BSSS Journal of Computer: ISSN (Print) - 0975-7228, E-ISSN - 2582-4880, Vol. XIII, Issue-I (2022), pp 23-38
PujhaShree.S.B, Lekhasree . R , Logasri. P, Darling Jemima .D: Student Placement Prediction
https://doi.org/10.51767/jc1303

STUDENT PLACEMENT PREDICTION

PujhaShree.S.B, Lekhasree . R , Logasri. P, Darling Jemima .D
Assistant Professor in Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore

ABSTRACT

In an educational institution, the most important objective is the placement of students. For each and every student, the placement part is a very important one in college life because for some sets of students it is the future. The prediction of students will not be 100% accurate but it depends how the students perform in every part of the placement. So, to predict the placement package chance of current students, we can analyse the previous year's student's data. The data has been collected from the institution and certain pre-processing techniques are applied to the models. Different algorithms have different accuracy Depending on the type of issue and dataset to be solved; different algorithms have varied levels of accuracy. As a result, we decided to assess the accuracy levels of three methods, namely Logistic Regression, Decision Tree Classifier, and Random Forest Classifier, with respect to our challengeand dataset. The efficiency/accuracy of each model is visualized and tested and based on the performance analysis, the best model results are declared.

Keywords: Placement, Performance, Machine learning algorithms, Logistic Regression, Decision Tree, Random Forest classifier.

I. INTRODUCTION

Placement plays a very significant role in every student's career as well as it is considered to be one of the important objectives for educational institutions. The students will take up the admission in college by analyzing the placement records of the educational institutions. Now in the present scenario, educational institutions handle the manual method for predicting the student placement packages. So to overcome this way of manual method, our project will help to predict the placement package of students using machine learning technique.

The major goal of this study is to determine whether or not the student will be placed. with a particular package in the placement. Random forest algorithm is used to predict this model. This

https://doi.org/10.51767/jc1303

model predicts the placement package by considering the student's academic and some of the other activities. Some of the main criteria considered like CGPA and Arrears.

II. SIGNIFICANCE OF THE STUDY

One of the most essential goals of any educational institution is to place graduates. Institutions make great efforts to achieve placements for their student's .This will always be helpful to the institution. The objective is to predict the students getting placed with package for the current year by analyzing the data collected from previous years students.Every student seeks admission to colleges by analyzing the college's placement record.The major goal is to anticipate whether a student will be placed in campus recruitment with a specific package or not. For this, the data considered is the academic history of students CGPA, Arrear history.

III. OBJECTIVES OF THE STUDY

- The major goal of this model is to determine whether or not a student will be placed in campus recruitment. (USING MACHINE LEARNING ALGORITHMS)
- A good placement leads to a good reputation as well as the good future of the organization as well for the wards..
- One of the biggest challenges that higher learning establishments face nowadays is to boost the placement performance of their scholars.
- One of the effective ways to address the challenges for improving the quality is to provide new software related to the educational institution about the students performance in placements.

IV. POPULATION AND SAMPLE

The population includes Activity of Students from a University. We have used random forest classifier techniques and based on their CGPA and Arrears randomly selected 128 Students Data from a university have been put on to test.

V. STATISTICAL TECHNIQUES USED IN THE PRESENT STUDY

https://doi.org/10.51767/jc1303

Machine learning model (Random Forest Classifier) for evaluating the student's eligibility was developed and validated by Lekhasree.R, Logasri.P and Pujha Shree S.B (2021). Averages of the Arrear and CGPA were used to analyze the data

VI. MACHINE LEARNING

The objective of any machine learning problem is to identify a single model that will best predict our desired outcome. The system proposed uses three different machine learning models to train and test the students' data.

The models that are trained with the dataset are,

- Random Forest Classification
- Logistic regression
- Decision tree

These models are trained and the model with the highest accuracy is observed by testing it with the dataset. This is an iterative process carried out that increases the consistency and accuracy of the models.

VII. ACCURACY SCORE:

Accuracy is one of the metrics for evaluating the classification models. Informally, accuracy is the fraction of predictions from our model that got right. And moreover, accuracy has the following definition: Accuracy = Number of correct predictions / Total number of predictions.

VIII. LOGISTIC REGRESSION

- Decide on one dependent variable, which in our case is the placement status variable and independent variables which are student regna, name, department, number of arrears, and CGPA.
- Importing the required libraries read the dataset and pre-processes it.

https://doi.org/10.51767/jc1303

- Splitting the sample dataset into training and testing sets.
- Now we will train the model using the training set. For providing training or fitting the model to the training set, we will import the Logistic Regression class of the sclera library. After importing the class, we will create an object and use it to fit the model to logistic regression that we have trained.
- With the developed model we now start to predict the values with the data set.
- The logistic regression model delivers the probability values as in ones or zeros, and then we develop a confusion matrix so that we can visualize the predicted values.

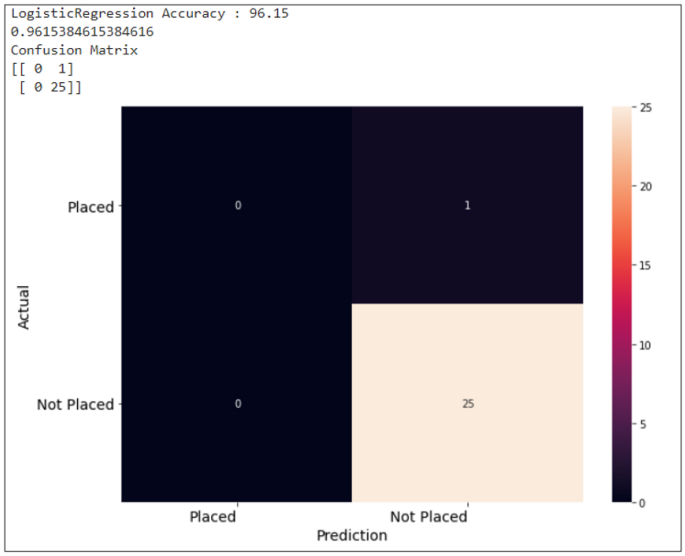

Figure.1 Confusion Matrix (Logistic Regression)

From the above visualization image, we can interpret that the confusion matrix has **1+0=1 incorrect predictions** and **25+0=25 correct predictions** out of 26 samples that were taken for prediction.

IX. DECISION TREE

- Determine the dependent variable, in our case the variable of the placement status as well as independent variables such as student regno, name, department, number of arrears, and

https://doi.org/10.51767/jc1303

CGPA.

- We will pre-process and prepare the data so that we can use it in our code efficiently and effectively.
- Divide the sample dataset into training and test sets.
- Now we will train the model using the training set. For providing training or fitting the model to the training set, we will import the Decision Tree Classifier of the sklearn library. After that we fit the model with the help of the classifier object and start the predictions.
- With the evolved model we now begin to predict the values with the data set.
- If we want to know the number of correct and incorrect predictions, we need to implement the confusion matrix.

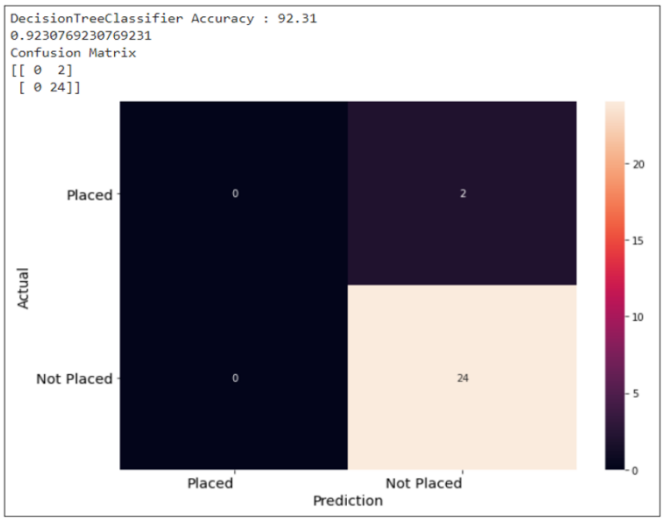

Figure.2 Confusion Matrix (Decision Tree)

As we can see, the confusion matrix has **2+0=2 incorrect predictions** and **24+0=24 correct predictions** out of 26 samples taken for predictions**.** Hence this model is not more accurate when compared with Logistic Regression.

X. RANDOM FOREST CLASSIFICATION

https://doi.org/10.51767/jc1303

- Define the dependent variables, in our case the placement status variable and independent variables such as the student regna, name, department, number of arrears, and CGPA.
- We will find the missing data and assess them with the suitable values by the process of pre-processing.
- Now we proceed to import the Random Forest Classifier class from the sklearn.Ensemble library. And predict the placement results effectively. Starting with the selection of random samples from the sample dataset. Next, the algorithm will construct a decision tree for every sample data. Then it will get the prediction result for every decision tree.
- In this step, voting will be performed for all the predicted result by the model.
- At last, selecting the most voted prediction result will be the final prediction result.
- If we wish to grasp the number of correct and incorrect predictions, we need to use the confusion matrix.

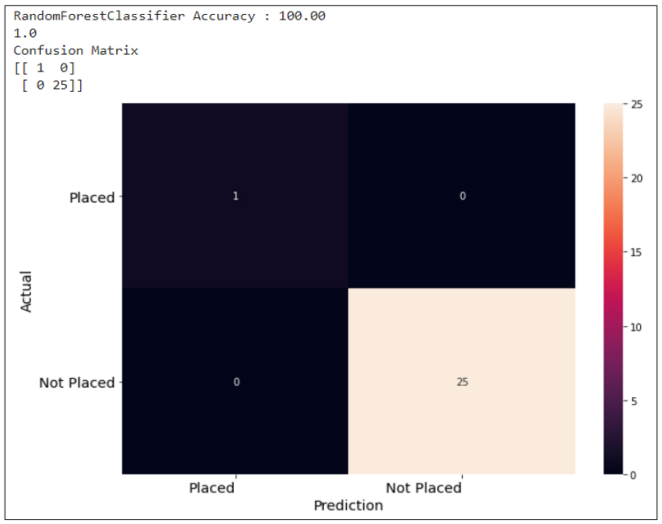

Figure.3 Confusion Matrix (Random Forest)

From the above visualization image, we can see the confusion matrix, which has **0+0=0 incorrect predictions** and **25+1=26 correct predictions** out of 26 samples that were taken for predicting**.**

https://doi.org/10.51767/jc1303

Therefore, we can say that compared to other classification models, the Random forest classifier produced good results compared to other models.

XI. REASON FOR IMPLEMENTING THE RANDOM FOREST:

The reason for implementing our project with this model apart from its accuracy, the other reason is that these predicted values never fall outside our highest or lowest values in our dataset which is because random forest always predicts the average values in the dataset.

XII.DATA BACKGROUND:

We have a sample of 128 observations of students and 2 attributes related to the observations here. We would like to predict and measure the relationship between placed Vs not placed along with their packages .

XIII. OUR GOAL IS:

To predict whether future students would be placed or not with their annual package by extracting the information from the real time data.

Our Placement Prediction model has been developed by extracting the features from students' qualifications, deploying machine learning models to train from the features to make the prediction and has been tested to observe the accuracy of the model. The table **Table.5** below evaluates the performance in terms of accuracy metrics, showing that the model is trained consistently and has been validated in order to accept a large set of data.

XIV. DATA ANALYSIS AND INTERPRETATION

https://doi.org/10.51767/jc1303

Table.4 Performance Measure Tabulation

ML Algorithm	True Positive	False Positive	False Negative	True Negative	Accuracy
Logistic Regression	0	1	0	25	96.15
Decision Tree Classifier	0	2	0	24	92.31
Random Forest Classifier	1	0	0	25	100.00

The evaluation of the models is made more capable to understand by visualizing the performance measures in the form of a bar chart. The accuracy obtained by using the Logistic Regression is 96.15%, Decision tree classifier is 92.31% and Random Forest classifier is 100.00%. It is observed that the Random Forest classifier provides a better result in predicting the status of the student's placement packages, since it works well for the large datasets. Hence, from the above analysis and prediction, it is preferable to utilize the Random Forest classifier to determine placement. Annual package results.

https://doi.org/10.51767/jc1303

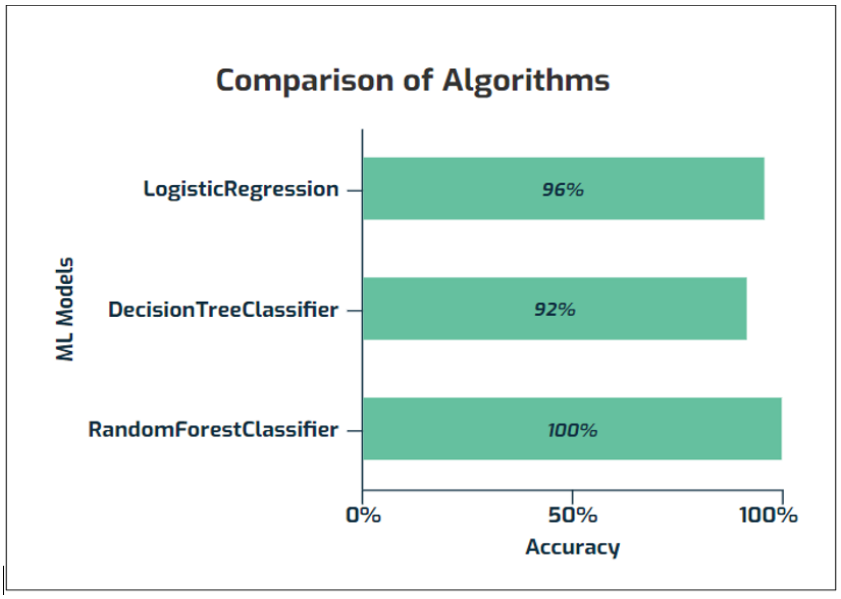

Figure.5 Visualization of the Algorithm

XV. RELATED WORK

C K Srinivas et al. (2020) [1] proposed the Student placement prediction using Machine Learning. In this paper, the algorithm used for prediction is logistic regression. The algorithm proposed here will check the eligibility of candidates on the basis of percentage & other technical knowledge. This algorithm trains the model to predict the probability of the student getting placed by the training dataset provided in their placements.

Varsha k. Harsher et al. (2020)[2] proposed the Student Placement Prediction System using Machine learning. The system proposed here is used to predict the probability of an undergraduate student getting placed. The author implemented different classification algorithms like multilayer perceptron (MLP), Logistic model tree (LMT), Sequential Minimal Optimization (SMO), and Simple logistic and logistic classifiers for prediction of the dataset. After comparing the performance measurement of each algorithm, the author concluded that MLP and LMT yield the maximum accuracy when compared to other algorithms for the given dataset.

Abhishek S. Rao et al. (2019)[3] proposed the Student Placement Prediction Model: A Data Mining Perspective for the Outcome Based on Education System. In this paper, the author had

https://doi.org/10.51767/jc1303

used data mining algorithms such as Support Vector Machine (SVM), K- Nearest Neighbor (KNN) and Artificial Neural Networks (ANN) for predicting the student placement. The performance measurement of the model was evaluated and the author concluded that ANN has given best results with highest accuracy. The model also suggests the preparation required to excel in placement, which helps the student to analyze them and get placed in a company.

Senthil Kumar Thangavel et al. (2017)[4] proposed the Student Placement Analyzer: A Recommendation System Using Machine Learning. In this paper, the author used Decision tree classifier for predicting the students placement. It presents a recommendation system that predicts Dream Company, Core Company, and Mass Recruiters, Not eligible and not interested in Placements are the five placement statuses that students must have. This model helps the placement cell within an organization to identify the prospective students, their future plans and pay attention to them to improve their technical as well as interpersonal skills.

Mangasuli Sheetal B et al. (2016)[5] proposed the Prediction of Campus Placement Using Data Mining Algorithm-Fuzzy logic and K nearest neighbour.This paper deals with Fuzzy logic and K nearest neighbor(KNN) algorithm for predictions. The validation for the two algorithms are compared and checked based on their performance and accuracy. The authors concluded that KNN algorithm results in highest accuracy of 97.33%.

Ravina Sangha et al. (2016)[6] proposed the Student's Placement Eligibility Prediction using Fuzzy Approach. In this paper, a rule based classification is proposed to predict the eligibility of students. The author had emphasized an efficient algorithm with the technique Fuzzy for placement prediction. The system is designed in such a way that will help the student in getting placed in a company which he/she is capable of and also help the institute to improve student placement record rate.

T.Jeevalatha et al.(2014)[7] proposed the Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms. In this paper, it describes how different Decision tree algorithms are used to predict students' performance in their placement. It deals with Decision tree algorithms such as ID3, CHAID, and C4.5 which were implemented using Rapid

https://doi.org/10.51767/jc1303

Miner tool. Validation is checked for three algorithms and accuracy is found for them. The authors concluded that ID3 algorithm is best in predicting the placement results with accuracy 95.33%.

Ajay Kumar Pal et al. (2013)[8] proposed the Classification model of Prediction for Placement of Students. In this paper, they have investigated different classification algorithms using data mining tools such as WEKA. They have made a comparison with three different algorithms for predicting the placement of students. The algorithms are Naive Bayes classification, Multilayer perceptron (MLP) and C4.5 tree. It was reported that Naive Bayes classification algorithm had the highest predictive accuracy of 86.15% with the lowest average errors.

V.Ramesh et al. (2011)[9] proposed the Performance Analysis of Data mining Techniques for Placement Chance Prediction. In this paper, they have dealt with data mining methodologies to study students' performance. Overall idea of this paper illustrates how well different classification techniques are used as predictive tools, after comparing the performance of every algorithm. The result is concluded as the MLP algorithm gives the greatest accuracy when compared to other algorithms.

XVI. MODELING ATTRIBUTES

a. DATASET AND ATTRIBUTE SELECTION MODULE

Student data is the performance of a student when in a class or an activity. This is primarily used as a means of assessing students' progress towards learning. The collected dataset has the results of the students, which has instances and 2 attributes. The dataset is in the 'CSV' file format.

b. PREPROCESSING MODULE

In the Data preprocessing we prepare the raw data and make it suitable for a machine learning model that we have trained. It is mandatory to clean the data and put it in a formatted

https://doi.org/10.51767/jc1303

way for the model. Data collection, data transformation, data integration and data cleaning takes place.

c. CLASSIFICATION MODULE

Now, the required students' data is extracted and classification rules are being implemented. The algorithms are applied that classifies the students to be placed to the given placement package statuses. The model here implemented is Random Forest Classifier.

d. ARCHITECTURE

At first, we collected the real time dataset (i.e., taken from our college pass outs. Then we trained our model with the real time data. Later, we can feed our data to the model and predict the output.

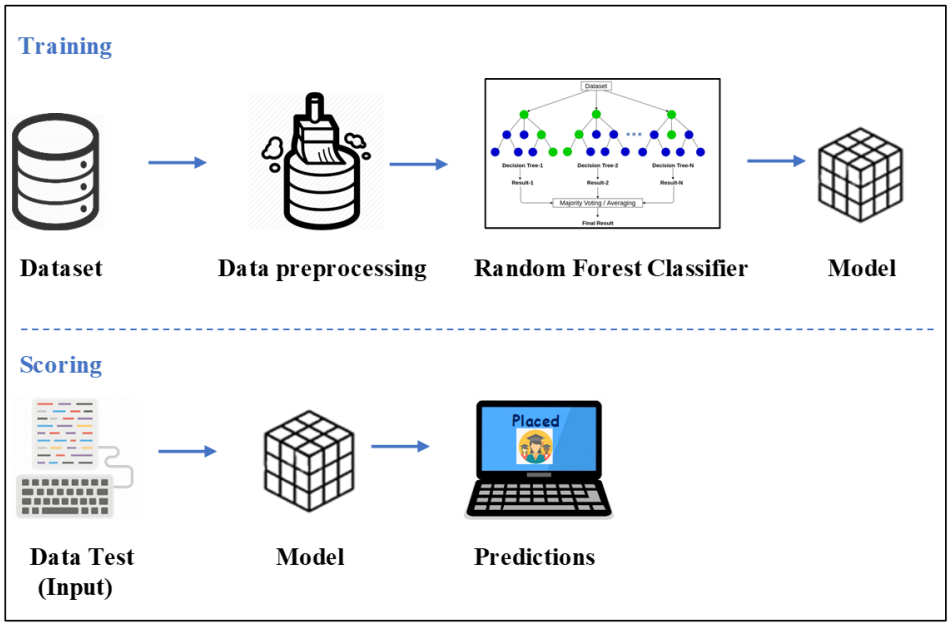

Figure.6 Architectur

https://doi.org/10.51767/jc1303

Sl. No.	dep	regno	studentname	cgpa	arrear	Name of the Employeer	Package	Annual Package	Results
1	CSE	1601019	ASWATH .S	8.34	0	ZOHO CORPORATION	4.6	460000	1
2	CSE	1601022	BALAHARISH .M	8.01	0	ZOHO CORPORATION	4.6	460000	1
3	CSE	1601024	BHARATH KUMAR. M	8.56	0	ZOHO CORPORATION	4.6	460000	1
4	CSE	1601063	SUVAB BARAL	9.18	0	RENTLY	8	800000	1
5	CSE	1601124	MOHAN RAM .M	8.68	0	ZOHO CORPORATION/ HACK WITH INFY	4.6	460000	1
6	CSE	1601209	SANJAY BHARATHI. S	8.7	0	ZOHO CORPORATION	4.6	460000	1
7	CSE	1601252	VIMAL ADIT .V	8.53	0	ZOHO CORPORATION	4.6	460000	1
8	CSE	1601201	RUBESH.C.V.	8.56	0	DELTAX/HACK WITH INFY	5	500000	1
9	CSE	1601015	ARAVIND .K.R	8.3	0	TRIMBLE	5.5	550000	1
10	CSE	1601008	AJITHAN .S	7.91	0	GO FRUGAL	4.2	420000	1
11	CSE	1601241	SWATHI .P	7.65	0	GO FRUGAL	4.2	420000	1
12	CSE	1601253	VISHAL. N	7.73	0	GO FRUGAL	4.2	420000	1
13	CSE	1601032	DHANDAPANI .B	7.8	0	QUBER COMM	3	300000	1
14	CSE	1601049	HARIHARAN .A	7.67	0	ODESSA TECHNOLOGIES/RYTHMOS	4	400000	1
15	CSE	1601148	PRAVEENKUMAR .S	8.18	0	ODESSA TECHNOLOGIES	4	400000	1
16	CSE	1601053	HESEN NIVAS .M	8	1	ODESSA TECHNOLOGIES/BNY MELLON TECHNOLOGY	4	400000	1
17	CSE	1601014	ANISH TIWARI	8.01	0	ODESSA TECHNOLOGIES	4	400000	1
18	CSE	1601002	AASHIQ .M	8.16	0	DIVUM	3.6	360000	1
19	CSE	1601208	SANDHYA .R	7.56	0	DIVUM	3.6	360000	1
20	CSE	1601214	SARUN .S	7.66	0	DIVUM	3.6	360000	1
21	CSE	1601009	AKHILA A	8.87	0	SOPRA STERIA	4	400000	1
22	CSE	1601140	POOPALAKANNAN .M	8.21	0	SOPRA STERIA	4	400000	1
23	CSE	1601225	SHRI HARINI G	8.11	0	SOPRA STERIA/HACK WITH INFY	4	400000	1
24	CSE	1601240	SWARNALATHA .N	8.03	0	PAYODA	4.5	450000	1
25	CSE	1601003	ABARNA .D	8.49	0	RFPIO	5	500000	1
26	CSE	1601144	PRASANNA .S	8.38	0	RFPIO	5	500000	1
27	CSE	1601020	AZHAGU JANAGI .A.L	8.87	0	DUCEN IT	4.2	420000	1
28	CSE	1601160	ROOPASRI .K	8.42	0	DUCEN IT	4.3	430000	1

Figure.7 Real time Dataset

Here in **Figure.7** we have mentioned the attributes like CGPA and Arrears which were the main aspect in the dataset.

XVII. BLOCK DIAGRAM

As the web application is visited by the user, a simple get request is sent from the streamlit front-end to backend. This request signals the backend server to run the model file, where the dataset is being fed.

Once this data is provided to the model, it starts the calculation, compares with the fed in dataset and provides the necessary output with the help of the pickle file.

When the model provides the output, the server sends these outputs as the request and these values are displayed in the User Interface.

https://doi.org/10.51767/jc1303

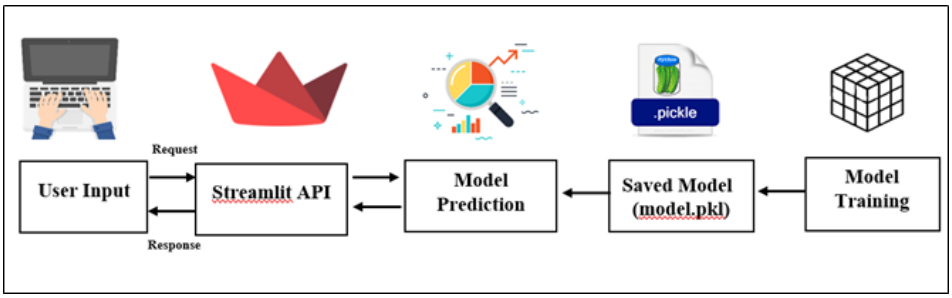

Figure.8 Block Diagram

XVIII. RESULT

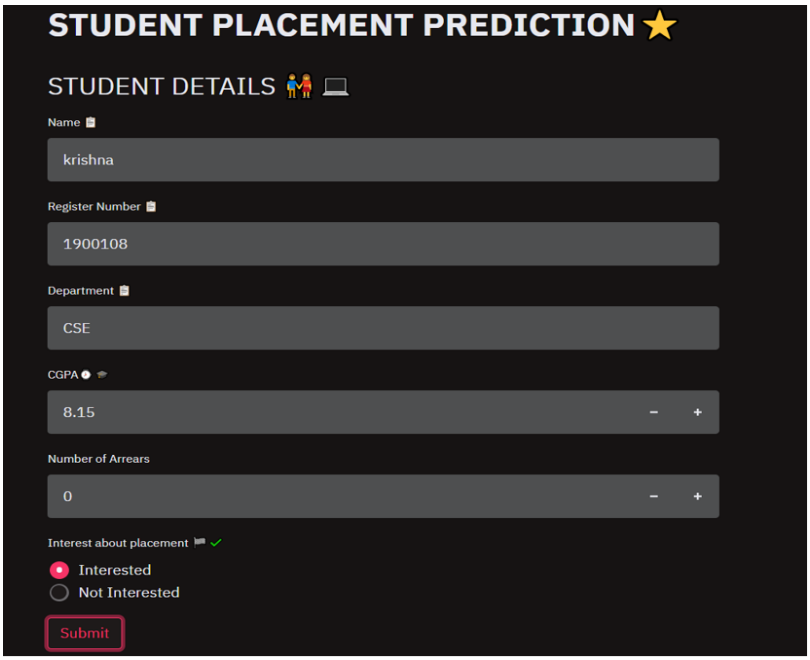

Figure.9 Front end of the project

Here in **Figure.9** represents the Front end of the project where the users provide the required data to fetch the predicted details.

https://doi.org/10.51767/jc1303

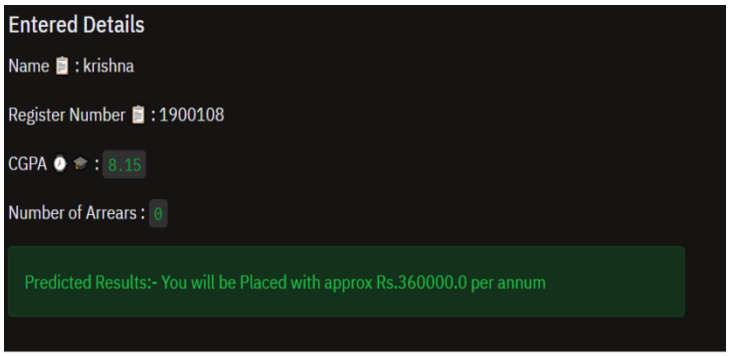

Figure.10 Result

The **Figure.10** This screen represents the final outcome of the project wherein it shows the predicted results for the above **Figure.9** given data with the package details.

XIX. CONCLUSION

In college, the placement and their results are very important to the university. More so, the developments of these models are bringing a shift in the way activities are accomplished by the students. As a result, the students are able to analyze their skills with these new models. At the same time the placement officers could look into their wards eligibility. It may increase the managerial skills of the students.This will also help teachers to pay special attention to the progress of their wards.

XX. FUTURE SCOPE

In future systems, if required will provide Admin and user login for the keen understanding of their performances. We may also create other models to adapt to the students of various course outcomes. We can also redesign this project for the convenient needs or methodology as it was developed to meet the specific goals for an organization.

REFERENCES

1. Mr. C K Srinivas, Nikhil S Yadav, Pushkar A S, R Somashekar, Sundeep K “Students Placement Prediction using Machine Learning”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), May 2020.

https://doi.org/10.51767/jc1303

2. Varsha K. Harihar, D.G.Bhalke"Student Placement Prediction System using Machine Learning", SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology Volume 12, Special issue 2, 2020.
3. Abhishek S. Rao, S V Aruna Kumar, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda "Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System", International Journal of Recent Technology and Engineering (IJRTE), ISSN: Volume–8 Issue–3, September 2019.
4. Senthil Kumar Thangavel, DivyaBharathi, AbijithSankar "Student Placement Analyzer: A Recommendation System Using Machine Learning", International Conference on advanced computing and communication systems (ICACCS-2017), Jan 2017, Coimbatore, INDIA.
5. MangasuliSheetal B and Prof. Savita Bakare "Prediction of Campus Placement Using Data Mining Algorithm-Fuzzy logic and K nearest neighbor", International Journal of Advanced Research in Computer & Communication Engineering (IJARCCE) Volume. 5, Issue 6, June 2016.
6. Ravina Sangha, AkshaySatras, LishaSwamy, Gopal Deshmukh "Student's Placement Eligibility Prediction using Fuzzy Approach", International Journal of Engineering and Techniques - Volume 2 Issue 6, Nov - Dec 2016.
7. T.Jeevalatha (M. Phil Scholar), N.Ananthi, D.Saravana Kumar, "Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms", International Journal of Computer Applications (IJCA) (0975 –8887), Volume 108 –Number 15, December 2014.
8. Ajay Kumar Pal (Research scholar) and Saurabh Pal"Classification Model of Prediction for Placement of Students", International Journal of Modern Education and Computer Science, 2013, 11, 49-56.
9. V Ramesh, P Yasodha "Performance analysis of data mining techniques for placement chance prediction", International Journal of Scientific & Engineering Research 2(8), 1, 2011.