INTRODUCTION

In the educational institutions lots of data are collected to monitor their regular performance. In which their personal details like father name, occupation, yearly income, communication address, phone number, mother name, occupation, phone number, and passport size photo are stored, Academic details like their semester, course code, course name, grade, year of passing, co-curricular and extracurricular activity details are stored for understanding the performance of the student. Based on these details Faculty Counselor can able to counsel them for their improvement. These counseling methodologies are useful only after getting their actual data. But there is no method to predict their future performance based on their present performance. Thus in this paper we tried to apply the machine learning algorithms kNN and naïve Bayes to predict their future performance and this will be useful for the students to improve themselves to get placement through campus.

Research Purpose

Most of the current work is to predict the students’ grade in the future semester. Abeer Badr El Din Ahmed et.al. Constructed decision tree which is applied on student’s database to predict the student’s performance on the basis of student’s database. Princy Christy et.al., extensively covers recent studies in predicting student performance by implementing data mining tasks and employed the supervised learning tasks classification, regression and recommender systems. Each of the techniques in their own ways influenced the outcomes of the prediction task.

Anuradha and Velmurugan applied the classification techniques for the prediction of the performance of students in end semester university examinations. Particularly, the decision tree algorithm C4.5 (J48), Bayesian classifiers, k Nearest Neighbor algorithm and two rule learner’s algorithms namely OneR and JRip used for classifying the performance of students.

Vivekananthamoorthy, et. al., they conducted to understand how young university students use the Social Networking Site LinkedIn and the responses were used to frame a questionnaire. Exploratory Factor Analysis was conducted based on their survey responses to identify the hidden factors associated with the indicator items in the data set. Subsequently, a theoretical model was constructed, the model helped to predict enhancement of student empowerment by using Social Media.

DISCUSSION

The large amount of data collected by the educational institutions to monitor their regular performance can be used to analyze their future performance. This can be achieved by...
the popular machine learning algorithms like kNN and naïve Bayes, shown in Fig.1.

Figure 1: Flow diagram of the machine learning process

Machine Learning Algorithms

In machine learning, build predictors that allow classifying things into categories based on some set of associated values. Many algorithms are available for automated classification, includes random forests, support vector machines, Naïve Bayes classifiers, and some types of neural networks. Simple and easy to implement is k-Nearest Neighbours (kNN) algorithm. kNN classifies new instances by grouping with the existing instances with the most similar group instances. The kNN task is divided into three major functions:

1. Distance between any two points has to be calculated.
2. Finding the nearest neighbours based on the distance value.
3. Assign the class labels based on the nearest neighbour value.
4. This kind of excel data is taken as input, verified for blank values, non-numeric values, invalid values and converted into csv format.

R programming language is used to read the csv data file as input for kNN.

5. >GVind<-sample(2,nrow(GV sample), replace =TRUE, prob =c(0.67, 0.33))

6. The index value was calculated with the probability of 0.67, 0.33.

> GVsample.training <- GV sample[GVind==1, 1:250]
In the above command first 250 rows are taken for training set.

> GVsample.test <- GVsample[GVind==2, 1:250]
In the above command first 250 rows are taken for test dataset.

> GVsample.trainLabels<-GVsample[GVind==1, 59]
Column 59 is having the detail of ‘Placed or not’ in terms of ‘1/0’. Those values are given as trainLabels.

> GVsample.testLabels <- GVsample[GVind==2, 59]
Test labels will be calculated for the ‘testdataset’ and the result can be stored in column 59.

> GV_pred <- knn(train = GVsample.training, test = GVsample.test, cl = GVsample.trainLabels, k=10)

> GV_pred
[1] 1 1 1 0 1
Levels: 0 1
The sample output shows whether the student will get placed or not 1is for Yes/0 is No, using kNN.

Another algorithm Naïve Bayes is used for the same kind of work, which is based on conditional probabilities. It uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values.

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and A represents the prior event, Bayes' theorem can be stated as follows:

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

Where,

$$P(H|X)$$ is the posterior probability, or a posteriori probability, of H conditioned on X.
P (H) is the prior probability, or apriori probability, of H.

P (X|H) is the posterior probability of X conditioned on H.

Naive Bayes algorithm in question classification proceeds by finding out the feature associated with each category. Features are the Headword occurring in questions pertaining to one particular category expressed in terms of their relevance in particular question.

>train data <- as.data.frame (sample [1:59,])


D,D,C,D,C,A,S,S,A,A,D,D,D,C,D,C,B,B,B,S,C,D,D,D,D,D,B,

In the above dataset, the input values are shown as grades in letters, and the last value is whether student is placed in terms of Yes/No.

> test data <- as. data. Frame (sample [16,])

Test data is created for the record number 16 along with all 59 attributes.

> Model <- naïve Bayes (Placed ~ IT1001 + ME1005 +
LE1001+MA1002+ME1001+EE1001+IT1002+.....+IT1049, train data)

Naive Bayes model is created based on the ‘Placed’ attribute associated with 59 attributes.

> Results <- predict(model, test data)

Test data has been stored with the dataset to be verified, that has been given as naïve bayes model to predict the result.

> Results

[1] Yes

Based on the model the placement opportunity for the given test data is ‘Yes’, means the student will get the employability opportunity. If the result shows ‘No’ then that student has to be monitored and given counseling to improve his performance.

**CONCLUSION**

The students’ employability is very much important as institution point of view as well as student point of view. In this regard to improve the students’ performance, the academic performance has been analyzed and predicted using the algorithms KNN and naive Bayes. The algorithms are applied on the data set of 250 students with 59 attributes. The accuracy obtained after analysis for KNN is 95.33% and for the naïve Bayes is 97.67%. Hence, from the above said analysis and prediction it would be better if the naïve Bayes is used to predict the student’s employability results.

**Acknowledgement:** We acknowledge with thanks Management and faculty of SRM University for the support and guidance and for providing opportunity. We also acknowledge the student community for their enthusiasm and overwhelming response to this project.

**REFERENCES**


Source of Support: Nil, Conflict of Interest: None.