Placement Prediction Decision Support System using Data Mining

Joshita Goyal[1], Shilpa Sharma[2]
M.tech Student, Department of Computer science and Engineering, Lovely Professional University, Phagwara, India[1]
Assistant Professor, Department of Computer science and Engineering, Lovely Professional University, Phagwara, India[2]

***Abstract*—Placement Prediction is important factor now a days for every Institute. Placement prediction of every student involves multi - level analysis that what level of the interview process a student is likely to clear. This will help Institutes to increase placements and encourage more number of admissions. For this we are proposing a placement prediction decision support system using Naïve Bayes classifier. We will be improving Naïve Bayes by embedding relief feature selection technique for better predictions. With this system, Institutes will be able to know that whether a student is eligible for placements then it will tell that what level of the interview process such that aptitude, programming, group discussions and HR or technical interview a student is likely to clear. The dataset used for analysis contains registration number, name, program, 10th marks, 12th marks, cgpa, standing arrears, logical reasoning, quantitative ability, English competency, coding ability and many other parameters for performance evaluation.**

***Index Terms*—Data mining, Knowledge discovery process, Naïve Bayes, Relief algorithm**

I. INTRODUCTION

Data mining has widened its area and is becoming essential in each and every part of software development. Data mining refers to exploring huge databases for hidden information. It is important process to undertake for future progress. As data mining is application in almost all areas like forensics, crime, health and agriculture. Likewise it has shown its importance in education also[1]. In this paper we will be applying the concept of data mining in placement prediction of students.

Data mining in student performance analysis: For institutes and students to know about what could be there placement status without giving any special placement related inputs require proper data analysis that can only be implemented with the use of techniques and tools used in education data mining. This is an attempt to improve the quality of placements and Institutes value. In this work we proposed a methodology which can prove fruitful for our intended goal. We have collaborated classification and association for this. We have used Naïve Bayes classification embedded with relief feature selection technique to get the analysis. Further rules will generated by the novel technique of association rule mining called apriori algorithm from output of improvised Naïve Bayes. For these things to be implemented, Data mining follows a general procedure called knowledge discovery process.

A. Knowledge discovery process

It form the basis of data mining. Its process which includes extraction of useful information from the data available. Following is the procedure:

1. Cleaning: In this phase, we remove unwanted data from the dataset and fill the missing data.
2. Integration: From various databases, data is integrated from data mining process.
3. Selection: Data is selected for analysis.
4. Data transformation: In this phase, data is transformed to a form which is applicable for data mining process.
5. Mining Operations: Various data mining techniques are applied so as to get hidden knowledge.
6. Result evaluation and presentation: Now the result is evaluated from previous operation and presented in graphical form.

B. Algorithms used in Data Mining

There are various algorithms used in data mining techniques. Few of them are listed below:

1. Linear regression: It is used to evaluate the real value based on continuous variables. Relation between dependent and independent variable are found using this.
2. Logistic regression: Discrete values are evaluated based on independent variable in this.
3. Decision tree: It is supervised learning algorithm. It splits the variable into two or more homogenous sets.
4. Support vector machine: Its classification algorithm in which each attribute of every data item is plotted on n dimensional graph.
5. Naïve Bayes: This classifier assumes that the presence of particular feature in a class is unrelated to the presence of any other feature.
6. K- Nearest neighbor: It a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbors. The case being assigned to class is the most common amongst its k nearest neighbors measured by distance functions.
7. K-means: It's unsupervised algorithm which is a way to classify a given data set through a certain number of clusters.
8. Random forest: It's another form of decision tree. It contains collection of decision trees. To classify a new object based on attributes, each tree gives a classification and we say a tree "votes" for that class. The forest chooses the classification having the most votes.
9. Apriori: It's an association technique which finds frequent item sets from large databases. It helps to improve businesses.

II. LITERATURE SURVEY

Following are the various studies and research work done in placement prediction.

Senthil Kumar, Divya Bharthi and abhijith work on to form placement analyzer using decision tree classifier. They used sci-kit libraries. They compared the results using weka and data miner and found that their results obtained using sci-kit was better. Accuracy of decision tree was

also compared with other data mining algorithms and results favored decision tree than other classifiers as accuracy was better with decision tree[2]. Ravi kumar and Jayanthi formed a system to find students' eligibility for placements. They took data of students containing academic details. They used fuzzy inference system for this[3]. K.Nasaramma, M.Bangaru Lakshmi, D.kiranmayi worked on to find which algorithm is better in predicting students' performance. They used feature selection on their dataset and compared the results of C4.5 and C5.0 in weka. It was concluded that C5.0 proved to be better than the former[4]. Ankita Kadambande, Snehal Thakur, Akshata Mohol, Prof A.M.Ingole collaborated to build a student performance prediction system in which they created three modules student, staff and admin. Wherein students will be appearing for test and can view their performance, staff can view the performance of students and admin can upload the events like test etc. This research was to perform a new system over an existing system which was that data was processed in weka using j48 algorithm. In proposed system the data stored in apache database was processed using support vector machine. And final result was evaluated and students performance was predicted[5]. Ravneet Kaur and Krishan Bansal, these authors collaborated to create a robust placement prediction system using scrum analysis. They gathered placement data and process it using k-mean and neural network classifier. Final they done re-sectioning of students[6]. A. Parkavi, K. Lakshmi worked to find that whether a student needs remedial classes for a particular subject or not. They did analysis to find out course level knowledge of the student. For this they used blooms taxonomy to for assigning ranges for evaluation of particular question. They used association rule mining and regression analysis. Regression analysis helped in too find that whether students' performance varied from the blooming level or not for the two tests conducted . It was concluded that it is good approach for course level evaluation[7]. Tribhuvan A.P. , Tribhuvan P.P. and Gade J.G. worked on to predict student performance using Naïve Bayesian classfier. They have used weka as data analysis tool. Different classifiers were used which were naïve bayes, ID3 and decision tree. ID3 proved to be the best algorithm for the dataset they took which gave higher accuracy than naïve bayes. Naïve bayes gave better results than decision tree[8]. Fahad Razaque, Nareena Soomro, Shoaib Ahmed, Safeeullah, Javed Ahmed, Natesh and Huma did student academic performance analysis using clustering and classification. They applied simple K-mean to form clusters of different students and then applied naïve bayes for classification. With this they rated students based on academic performance, attendance, extra coaching and performance analysis[9]. Amanpreet kaur and Amanpreet kaur Batth collaborated to predict placed and non-placed students based on their placement status. They grouped the students into two dissimilar clusters and classified them using SVM. They proposed KNN classifier tho predict placed and non-placed students[10].Tansen Patel and Anand Tamrakar implement various clustering algorithms on student performance dataset. They results showed that simple k-mean, Farthest first and filtered cluster algorithm took similar execution time whereas hierarchal clusters took maximum time for faster execution[11].

III. PROBLEM FORMULATION

Student performance prediction has been done using various data mining techniques. By applying data mining techniques on student data, the extracted knowledge from data describes the student performance. This knowledge will help to improve the education quality, student's performance and to decrease failure rate. In the existing work, the analysis of student's placement data has been done using naïve Bayes algorithm. In this work, only the individual naïve Bayes algorithm is analyzed for the placement performance prediction which contributes less accuracy rate. For achieving the better accuracy of naïve Bayes algorithm, feature selection technique is embedded with the classifier in order to reduce the dimensionality of the data and to enhance the accuracy. Also, the predicted data is then used for extracting the patterns of placement performance using apriori rule mining algorithm.

Proposed method consists of 2 Phases:

1. Finding the class of the test set: Firstly, the enhanced naïve Bayes algorithm is trained using the training dataset. This trained data model is used to predict the class label for the testing set.
2. Generating the association rules using Apriori algorithm: The predicted test set is used as an analysis purpose for apriori algorithm. This will generate the frequent patterns and the rules which will help to predict the future analysis of placements of pre-final year students.

Enhanced Naïve Bayes Algorithm

Proposed a novel Enhanced Naive Bayes classification algorithm based on the all the features for determining the dependent attributes in a dataset and removing those dependent attributes, thereby reducing the attribute set to increase the classification accuracy and reduce the computational time. In order to reduce the dimensionality, Relief algorithm is used.

Naive Bayes implements probabilistic naive bayes classifier. Naive means conditional independence among attributers of features. The naive assumption greatly reduces computation complexity to a simple multiplication of probabilities. The Naive Bayes handles numeric attributes using supervised discretization and uses kernel density estimators that will improve the performance. It needs only small set of training data to develop accurate parameter estimations because it requires only the calculation of the frequencies of attributes and attribute outcome pairs in the training data set.

In order to reduce the dimensionality, Relief algorithm is used. It selects the best attributes out of all the attributes that contributes more to the prediction and ranks the attributes according to the weights calculated.The results is a weightage assigned to attributes between −1 and 1 The attributes with positive weightage is having more predictive accuracy.

The random sample is selected from the dataset then find the nearest neighbors that belongs to the same class this case is known as nearest hit and the nearest neighbor that belong to the opposite class is known as nearest miss. A change in attribute value due to change in class will increase the weight of attribute and a change in attribute value by no change in class will decrease the weight of attribute. This procedure is performed on each and every sample of the data. These all weights are then averaged in order to obtain the final resultant weightage.

Tool to be used

Weka a data analysis tool is used for data analysis. WEKA stands for Waikato environment for knowledge analysis. It contains suit of various machine learning algorithms written in java and it was developed by University of Waikato[12]. Weka is used in many applications because is freely available.

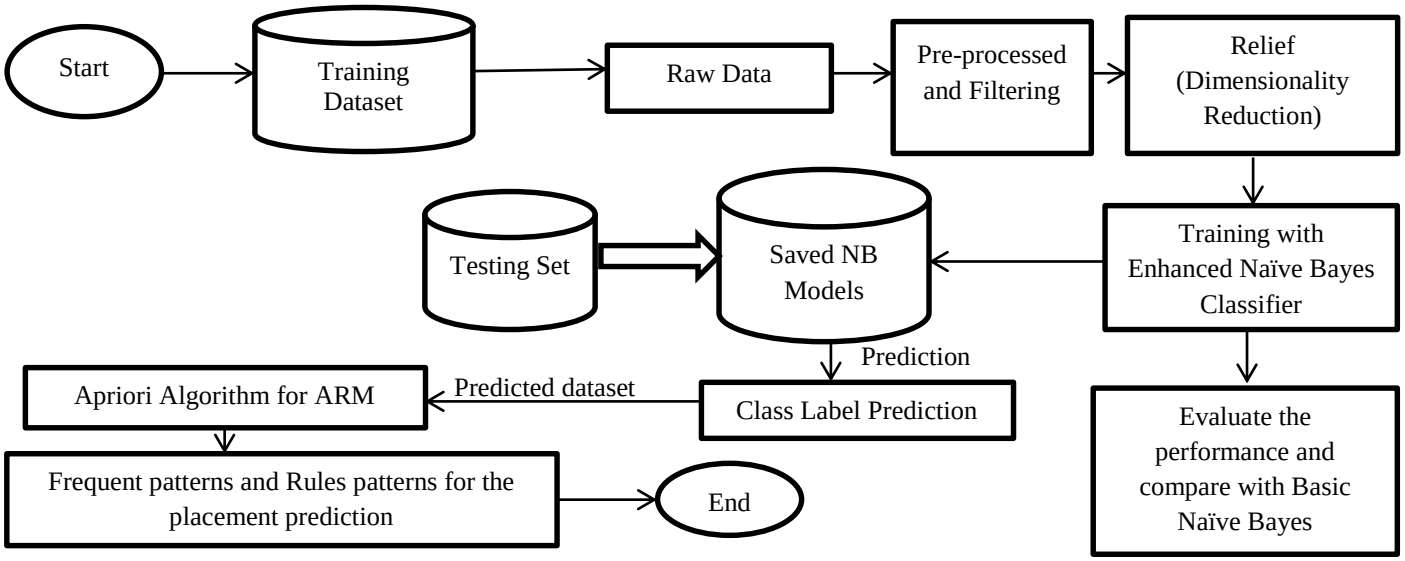

Fig 1. Flowchart for proposed system

IV. CONCLUSION

The system proposed above will let Institutes' know that whether a student is eligible for placements and also what level of the placement process a student is likely to clear. Considering Level 1 of the interview process as aptitude, Level 2 as Coding, Level 3 as Group Discussion and Level 4 as HR and technical interview. Finally the proposed technique will tell that these students are able to get placed after clearing all the level. This data will be exported to excel sheet. Based on the dataset, these values will be predicted with better accuracy because of the proposed algorithm.

REFERENCES

[1] Joshita Goyal and Shilpa Sharma,"Education Data Mining: A Review,"International Journal of Innovative Research in Science, Engineering and Technology,Vol. 6, Issue 7, July 2017

[2] Senthil Kumar Thangavel , Divya Bharathi P and Abijith Sankar," Student Placement Analyzer: A Recommendation System Using Machine Learning," 2017 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), Jan. 06 – 07, 2017, Coimbatore, INDIA

[3] Ravi Kumar Rathore and J. Jayanthi," Student prediction system for placements training using fuzzy inference system. ICTACT Journal of Soft Computing, April 2017, VOLUME: 07, ISSUE: 03

[4] K.Nasaramma, M.Bangaru Lakshmi, D.kiranmayi," Prediction and Comparative Analysis of Students Placements Using C4.5 &C5.0", K.Nasaramma et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 8 (3) , 2017,367-368

[5] Ankita Kadambande, Snehal Thakur, Akshata Mohol, Prof A.M.Ingole," Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 05 | May -2017

[6] Ravneet Kaur and Krishan Bansal,"Student's Performance Analysis Using Scrum Practices," International Journal of Engineering Research & Technology (IJERT), Vol. 6 Issue 04, April-2017

[7] A. Parkavi, K. Lakshmi," Predicting the Course Knowledge Level of Students using Data Mining Techniques," 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017, Pages: 128 – 133

[8] Tribhuvan A.P., Tribhuvan P.P. AND Gade J.G.,"Applying naive bayesian classifier for predicting performance of a student using weka,"Advamces in computational research journal, 2015

[9] Fahad Razaque, Nareena Soomro, Shoaib Ahmed Shaikh, Safeeullah Soomro, Javed Ahmed Samo, Natesh Kumar, Huma Dharejo;" Using naïve bayes algorithm to students' bachelor academic performances analysis,"2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS),2017,Pages: 1 - 5

[10] Amanpreet Kaur , Amanpreet Kaur Batth,"Classification Technique for Scrum Analysis in Data Mining,"International Research Journal of Advanced Engineering and Science,Volume 2, Issue 3, pp. 35-37, 2017

[11] Tansen Patel and Anand Tamrakar,"A data mining techniques for campus placement prediction in higher education," International Journal of science and research, 467-471, 2017

[12] Getaneh Berie Tarekegn and Dr.Vuda Sreenivasarao,"Application of Data Mining Techniques to Predict Students Placement in to Departments,"International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), Volume 3, Issue 2, 2016, PP 10-14