A Quality Based AutomatedAdmission System for Educational Domain

Sushruta Mishra[1],Soumya Sahoo[2]
C.V Raman College of Engineering,
Bhubaneswar , India
mishra.sushruta@gmail.com
ritun_08@yahoo.co.in

Brojo Kishore Mishra[3],Santosh Satapathy[4]
C.V Raman College of Engineering,Bhubaneswar,India
Gandhi Engineering College,Bhubaneswar,
India
brojokishoremishra@gmail.com
santosh.satapathy@gec.edu.in

Abstract— **In last two decades several educational institutes have started gaining momentum while many of them are in self financing mode. Every institute wants to have good student strength to allow a smooth academic session. This paper proposes the use of machine learning techniques in educational domain to enhance the quality of student admissions in any higher educational institute. The focus of this paper is to identify those admissions inquires which most likely to turn into actual admissions. The result of analysis will assist the academic planners to focus their efforts on the set of students that are likely to take admission in the institution after initial enquiry.**

Keywords—Machine learning; Clustering; K–Means algorithm; Euclidean distance component;

I. INTRODUCTION

Quality based education is the right of every individual. Nowadays everywhere educational institutions have gathered momentum in large numbers. Due to the sudden rise of such institutions student admission system has been affected. Good quality students need to select universities that impart quality education at a reasonable cost. Today students have plenty of options as to which university they want to get into. Even below average students can get an entry in an educational institute with ease. As a result the education system has become more vulnerable in today world. Some institutes have the provision of entrance tests and two level interviews to filter good students but still many institutes allow direct admission too. Machine learning is an active area that may be suitable to implement in such scenario. The entire student admission system can be automated with the help of machine learning applying knowledge mining. This paper is concerned with the implementation of knowledge mining to enhance the admission process in any institute motivated upon direct admission and tries to identify those admission inquiries which most likely turn into actual admissions in future. With the help of knowledge mining educational institutes can focus on the most important information in the data they have collected when admission inquires forms are filled. It discovers information within the data that queries and reports can't effectively reveal. After gathering data from the admission forms filled by students seeking admission collected over years, this technique need to be applied to determine set of patterns of students seeking admission in college. Various techniques such as clustering and association rule mining can be of great help to discover the main features from the admission details of students and possibly use them for future prediction. Our research focuses on analysis of k-means clustering algorithm as a simple and efficient tool to analyze the admission taken by the students in previous years [1].

II. RELATED WORK

A CET (common eligibility test) based decision support system was proposed by Miren Tanna[2] which introduced a new concept. In [3] Sujit Kumar Yadav et.all and his colleague implemented a decision tree architecture to forecast performance of students and computed students's eligibility for MCA course. Malaya Dutta Borah, Rajni Jindal and Daya Gupta [4] had proposed a heuristic function of C4.5 algorithm and used it to predict the branch of admission for students. Researcher have also build hybrid recommender system which can be used for the admission process for students[5]. Malaya Dutta Borah, Daya Gupta and Gokul Pandey introduced two modification on C4.5 algorithm and compared its results with original C4.5 algorithm. They have done branch prediction and future grade prediction in their work. Simon Fong, Yain-Whar Si, Robert P. Biuk Aghai had used back-propogation algorithm and C4.5 algorithm for the student admission process.

978-1-5090-4620-1/16/$31.00 ©2016 IEEE

III. METHODOLOGY

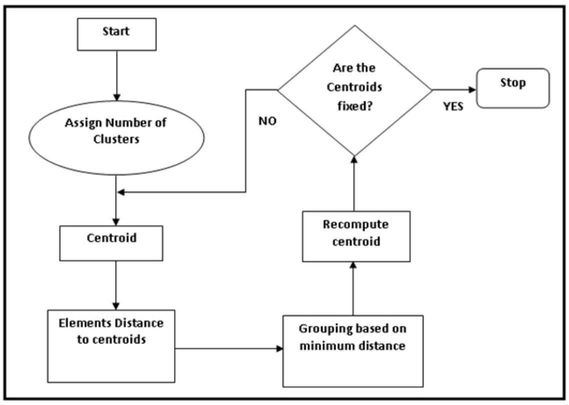

Fig 1. K-means Clustering technique

Clustering is the process of grouping elements so that elements in the same group called a cluster are more identical to each other than to those in other groups. Clustering is the process of partitioning a group of data points into a small number of clusters. For instance, the items in a supermarket are clustered in categories (butter, cheese and milk are grouped in dairy products). This is a qualitative kind of partitioning. A quantitative approach would be to measure certain features of the products, say percentage of milk and others, and products with high percentage of milk would be grouped together. In general, we have n data points xi,i=1...n that have to be partitioned in k clusters. The goal is to assign a cluster to each data point.

IV. K-MEANS CLUSTERING METHOD

A very simple yet popular unsupervised clustering technique used in many real time applications is K-Means clustering. A predefined number of clusters are denoted at prior to classify a given dataset. The main concept is to determine a unique centroid for every cluster. These centroids are arranged far away from each other. All points that belong to a particular dataset are associated with the nearest centroid. The first step is completed when no point is left such that initial grouping is done. Then new centroids are recomputed as bar centers of the clusters that result from the previous round. A new binding is to be done between the identical data set points and the closest centroids after the new k centroids are determined. It is observed that at every round the k centroids alter their position through a loop till no more variations are noticed which implies there is no changes in position of centroid. At the end the prime motive of K-Means algorithm is to reduce a squared error function called an objective function. The objective function is denoted as:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 ,$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centres.

A. Algorithm Steps

Consider X=X$_1$*,X*$_2$*,......Xn be the set of data points and V=V*$_1$*,V*$_2$*,......V*$_n$ *be the set of data centers.*
1.Let ' C' be the number of cluster centres selected randomly.
2.Calculate the eucledian distance between each data set and cluster centre.
3.Assign each data point to its respective nearest cluster centre.
4.Recalculate the new cluster center by using the following formula: $Vi=(1/Ci)\sum_{j=1}^{Ci} Xi$
Where Ci represents the number of data points in ith cluster.
5.Recalculate the distance between each data point and all the newly obtained cluster centers.
6.If no data point was re-assigned then stop the process, else go to step 3.

B. Benefuts of K-Means Algorithm

- Latency is low and is more robust and easily understandable.
- Comparatively simple and efficient than other algorithms.
- It yields optimum result when dataset provided are distinct and quite separated from each other.

V. RESULTS

A In our study K-Means algorithm is implemented in developing a quality based admission system for students. The entire work is conducted in WEKA 3.7.12 software. Our dataset comprises 13 different features collected during student admission process. A collection of 129 records were aggregated by consulting different parents. The dataset was transformed into .arff format to be implemented in WEKA software. K-Means algorithm follows a very simple yet effective method in classification of a given dataset samples into predefined number of clusters. Three parameters are involved in this technique which includes number of clusters (k), initialization of cluster and distance metric. Selecting the number of clusters is the most vital task. In our work two distinct clusters are created i.e cluster 0 with 68 admission inquiries and cluster 1 with 61 admission inquiries. It has been clearly indicated that at cluster 0, 57 out of 61 inquires are being converted into admission while at cluster 1, 51 out of 61 inquires are converted to admission. It has been clearly inferred in figure 3 that the model developing time is very less which is 0.02 seconds while completing a complete 4 iterations. Figure 6 depicts the success rate of student's admission in both clusters. Although the success ratio of both clusters (83.82 and 83.60) have a minor difference but there is need to focus more on the first cluster to improve chances of admission in institute.

Time taken to build model (full training data) : 0.02 seconds
Total Number of instances: 129 [Cluster 0/Cluster 1 = 68/61]
Number of iterations: 4
Within cluster sum of squared errors: 669.0

Fig 2. Result details after implementing K-means

Features	Domain Range
Friends-Feedback	Poor, Average, Good, Very Good
Motivation	Good_Job_Opportunities, College_Reputation, Your_Interest, Reference
Awareness	Newspaper, Friends, Senior, Any_publicity_hording, Internet
College-Infrastructure	Poor, Average, Good, Very_Good
Entrance-Rank	1-20000, 20001-50000, 50001-100000, 100000+
Graduation %	50-59, 60-74, greater_than_74,
Family-Income	300000-700000,Above_700000
Family-Member	3,4,5
Father-Occuption	Domestic_Buss, International_Buss Center_Govt_Emp, State_Govt_Emp, Priv_Emp
Mother_Occuption	Home_Maker, Govt_Job, Private_Job
Father_Qualification	12th, Graduation,Post_Graduate, Post_Graduate_Professional
Mother_Qualification	12th,Graduation,12th,Post_Graduate_Professional ,Post_Graduate
Admission_Taken	Yes, No

Fig 3. Student Admission system dataset

Classes to Clusters:
0 1 <-- assigned to cluster
57 51 | yes
11 10 | no
Cluster 0 <-- yes
Cluster 1 <-- no

Clustered Instances		
0	68	(53%)
1	61	(47%)

Fig 4. Detail Cluster classes and instances

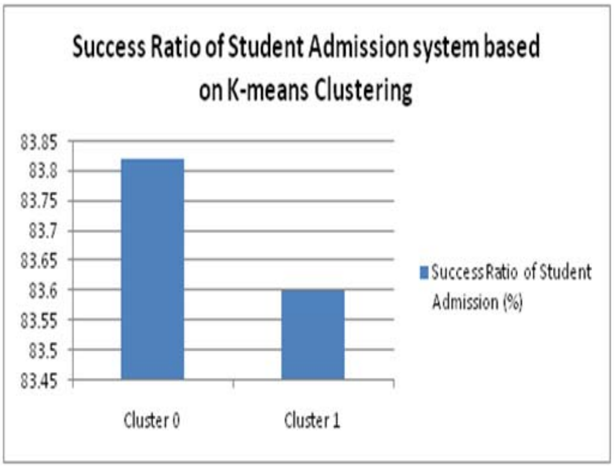

Fig 5. Student admission system success rate based on K-means clustering

VI. Conclusion

The main focus of this paper is a simple implementation of K-Means clustering algorithm where data gathered from the admission form filled by candidates are examined and knowledge is extracted. Here two different clusters are created with 68 and 61 admission inquiries respectively. It is noticed that in the first cluster 57 out of 68 inquiries are transformed into admission while in the second cluster a total of 51 out of 61 are getting transformed into admission. Though the success ratio suggests that there is a marginal difference in both clusters, more emphasis is to be given to the second cluster to enhance more chances of admission in the institute. This mechanism will be a boost to many organization to oversee student's admission details that are willing to take admission over the years. This system model will also be helpful in finding the faulty areas of institute due to which admissions have reduced over time and hence will be beneficial in formulating steps to improve admission performance in the subsequent sessions. Thus our proposed model is helpful in listing out students on which main focus is to be drawn to transform inquiry into admission.

References

[1] Arora K. Rakesh, BadalDharmendra," Evaluating Student" s Performance Using k-Means Clustering", IJCST Vol. 4, Issue 2, April - June 2013.

[2] MirenTanna, "Decision Support System for Admission in Engineering Colleges based on Entrance Exam Marks", IJCA, vol 52-no 11, August 2012.

[3] Surjeet Kumar Yadav, Saurabh Pal, " Data Mining Application in Enrollment Management: A Case Study", IJCA, vol 41- no 5, March 2012.

[4] Malaya Dutta Borah, Rajni Jindal, Daya Gupta, "Application of Knowledge based decision technique to prediction student enrollment decision", IEEE, 2011, pg. 180 - 184