International Research Journal of Engineering and Management Studies (IRJEMS)
Volume: 03 Issue: 05 | May -2019 ISSN: 2395-0126

STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING ALGORITHMS

Deeksha D[1] ,Gnanavi N[1], Krupa N[1],Navyashree S L[1], Dr.Meenakshi Malhotra[2]

[1]Navyashree S L,[2] Gnanavi N, [3] Deeksha D , [4] Krupa N, [5] Dr.Meenakshi Malhotra
[1,2,3,4,5]Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.
Email : [1] navya21aug@gmail.com , [5] meenakshi.malhotra@jyothyit.ac.in
Contact : [1] 9483102684 , [5] 9880858358

Abstract - ***Machine learning has its broad research in the field of education. The goal of this work is to use Machine learning algorithms to predict student placement. The placement of students is one of the very important aspect in educational institutions. In this work ML algorithms namely KNN, Random Forest and Bayesian algorithms are used to predict student placement during campus placement. The objective is to analyse previous year students data, train ML algorithms using this data and predict placement of current batch students. The experimentation involve 150 training data from previous batches and the machine learning model is tested for 57 test data for current batch students. The bayesian model provides the highest efficiency of 92%. The prediction results can be used by students and college authorities in advance to improve the placement of students.***

***Key Words*: Machine Learning, Student Placement, KNN, Random Forest,Bayesian**

1.INTRODUCTION

Students studying in final year of the engineering always feel the pressure of placement. Also, at timeS there are not much placements in good companies. As well as, Institutions are concerned about their overall on-campus placement of the students for each academic year. Based on the percentage of campus placements and overall reviews of the college, upcoming students who desire for joining engineering colleges will make a choice in joining a college based on previous year placements. The institutions provide seminars, workshops, guest lectures, sof- skill training sessions and conferences, to enhance the skills of the students.

This work deals with development of the student placement prediction system that trains over the previous year data and predicts the the placements for the test data.The student placement prediction system can help in improving the placement results for colleges followed by the recommendations to place them in better jobs and companies. Machine learning algorithms have provided a mechanism to predict the placement of a student in accordance to the current skill before a student is placed. The system architecture was proposed by the authors in their earlier work[7].

The figure 1 depicts the structure for creating Machine learning model. Previous year data is used for training the model then feature extraction is done for that data. The data will be sent to train the model using KNN, Decision Tree, Bayesian. Once the model has been created it will be evaluated. At the end current final year data has sent for feature extraction and then placement prediction is done for students.

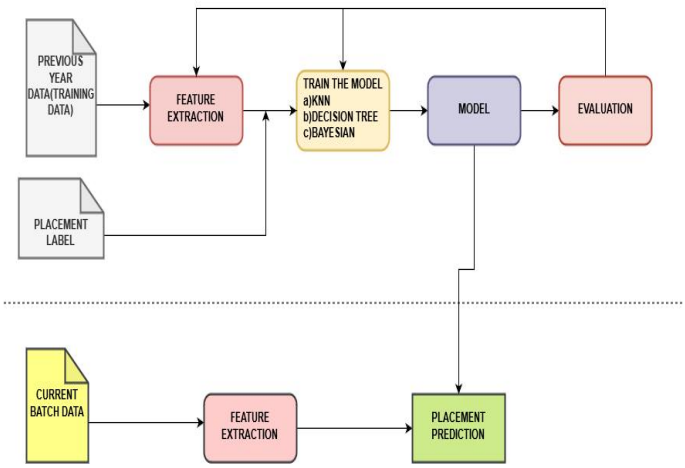

Figure 1: Structure for creating Machine learning models.

In these paper three classification algorithms have been used namely KNN, Decision tree and Bayesian. These algorithms are trained using preprocessed data and predict placement of current batch students and the details are included in section 3. Section 4 provides the results of each algorithm are compared and the one with more accuracy will be used for prediction of student placement.

2. LITERATURE SURVEY

This part of the paper includes the overview of survey conducted on student placement prediction. The literature survey highlights the existing methodologies and their drawbacks to provide better solutions.

C4.5 that is a Successor of ID3, Classification and Regression Tree (CART) and CHAID [2]. The above-mentioned decision tree algorithms work with supervised data and provide a decision tree for the classification problem of placement prediction. The prediction involved training the various algorithms with dataset that included data for 1342 students. The ID3 algorithm provided the maximum accuracy of 95.33 % [1].

Decision tree learning, Sci-Kit learn and Logistic Regression are the techniques that has been used. Decision tree works with the supervised data and uses logistic regression to find the probability of the student getting placed belonging to different departments. Thangavel, Bkaratki and Sankar have used decision tree approach to train the model dataset of size of 2205 tuples as training data and test data of size 289. The implementation provided accuracy of 71.66% with tested real life data[2].

Data mining and Classification Association Rules (CARM) are the two techniques that has been used by Badr, Algobail, Hanadi and Manal. Data mining is used in higher education to understand the student learning activities . Whereas CARM is used to create classifier using the rules called Classification Association Rules (CARs). Thus, instead of using dataset to train the model the authors used the courses and divided each course into four rules. The accuracy of Data mining and CARM is 67.33%[3].

Guleria and M. Sood implemented prediction system using Data mining and Naive Bayes Algorithm. Data mining is used to analyze educational data and extract useful information from large amount of data. Naive Bayes Algorithm is preferred when the dimensionality of the input is high and requires small amount of training data. The dataset of students were divided into tables for easy computation [4].

Bayesian Belief Network (BBN) machine learning techniques are used to predict the student placement using their competence and skills. Slim, Heileman, Kozlick and Abdallah took the dataset of 400 from different branches. The implementation of BBN obtained as marginal error of 0.16 [5].

Nguyen and et. Al implemented techniques that are used are Collaborative filtering and Matrix Factorization. Collaborative filtering is used on assumption that similar users like similar things.Matrix Factorization is used as it is one of the most successful methods for predicting. This was implementing using dataset of size 1,416,473 entries [6].

3. SYSTEM ARCHITECTURE

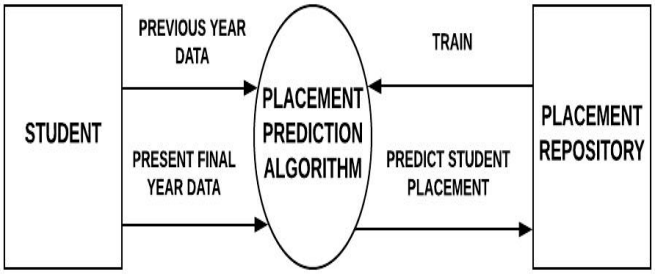

Figure 2: System Overview

Figure 2 depicts the system overview that gives the representation of the prediction system being modeled. Previous year placement data will be given as training data and current final year data as a test data for algorithm. Algorithm will be trained based on previous year data, so it will predict the placement of a student.

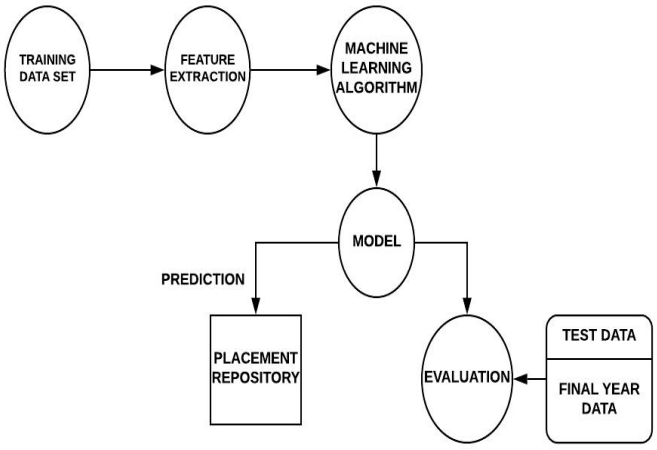

Figure 3: Detailed Structured Overview

Detailed structured overview has more text to reach the necessary level of detail about the system's functioning. In the figure 3 feature extraction process is done from training data set from which machine learning algorithm will create a model that model will evaluate final year data i,e test data and placement repository containing details about placements of previously passed out students will be used for prediction.

The components required to do prediction is as shown in the figure-4 :

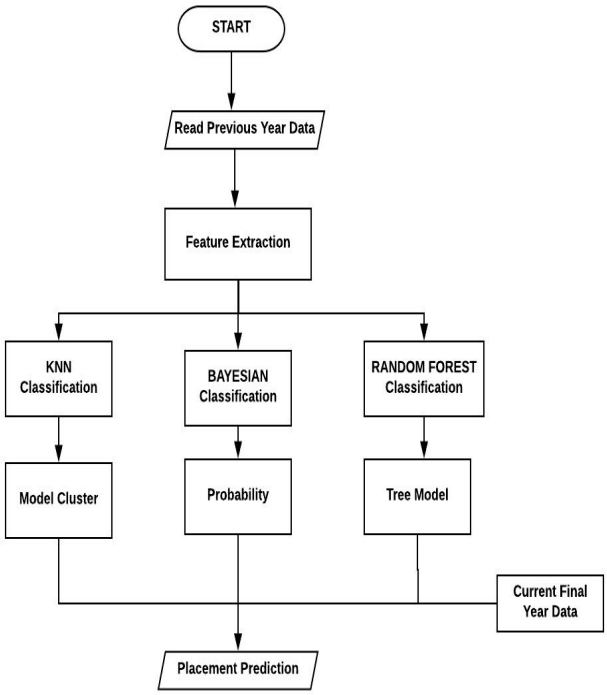

Figure 4: Flow Diagram

4. EXPERIMENTATION

For predicting the student placement previous year data is taken as training data, current final year data is taken as

test data.The previous year data is collected using the google forms. The google form is circulated among the students of previous year batches. On receiving the response that data is stored in CSV file. The figure-5 shows the snapshot of the google form created to collect data from previously passed out students.

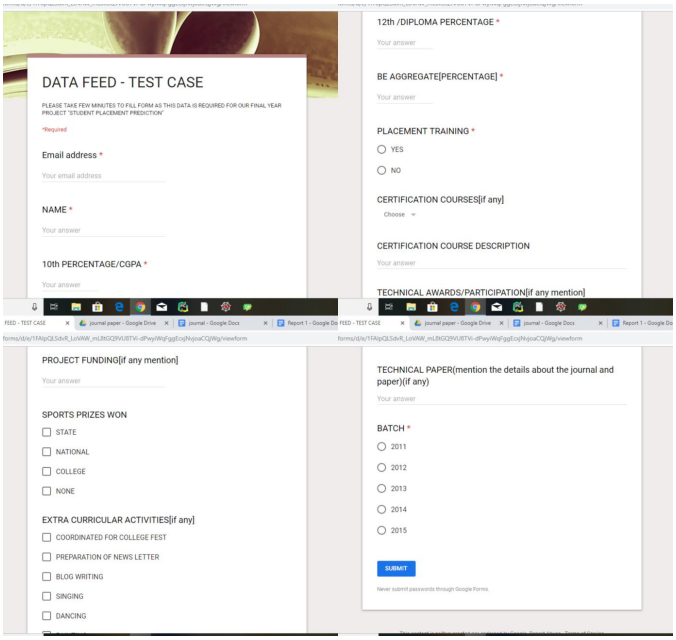

Figure 5: Screenshots of google form

Normalization of Data:

Once the data is received normalization is done using label encoder class from sklearn library. The following function shown in figure below is used to do label encoding:

```
from sklearn.preprocessing import LabelEncoder
labelencoder_X = LabelEncoder()
X[:, 4] = labelencoder_X.fit_transform(X[:, 4])
X[:, 5] = labelencoder_X.fit_transform(X[:, 5])
X[:, 6] = labelencoder_X.fit_transform(X[:, 6])
X[:, 7] = labelencoder_X.fit_transform(X[:, 7])
X[:, 8] = labelencoder_X.fit_transform(X[:, 8])
X[:, 9] = labelencoder_X.fit_transform(X[:, 9])
```

```
# Encoding the Dependent Variable
labelencoder_y = LabelEncoder()
y = labelencoder_y.fit_transform(y)
```

Figure 6: code snippet of label encoder function

The above figure 6 shows the function used to convert the non-categorical data into categorical data using label encoding technique.

This data will be provided as the input for the three algorithms namely KNN, Random Forest, Bayesian Network computing accuracy and confusion matrix . The below code snippets shows the function for these algorithms:

K-Nearest Neighbors:

```
# Fitting classifier to the Training set
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=5, metric='minkowski', p=2)
classifier.fit(X_train, y_train)

KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
        metric_params=None, n_jobs=1, n_neighbors=5, p=2,
        weights='uniform')
```

Figure 7: code snippet of K-NN Classifier function

Random Forest:

```
# Fitting Random Forest Classification to the Training
set
from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(n_estimators = 10,
criterion = 'entropy', random_state = 0)
classifier.fit(X_train, y_train)
```

Figure 8: code snippet of Random forest Classifier function

Bayesian:

```
# Fitting Naive Bayes to the Training set
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)
```

Figure 9: code snippet of Bayesian Classifier function

5. RESULTS AND DISCUSSION

From the experiment the results obtained are specified below

a) Prediction accuracy

The prediction is done using three different algorithms Random forest, KNN, Bayesian. The below table shows the result and comparison between algorithms

SL.NO	ALGORITHMS USED	RESULTS(in percentage)
1	RANDOM FOREST	84
2	KNN	86
3	BAYESIAN	92

Table I: Comparison between algorithms

b) Factors that contribute for prediction

In this experiment all the features contribute to the prediction factor. The Table II shows the confusion matrix for all the three algorithms that are evaluated:

SL.NO	ALGORITHM	CONFUSION MATRIX
1	RANDOM FOREST	[[2 1] [5 30]]
2	KNN	[[1 2] [3 32]]
3	BAYESIAN	[[1 2] [1 34]]

Table II: Confusion matrix for all 3 algorithms

CONCLUSION

This work contributed towards demonstrating the attainment of the machine learning algorithms for student placement prediction. This paper emphasizes the importance of the proper data collection, converting the non-categorical data into categorical data which required at most focus and normal verification. Algorithms namely KNN**,** Bayesian and Random Forest are used to calculate accuracy and confusion matrix for each model. From the accuracy calculated, student placement can be predicted. It offers a great advantage to the institution as well as the students. Our future work includes recommender system which involves recommending skillset and course certifications.,etc so that the current final year students can be recommended with the special skills that are needed for the good placements.

REFERENCES

[1] Jeevalatha, T & N.Ananthi, N.Ananthi & Kumar, Saravana. (2014). "Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms". International Journal of Computer Applications. 108. 27-31. 10.5120/18988-0436.

[2] S.K.Thangavel,P.D.Bkaratki and A. Sankar, "Student placement analyzer: A recommendation system using machine learning," *2017* 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017, pp. 1-5.doi:10.1109/ICACCS.2017.8014632.

[3] Ghada Badr, Afnan Algobail, Hanadi Almutairi, Manal Almutery,"Predicting Students Performance in University Courses: A Case Study and Tool in KSU Mathematics Department",Procedia Computer Science,Volume 82,2016,Pages 80-89,ISSN 1877-0509.

[4] P. Guleria and M. Sood, "Predicting student placements using Bayesian classification," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 109-112.doi: 10.1109/ICIIP.2015.7414749.

[5] A. Slim, G. L. Heileman, J. Kozlick and C. T. Abdallah, "Predicting student success based on prior performance," 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Orlando, FL, 2014, pp. 410-415. doi: 10.1109/CIDM.2014.7008697.

[6] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, Lars Schmidt-Thieme,"Recommender system for predicting student performance", Procedia Computer Science, Volume 1, Issue 2,2010,Pages2811-2819,ISSN 1877-0509.

[7] Dr.Meenakshi Malhotra,Gnanavi N,Deeksha D,Krupa N,Navyashree S L,"Evaluating supervised machine learning algorithm for student placement prediction"February 2019,Volume 06,Issue 2,pages 1367-1370,ISSN 2395-0056.