost-effectiveness of the return on investment and investment in education and other aspects is evaluated and analyzed.

• We only removed the Gates Foundation and the investment schools of the foundation, and did not consider the schools invested by other large foundations.

• Use the log function to solve the RGR. Although the solution idea is simple, the result is not high enough. Due to the limited time, we do not have enough energy to collect the academic and economic score data of the school over time. Changes in indicator scores, so the term of the investment obtained may not be accurate enough.

8. Application prospects

In the era of big data, the established PCA model can be used for multi-level indicators and extract main information. Flexible normalization methods can be applied to data with multiple attributes, and then use the data as a starting point to objectively extract from massive data. Produce the most valuable information and score samples instead of subjective imagination.

Taking bank credit risk assessment as an example, in the public data set German Credit Data (https://onlinecourses.scie nce.psu.edu/stat857/node/215), we applied the principal component analysis method to 21 indicators in the original data Twelve new main components were extracted from it, and the bank credit of new users was evaluated, and an accuracy rate of 82% was obtained.It can be seen that the principal component analysis can not only be applied to evaluation scores, but also to delete useless information.After extracting the main information, Make decisions with better results.

9. Conclusion

In the data preprocessing stage, the indicators are first divided into first-level indicators and second-level indicators, and the irrelevant first-level indicators are subjectively removed, and then the removed indicators are tested to prove the correctness of the removed indicators.

In the stage of planning investment plans, first, the principal component analysis method is used to extract important indicators and assign weights. The data shows that the characteristics of institutions, degree award rates, student retention rates, and other indicators representing school strength have the highest weight; second, each school is calculated Comprehensive scores and select 10 schools for investment, collect the scores of multiple indicators such as the economic return rate of the investing school, the graduation rate and academic impact, and borrow the idea of generalized regression to bring the scores into the established equation, and determine the investment period as 6 years, and the results show that the returns and benefits of schools will gradually decline. Therefore,

investment in schools and education requires long-term persistence.

And investment in education is the most valuable investment, and it is also responsible for the future of the country. Finally, at the end of the investment, we evaluated the cost-effectiveness of each school's investment, and the results show that the University of Central Florida school's effectiveness Optimal.

References

[1] Qu Qiongfei, Shu Zhongmei. Research and enlightenment of the management and operation mechanism of ipeds in the US post-secondary education data comprehensive system [j]. Higher Education Exploration, 2013 (05): 82-87.

[2] Guo Congbin, Sun Qiming. A Comparative Analysis of Higher Education Institutions in Mainland China and World-Class Universities——From the Perspective of University Rankings [j]. Educational Research, 2015, 36 (02): 147-157.

[3] Wang Xiaochun. Assessing the return on investment (roi) measurement of human resources projects [j]. China Human Resources Development, 2008 (11): 32-35.

[4] Yang Suhua. Investment income (roi) seeks the key to the most profitable franchise company [j]. China market, 1997 (11): 69.

[5] https://www.payscale.com/college-roi?page=130

[6] https://www.topuniversities.com/student-info/qs-guides/qs-world-university-rankings-2018-supplement

[7] https://www.topuniversities.com/student-info/qs-guides/qs-world-university-rankings-2018-supplement

[8] https://www.topuniversities.com/student-info/qs-guides/qs-world-university-rankings-2018-supplement