A Detailed Survey on Performance Analysis and Ranking of Educational Universities with Big Data

Rasheed Mansoor Ali S
Research Scholar,
Department of computer Science,
Vels Institute of Science, Technology and Advanced Studies (VISTAS),Chennai,India
Email - srasheedmansoor@gmail.com

Dr.S.Perumal,
Associate Professor,
Department of computer Science,
Vels Institute of Science, Technology and Advanced Studies (VISTAS),Chennai,India
Email – perumal.scs@velsuniv.ac.in

Abstract - The growing world population increases the need of identifying specific top-ranked institutions for the future development of students. There exist numerous universities present in any country but the selection of institutions towards enrollment has been identified as a challenging task. Towards this, this paper presents a detailed analysis of ranking universities in several ways and how the existing approaches perform the task. Also, the performance of any university has been identified and ranked according to the performance of the students. According to this, different approaches are analyzed in detail and how they have been ranked is detailed. Further, a detailed comparative study has been presented in the end. The performance of different approaches are measured and compared on specific performance metrics in detail.

Keywords: Institutional Ranking, Performance Analysis, Big Data, Student Performance, Academic Data.

I. INTRODUCTION:

The world population has been growing each day and the number of students comes out of secondary education are identified as higher. To move towards their future, human society looks for different institutions and universities which are more efficient and more supportive of their future development. Similarly, there are a number of organizations that places the students comes out from different universities in their concern. But, it is really important for them to choose more efficient students from reputed institutions. So, the performance of any university is really depending on the performance of the student who comes out from there. Educational universities have a higher impact on student performance in their growth. The organizations rank the universities according to student performance. The student performance has been ranked according to many factors like academic, technical, and others. However, monitoring the student's performance and ranking them has been identified as a challenging task.

The organizations would rank the institutions according to their activities in academic, technical and other areas. Still, by measuring their behavior in various spectrums would support the ranking of institutions. Towards this, there are many articles being published and many techniques are recommended by the researchers. Each uses different techniques

and differs with the feature and method considered. This article analyses such approaches in detail and according to that different methods are analyzed for their performance and compared.

The big data are used in several problems now a day as the volume and size of data belongs to any entity that gets increased. By considering the Big data, the result of analysis on the big data would support various problem solutions. Bigdata has been used for numerous analysis and can be adapted for many problems. The same can be used for the analysis of the performance of Universities and would support ranking them. The ranking of universities is performed in several ways. As an example, they have been ranked according to the number of students being enrolled in every year. But it will not be effective and it cannot be said as it has a higher admission rate, its performance is higher. Similarly, the placement of students in different organizations can be used in ranking the institution. Also, this will produce a higher false ratio as the placement rate is depending on the state of organizations that recruit them and current employment conditions of any country or region. Similarly, there are a number of approaches that can be considered in ranking the institutions.

Still, the performance of students has a higher impact on ranking the institutions. The student would get involved in different behaviors and activities. In order to rank the institution, it is necessary to consider different activities of students in studies, curricular activities, sports and so on. By considering all these the performance of ranking the institutions can be performed effectively. Before this, this article analyzes the different approaches in the next section.

Literature Survey:

The approaches of ranking institutions have been classified under several categories and presented in this section:

II. MACHINE LEARNING AND DATA MINING TECHNIQUES IN INSTITUTIONAL RANKING:

The machine learning algorithms are used in several problems. By adapting the machine learning and data mining techniques, the problem of institutional ranking is solved. Such methods are discussed in detail in this section.

(Chew Li Sa, et. al 2015) presented a Student performance analysis system (SPAS) which maintains the performance track details of different students. The model is designed to predict the student performance to support the lecturers in identifying the poor performing students. The model maintains a set of rules based on the previous records and based on that data mining techniques are adapted towards classifying the students.

(V. Shanmuga rajeshwari and R. Lawrance 2016) presented a performance analysis model to measure the students performance according to classification techniques. The method uses C5.0 algorithm in classifying the performance of students as well as teachers. Based on the result of classification, a warning or alert has been produced to the students.

(Anoushka Jain, et. al 2017) presented an Intellectual performance analysis model that compares the performance of different data mining techniques used in analyzing the performance of students. The model considers different social factors which have a higher impact on the education of students. According to that, the model classifies the

students and measure or predict the performance of students.

(Ishwank Singh, et. al 2016) presented a clustering-based performance analysis algorithm that considers the factors of admission and placement. The model groups the students according to the above-mentioned factors under different categories. Based on that the performances of students are improved.

(Mehil B Shah, et. al 2019) presented a machine-learning algorithm towards the prediction of student performance and support ranking of students. The method considers the features like the interest of users, attributes, and variables of their lifestyle. Deep learning techniques are used for prediction and correlation measures are used.

(Maryam Zaffar, et. al 2017) claims that the performances of student performance prediction algorithms are depending on the feature selection algorithm being used. For any classification algorithm, the feature selection algorithms are more essential and their performance is highly dependent on the performance of the feature selection algorithms. Towards this, various feature selection schemes are analyzed for their performance in predicting student performance and the quality of education.

(T.Rajesh Kumar, et. al 2019) presented analysis on different data mining techniques towards predicting the performance of students. The model considered factors like scholastic accomplishments (CGPA), sex, class test grade, condition of class, Fund/Scholarships/Private and so on. According to the above-mentioned features, the classification is performed with Naïve Bayes, Knn, and neural networks. According to the result of classification, the rating and ranking of students are performed.

(Pooja Thakar, 2015) present a comprehensive survey on the methods of data mining being recommended by different researchers towards predicting the performance of different students. Similarly, (Amirah Mohamed Shahiri, 2015) presented a detailed review of the data mining techniques on performance evaluation of different academic sectors. The analysis is focused on identifying the factors which affect the performance of students as well as institutions.

(Juan L. Rastrollo-Guerrero, and Juan A. Gómez-Pulido , 2020) presented a student performance analysis and prediction scheme which is focused to identify the reason for dropouts in degree courses. Different methods are adapted for the performance analysis and based on that set of methods are identified as more effective in predicting student performance.

1) Ensemble-Based Prediction Model:

Ensemble-based approaches are used in different classification problems. The ensembles are nothing but a prototype and based on that the student performance can be predicted. Such methods are discussed in this section.

(Han, M. et. al 2017) presented an ensemble-based approach towards the prediction of student performance where the AdaBoost algorithm has been used for prediction. The performance of the method has been compared with the performance of other approaches.

(Jie Xu, et. al 2017) presented a machine-learning based prediction model for student performance which works on a bilayered model and ensembles. According to the ensembles, the probability value has been generated to detect the relevancy in course.

2) Behavior-Based Prediction Model:

The performance of student is predicted in several ways. However, the behavior of students can be used in predicting the performance of students more effectively. Towards this, several techniques are discussed in literature. Such methods are discussed in detail in this section.

(Villagrá-Arnedo, C. et. al 2016) presented a prediction model that uses different contextual data which include the behavior of different users. The model considers both behavioral data and the way of learning. The learning process considers both of them to perform prediction where the classifier is used with SVM. The method estimates the probability value and classifies into three categories.

3) Clustering-Based Prediction Model:

Any educational institutions have a number of students. The performance of them would vary according to different features. However, they can be grouped according to their performance in previous semesters. By grouping the students accordingly, the prediction process can be performed efficiently.

(Rana, S. et. al 2017) presented a clustering and classification approach towards the performance prediction of Punjab university. The method first group the students with K means hierarchical clustering algorithm. Further, the M5P algorithm is used for classification.

(Jayaprakash, S. 2018). Presented a detailed review on the progress of students with different classification algorithms. The article analyzed different works in identifying a solution to improve the future research performance.

4) Decision-Tree Based Prediction Model:

The decision tree is more effective in the classification problem. According to the application of decision trees, different methods are discussed in literature towards the prediction of student performance. Such methods are discussed in detail in this section.

(Al-barrak, M. A., & Al-razgan, M. 2016) presented a decision tree-based prediction scheme towards the value of GPA achieved by students. The method considers the value of GPA obtained in previous courses. The classification is performed with the J48 decision tree method and identifies classification rules.

(Hamoud, A. K., Hashim, 2018) presented a decision tree analysis scheme towards the prediction of student performance. The author considered different classifiers like J48, Random Tree and REPTree which are applied with questionnaires filled by students.

5) SVM-Based Prediction Model:

The support vector machines are used in different scientific problems. The same can be applied in predicting the student performance. Such methods are presented in this section.

(Huda Al-Shehri, et. al) presented a SVM and KNN-based student performance prediction model which considers the relationship between the preadmission academic profile and final academic performance.

(Ceasar Ian P. Benablo,Evangeline T Sarte 2018) presented a predictive analytics model that uses the Support Vector Machine (SVM). The method uses the profiles of earlier students to classify into a number of categories like performing and under-performing. According to that the SVM algorithm classifies the user performance.

6) Fuzzy Logic-Based Prediction Model:

The fuzzy logic is used in different scientific problems and the same can be used in the prediction of student performance. According to this, there are a number of articles presented in the literature. This section details such methods and presented in detail.

(Kousik Das, 2020) presented a fuzzy logic-based mathematical model towards ranking the educational institutions. The method uses linguistic variables to perform ranking.

(Rashid, K.A., Amin, H.U., Rehman, Z.U. 2011) presented the application of fuzzy logic in evaluating the performance of teachers. The method considers the sketch of students and researchers which has been represented in form of fuzzy values. Using them the method computes the weight measures to classify their performance.

(Alam, J., Pandey, M.K.: 2017) presented a soft computing model towards the performance measure of teachers. The model measure the performance of teaching using two different models namely TOP-M1 (Teacher overall performance model 1) which computes the teaching performance and (TOP-M2) teachers' overall performance module-2 calculates academic and administrative performance.

(Thakre, T.A., et. al 2017) presented a fuzzy logic approach to the performance evaluation of teachers. The method takes multiple inputs and introduces crisp data depending on extreme values. The model analyzes the performance according to numeric values.

(R.Srinivasan, et. al 2020) presented a soft computing model towards institutional ranking which considers multiple parameters like student intake, faculty strength, the expenditure of the institution, research paper published per faculty, placements and

perception. According to the features, the method performs fuzzy inference.

7) Other Methods:

Further there are a number of approaches are presented towards prediction of student performance. Such methods are classified in this section.

(Sultana, S.; et. al 2017) presented a performance prediction model that uses cognitive and non-cognitive features towards the detection of drop out. The method applies both cognitive and non-cognitive features in analyzing the performance of students. The inclusion of non-cognitive feature improves the performance.

(Verhun, V. et. al 2018) analyze the performance of different models of performance prediction. Current capabilities of software systems related to the education domain allow analyzing and leveraging outcomes of the data mining process and machine learning algorithms in order to make decisions and justifications of educational approaches.

(Eyman Alyahyan & Dilek Düştegör, 2020) presented an academic success prediction scheme, which finds the risk at the early stage by providing guidelines for educators by applying data mining techniques towards predicting student success.

(Adekitan, A. I., & Salau, O.2019). presented a predictive analysis towards predicting the final CGPA in the final year students. Similarly, (Lubna Mahmoud Abu Zohair, 2019) presented the student performance prediction model which works based on a small data set. The method uses virtualization and clustering algorithms. The method is compared with the performance of SVM and LDA schemes.

(Domingo Docampo & Lawrence Cram, 2017) presented a cross country analysis in ranking the universities. The analysis is performed according to the number of research articles published per year, number of citations, number of students ,and total pass out per year.

(Vicente Safón, 2019) presented analysis model on ranking different universities according to inter ranking with ARWU and THE. The ranking is performed according to the reputation of different countries. Similarly, (Manuel Salas-Velasco, 2020) presented an analysis model which uses several inputs and outputs at the institutional level and use a bootstrap algorithm to perform classification and ranking.

(Lili LinZhuoming XuYuanhang ZhuangJie Wei, 2014) presented factor graph-based institution ranking model which assesses the institutions according to the ground trust. Based on that the ranking and case study is conducted in US institutions.

(Fernandes, Eduardo, et. al 2019) presented a predictive analysis model towards the prediction of student performance in the federal district of Brazil. The method uses the data set contains information's like preschool data, performance after two months, and so on. Using them, the method uses Gradient Boosting Machine (GBM) for classification.

TABLE 1: COMPARATIVE STUDY

       The comparative study on different methods according to the method and measures used are presented in Table 1.

Author Year	Method and Feature Used	Advantages	Disadvantages
Chew Li Sa, et. al 2015	Student performance analysis system (SPAS). Maintains student track. Generates rules to predict the performance.	Supports the lecturer in predicting the student performance. Moderate performance in prediction	Higher time complexity
V. Shanmugarajeshwari and R. Lawrance 2016	uses C5.0 algorithm in classifying the performance of students	Moderate performance in prediction.	Noticeable false ratio.
Anoushka Jain, et. al 2017	Intellectual performance analysis model. Considers different social factors.	Moderate performance in prediction	Higher time complexity.
Mehil B Shah, et. al 2019	Deep learning technique is used Like the interest of user, attributes and variables of their lifestyle are considered.	Noticeable performance in prediction.	Higher time complexity.
Han, M. et. al 2017	Ensemble-based approach is used and Adaboost algorithm for classification.	Higher classification performance.	Interest prediction accuracy is poor.
Villagrá-Arnedo, C. et. al 2016	Behavior-based model and uses SVM as classifier	Higher classification performance	Time complexity is higher.
Al-barrak, M. A., & Al-razgan, M. 2016	Decision Tree-based model. J48 decision tree for classification.	Moderate classification and prediction performance.	Higher time complexity.

Huda Al-Shehri, et. al, 2018	SVM and KNN prediction Model	Introduces moderate prediction accuracy	Higher time complexity.
Kousik Das, 2020	Fuzzy Logic Mathematical Model uses linguistic variables in ranking.	Introduces moderate ranking performance	Higher time complexity.
Thakre, T.A., et. al 2017	Fuzzy logic Approach and considers numeric values	Prediction performance is moderate.	Introduces poor performance in classification.
Lili LinZhuoming XuYuanhang ZhuangJie Wei, 2014	Factor Graph-Based Ranking	Uses Ground Trust in ranking. Introduces moderate ranking performance.	Higher time complexity.

III. RESULTS AND DISCUSSION:

The methods of performance prediction of students and institutional ranking are implemented and evaluated for their performance. Obtained results are compared with the results of other methods. The performances of the methods are analyzed in various parameters in this section.
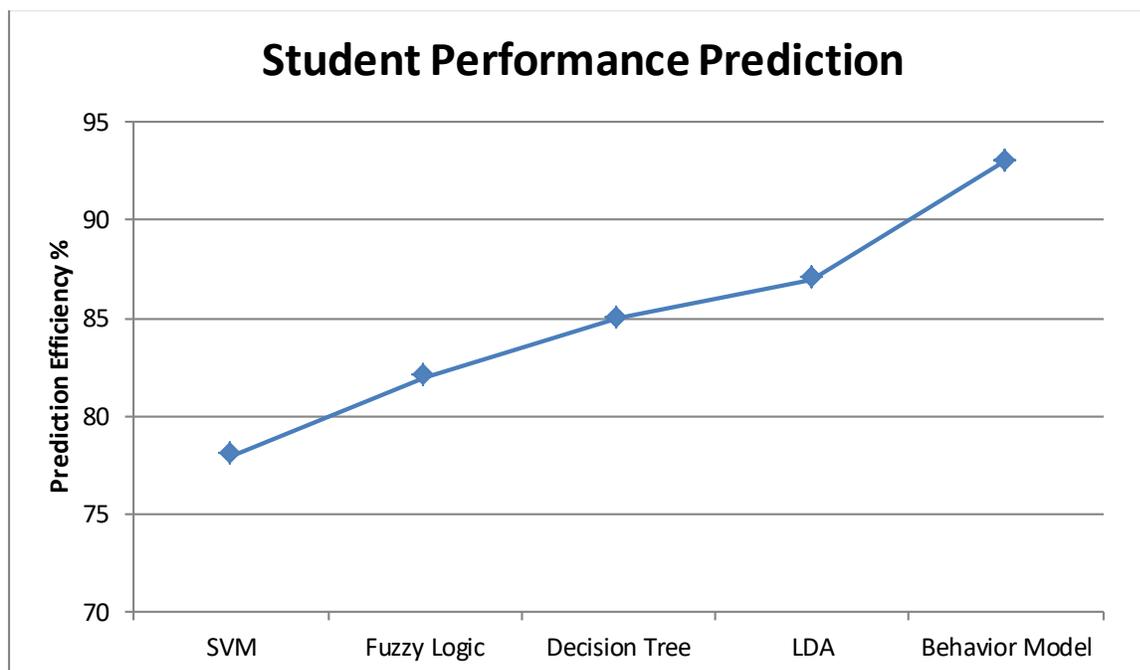

Figure 1: Analysis on Student Performance Prediction Accuracy

The analysis on student performance prediction accuracy produced by different methods are measured and presented in Figure 1.
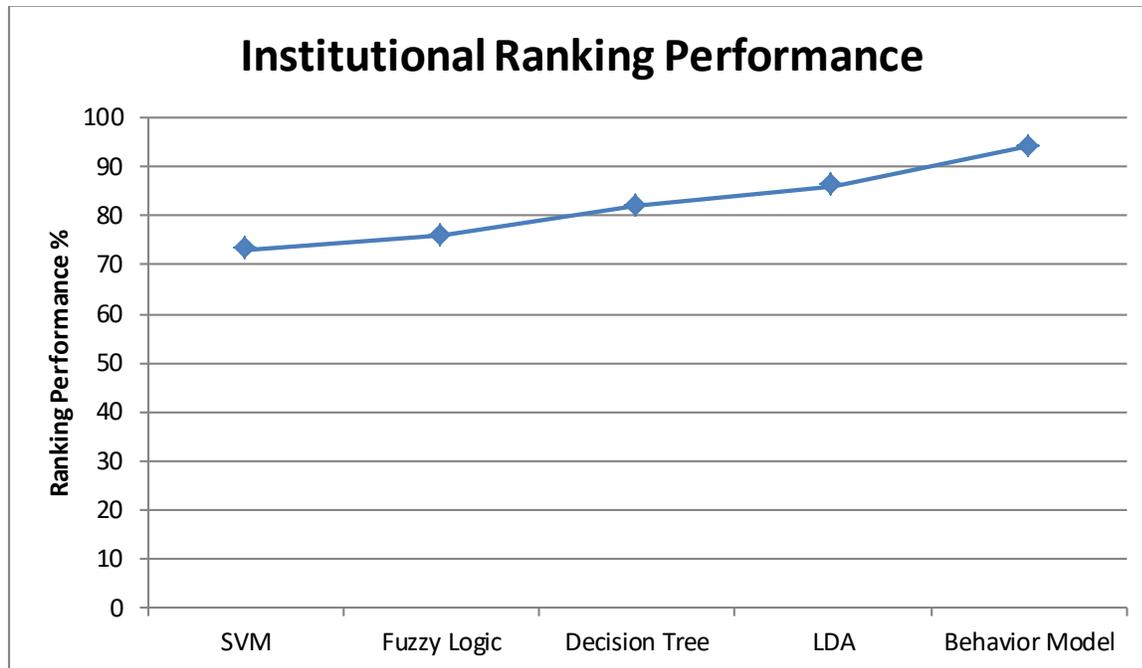

Figure 2: Analysis on Ranking Performance

The ranking performance produced by different methods is analyzed and presented in Figure 2. Behavior analysis models have higher performance than other methods.

IV. CONCLUSION:

The analysis of different institutional ranking and student performance prediction schemes are analyzed. Each method has considered different features and they used various classification algorithms. This article analyzed different methods being used by various researchers. The results produced by them are well analyzed and their performance is evaluated under various constraints. The detailed analysis is presented with a comparative study.

V. REFERENCES:

1. Chew Li Sa ; et.al , Student performance analysis system (SPAS), IEEE , ICT4M, 2015.

2. V. Shanmugarajeshwari ; R. Lawrance, Analysis of students' performance evaluation using classification techniques, ICCTIDE, 2016.

3. Anoushka Jain; et.al Intellectual performance analysis of students by comparing various data mining techniques, International Conference on Applied and Theoretical Computing and Communication Technology , 2017.

4. Ishwank Singh, et.al , Student performance analysis using clustering algorithm, International Conference - Cloud System and Big Data Engineering (Confluence), 2016.

5. Mehil B Shah ; et.al Student Performance Assessment and Prediction System using Machine Learning, International Conference on Information Systems and Computer Networks (ISCON), 2019.

6. Maryam Zaffar ; et.al Performance analysis of feature selection algorithm for educational data mining, IEEE Conference on Big Data and Analytics (ICBDA), 2017.

7. T. Rajesh Kumar ; et. al Students Performance Prediction Using Data Mining Techniques, International Conference on Intelligent Sustainable Systems (ICISS), 2019.

8. Pooja Thakar, Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue, International Journal of Computer Applications (0975 – 8887), Volume 110 – No. 15, 2015.

9. Amirah Mohamed Shahiri, A Review on Predicting Student's Performance Using Data Mining Techniques, Science Direct, Procedia Computer Science, Volume 72, 2015, Pages 414-422.

10. Jie Xu ; et. al A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs, IEEE Journal of Selected Topics in Signal Processing (Volume: 11 , Issue: 5 , 2017)

11. Juan L. Rastrollo-Guerrero,Juan A. Gómez-Pulido , Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review, MDPI, 2020.

12. Sultana, S.; et. al , M. Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. Int. J. Electr. Eng. Educ. 2017, 54, 105–118

13.   Verhun, V et. al Learning Analysis as a Tool for Predicting Student Performance. International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2018; Volume 2, pp. 76–79.

14. Han, M.; et. al Application of Ensemble Algorithm in Students' Performance Prediction. International Congress on Advanced Applied Informatics (IIAI-AAI), 2017; pp. 735–740

15. Villagrá-Arnedo, et. al Predicting academic performance from Behavioral and learning data. Int. J. Des. Nat. Ecodyn. 2016, 11, 239–249.

16. Rana, S.; Garg, R. Prediction of students performance of an institute using Classification Via Clustering and Classification Via Regression. Adv. Intell. Syst. Comput. 2017, 508, 333–343.

17.   Eyman Alyahyan & Dilek Düştegör, Predicting academic success in higher education: literature review and best practices, International Journal of Educational Technology in Higher Education volume 17, : 3 (2020).

18. Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. Heliyon, 5(2), e01250.

19. Al-barrak, et. al . Predicting Students Final GPA Using Decision Trees : A Case Study. International Journal of Information and Education Technology, 6(7), 528–533, 2016.

20. Hamoud, A. K., Hashim, Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. International Journal of Interactive Multimedia and Artificial Intelligence, 1, 2018.

21. Jayaprakash, S. A Survey on Academic Progression of Students in Tertiary Education using Classification Algorithms. International Journal of Engineering Technology Science and Research IJETSR, 5(2), 136–142, 2018.

22. Lubna Mahmoud Abu Zohair, Prediction of Student's performance by modelling small dataset size, International Journal of Educational Technology in Higher Education, 2019.

23. Huda Al-Shehri, et. al Student performance prediction using support vector machine and k-nearest neighbor. In International Conference on Electrical and Computer Engineering (CCECE). IEEE, 1–4, 2017.

24. Kousik Das, Sovan Samanta , Ranking of educational institutions using fuzzy logic: a mathematical approach, Springer, Afrika Matematika (2020),

25. Rashid, K.A.,   Application of expert system with fuzzy logic in teachers performance evaluation. Int. J. Adv. Comput. Sci. Appl. 2 (2011)

26. Alam, J. A soft computing model for evaluating teachers' overall performance using fuzzy logic. J. Inf. Technol. Softw. Eng. 7(2), 1–9 (2017)

27. Thakre, T.A., et. al A fuzzy logic multi criteria approach for evaluation for teachers' performance. Adv. Fuzzy Math. 12(1), 12–145 (2017)

28. Vicente Safón, Inter-ranking reputational effects: an analysis of the Academic Ranking of World Universities (ARWU) and the Times Higher Education World University Rankings (THE) reputational relationship, Springer, Scientometrics volume 121, pages897–915(2019).

29. R. Srinivasan, et. al Perception-based performance analysis of higher education institutions: a soft computing approach, Springer, Soft Computing volume 24, pages513–521(2020).

30. Manuel Salas-Velasco, The technical efficiency performance of the higher education systems based on data envelopment analysis with an illustration for the Spanish case, Springer, Educational Research for Policy and Practice volume 19, pages159–180(2020

31. Lili LinZhuoming Xu, et. al Evaluating the Academic Performance of Institutions within Scholarly Communities, International Conference on Asian Digital Libraries ICADL 2014: The Emergence of Digital Libraries – Research and Practices pp 76-86.

32. Fernandes, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," Journal of Business Research, Elsevier, vol. 94(C), pages 335-343, 2019.

33. Ceasar Ian P. Benablo,Evangeline T Sarte , Higher Education Student's Academic Performance Analysis through Predictive Analytics, ICPS Proceedings, 2018.

34. Domingo Docampo, Academic performance and institutional resources: a cross-country analysis of research universities, ACM, Scientometrics, 2017.