PREDICTION OF STUDENT PLACEMENT USING MACHINE LEARNING ALGORITHM

[1]G.GAUTAMI,[2]MONAM HAYAT,[3]PALLAB BANERJEE,[4]BIRESH KUMAR

[1,2]B.Tech Scholar,[3,4]Assistant Professor

[1]Department of Computer Science and Engineering,

[1]Amity University, Ranchi Jharkhand,India

Abstract: **Traditionally, data analysis was trial and error-based, an approach that becomes impossible when data sets are large and heterogeneous. Machine Learning being a part of data science provides quick-witted substitutes for exploring enormous mass of data. By growing rapid and well organized algorithms and data-driven models for actual-time processing of data. The purpose of institutions is to provide golden opportunities to their students. Linear Regression, a tool of supervised machine learning can contribute its features to serve this purpose. It is an effective method helping in predicting future trend of student placement based on advance placement practice test marks. The result obtained from this will help the students to better understand their weak areas to work upon. Working on these areas will let students achieve higher number of placements in an institution.**

Index Terms - **Machine Learning, Student Placement, Predictive Analysis, Supervised Learning**

I. INTRODUCTION

College days get cohesive as students pioneer on to the road to their Higher Education degree programs. The key role that the educational institution plays in placements is to prepare and guide the students well through the process of placements by providing them the needed training for getting placed in high-end companies. They provide the platform to the students to showcase their skills and abilities to the prospective employers by inviting them for placement drives. With the proper placement training, students learn how to put forth their knowledge and abilities in the right way to fetch the best of jobs.

One aspect the placement cell can opt is to analyse their students graph based on placement practice tests before presenting them infront of companies. This graph will be generated on the final year student dataset and this will help students realise which field they should work upon more. The fields could be aptitude, reasoning, verbal, technical, communication skills and so on. This method will be effective for both- the students as well as the institution for holding up a good placement record.

I. LITERATURE SURVEY

K. Sreenivasa Rao presented a paper in which using education data mining students performance is predicted based on various parameters[1]. Predictions are made using machine learning algorithm in weka tool and R studio.

Syed Ahmed propose a TPO management system from an existing system[2]. The objective is to analyze previous year's student's historical data and predict placement eligibility of the current students and the percentage placement chance of the institution. They used Decision tree C4.5 Algorithm.

Equity Action Plan (EAP): The objective of the Equity Action Plan is to ensure that all students and faculty in the project institutions have equal opportunity to avail of the benefits of the project with substantial improvement in the performance of weak students[3].

Satish Kumar's objective is to study the nature of campus placements which is useful for both Students and Institution[4].

Also to Build a model that can be used to predict the probability that a randomly chosen student will be placed or not.

To Identity the factors that are influencing the placement chance of a student in technical education. They used binary logistic regression.

R. Rajalaxmi's objective of the paper is to use linear regression techniques to build a model which predicts the performance of the students in Engineering Discipline[5]. The output or dependent variable is the prediction of end semester examination grades i.e. CGPA (Cumulative Grade Points).

Pothuganti Manvitha's the objective is to analyze previous year's student's data and use it to predict the placement chance of the current students[6]. This proposed model is also compared with other traditional classification algorithms such as Decision tree and Random forest with respect to accuracy, precision and recall.

Ajay Kumar Pal's study presents a proposed model based on classification approach to find an enhanced evaluation method for predicting the placement for students[7]. This model can determine the relations between academic achievement of students and their placement in campus selection.

Shreyas Harinath's paper presents a recommendation system that predicts whether the current student will be placed or not, if the student is placed the company is also predicted based on the data of previously placed students[8]. Here two different machine learning classification algorithms, namely Naive Bayes Classifier and KNearest Neighbors [KNN] algorithm are used.

II.   RESEARCH METHODOLOGY

The placement prediction portrayed in this paper is according to the following diagram. The collected dataset of final year students consist of parameters like stud_id, name, course, performance in class, cgpa, marks of advance placement practice test etc. Out of which the data's which are not contributing to this paper had been abolished. Futher processing on data is performed using machine learning algorithm- linear regression and it is plotted with the help of scatter plot. It is performed in jupyter notebook via python and the results are drawn.
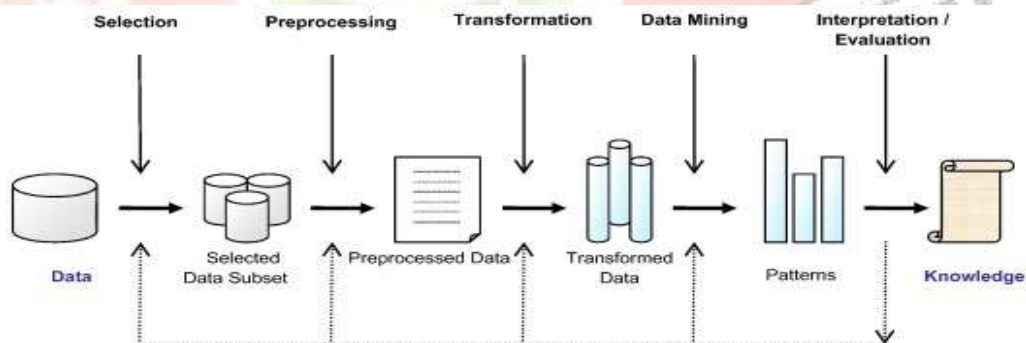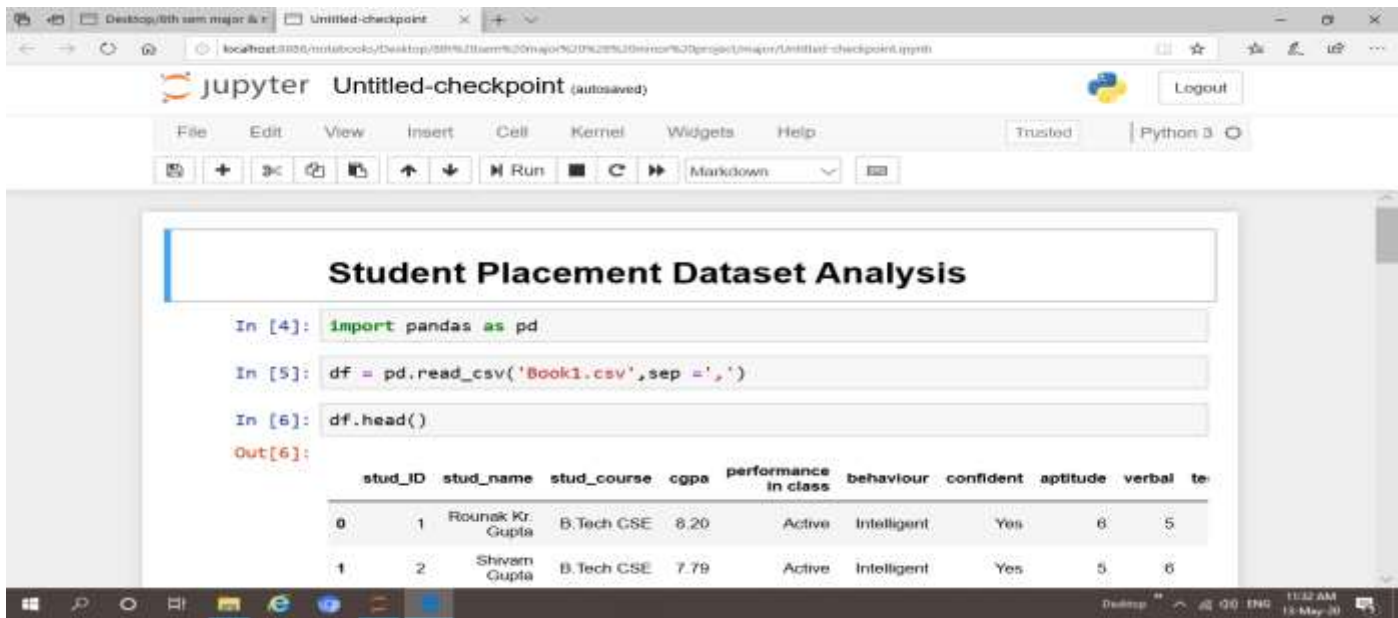
Overview of Regression and Scatter plot:

A Regression Analysis involves graphing a line over a set of data points that most closely fits the overall shape of data or regression shows the changes in a dependent variable on the y-axis to the changes in the explanatory variable on the x-axis. Regression determines the strength of prediction-can be anything like future trend or values.

Linear Regression selection criteria:-

- **Classification and regression capabilities**

- **Data quality**

- **Computational complexity**

- **Comprehensible and transparent**

A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

When a scatter plot is used to look at a predictive or correlational relationship between variables, it is common to add a trend line to the plot showing the mathematically best fit to the data. This can provide an additional signal as to how strong the relationship between the two variables is, and if there are any unusual points that are affecting the computation of the trend line.

I. WORKING ON JUPYTER NOTEBOOK

At first python libraries- numpy, pandas, and matplotlib were imported followed by reading of dataset (.csv file).
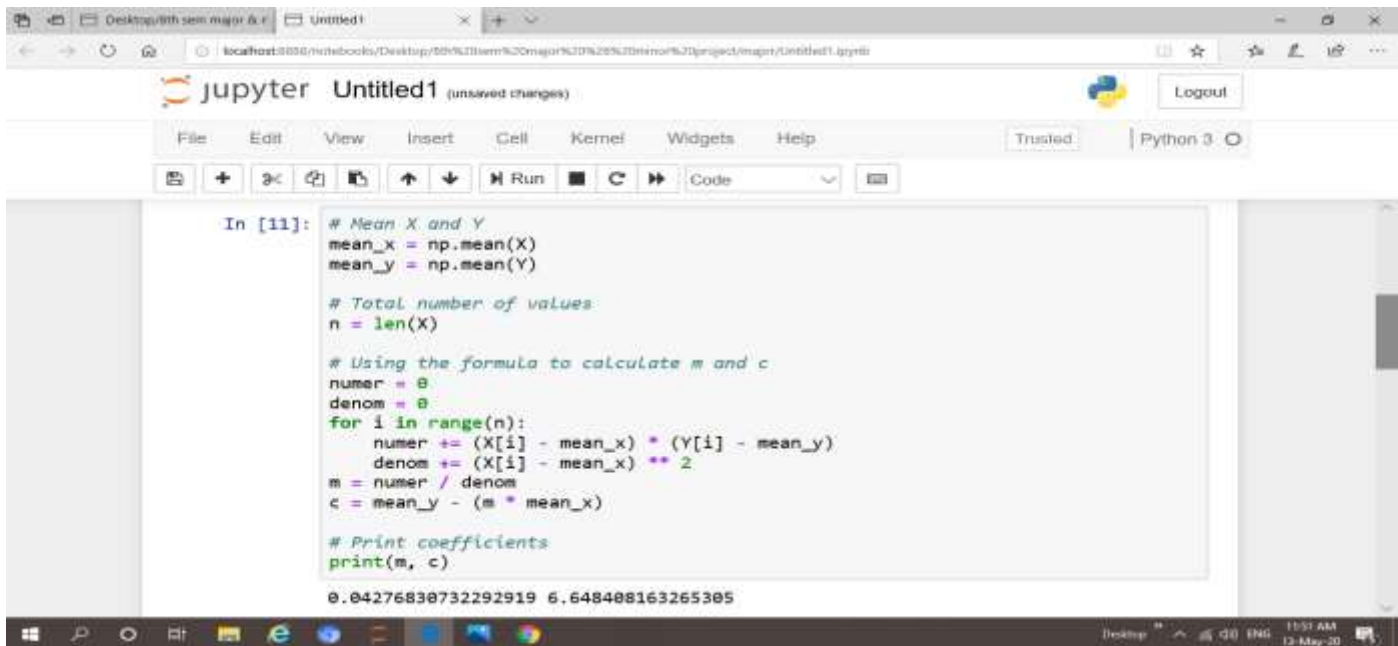
Then values of X(stud_id) and Y(marks of verbal ability) were gathered and mean was calculated.

R Min was plotted (mean_x, mean_y)

In the formula y = mx +c, m(slope of line) and c(y-intercept of line) were calculated.

$$m = \sum \frac{(x - \bar{x})(y - \bar{y})}{(x - \bar{x})^2} = 0.04$$

c = 6.64

Calculation of line x and y was done.

Plotting of regression line and scatter point was performed.

Lastly, standard error rate was calculated by using R-squared method.

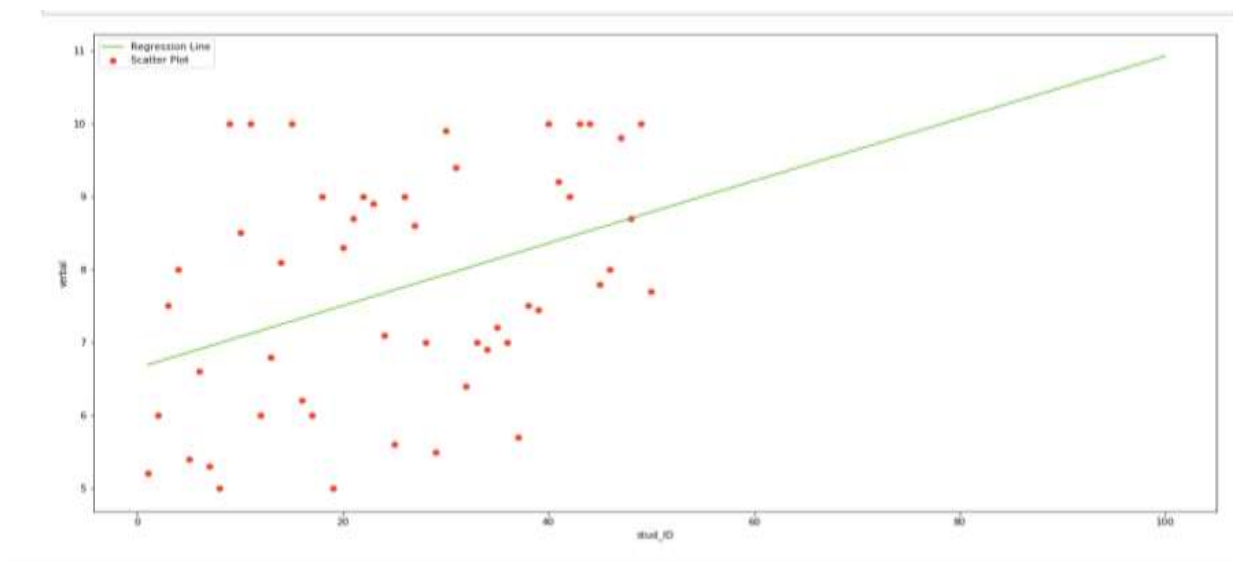$$R^2 = \sum \frac{(y_p - \bar{y})^2}{(y - \bar{y})^2} = 0.32$$

IV. RESULTS AND DISCUSSION

It performs n no of iterations for different values of m and will calculate equation of line i.e y = mx +c.

As the value of m changes the line will change. After every iteration it will calculate the predicted value according to the line and compare the distance of actual value to the predicted value and the value of m for which the distance between them is minimum will be selected as the best fit line.

Also the relationship between the scatter points and regression line is moderate linear correlation

After finding the best fit line it checks the goodness of fit.

R^2 value is a statistical measure of how close the data are to the fitted regression line. In general it is considered that high r-squared value model is a good model but we can also have a low r-squared value for a good model as well as a higher r-squared value for a model that does not fit at all.it is also known as coefficient of determination or coefficient of multiple determination.

Here, R^2 = 0.32 this is not a good fit. So it suggests that data points are far away from regression line. As the value of R (standard error rate) will increase we will see that the actual value will move closer to the regression line.

II. CONCLUSIONS

The main reason why college placements are becoming the major criteria for the students for choosing an educational institution is that they ease barriers to employment for graduates, apart from holding considerable value for students, in terms of personal and professional development. Moreover, college placements not only act as a platform to enhance their employability skills but also serve to better their academic performance and practical knowledge through intensive training and workshops.

Since, compared to the outside world, the competition remains confined in campus placements, their probability of getting placed goes relatively higher. All of these things collectively bring in a range of benefits and positive impacts on students.

This technique will eventually contribute its features in the process of placement and will help the students as well as the placement team to analyze students performance and let them know their standard error rate so that it can be reduced in order to come close to the best fit line to achieve more number of placements. Also the institution will have a good placement record. It is beneficial in predicting a future trend- to work upon today for a better tomorrow.

REFERENCES

[1]   K. Sreenivasa Rao, N. Swapna, and P. Praveen Kumar, Educational Data Mining, International Journal of Engineering & Technology, 7 (1.2) (2018) 43-46.

[2]   Syed Ahmed, Aditya Zade, Shubham Gore, Prashant Gaikwad and Mangesh Kolhal, TPO Management System, (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887, Volume 5 Issue XII December 2017.

[3]   Equity Action Plan (EAP), Technical education quality improvement programme(TEQIP) Phase-II.

[4]   D. Satish Kumar, Zailan Bin Siri, D.S. Rao and S. Anusha, Students campus placement probability using binary logistic regression, (IJITEE), ISSN: 2278-3075, Volume-8 Issue-9, July, 2019.

[5]   R.R.Rajalaxmi , P.Natesan , N.Krishnamoorthy and S.Ponni, Regression model for predicting students academic performance, (IJRTE), ISSN: 2277-3878, Volume-7 Issue-6S3 April, 2019.

[6]   Pothuganti Manvitha and Neelam Swaroopa, Campus placement prediction using supervised machine learning techniques, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 9 (2019) pp. 2188-2191.

[7]   Ajay Kumar Pal and Saurabh Pal, Classification model of prediction for placements of students, I.J.Modern Education and Computer Science, 2013, 11, 49-56.

[8]   Shreyas Harinath, Aksha Prasad, Suma H S, Suraksha A and Tojo Mathew, Student placement prediction using machine learning, (IRJET), e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 06 Issue: 04 | Apr 2019.

[9]   Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B and Prashanth Gouda, Student placement prediction model: Perspective for Outcome based Education System, (IJRTE), ISSN: 2277-3878, Volume-8 Issue-3, September, 2019.

[10] Mohana Bangale, Shubham Bavane, Akshay Gunjal, Rohit Dandhare, Sudhir D. Salunkhe, Survey on placement prediction using machine learning, ISSN [ONLINE]: 2395-1052, IJSART - Volume 5 Issue 2 –FEBRUARY 2019.