International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 10 (2015) pp. 25347-25354

http://www.ripublication.com

Evaluation of Student Performance Using Z-Score and K-Means

Mr. Vinayak Hegde[1], Ms. Deepika N C[2] , Ms. Sushmitha K[3]
Amrita Vishwa Vidyapeetham, Mysore Campus, #114 Bogadi 2nd Stage Mysore. Karnataka
India, vinayakhegde92@gmail.com, Mob No: 9886484778

Abstract

In India ranking of students is mainly done on basis of marks obtained in exams. Majority of the Indian college obey the typical pattern of grouping the students in the category of pass with distinction for students scoring 75% and above, first division for the scores within 60% to 75%, second division for the scores within 50% to 60%, third division for those within 40% to 50%. Most of the Universities set a minimum set of marks that should be maintained in order to continue in the degree program. Therefore, marks obtained in examination remains the most common factor used by the academic planners to evaluate progression in the domain of academics. It is onerous to judge whether course appraisal reflect how much students really learn from a class because valid measures of learning are hardly available. This paper introduces a straightforward methodology based on Z-Score, k-means clustering algorithm and deterministic model to evaluate the performance of students in higher institutions. This methodology will aid in academic planning, accordingly the instructors can supervise student’s performance during every semester. Therefore this method will play an important role for academic planners to determine the reasons for decline in performance of students during particular semester and steps that need to be taken to improve performance from next academic session.

Key Words: Mean, Standard Deviation, Z-Score, Student performance, Statistical analysis, K-Means.

Introduction

There is an impending crisis in the administration of higher education due to rapid opening of educational institutions. One among the most usual evaluation tools in higher education available today is the Student evaluations of courses. Most of the university stakeholders use scores on course evaluation forms as a measure of the transmission of knowledge in a course. In order to have supremacy and good count of students it is important to possess a resilient strategic management tool in the educational sphere, that aims at a finer comprehension of the students learning techniques and ascertain the students learning settings to acquire the improved educational outcomes and to gain insight into it handling the huge amount of student data.

The problem is many educational institutions is that student performance evaluation. Analysis and identification of only well performing students are done and weaker students are ignored. All these problems have been overcome by concentrating on the ignored areas aiming at simplified and economic data management, admittance and analysis of student data.

Literature Survey

Several technique has been defined using data mining to evaluate the student performance in the educational domain. The overview of existing experimentations reported in the literature is discussed below.

Rakesh Kumar Arora et.al [1], has proposed one such simple and easy group the student data is K-Means clustering technique. Using K-Means clustering the performance of the students are easily differentiated based on the number of clusters given. Brijesh Kumar Bhardwaj et.al[2], has proposed a technique in predicting the student performance by using Bayesian classification method. Where the previous year performance is used to predict the current year performance. The resulting approach well be upon the weaker students who require special attention. Anwar M. A et.al[3], has proposed the analysis of supervised and unsupervised assessment of students data, using association rule technique. Using association rule, where the it gives clear difference between the performance in the supervised assessments and unsupervised assessments of the students. By this it is easy to identify where exactly the students are lacking. Ajay Kumar Pal et.al[4], proposed a methodology using classification model. Where in this students previous year's performance data and placement training data has been taken for the analysis. Using the weka tool, classification has been done. After the classification in the weka tool based on previous year's performance and placement training performance, it is a prediction that how many can get placed in the placements. Rakesh Kumar Arora et.al[5], has proposed a technique of association rule to improve the student performance using weka tool. Where the students whole semester performance grades has been taken for the analysis. The resulting helps the academic planners to compare the students' performance with well performing students.

Most of the methods in the literature focused on identifying students' performance in without comparing to the others students performance. In some of the papers they used tools like weka tool for analysing the students' performance. The proposed methodology gives the individual performance of the students with comparing to the others students' performance as well as group performance of the students.

Proposed Methodology

The application of analytics in the field of education has fatten in the late years for four primary reasons: a substantial increase in amount of data, refined data formats, move forwards in computing, and lastly the highly knowledgeable tools available for analytics. The relationship of a score with the mean of the group of scores gives a statistical measurement called Z-score. Sometimes we want to do more than summarize a bunch of scores. Sometimes we want to talk about particular scores within the bunch. We may want to tell other people about whether or not a score is above or below average. We may want to tell other people how far away a particular score is from average. We might also want to compare scores from different bunches of data. We will want to know which score is better. Z-scores can help with all of this. A Z-score value can be a positive or a negative. The value of 0 Z-score indicates that the score is equal to the mean value obtained. The positive or negative Z-score value indicates whether it is above or below the mean value and by how many *S* (standard deviation) values.

In the proposed system the student performance is evaluated using Z-Score. Where Z-Score values are used to assess individual performance of students in a particular subject and also reveals whether the students stands above or below average in performance. The application also visualizes the interpreted data and where it helps to understand the overall performance of all students in a particular subject. By this the instructor of the particular subject can concentrate on weaker students and provide right instruction in order to improve their performance.

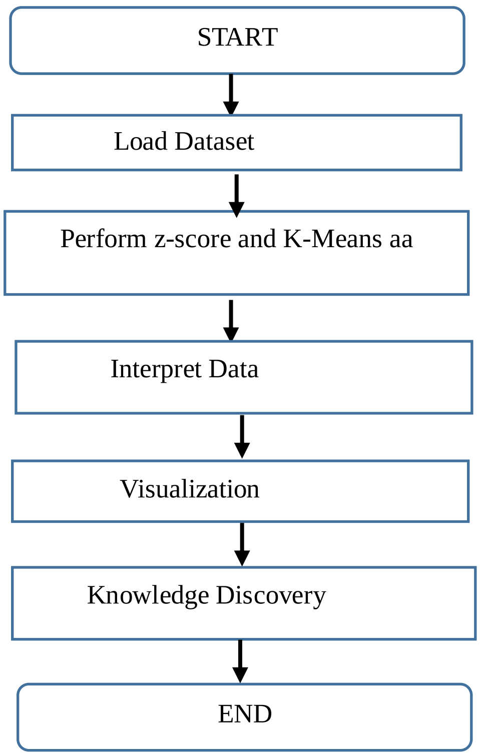


Figure 1: Process Methodology

$$z = \frac{x - \overline{x}}{s}$$

x_i is each value of dataset
x is the mean value of data
s is the standard deviation

The absolute value of the z-score tells you how many standard deviations you are away from the mean. If a z-score is equal to 0, it is on the mean. If a Z-Score is equal to +1, it is 1 Standard Deviation above the mean. If a z-score is equal to +2, it is 2 Standard Deviations above the mean. If a z-score is equal to -1, it is 1 Standard Deviation below the mean. If a z-score is equal to -2, it is 2 Standard Deviations below the mean. Some z-scores will be positive whereas others will be negative. If a z-score is positive, its' corresponding raw score is above (greater than) the mean. If a z-score is negative, its' corresponding raw score is below (less than) the mean.

K-Means Clustering:

The main idea is to define k centers, one for each cluster. The better choice is to place the centers as much far as possible from one another. The following step is to take every point belonging to a given data set and associate it to the adjacent center. And when no point is undone, the initial step is completed and an early grouping is performed. At this stage we need to re-compute k new centroids as barycenter of the clusters resulting from the previous step. Once we have these k new centroids, a contemporary binding has to be done between the same data set points and the adjoining new center. Now a loop has been created. Due to the generation of this loop we may observe that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Lastly, this algorithm aims at minimizing an objective function known as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\|x_i - v_j\| \right)^2$$

where,
'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.
'c_i' is the number of data points in i^{th} cluster.
'c' is the number of cluster centers.

The algorithm comprises of the successive steps:

1. Locate K points into the area represented by the clustered objects. Each of these points represent initial centroids for the respective group.
2. Allocate each object to the group that has the nearest centroid.
3. Once all assignments are done, recompute the positions of the K centroids.
4. Repeat Step 2 and 3 until the centroids no longer move. This yields a parting of the objects into groups from which the metric to be minimized can be computed.

Experimental Analysis

The experimentation of student performance evaluation, with respect to the proposed methodology is conducted with around 70 student's marks. Which includes Z-Score, where it determines the individual performance of the students. In order to get the overall performance of the students K-Means algorithm is used. In K-Means algorithm consider total marks scored by the student and based on the number of clusters as user choice, the marks are divided into those many clusters. After clustering, individual performance of each cluster is taken. Then cluster size and performance of individual cluster is considered and plotted in a graph.

Name	Roll No	Total	Z-Score
Ravi	109	75.56	0.4393
Anusha Raj	110	65.35	-0.320
Bajan k	111	90.7	1.566
Disha B.N	112	75	0.3977
Eshan Khan	113	73.78	0.3069
Gulshan Kumar	114	56.1	-1.008
Harsha	115	65.43	-0.3144
Jithin	116	84.2	1.0823
Karthik G	117	66.99	-0.1983
Mohan R	118	38.1	-2.348
Rithwik	119	75	0.39771

Figure 2: Calculated Z-Score of Students Marks

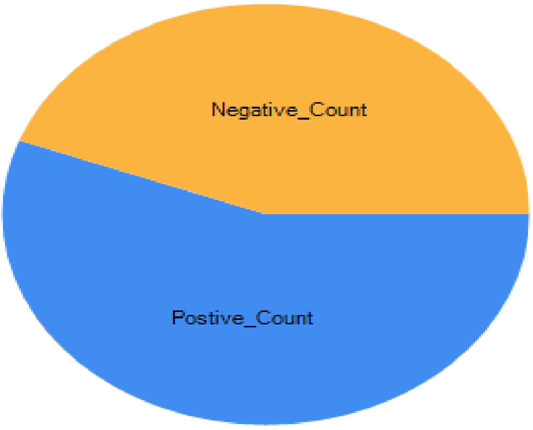
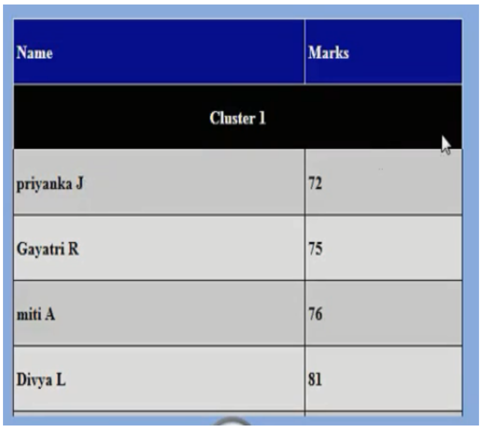
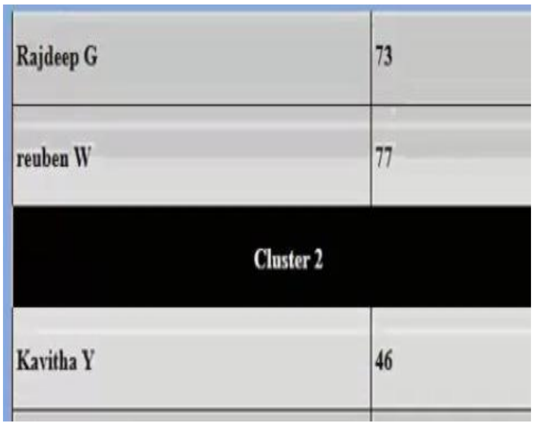


Figure 3: Pie chart visualization of Z-Scores



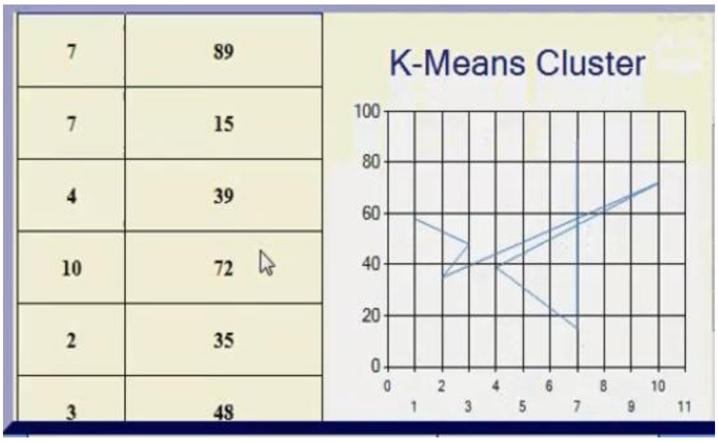
Name	Marks
Cluster 1	
priyanka J	72
Gayatri R	75
miti A	76
Divya L	81

Figure 4: Cluster 1 of students marks



Rajdeep G	73
reuben W	77
Cluster 2	
Kavitha Y	46

Figure 5: Cluster 2 of student marks



7	89
7	15
4	39
10	72
2	35
3	48

Figure 6: Cluster size, each cluster performance and visulization

Conclusion

Data mining is a powerful analytical tool that enables educational institutions to better allocate resources and staff, providently manage student outcomes, and improve the effectiveness of student performance. With the ability to uncover hidden patterns in large databases, colleges and universities can build models that predict—with a high degree of accuracy—the behavior of population clusters. By acting on these statistical models, educational institutions can effectively address issues ranging from student performance.

References

[1] Rakesh Kumar Arora, DharmendraBadal, "Evaluating Student's Performance Using K-Means Clustering" IJCST Vol. 4, Issue2, April - June 2013 ISSN : 2229-4333.

[2] Brijesh Kumar Bhardwaj and Saurabh Pal", Data Mining: A prediction for performanceimprovement using classification". (IJCSIS) International Journal of Computer Science and Information Security,Vol. 9, No. 4, April 2011.

[3] Anwar M. A., Naseer Ahmed" Information Mining in Assessment Data of Students' Performance".ISSN: 2319-5967 ISO 9001:2008 Certified International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 1, Issue 2, November 2012.

[4] Ajay Kumar Pal, Saurabh Pal" Classification Model of Prediction for Placement of Students".I.J.Modern Education and Computer Science, 2013, 11, 49-56 Published Online November 2013 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijmecs.2013.11.07.

[5] Rakesh Kumar Arora, Dr. Dharmendra Badal," Mining Association Rules to Improve Academic Performance", ISSN 2320–088X, IJCSMC, Vol. 3, Issue. 1, January 2014.

[6] Brijesh Kumar Baradwaj ,Saurabh Pal" Mining Educational Data to Analyze Students Performance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.

[7] Rakesh Kumar Arora, DharmendraBadal, "Admission Management through Data Mining using WEKA",IJARCSSE-International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 10, October 2013.

[8] Umesh Kumar Pandey, Surjeet Kumar Yadav, Saurabh pal," Data Mining Application to Attract Students in HEI", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN : 0975-3397 Vol. 4 No. 06 June 2012.

[9] Simon Fong and Robert P. Biuk-Aghai," An Automated University Admission Recommender System for Secondary School Students",the 6th International Conference on Information Technology and Applications (ICITA 2009),ICITA 2009 ISBN: 978-981-08-3029-8.

[10] Umamaheswari. K, S. Niraimathi "A Study on Student Data Analysis Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013 ISSN: 2277 128X

[11] Bhise R.B, Thorat S.S, Supekar A.K" Importance of Data Mining in Higher Education System", IOSR Journal Of Humanities And Social Science (IOSR-JHSS), ISSN: 2279-0837, ISBN: 2279-0845. Volume 6, Issue 6 (Jan. - Feb. 2013)

[12] Manjunath T. N., Ravindra S. Hegadi, Umesh I. M., and Ravikumar G. K" Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain", International Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012.

[13] Monika Goyal, and RajanVohra" Applications of Data Mining in Higher Education", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012, ISSN (Online): 1694-0814

[14] Suman, Mrs.Pooja Mittal" A Comparative Study on Role of Data Mining Techniques in Education: A Review", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 3, May – June 2014 ISSN 2278-6856.

[15] Dr. Varun Kumar, Anupama Chadha" An Empirical Study of the Applications of Data Mining Techniques in Higher Education", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.3, March 2011.

[16] Manmohan Singh, Dr. Anjali Sant" Education using Computational Data Mining Models", International Journal of Advance Research in Computer Science and Management Studies, ISSN: 2321-7782 (Online) ,Volume 1, Issue 7, December 2013

[17] Abdul Hamid M. Ragab, Abdullah S. AL-Ghamdi ,Amin Y. Noaman ,Ayman I. Madbouly" A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining".

[18] Dr. Vijayalakshmi M N, S. Anupama Kumar, Kavyashree BN" Investigating Interesting Rules Using Association Mining for Educational Data", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 2, February 2014.