Prediction of an Employment for a Graduate based on Academic Performance using Logistic Regression - A Machine Learning Approach

Gurusiddappa Hugar

Assistant Professor
Computer Science and Engineering
A.G.M. Rural College of Engineering and Technology, Varur, Hubli, India

Abstract—**There are several areas in which organizations can adopt technologies that will support decision-making: Machine Learning is one of the most innovative research areas that consists of different algorithms and are widely used to develop different applications such as medical diagnosis and healthcare, image and speech recognition, business strategies, organizational aspects and people management. In this research work, Logistic Regression is used to predict the employment chances by the Institute. The student age, overall percentage and number of projects done during academic years are considered as objective factors to the work. The main goal of this research work is to analyze how objective factors influence the people while choosing the best Institution. The obtained model for the prediction of Employment, which is trained with the help of data-set collected from our Institution and nearby Institutions.**

Index Terms—**Machine Learning, Prediction, Performance**

I. INTRODUCTION

In today's world, organizations are able to support decision making in various areas that includes technologies like Machine Learning. Machine Learning is one of the most innovative research areas that consist of different algorithms and are widely used to develop different applications such as Medical Diagnosis and healthcare, image and speech recognition, business strategies, organizational aspects and people management [1]. The main goal of this research work is to provide the opinion to the people while choosing the best Institution. Since there is an increase in engineering graduates produced by the higher education institution in every year, there is a rise in competition for getting employment [2,3]. It's the prior goal of students to achieve their dream job before getting graduated and every institution plans for 100% placement for their students every year. In order to predict the student's employment, a Machine Learning technique called Logistic Regression is used [5]. Logistic regression is a statistical analysis method to predict a binary outcome, like yes or no, supported prior observations of a data-set. A logistic regression model predicts a dependent data variable by analyzing the link between one or more existing independent variables. Logistic regression is simpler to implement, interpret, and extremely efficient to coach and it's in no time at classifying unknown records. Hence, this offers the simplest result which avoids the confusion in student's mindset. This research work also provides a second opinion while choosing best institution based on the placement prediction. The obtained model for the prediction of Employment, which is trained with the help of data-set collected from our institution and nearby institutions. The results will also be displayed in charts for better comparison. Here, it is preferred Doughnut chart, Bar chart and Line chart. Students can choose the most effective institution as per the very fact shows in our analysis. The research work also keeps track on the records which is being stored in MongoDB database. This research work avoids fear, confusion and stress level of students in choosing the best institution and shows the true fact between institutions. It is planned to hold the record of testing done by the students along with the respective results for further analysis and thereby improving the performance of the system.

II. DESIGN AND METHODOLOGY

The workflow of the research work is depicted in the figure 1.
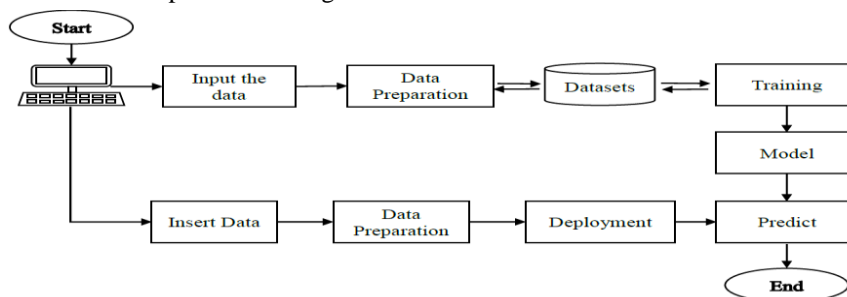
Fig.1. Workflow

There are two major phases, training phase and testing phase. Training phase includes input data, data preparation, model development and prediction, whereas testing phase includes input data and prediction. In training phase, input data refers to the collection of data done from the Institutions, the collected data is preprocessed to select the essential objective factors. In testing

phase, user has to enter the objective factors age, overall percentage and number of projects done. With the help of Flask framework of python, the prediction result is displayed in terms of either 0 or 1.
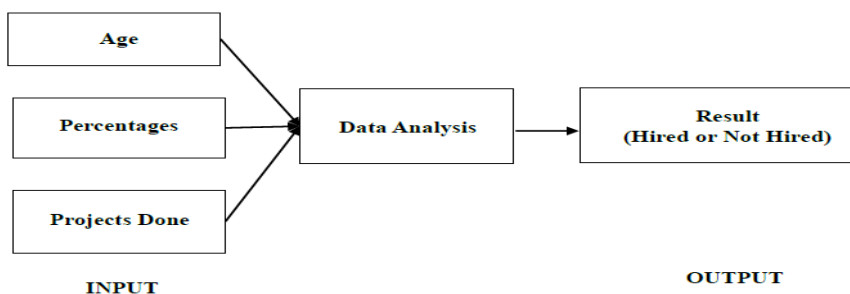
Fig.2. Deployment flow

The figure2 shows the formal deployment process, which includes input of three attributes and analysis is done using the Logistic Regression. Here, 0 refers to the student who can get a job or employable or hired, whereas 1 refers to the student who is not hired or may not be able to get a job. The crucial conditions for Logistic Regression are discussed below:

Age: Age refers to the learning period of the student, and gets to know the prediction of being employable on the basis of age. Here, student age should follow 21 or 22 to in order to reach employment resulting as 0, or the prediction will be 1 which predicts chances of unemployable.

Percentage: Percentage refers to the overall performance of the students scored in their academics. Here, the student percentage has to be at-least 60 or above 60 in order to meet the eligibility.

Projects Done: As this shows the practical skills of the students in the form of projects performed in their academics. Hence, this study is not only related to final year students but also helps the pre-final year students as well. Projects done should be at-least 1 or more than 1 in order to meet the eligibility.

Therefore, a student gets the eligibility status if and only if all the above mentioned conditions are satisfied, or it is also helpful in improvising their skills or performance. Data analysis performs the testing part of the input data using Logistic Regression and results in 0 or 1.

The methodology of a overall research work is shown in Figure 3 shows the process of training and testing phases in which data acquisition contains the real world data-set which may contain other attributes also such as Gender, Age , Institute Name, Branch, Percentage, Projects Done, Employment Status   and Salary (LPA). Preprocessing of data is carried out. Feature extraction is done to select the required objective factors, institute classification is followed for the comparison between different institutes. The analysis of data is done by the Logistic Regression to predict.
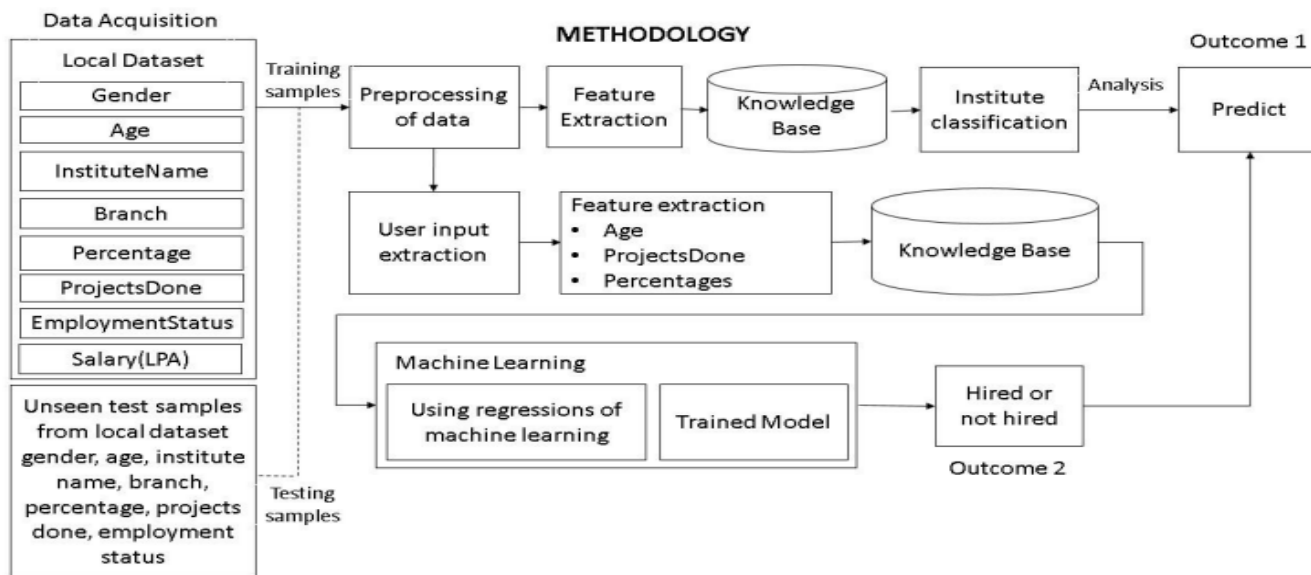
Fig.3. Methodology

Students can check their chance of employment by inputting the attributes such as age, overall percentage and number of projects done. These 3 attributes helps to decide skills and performance of a student. As Logistic Regression measures prediction in binary outcome, such as yes or no, based on prior observations of a data-set. Hence 0 predicts for chance of employment and 1 for chance of unemployment. On the basis of data-set of different Institution comparison is made with the help of chart.js.

Fig.4. Data-set showing different attributes

The Fig.4. shows the data-set that is collected from the institutions which contains many attributes. The essential objective factors are extracted from this database and used for the analysis process.

III. RESULTS

This section will present the various results and analysis obtained from the training and experiments conducted. Here, graphical user interfaces are designed to easily interact with the developed work. To enter the input data for objective factors the following form is designed. The form will work in the local host system as illustrated in figure 5.

Fig.5. Local host running to enter numeric values

Now, the user have to enter the objective factors value related to a student such as Age, Overall percentage and Number of projects done during their academic years. The user has to enter numerical values as illustrated in figure 6.

Fig.6. Input numeric values to the system

After entering numeric values, user has to click the predict button to get the prediction result. The sample result is illustrated in Figure 7. The developed research work takes the values of attributes, based on the data provided by the user, predicts the chance of employment and unemployment.
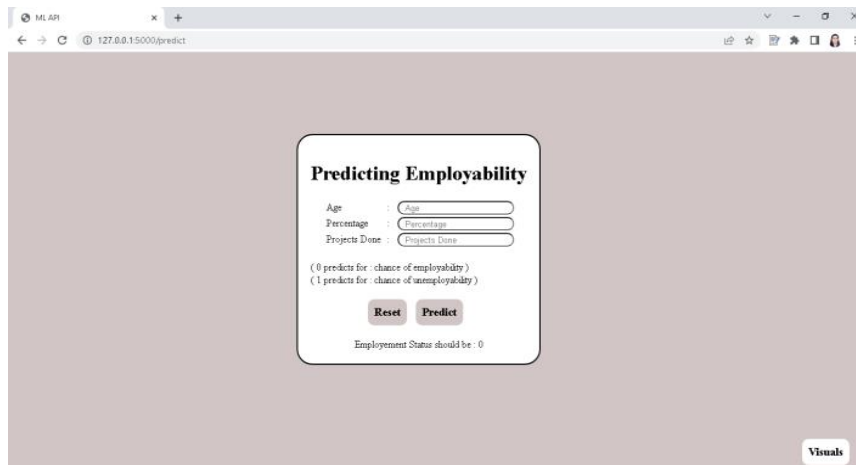

Fig.7. Showing predicted result

The tested sample details are stored in the database for further analysis. The snapshot of the running database is shown in the Figure 8.
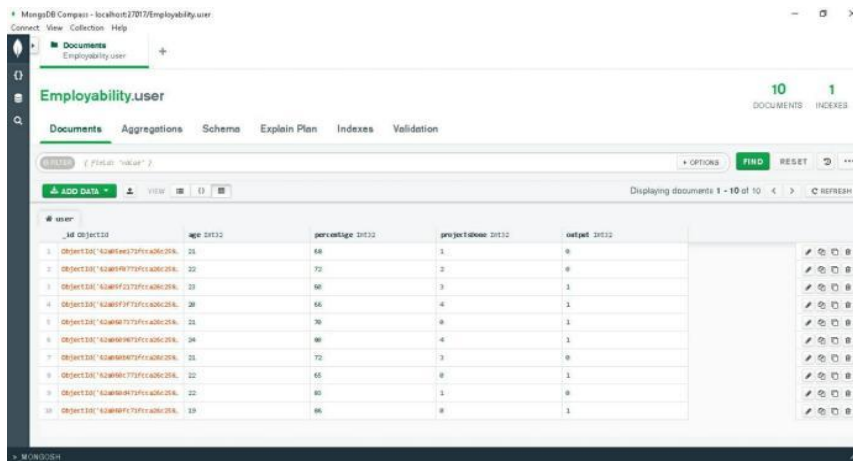

Fig.8. Results of MongoDB database

Based on the available results in the database, it is further represented branch wise which shows how many students of each branch are tested. This is done with the help of Doughnut Chart as illustrated in the Figure 9.


Fig.9. Doughnut Chart showing branch wise test status

The stored results are also represented in segments with the help of bar chart as illustrated in Figure 10.
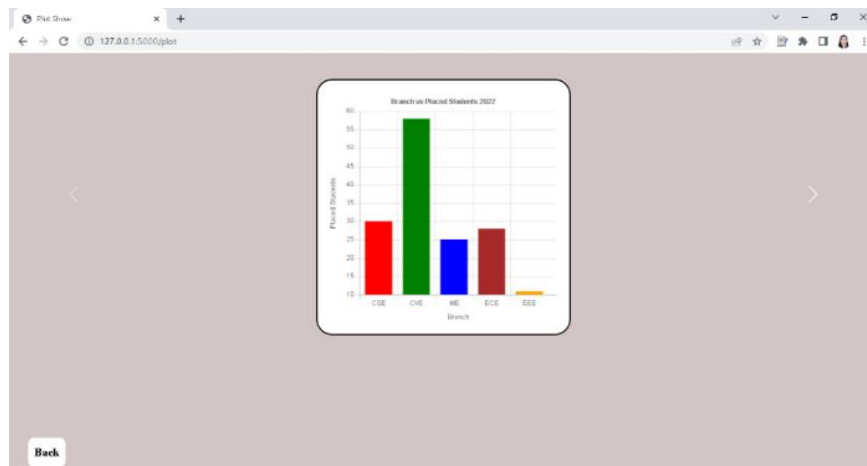
Fig.10. Bar chart showing different segments

The results are also represented institute wise to compare the performance of the different institutions. This gives the second opinion to the students. This process is illustrated represented by Line Chart in the Figure 11.
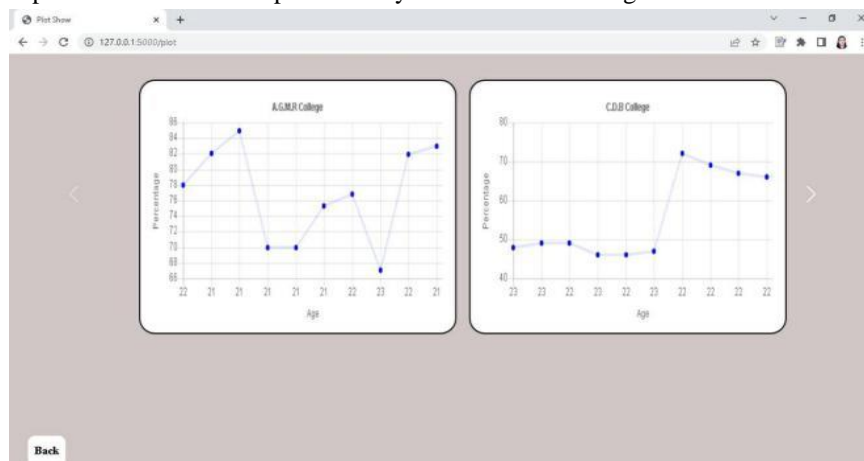
Fig.11. Line chart showing the performance of different institutions

IV. CONCLUSION

Since there is an increase in graduates produced by the higher education institution in every year, there is a rise in competition for getting placed. This research work predicted employment of a graduate based on their academic performance. It's the prior goal of students' to achieve their dream job before getting graduated and every institution plans for 100% employment for their students every year. In this regard to improve the students' chance of employment and their academic performance, this research work aids them in reaching it. The Logistic Regression technique is successfully applied to the collected data-set for the prediction. Hence Logistic Regression provides good results in statistical analysis process.

V. ACKNOWLEDGMENT

REFERENCES:
[1] Tom M Mitchell, Machine Learning. McGraw Hill Education, 1997.
[2] Cherry D. Casuat; Enrique D. Festijo, "Predicting Students' Employability using Machine Learning Approach," IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), pp. 1-5, 20-21 December 2019, doi:10.1109/ICETAS48360.2019.9117338.
[3] Cherry D. Casuat; Enrique D. Festijo, "Identifying the Most Predictive Attributes Among Employability Signals of Undergraduate Students," 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020), Langkawi, Malaysia, pp. 1-4, 28-29 Feb. 2020, doi:10.1109/CSPA48992.2020.9068681
[4] Casimiro A. Curbelo Montañez and William Hurst, "A Machine Learning Approach for Detecting Unemployment Using the Smart Metering Infrastructure," VOLUME 8, pp 22525-22536, 2020, doi:10.1109/ACCESS.2020.2969468
[5] Nesrine Mezhoudi Rawan Alghamdi Rim Aljunaid Gomathi Krichna Dilek Düştegör "Employability prediction: a survey of current approaches, research challenges and applications", Journal of Ambient Intelligence and Humanized Computing 16 June 2021. doi: 10.1007/s12652-021-03276-9

[6] K Vinutha H K Yogisha, "Prediction of Employability of Engineering Graduates using Machine Learning Techniques", 8th International Conference on Computing for Sustainable Global Development (INDIACom),New Delhi, India 17-19 March 2021.

[7] Mustafa Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms", Yağcı Smart Learning Environments (2022) 9:11https://doi.org/10.1186/s40561-022-00192

[8]   Ali Salah Hashim et al, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms", 2020 IOP Conf. Ser.: Mater. Sci. Eng. 928 032019

[9] Lamees Al-Alawi, Jamil Al Shaqsi, Ali Tarhini & Adil S. Al-Busaidi "Using machine learning to predict factors affecting academic performance: the case of college students on academic probation" Education and Information Technologies https://doi.org/10.1007/s10639-023-11700-0

[10] Akanksha Pandey, L S Maurya "Career Prediction Classifiers based on Academic Performance and Skills using Machine Learning, SSRG International Journal of Computer Science and Engineering Volume 9 Issue 3, 5-20, March 2022 ISSN: 2348 – 8387 / https://doi.org/10.14445/23488387/IJCSE-V9I3P102