e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

# CAMPUS PLACEMENT PREDICTION SYSTEM USING MACHINE LEARNING

**Prof. S.S. Kashid[*1], Ashish Badgujar[*2], Vishal Khairnar[*3], Anurag Sagane[*4], Nishant Ahire[*5]**

[*1]Assistant Professor, Smt. Kashibai Navale College Of Engineering, Pune, India.

[*2,3,4,5]Smt. Kashibai Navale College Of Engineering, Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS36827

## ABSTRACT

A campus placement prediction system is developed to calculate the possibility of a student getting jobs in a company through campus placements. The model takes many parameters which can be used to get an idea about the skill level of the student. While some data are taken from the college level like academic performance, CGPA, pointers, attendance etc others are obtained from tests conducted in the placement management system. Combining these data points, the model is to accurately predict if the student will or will not be placed in a company. Also, Data from past year students are used for training the model. We are using educational data mining by which we can get real past year's students data of that specific college. This will help the machine learning model to be more effective about the predictions of that particular college.

The purpose behind this placement prediction system is to help students improve their results, academic performance and also develop another soft skill that will help them to maximize their chance of getting placed. Such a study will help the faculties of the college to train the students accordingly and improve the placement department of their institutions. This will give an idea about how students are doing and will ensure their institution can satisfy the needs of recruiters.

For this supervised machine learning especially logistic regression is better, Logistic model designing plays a key role to get correct predictions. This process includes the selection of tuples for training data and their pre-known outcome often known as real data.

**Keywords:** Data Mining, Campus Placement, Logistic Regression, EDM, Supervised Learning, Logit Function, Machine Learning, Accuracy, Prediction.

## I. INTRODUCTION

Academics is one of the most important things in any student's life. A good academic result shows how a student is performing and is he/she good in his/her studies or not. Nowadays many companies are recruiting freshers based on their academic performance in college.

Placements are very important for each and every college. The success of the college is measured by the campus placement of the students. Our approach is about the prediction and analyses for the placements which are important in the colleges and which will help students to improve their placement chances. The main objective of this model is to predict whether the student gets placed or not in campus recruitment. For this we are considering the academic history of students' marks in various important core subjects like Data Structures and Algorithms, Object Oriented Programming, Database Management System, Operating System, Computer Network Security ,Cloud Computing ,etc. which are very important in the perspective of campus placement.

Hence, this will help students to improve their academics and acquire more skills for better placement. For institutions, it will help to build better placements and academic records as well to attract new admissions as well as to get reputation.

## II. RELATED WORK

In this section, a literature survey is done on some research papers on Campus placement using Supervised machine learning.

In [1] Irene Treesa Jose, Daibin Raju, Jeebu Abraham Aniyankunju, Joel James, Mereen Thomas Vadakkel discussed various machine learning models and their efficiency comparison. Among the supervised machine

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

learning algorithms, KNN, SVM, Logistic regression and Random forest also showed their advantages and disadvantages. They trained and predicted the placement status of students based on the same dataset and found the True Positive, False Positive, False Negative, True Negative, and accuracy of each algorithm.

| ML Algorithm | True Positive | False Positive | False Negative | True Negative | Accuracy |
|---|---|---|---|---|---|
| KNN | 49 | 2 | 2 | 30 | 95.18 |
| Logistic Regression | 51 | 0 | 2 | 30 | 97.59 |
| Random Forest | 50 | 0 | 3 | 30 | 96.38 |
| SVM | 51 | 0 | 0 | 32 | 100 |

**Fig 2.1** The final result of performing various machine learning algorithms [1]

As in the fig 2.1 SVM has 100% accuracy but also it has 0 false positive and false negative means its precision is 1 and recall is also 1. Its F-measure is also 1 then we can say that the model is overfitted. Logistic regression had 97.59% accuracy with precision 1 but recall 0.96 means the efficiency of the model is better than the other algorithms.

In [2] D. Satish Kumar, Zailan Bin Siri, D.S. Rao, S. Anusha used UG and PG both student data like MBA CGPA, UG CGPA, soft skills. Their model results from the odds of a student being placed were positively related to his PG CGPA (2.83(p=0.001)), Gender, PG specialization, and UG specialization. 2. The higher the PG CGPA the more likely it is that a student is being placed. Based on the ROC analysis we may say that the model with the six predictors considered has a 60% accuracy in predicting whether a student gets placed or not. Also, Given the same CGPA students belonging to Marketing and Finance are more likely to be placed.

In [3] Manoj K Shukla, Pranay Rambade, Jay Torasakar, Rakesh Prabhu, Prof. Deepali Maste discussed Students Placement Prediction Model Using Logistic Regression.

They used previous year placement data as training data which they applied on a logistic mathematical equation. After mapping this data the final logistic model was used to predict the placement chances of next year students with help of thirteen same parameters which were used during model processing. They have also shown the pseudo-code of the whole process.

In [4] Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar discussed PPS with Logistic regression. Placement and academic data provided by GNDEC are used in this model. Data is cleaned and converted into numbers to use for number crunching. Machine learning technique is used to design and implement a logistic classifier that predicts the probability of the student getting placed. This predictive model predicts the future outcomes of each student in future sessions of jobs. The parameters are learned by running a gradient descent algorithm on training data that provides an idea about the most necessary features that are responsible for a positive outcome.

In [5] Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi, Naini Dubey proposed placement predictions using the K-nearest neighbors algorithm. They also compare the results of the same against the results obtained from other models like Logistic Regression and SVM. For that, they used students data like the academic history of the student as well as their skill set like, programming skills, communication skills, analytical skills, and teamwork, which are tested by the hiring companies during the recruitment process.

In [6] Shawni Dutta1 and Samir Kumar Bandyopadhyay applied Data mining and knowledge discovery processes to the academic career of students. Which uses an ensemble approach-based voting classifier for choosing best classifier models to achieve better results over other classifiers. Results of the experiment have shown 86.05% accuracy of ensemble-based approach which is better than other classifiers. This prediction has an accuracy of 86.05%, F1-Score 0.86, CohenKappa Score is 0.72, and MSE 0.14 which is better than other classifiers.

In [7] Vijay N. Kalbandea , Dr. Chandrahas. C. Handa, they developed an Artificial Neural Network (ANN) model for predicting the probability of placement based on employability skills of students. They identified 22 employability skills that were validated by stakeholders and grouped them into four major employability skills. They used Aptitude, Communication, Technical & Personality skill as input variable, and success/failure in campus placement considered as output variable. They have shown the potential of the artificial neural

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

network for enhancing predicted accuracy to measure performance of engineering students in campus placement for IT Sector. The model achieved an accuracy of 80.19 %, which shows the potential efficiency of the Artificial Neural Network as a prediction tool.

## III. PROPOSED SYSTEM

Firstly the project object is designed and the data is collected by the college, after collecting the data we prepare the data into training and testing datasets. The data is separated into dependent variables and independent variables and then the outliers are removed.

1) Model is trained with training data.

2) The model is fine-tuned and the training data is further classified according to the parameters specified in the model

3) After training the model the testing dataset which was stored earlier is processed to the model for the prediction.

4) If the testing values come out to be true the values are finally set as the prediction.

5) If the testing values comes out to be false means not satisfactory then the values are again sent back for the fine-tuning of the model.

**Algorithms used :**

1. Logistic Regression
2. XGBoost

**1. Logistic Regression (Binary classification)**

Logistic regression is a common statistical method used in predicting binary outcomes, such as whether or not a person will be placed in a job after completing a training program.

To use logistic regression for placement prediction, we typically started by collecting data on individuals who have completed the training program, including information such as their education level, work experience, and performance in the training program

To make predictions using the logistic regression model,we simply input the values of the predictors for a new individual and the model would output the estimated probability of that individual being placed in a job. If the probability is above a certain threshold (e.g., 50%), the system would predict that the individual will be placed in a job; otherwise, not.

**2. XGBoost (eXtreme Gradient Boosting)**

XGBoost is a popular machine learning algorithm that can be used for multiclass classification problems, including predicting the package range of students in campus placement based on their performance and other parameters. Here's a high-level overview of problem solving approach using XGBoost:

**1) Data Preparation:** Collected data on students' academic performance, skills, work experience, and other relevant parameters. Preprocessing the data by handling missing values, handling outlier, encoding categorical variables, and scaling numerical variables.

**2) Feature Selection:** Identify the most important features that are highly correlated with the target variable (i.e., package range). You can use techniques such as correlation analysis, mutual information, or feature importance ranking provided by XGBoost.

**3) Train-Test Split:** Split the data into training and testing sets. Typically,80-20 split is used.

**4) XGBoost Model Training:** Train an XGBoost classifier on the training data. XGBoost is a gradient boosting algorithm that builds an ensemble of decision trees iteratively to minimize the loss function (e.g., softmax for multiclass classification).

**5) Hyperparameter Tuning:** Optimize the hyperparameters of the XGBoost model using techniques such as grid search or random search. Important hyperparameters include the learning rate, number of trees, maximum depth, and regularization parameters.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

**6) Model Evaluation:** Evaluate the performance of the XGBoost model on the testing data using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. You can also visualize the feature importance and decision boundaries of the model.

**7) Deployment:** Model is packaged into joblib file and further used in flask API.

Here , Comparison of the 3 Algorithms for multiclass classification is done through which the best one is selected.

1. LogisticRegression
2. Random Forest
3. XGBoost

**Table 3.1:** Comparison of Algorithms

| Algorithm | precision | recall | f1-score | accuracy |
|---|---|---|---|---|
| Logistic Regression | 0.522 | 0.522 | 0.522 | 0.522 |
| Random Forest | 0.819 | 0.819 | 0.819 | 0.819 |
| XGBoost | 0.831 | 0.831 | 0.831 | 0.831 |

As XGBoost shows higher accuracy among the listed algorithms , hence XGBoost is used.

**Overall Architecture Diagram**

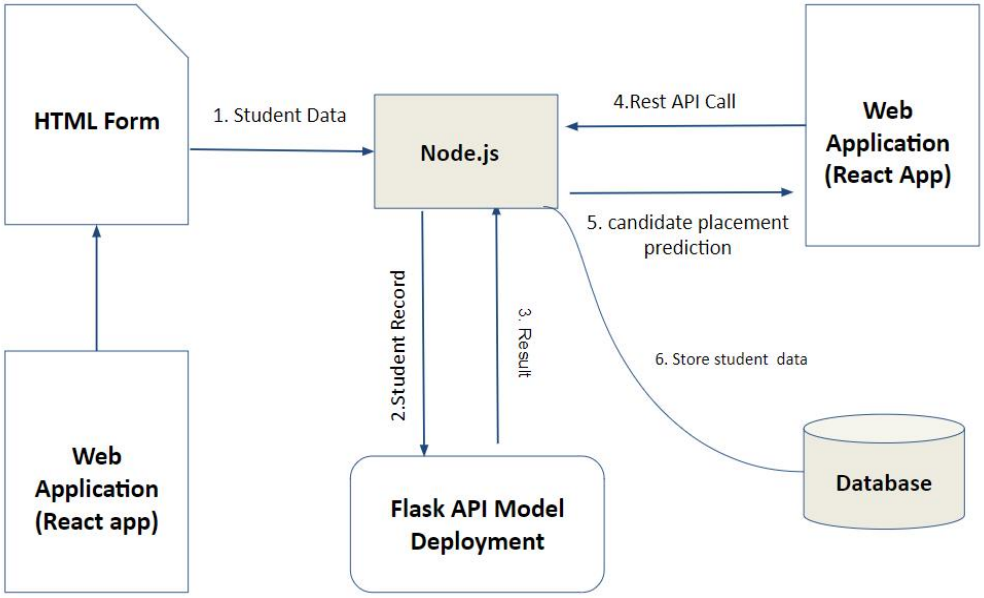


**Fig 3.1** Overall Architecture of System[8]

Fig.3.1. shows the overall architecture of the system , there are two spaces: user space and server space. In user space there is a front end where user enters marks as input parameter and rest api call is made. In server space there is Python Flask API and ML Model. after processing the input data the output is sent to the user space from server space. Users can see the output on Web Application Frontend.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

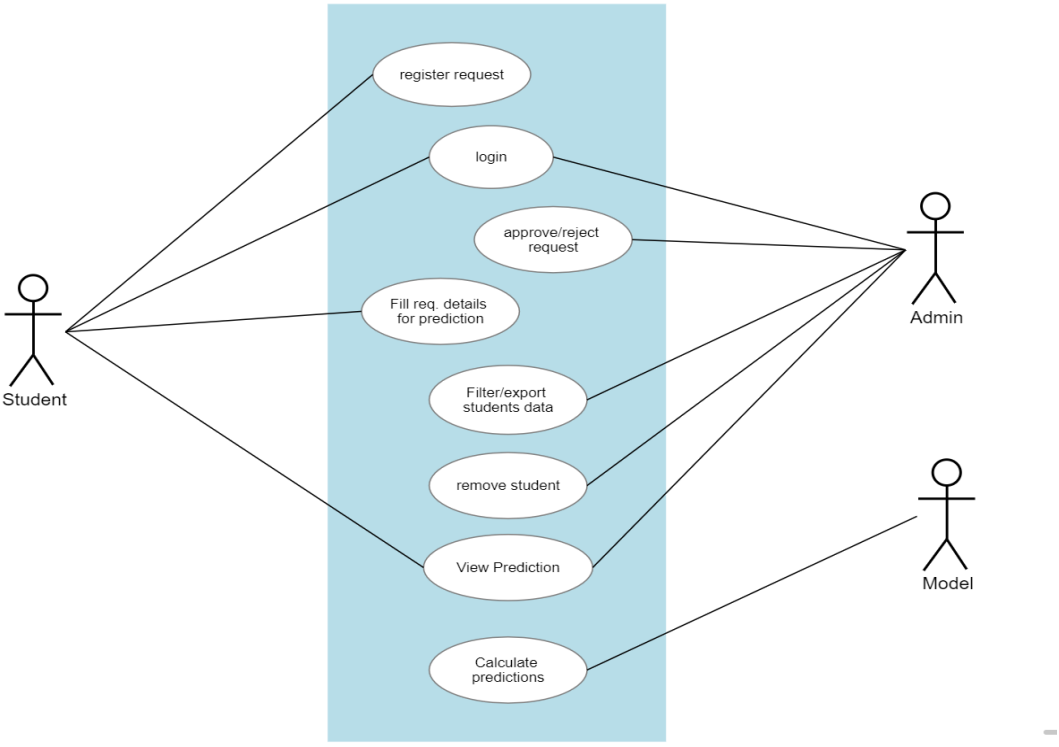


**Fig 3.2:** Use-Case Diagram [9]

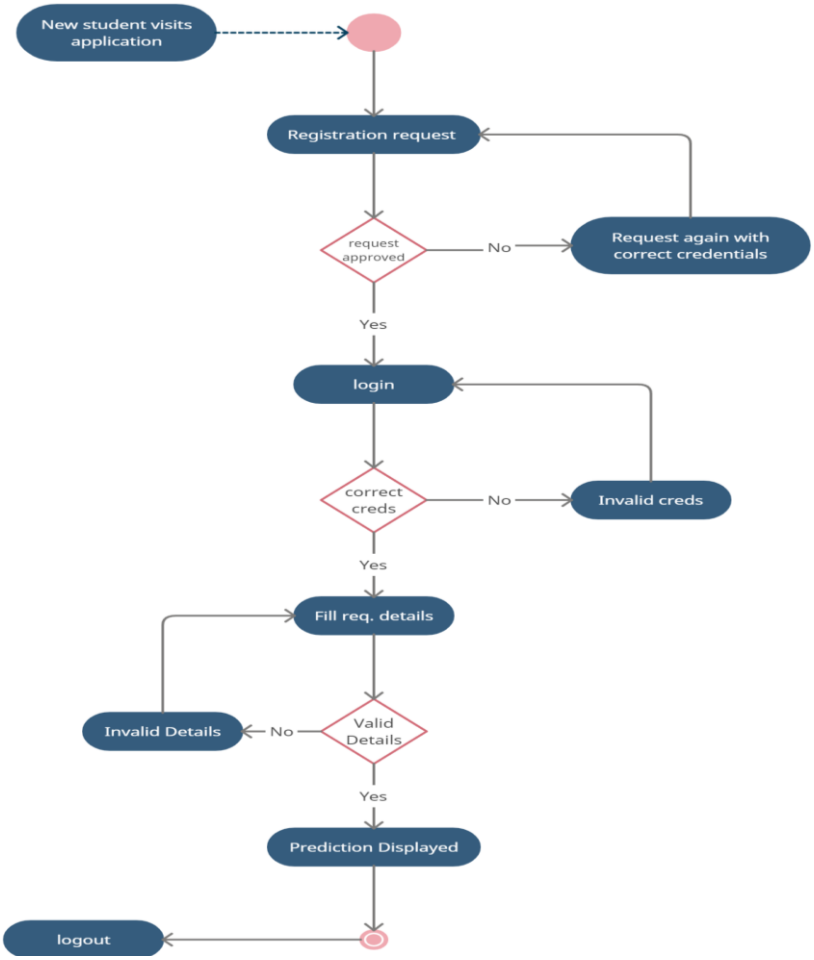


**Fig 3.3:** Activity Diagram (student) [10]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

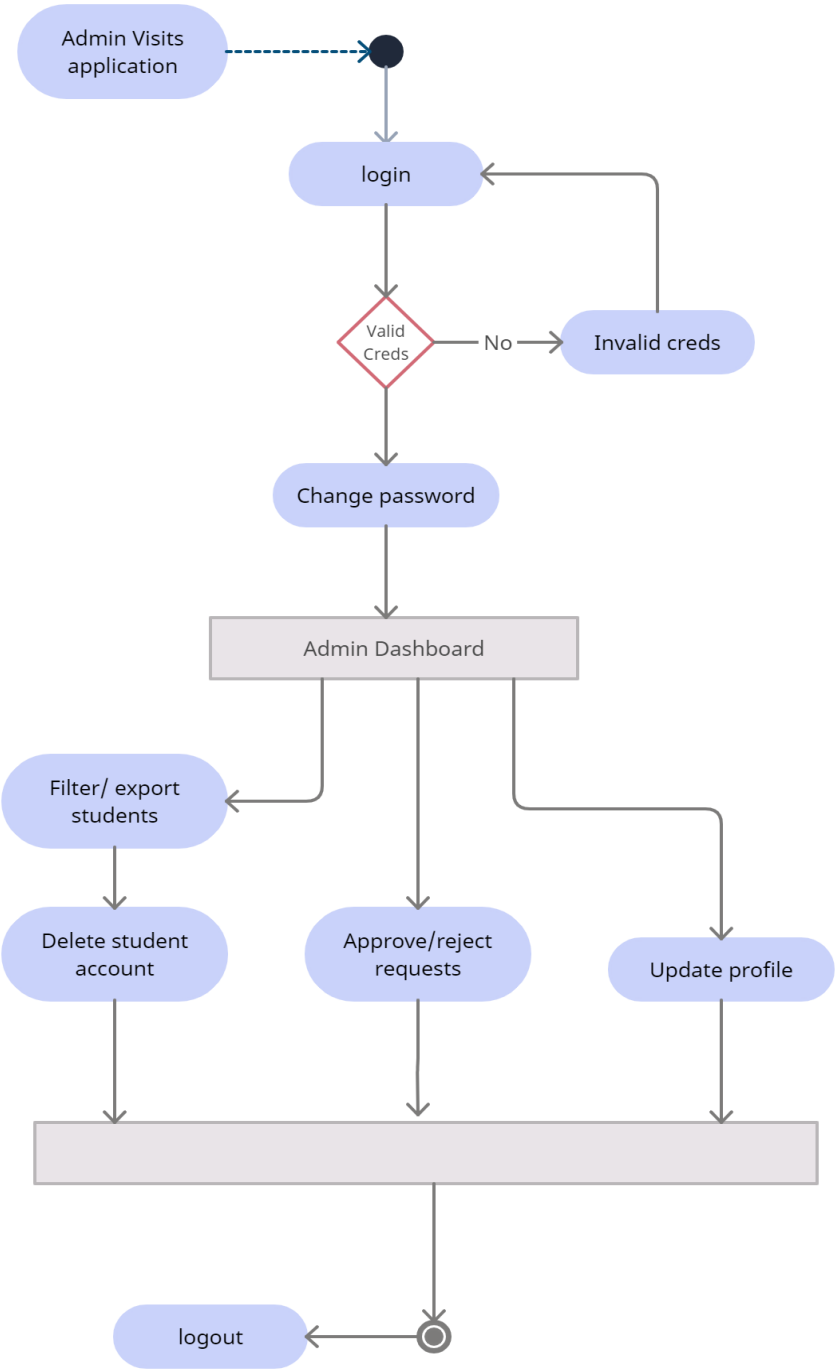


**Fig 3.4:** Activity Diagram (Admin) [11]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

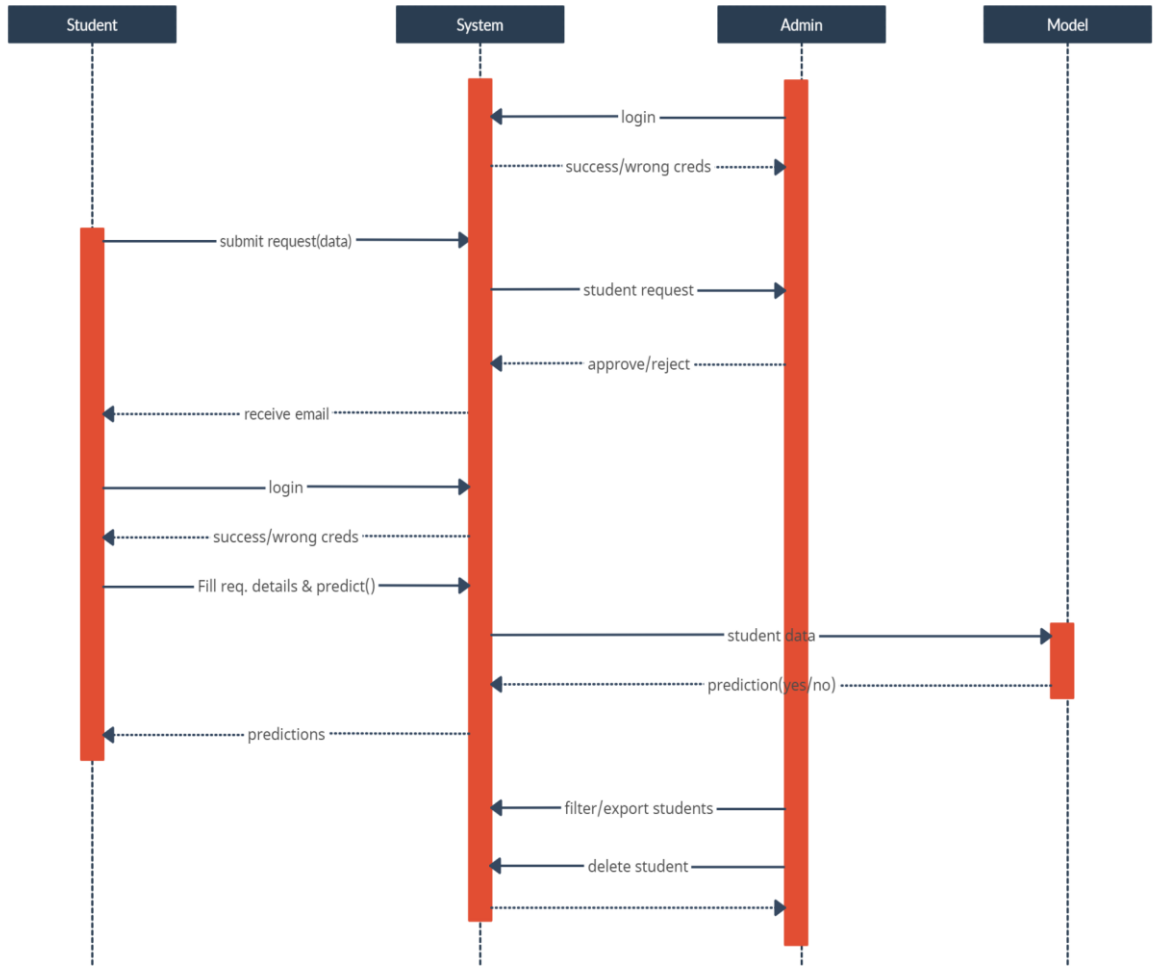


**Fig 3.5:** Sequence Diagram [12]

## IV. CONCLUSION

In this way, logistic regression from a machine learning method in predicting the campus placement of students based on his/her performance in academics and other skills and exams. The model is trained using the previous year's data like students marks in all semesters, personal interests, internships, other activities, etc.

The limitations of the Project is that the student should not do any malpractice while appearing for the general aptitude exams which we have asked them to solve, in order to test their knowledge and he/she should fill in their correct academic details .

This seminar report is undertaken to explain machine learning prediction approaches and improve the efficiency of Machine Learning Model to predict the campus placement of students.

## V. REFERENCES

[1] Irene Treesa Jose, Daibin Raju, Jeebu Abraham Aniyankunju, Joel James, Mereen Thomas Vadakkel "Placement Prediction using Various Machine Learning Models and their Efficiency Comparison" Volume 5,P 1007,1008 Issue 5, May – 2020

[2] D. Satish Kumar, Zailan Bin Siri, D.S. Rao, S. Anusha "Predicting Student's Campus Placement Probability using Binary Logistic Regression" Volume-8 Issue-9, July, 2019

[3] Manoj K Shukla, Pranay Rambade, Jay Torasakar, Rakesh Prabhu, Prof. Deepali Maste "Students Placement Prediction Model Using Logistic Regression" Volume 5, Issue 01

[4] Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar "PPS - Placement Prediction System using Logistic Regression" P 338,339 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE)

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

[6] Shawni Dutta1 and Samir Kumar Bandyopadhyay "Forecasting of Campus Placement for Students Using Ensemble Voting Classifier" Asian Journal of Research in Computer Science 5(4): 1-12, 2020; Article no.AJRCOS.57125 ISSN: 2581-8260

[7] Vijay N. Kalbandea , Dr. Chandrahas. C. Handa "Predicting the Performance of Engineering Students in Campus Placement for It Sector by Using ANN" Volume III, Issue II, February 2016 IJRSI ISSN 2321 – 2705.