

Employability Prediction on The Basis of Academic Factors Influencing Placements

Puja Chaudhari¹, Gauri Patil², Kinjal Patil³, Poonam Patil⁴, L.M.Kuwar⁵

^{1,2,3,4,5} Dept of Computer Science

^{1,2,3,4,5} Computer Science

Abstract- Various factors affect the placement of student from campus. Most of the factors are related to academics. This project takes into consideration academics factors of students such as SSC HSC percentages, Employability test marks, etc. and predicts whether he will be placed or not. Placement of students is one of the very important activities in educational institutions. Admission and reputation of institutions mainly depends on placements. Hence all institutions strive to strengthen placement department. In this study, the objective is to analyse previous year's student's historical data and predict placement chance of the current students and the percentage placement chance of the institution. A model is proposed along with an algorithm to predict the placement chance of students.

Keywords- Prediction, Placement, Data Mining, KNN, Naïve Bayes.

I. INTRODUCTION

Various factors affect the placement of student from campus. Most of the factors are related to academics. This project takes into consideration academics factors of students such as SSC HSC percentages, Employability test marks, etc. and predicts whether he will be placed or not. It's a well-known fact all round the world that admission of students in an educational institution depends on the placements. Placement is one of the factors considered for determining the quality of the institution. Hence every institution strives hard to provide better placements to their students.

An educational institution contains a large number of student records. This data is prosperity of information, but is too large for any one person to understand in its entirety. Finding characteristics in this data is an essential task in education research. It does not make sense to find the placement possibility of all the students in the institution as all the students will have not have good KSA(knowledge, skill and attitude) score. Hence there is a need for identifying those students among the whole set of students who have good KSA score and finding placement chance for them would help us achieve the objective and thus save lot of time. Hence input for the study is the best cluster of students having better KSA

score who will have good chance of placement which is obtained by applying clustering algorithm and other necessary data preprocessing techniques.

II. IDENTIFY, RESEARCH AND COLLECT IDEA

We concern about the challenges face by any institute regarding the placement. The placement prediction is very complex when the number of the entities increases in any institute. With the help of machine learning this complex problem of prediction can be easily solved. In this paper all the academic record of student is taken into consideration. Various classification and data making algorithms are used such as Naïve Bayes, KNN and Logistic Regression. After the prediction of the students can be placed in of the given category that is core company, dream company or support services.

III. STUDIES AND FINDINGS

"A Placement Prediction System Using K-Nearest Neighbors Classifier", Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi, Naini Dubey, Second International Conference on Cognitive Computing and Information Processing (CCIP), 2016.

The placement prediction system predicts the probability of students getting placed in various companies by applying K-Nearest Neighbors classification. The result obtained is also compared with the results obtained from other machine learning models like Logistic Regression. The academic history of student along with their skill sets like programming skills, communication skills, analytical skills and team work is considered which is tested by companies during recruitment process. Data of past two batches are taken for this system.

Existing Systems

It's a well-known fact all round the world that admission of students in an educational institution depends on the placements. Placement is one of the factors considered for

determining the quality of the institution. Hence every institution strives hard to provide better placements to their students. An educational institution contains a large number of student records. This data is a prosperity of information, but is too large for any one person to understand in its entirety. Finding characteristics in this data is an essential task in education research. It does not make sense to find the placement possibility of all the students in the institution as all the students will have not have good KSA(knowledge, skill and attitude) score.

IV. PROPOSED SYSTEM AND METHODOLOGY

Placement of students is one of the very important activities in educational institutions. Admission and reputation of institutions mainly depends on placements. Hence all institutions strive to strengthen placement department. In this study, the objective is to analyse previous year's student's historical data and predict placement chance of the current students and the percentage placement chance of the institution. A model is proposed along with an algorithm to predict the placement chance of students.

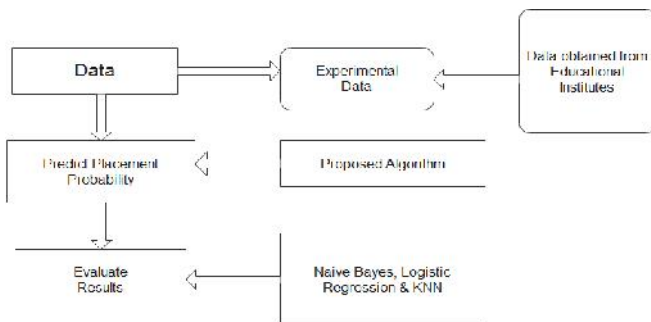


Fig 1. System Architecture

Modules of the System –

1. Dataset and EDA –

Gathering a dataset is one of the most important steps in machine learning. Obtaining the right data set for our project is very crucial. The data set used for this project is obtained from Kaggle. It is provided by a particular college for research purposes. The data set can be in any form whether it is .csv or .arff. Data set consists of many attributes. This dataset was built with the purpose of understanding the reason behind the selection of students for a particular job. The predicted output gives us whether the student will be placed or not.

2. Understanding Dataset and Generating train test -

The next step is understanding the data. For that, we display attributes details whether it is nominal or numeric. Then checking if it has any missing values or not. How many distinct values it has. After that, we display min, max, mean, mode values of numeric values of the attribute. The next step is creating a test and train dataset. For that, we will be dividing our main dataset file into train and test files with proportions of 70% and 30% respectively. This is done programmatically. The dataset file is selected and randomly 70% of instances are stored in the training file and the remaining instances are stored in the test file.

3. Creating Machine learning models -

After cleaning the data and generating the test and train files, the next step is to create the machine learning model that suits our data set and gives us maximum accuracy. In this case, we want to predict whether the student will be placed or not. So this problem comes under the classification part of predictions. The three algorithms that are considered in this project are Naïve Bayes, KNN, and Logistic Regression algorithms. Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. Logistic Regression is a supervised machine learning algorithm that can be used for regression challenges. The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used for both classification or regression challenges. The k-nearest neighbors(KNN) algorithm is simple, easy to implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

4. Comparing Machine Learning Models -

After training machine learning models with three of the above-mentioned algorithms the result shows accuracy, precision, recall, and all the related summary of the algorithm. With that information, we can compare the three algorithms and select one of the algorithms for our prediction system which has the highest accuracy.

5. Predicting the Student Employability -

After creating the models and choosing the best suitable one for our data set the next step is to predict whether a student will be placed or not. When a user provides the inputs to GUI, the parameters are there in sent for prediction in the machine learning model, and on the base of the training data set the class value is predicted by the model.

6. UI/UX -

The UI is developed in Java FX as it is very clean and robust. To develop Client-Side Applications with rich features, the programmers used to depend on various libraries to add features such as Media, UI controls, Web, 2D and 3D, etc. JavaFX includes all these features in a single library. In addition to these, the developers can also access the existing features of a Java library such as Swing.

7. Algorithm implementation -

It is the most challenging task of this project. As mentioned above the three algorithms that are used are Naïve Bayes and KNN and Logistic Regression algorithms. All three algorithms are trained using a given data set, after performing data cleaning and data visualization. We will be choosing the best API for prediction.

V. FUTURE SCOPE

One of the most recent and biggest challenges that higher education faces today is making students skillfully employable. Many universities/institutes are not in a position to guide their students because of a lack of information and assistance from their teaching-learning systems. To better administer and serve the student population, the universities/institutions need better assessment, analysis, and prediction tools.

A considerable amount of work is done in analyzing and predicting academic performance, but all of these works are segregated. There is a clear need for a unified approach. Other than academic attributes, there are large numbers of factors that play a significant role in prediction, which includes non-cognitive factors (set of behaviors, skills, attitudes). Suitable data mining techniques are required to measure, monitor, and infer these factors for prediction. Thus enriching the input vector with qualitative values may increase the accuracy rate of prediction as well. Integrated Models/Frameworks are required for all the stakeholders of an Institution; hence ensuring sustainable growth for all (Management, Teachers, Students, and Parents).

VI. CONCLUSION

Machine Learning techniques applied to educational data are concerned with developing methods for exploring the unique types of data; in the educational domain, each educational problem has specific objectives with unique characteristics that require different approaches for solving the problem. In this study, a model was proposed along with an

algorithm. This was compared with three other classification algorithms such as KNN, naive Bayes, and logistic regression in terms of accuracy, precision, true positive rate (recall). The proposed model proved to be the best predicting model for solving placement chance prediction problems compared to all other algorithms. Hence, having the information generated through our study, institution would be able to design strategies to overcome lacunae and improve placements with the best chances of getting placed. Thus admission can be increased.

REFERENCES

- [1] "Data Mining Approach for Predicting Student and Institution's Placement Percentage", Professor. Ashok M Assistant Professor Apoorva A, 2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions.
- [2] "Student Placement Analyzer: A Recommendation System Using Machine Learning", Senthil Kumar Thangavel, Divya Bharathi P, Abijith Sankar, International Conference on Advanced Computing and Communication Systems (ICACCS -2017), Jan. 06 - 07, 2017, Coimbatore, INDIA.
- [3] "A Placement Prediction System Using K-Nearest Neighbors Classifier", Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi, Naini Dubey, Second International Conference on Cognitive Computing and Information Processing (CCIP), 2016.
- [4] "Class Result Prediction using Machine Learning", Pushpa S K, Associate Professor, Manjunath T N, Professor and Head, Mrunal T V, Amartya Singh, C Suhas, International Conference On Smart Technology for Smart Nation, 2017.