

Analysis on Student Admission Enquiry System

Ameya Joshi¹ Manjusha Singh²

¹M.Tech Scholar ²Assistant Professor

^{1,2}Chhatrapati Shivaji Institute of Technology, Bhilai, DURG

Abstract— Data mining referred to extracting the hidden predictive information from huge amount of data set. Recently, there are number of private institution are came into existence and they put their efforts to get fruitful admissions. In this paper, the techniques of data mining are used to analyze the mind setup of student after matriculate. One of the best tools of data mining is known as WEKA (Waikato Environment Knowledge Analysis), is used to formulate the process of analysis.

Key words: Data Mining, Clustering, KDD, WEKA

I. INTRODUCTION

Data mining define the process of extracting or mining the knowledge from large amount of data. The term data mining is formally known as Knowledge mining from data or Knowledge data mining. It is the process of analyzing data from different vision and detailing it into useful information. The overall goal of Data mining process is to extract hidden information which are potentially useful and structure it for further use.

Data Mining is an analytic process designed to explore large amounts of data typically business oriented or market related also known as "big data" in search of consistent patterns and/or appropriate relationships between objects, and then to validate the findings by applying the detected patterns to new subsets of data. The main goal of data mining is prediction and Classification. Data mining is the most common type of direct business applications. The fundamental and perspective view provided by data mining move beyond the past event of retrospective tools typical of decision support systems. Business questions can be easily answered by data mining tools that traditionally were too time taking to resolve.

Data mining techniques can be used rapidly on various existing software and hardware platforms to enhance the quality of existing information resources that can be integrated with upcoming products and systems as they are brought on-line. In this paper we are focusing on techniques of data mining: Clustering technique is the crucial task for distinguish the data in to sub sets, Classification technique is used for predicting the outcomes based on the given inputs. By combining these two techniques we are analyzing the activities of the student behavior. There are various applications of Data mining such as in marketing, web mining, scientific data analysis and many more.

A. Data Mining Consists Of Five Major Elements:

- 1) Extract, transform, and load data onto the data warehouse system.
- 2) Store and manages the data in various dimensions of schema
- 3) Provide data access to business analysts, executives and information technology professionals.
- 4) Analyze the data by application software.
- 5) Present the data in a split charts, diagrams, graphs or table.

Clustering is the task of nominating a set of objects into batch (called cluster) so that the objects in the same clusters are more identical to each other to those in other cluster.

Clustering is an ultimate task of explorative data mining, and a common technique for statistical data analysis used in various fields such as information extraction, pattern evaluation, machine learning techniques, image analysis. We are using WEKA tools which provide a better solution for analysis and help to pretend the behavior of the students. WEKA is a acquisition of machine learning algorithms for data mining activities.

B. The Key Features Of WEKA Are As Follows:

- it facilitate different algorithms for data mining and machine learning
- is is open source and implemented in java
- it is platform-independent and portable
- it provides flexible facilities for scripting experiments.



Fig. 1: WEKA GUI

II. LITERATURE REVIEW

Rakesh kumar arora, (krishna engineering college ghaziabad, up) Dr. Dharmendra badal (bundelkhand university, ghaziabad, up)(2013) has presented to analyze the data obtained from the admission form filled by admission seekers. Determining the reasons for decline in quality of admissions taken in the institute over the year. In this paper, author explained the tools and the methods applied for admission management procedure.[1]

Dr. Sudhir B. Jagtap SVITN (Udgir) M.H. Dr. Kodge B. G. SVITN (Udgir) M.H have searched Knowledge extraction from database is become one of the key process of each and every organization for their development issues and plays a vital role in e-governance for future planning and development issues.[2]

Narendra Sharma, Aman Bajpai , Ratnesh Litoriya Jaypee University of Engg. & Technology(2012) says that k-means clustering algorithm is simplest algorithm as compared to other algorithms. We can't required deep knowledge of algorithms for working in WEKA. That's why WEKA is more suitable tool for data mining applications.[3]

R. ROBU, C.HORA University “Politehnica” Romania (2012) analysis such as images, signals through data mining techniques aims to extract useful knowledge from the data. [4]

Suhem Parack, Zain Zahid, Fatima Merchant(2012) searched about the algorithms such as k-means or Apriori offers an effective way of profiling students which can be used in educational systems. We can easily group the students, identify hidden patterns about their learning style.[5]

Anand, Vaibhav, Priti Maharashtra, India(2014) data mining techniques is potentially useful and ultimately understandable. It involves task like anomaly detection, classification, regression, clustering etc.[6]

III. CLUSTERING

Clustering is the activity of composing objects into groups whose members are similar in some properties. A cluster is a group of objects which are “identical” between them and are “dissimilar” to the objects belonging to other clusters.

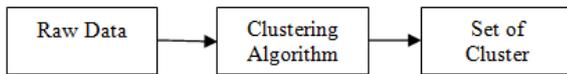


Fig. 2: Process of Clustering

A. K- Means Clustering:

K –Means clustering is one of the best algorithm in terms its simplicity and efficiency. K-means clustering aims to splitting n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.[12].

In the past work K-Means is being used for the analysis of the data obtained from the admission procedure. The data set is collected from the college of Ghaziabad. The number of student involved is 129 and is done on the basis of different parameters. The author concluded that this model will help to assist the management persons to filter good student who are willing to join the institute.[1]

B. Algorithm Steps for K-Means Clustering:

Input:

D = {d1,d2...dn} set of n data items.

K = number of desired clusters

Output: A set of k clusters.

Steps:

- 1) Enter the number of cluster K.
- 2) Find out the centroid.
- 3) Calculate the distance from object to centroids.
- 4) Check object moved to the specified group or not?
- 5) If yes move the object to the respective cluster.
- 6) If not, then recalculate the distance from object to centroid until the convergence criteria meet.

But each and every good thing come up with little disadvantages, K-Means clustering also facing difficulties while comparing quality of the cluster and it does not work efficiently with non-globular clusters.

IV. CLASSIFICATION

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of

classes or groups. It is considered as supervised learning in which we have a Training set containing data that have been previously categorized.

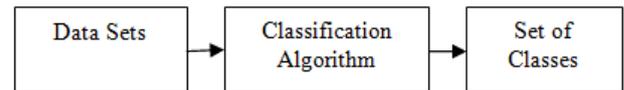


Fig. 3: Classification Strategy

A. Naïve Bayes:

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem with strong (naive) independence assumptions between the features. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem.

B. Bayesian Rule:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Posterior = $\frac{\text{Likelihood} \times \text{prior}}{\text{Evidence}}$

- P(C): independent probability of C: prior probability
- P(X): independent probability of D
- P(X|C): conditional probability of X given C: likelihood
- P(C|X): conditional probability of C given X: posterior probability.

On the basis of clustering and classification technique we can judge the attributes of the students. Previously in admission procedure K-means clustering is used to analyze and summarize the data and it will also assist the admission planner to monitor the activity of the admission seekers. We are expecting to merge this technique which will help to classify more efficiently

V. OBJECTIVE OF STUDY

- To Study of existing data analysis clustering technique.
- To Analyze complexity and outlier issue in Algorithm and to find point of complexity in Algorithm.
- To deploy the clustering methodology in the management process.
- To compare the results with K-Means algorithm in terms of efficiency and throughput.

Here efficiency is in terms of number of student converting their enquiries in to admission and throughput deals with the conversion rate of the students.

VI. RESULTS AND CONCLUSION

On the basis of this survey, we found that K-Means clustering is the easiest algorithm to implement. WEKA is the efficient tool of data mining for analysis which does not required deep knowledge about the algorithms. This will help in identifying the set of students that need to be focused to actually convert the enquiry into admission. Management will get beneficial inputs. We wish to develop the algorithm with the concept of clustering and classification, and to deploy the same in admission procedure to fulfill the requirement and to compare the result with existing algorithm and we try to observe in terms of accuracy and throughput.

REFERENCES

- [1] Rakesh Kumar Arora, Dr. Dharmendra Badal “Admission Management through Data Mining using WEKA” , International Journal, Volume 3, Issue 10, October 2013. Page No. [674-678].
- [2] Dr. Sudhir B. Jagtap and Dr. Kodge B. G., “Census Data Mining and Data Analysis using WEKA”, International Conference in “Emerging Trends in Science, Technology and Management-2013, Singapore. Page No. [35-40].
- [3] Narendra Sharma, Aman Bajpai, Mr. Ratnesh Litoriya “Comparison the various clustering algorithms of WEKA tools’ , International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, Volume 2, Issue 5, May 2012) Page No-[73-80].
- [4] R Robu and C. Hora, “Medical data mining with extended WEKA ” , IEEE 16TH International conference on intelligent Engineering Systems , June 13-15, 2012 page No. [347-350].
- [5] Suhem Parack, Zain Zahid Fatima Merchant., “Application of data mining in Educational databases for predicting academics trends and patterns. ”, IEEE-2012.
- [6] Anand V Saurkar, Vaibhav Bhujade, Priti Bhagat, Amit Khaparde “International Journal of Advanced Research in Computer Science and Software Engineering, Volume-4, Issue 4, April 2014 ISSN: 2277 128X . Page No-[98-101].
- [7] M Bhoomi, “Enhanced K-Means Clustering algorithm to reduce time complexity for numeric values ”, International Journal of Computer Science and Information Technologies, Volume 5(1), 2014, ISSN: 0975-9646 page No. [876-879].
- [8] <http://www.ijcsit.com/docs/Volume%205/vol5issue01/ijcsit20140501189.pdf>
- [9] http://www.iaeng.org/publication/WCE2009/WCE2009_pp308-312.pdf
- [10] <http://documents.software.dell.com/Statistics/Textbook/Data-Mining-Techniques>
- [11] <http://esatjournals.net/ijret/2013v02/i11/IJRET20130211019.pdf>
- [12] https://en.wikipedia.org/wiki/K-means_clustering
- [13] <https://www.google.co.in/search?q=flowchart+of+kmeans+clustering+algorithm&biw=1024&bih=639&tbm=isch&tbo=u&source=univ&sa=X&sqi=2&ved=0ahUKEwiy2fyh2KDJAhUBH5QKHxVGBKIQsAQIIA>