Teaching Portfolios Using Data Mining Based on WEKA Platform

Md. Alamgir Kabir¹Md Shahiduzzaman²

Abstract—The quality of education of a university depends on its teacher’s major strength and their performance. In a country there have several kinds of universities but their standards of educational qualities are not same. A university will be developed and well known around the world by its educational quality which is completely depend on the quality of an instructor’s teaching proficiency and his major strength as well as his motivation and devotion to the university and it’s students. Therefore, in this paper we compared and analyzed the faculty teachers’ major strength on the specific courses and their teaching proficiency and also their intimacy with the students by getting the feedback from the students of Computer science and engineering department in daffodil international university through online faculty evaluation form and then we do the data mining analysis using the data mining tools WEKA.

Key Terms: - Data collection, Data Preprocessing; Data Mining Application WEKA; Implementation; Different kinds of data mining algorithms; Classification analysis; Cluster analysis; Association rules;

1. INTRODUCTION

A teaching portfolio is a collection of documents which give us the information about the teachers’ major strength and his effectiveness as a teacher and also his teaching style as well as his teaching proficiency. Most of the well-known educational institutions in the world are taking their decision regarding faculty appointments on the base of following teaching portfolio. To demonstrate an effective teaching system is now becoming an important factor at all education institution in higher studies. A prudent and proficient instructor can lead a nation to the acme of its success. And to make a country as a developed country at first we have to assure the quality of the education system. It profoundly enhances the novel virtues of human beings and their prosperity. In a country there have thousands of educational institutes but the qualities of the institutes are varying one university to another. And these differentiations happen because of the teachers of the university who are governing the institutes and giving their precious lecture to the students. To keep pace with the modern age we have to give the students a standard quality of education which will give them a proper knowledge of any specific courses and make them confident to struggle with their life and so that they never be frustrated and they know how to persevere with challenges. And also we should make sure the satisfaction of the students about their education.

2. BACKGROUND

Teaching portfolio is as similar as an artist portfolio which does assessment of artist strength of imagination, his unique way of style and his individual qualities. A teaching portfolio enables a person to identify a proficient teacher and his way of pedagogy and also his devoutness and effectiveness to his profession. Every year thousands of graduates are receiving their degrees but they feel frustrated about their carrier and confidence. And they also suffer dilemma about their job. Because they didn’t get a proper education that couldn’t make them adept to their specific major and for that terrible situation the instructors of the specific institutions are responsible alongside the students because they couldn’t give them a proper knowledge of the specific courses.

A teaching portfolio can be used in other purposes as well. It can be used in other department to assess any employee or any member of the company or the team leader by making an effective
evaluation survey form which will help not only to allocate the right person to the right position but also to choose the harbinger and the greatest mentor to their students and their followers. And for that reason the data mining is very important to analysis the data’s so that it could get some prediction where the data’s are not smooth. Due to the exponential growth of data especially in areas such as business, some controversial theory data mining is very important tools. And it converts the large wealth of data into intelligence.

3. DATA COLLECTION

Data mining is used to predict and to group the data’s on the base of their features and it is also known as KDD (knowledge Discovery in Data) which deals with the useful and previously unknown information from the raw data. For doing data mining on teaching portfolio we need a lot of data’s to do analyze. And for that reason we have collected a sufficient amount of data’s from the students of the department of computer science and engineering in daffodil international university by creating an online teaching evaluation form where they can give their feedback to us and also they can do comment of their own ideology. To create this online teaching portfolio survey form we used the Google doc and we used 5-likert scale to define the values of each option with the range of 1 to 5 and we have created 23 very effective questions which will cover the expectation of the students about their teacher.

![Teaching Portfolio Data Mining](image)

And then we shared this evaluation form to our official Facebook page and our university computer science department groups and the university dormitories group. And we got 308 responses from the computer science and engineering department.

4. DATA PREPROCESSING

An efficient and effective result of data mining lies in the good quality of data. And for that reason we have to cleansing the data before analysis and it is called data preprocessing which makes the data’s in an understandable format that is the important task of data mining. In a large database there have different kinds of data’s like noisy data, corrupted data, and inconsistent data. To eliminate those data’s we have to follow a methodology to accomplish it. For data preprocessing at first we have to do

4.1 DATA CLEANSING

For cleansing the data’s at first we have to fill up the missing values and smoothing the noisy data’s and also resolve the inconsistency of the data’s into the database.

4.2 DATA INTEGRATION
Group the data’s on the base of their instances and resolve the conflicts data’s.

4.3 DATA TRANSFORMATION

In some data implementations it may require to normalize the data’s and also have to do aggregate and generalized.

4.4 DATA REDUCTION

Due to get some specific consequences it also may require to reduce the data’s from the wealthy of large data warehouse.

4.5 DATA DISCRITIZATION

Data discretization is the part of data reduction and it replaces the numerical values with nominal ones to get the unique data set.

After following this data preprocessing steps we get the very authentic and pure data’s from the large data warehouse and we can start to do our mining.

5. DATA MINING APPLICATION WEKA

There have different kinds of data mining tools like WEKA, Orange, Rapid Miner, R-Programming, KNIME, and NLTK. From all of this we used WEKA (Waikato Environment for Knowledge Analysis) data mining tools which were developed by the University of Waikato in New Zealand and it was first implemented in 1997[3]. And it is so much popular suite of machine learning software written in JAVA [3] [7].

6. IMPLEMENTATION

After getting all the responses from our Google doc we converted those responses into .csv format file to get the standard spreadsheet. And then we began to do our next step. In data mining it may require to replace all string data’s with numerical values. And for our project purpose we replaced all the string values with numerical ones and as we used the 5-likert scale in Google doc to make the evaluation form so we set the values for each option within the range of 1 to 5 where 1 is Below average, 2 is average, 3 is good, 4 is very good and 5 is excellent. And we also add a class set for each data item sets by using the binary values 0 and 1 according to the rating of the item sets where we defined 0 in the class set which has 1 and 2 ratings more than 3, 4 and 5 and we defined 1 in the class set where the item sets are 3, 4 and 5 more than 0 and 1 rating item sets.
There have different kinds of algorithms has been used for different purposes in data mining. And we applied classification, clustering, association rules and decision tree to find out which gives us the efficient result by using the WEKA tools.

7. DIFFERENT KINDS OF DATA MINING ALGORITHMS

7.1 CLASSIFICATION

Classification is a supervised learning technique and it is use for grouping the same instances on the based on its unique features where there have the training sets of data and response variables. And by classifying we can get prediction about the new instances as it is supervised training set. Classification can be classified into two types: Supervised and Unsupervised [4].When the data’s are known in advance and the classes are pre-defined then we call it supervised classification and when we don’t know the data’s or object then we can call it as unsupervised classification [2][4].

As we have the data’s in advance and we used the nominal and binary values so for that reason we used the algorithm named J48 -C 0.25 -M 2 and it is supervised classification.J48 is the optimized implementation of C4.5 algorithms [1] and the output of J48 is decision tree [1]. A Decision tree is similar to the tree structure having root node, intermediate nodes and leaf node[1]. Each node of the tree give inkling of the decision which lead us to the final result[1]. By using J48 classification algorithm we got a very efficient outcome where we have 308 instances and 24 attributes and we got correctly classified instances 274 which is 88.961% and we also got incorrectly classified instances 34 which is 11.039%
Clustering method is applied when the dataset are unknown and it is an unsupervised learning technique. It uses an iterative refinement technique [10]. In clustering it group the data's based on the sum of square of distances between the given data and cluster centroid and the elements which have the minimum distance [10]. It doesn't require the training set. As we have our training set so it didn't give a very efficient result. For clustering we used SimpleKmeans algorithm where it divide the whole data set into two cluster 0 and 1 and then we got the result where 262 instances are negative binary 0 which is total 85% and also we got the result where 46 instances are positive binary 1 which is total 15%.

To compute the minimum distances between the data objects there have two types of technique used by clustering method. One is Manhattan technique and the other is Euclidian technique. In Manhattan distance it considers the sum of the distances of the whole data objects in the data set. The Euclidian distance computes the shortest distance between two data objects [5]. The formula of the Manhattan and Euclidian distance are given bellow

\[ d = \sum |x_i - y_i| \]

\[ d = \sqrt{\sum (x_i - y_i)^2} \]
7.3 ASSOCIATION RULES

To find out the relationship between variables in the given dataset association rules need to be applied. It discover the relationship among the variables in the data warehouse and also it find out the minimum confidence level and the minimum support [9]. Association rules use the if else statement to disclose the relationship between the variables and for our data set we used Apriori algorithm and we got 62 instances with minimum support is 0.2 and minimum confidence is 0.9 and we also got three large item sets with the size of L(1): 49, L(2): 168, L(3): 87 respectively.

8. CONCLUSION

In this paper we compared different kinds of data mining technique on the base of students' feedback about the teachers evaluation survey form. Regarding the performances of classification, clustering and association rules which we used on our the data set we got different kind of experimental result which makes us aware about which algorithms is best for getting the efficient result for training data sets. We have shown that classification rules give us a very effective result but other’s do not. We also indicate that we have to make sure the quality of the instructor performances not the quantity of the instructors. And we have gotten that how a good quality data’s can effect on the instructor’s performances.

9. REFERENCES

[1] Yugal kumar1 and G. Sahoo2. 2012. Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA.


[5] Prakash Singh1, Aarohi Surya2. 2015. PERFORMANCE ANALYSIS OF CLUSTERING ALGORITHMS IN DATA MINING IN WEKA


AUTHORS

Email: alamgirm800@gmail.com

MD. Alamgir Kabir received his Bachelor of science degree in computer science and engineering from Daffodil international university and his research interest in Data mining, Image processing, computer networking, programming language, artificial intelligence, Image processing and algebra.

Email: shahiduzzaman.diu@gmail.com

MD. Shahiduzzaman received his Bachelor of science degree in computer science and engineering from Daffodil international university and his research interest in Data mining, Image processing, web developing and computer networking and in programming language and artificial intelligence.