VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590018

Project Report
on

"MACHINE LEARNING PREDICTIVE MODEL OF JOB PLACEMENTS FOR UNDERGRADUATE STUDENTS"

Submitted in partial fulfilment of the requirements for the VIII Semester

Bachelor of Engineering
in
COMPUTER SCIENCE AND ENGINEERING
For the Academic Year
2019-2020
BY

Pranav Manjunath	**1PE16CS114**
Renu Prasad K	**1PE16CS194**
Kavya N P	**1PE16CS070**
Pavan Kumar V	**1PE15CS102**

UNDER THE GUIDANCE OF
Prof. Prajwala TR
Assistant Professor, Dept. of CSE, PESIT-BSC

Department of Computer Science and Engineering
PESIT - BANGALORE SOUTH CAMPUS
Hosur Road, Bengaluru - 560100

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590018

Project Report

on

"MACHINE LEARNING PREDICTIVE MODEL OF JOB PLACEMENTS FOR UNDERGRADUATE STUDENTS"

Submitted in partial fulfilment of the requirements for the VIII Semester

Bachelor of Engineering

in

COMPUTER SCIENCE AND ENGINEERING

For the Academic Year

2019-2020

BY

Pranav Manjunath	**1PE16CS114**
Renu Prasad K	**1PE16CS194**
Kavya N P	**1PE16CS070**
Pavan Kumar V	**1PE15CS105**

UNDER THE GUIDANCE OF

Prof. Prajwala TR

Assistant Professor, Dept. of CSE, PESIT-BSC

Department of Computer Science and Engineering

PESIT - BANGALORE SOUTH CAMPUS

Hosur Road, Bengaluru - 560100

PESIT - BANGALORE SOUTH CAMPUS

HOSUR ROAD, BENGALURU - 560100

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project work entitled "Machine Learning Predictive Model of Job Placements for Undergraduate University Students" carried out by "Pranav Manjunath, Renu Prasad, Kavya NP, Pavan Kumar V bearing USN's 1PE16CS114, 1PE16CS194, 1PE16CS070, 1PE15CS102" respectively in partial fulfillment for the award of Degree of Bachelors (Bachelors of Engineering) in Computer Science and Engineering of Visvesvaraya Technological University, Belagavi during the year 2019-2020. It is certified that all corrections/ suggestions indicated for internal assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said Degree.

Signature of the Guide	Signature of the HOD	Signature of the Principal
Prof. Prajwala TR	**Dr. D Annapurna**	**Dr. Subhash Kulkarni**
Assistant Professor	**HOD, CSE**	**Principal, PESIT-BSC**

External Viva

Name of the Examiners	**Signature with Date**
1.	
2.	

Acknowledgements

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned my effort with success.

We are indebted to our Guide, **Prof. Prajwala T R**, Assistant Professor, Department of Computer Science and Engineering, PESIT - Bangalore South Campus, who has not only coordinated our work but also given suggestions from time to time.

We are also extremely grateful to **Prof. Animesh Giri** for providing us with the PESIT placement dataset from 2013-2019, our Project Co-ordinators, **Prof. Shubha Raj K B**, Assistant Professor and **Prof. Sheba Pari**, Assistant Professor, Department of Computer Science and Engineering, PESIT Bangalore South Campus, for their constant support and advice throughout the course of preparation of this document.

We are greatly thankful to **Dr. Annapurna**, Professor and HOD, Department of Computer Science and Engineering, PESIT Bangalore South Campus, for his able guidance, regular source of encouragement and assistance throughout this project.

We would like to express our immense gratitude to **Dr. Subash Kulkarni**, Director and Principal, PESIT Bangalore South Campus, for providing us with excellent infrastructure to complete our project work.

We gratefully acknowledge the help lent out to us by all faculty members of the Department of Computer Science and Engineering, PESIT Bangalore South Campus, at all difficult times. We would also take this opportunity to thank our college management for the facilities provided during the course of the project. Furthermore, we acknowledge the support and feedback of my parents and friends.

ABSTRACT

In this project, we aim to build an application to predict job placements for undergraduate students using PESIT BSC placement data from 2013-2019. The user has to input his student details and the application will display the predicted Firm CTC and Firm Type for the respected student.
A special feature of the application is the feedback mechanism. A survey was sent to alumni of PESIT, obtaining their feedback on placements. The survey consisted of questions regarding their employee satisfaction, potential career growth, and whether they felt it would be better to go to higher studies or directly go for placements. Based on the predicted CTC and Type of the firm, the model will predict and display the appropriate feedback from the PESIT alumni. This gives university students direct advice from their seniors and along with predicting where the student will be placed. The application uses two main sources of data, PESIT BSC placement dataset and senior's survey answers. A comparative analysis for various Machine Learning and Deep Learning algorithms have been implemented to decide the most accurate models. The final application uses Random Forest and KNN classification algorithms for classifying and predicting the Student's placeement. The application consists of a tab regarding exploratory data analysis of placements in PESIT. This tab consists of graphs and tables regarding the number of students who got placed via placements over the years, with respect to Tier, Branch of Engineering. This will help the students identify trends in the placements that occur in PESIT BSC. There are buttons on this page which takes the user to the college department's homepage.

Keywords: Placement Predictions, Classification Algorithms, K Nearest Neighbors, Random Forest, Artificial Neural Networks, Exploratory Data Analysis, Educational Data Mining, Machine Learning, Deep Learning, Decision Trees.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 College Placements

Campus placement is the program organized and conducted within educational institutions aiming to provide job opportunities to final year students. Industries visit campus to select students depending on their academic performance and skills. The main aim of campus recruitment is to identify the talented and qualified student before completing their course. A major plus point of college placement is that it reduces the time for an industry to select the candidates based on their requirements.

1.2 Motivation

Over a million engineers are produced in India every year. Unfortunately, the number of jobs available are not increasing at the same rate. Due to the intense competition, it is extremely necessary to understand the requirements needed to be placed in a company. Most of the undergraduate students join a course for securing a good job. Hence taking a wise career decision regarding placements after completing a particular course is important in a students life.

Using previous students' placement data, we can analyse the type of jobs offered and the students' profile that were placed. With this, we can provide insights to aspiring students to help them attain their dream job. Students should also be aware and have knowledge of the trend of job markets. This could also help in future academic planning of an university.

1.3 Proposed System

The end result of this project is a user friendly application that uses machines learning algorithms to predict a student's placement based on his/her profile. The

application will consist of a form that the user must fill. The form values will be sent to the back end machine learning model to predict various aspects of the student's placement. For the back end analysis, we will implementing Random Forest Algorithm, Artificial Neural Networks, Decision Tree, and K Nearest Neighbours Algorithm and identify which algorithms provide better results. The application will consist of a section dedicated for exploratory data analysis of PESIT placement data. It will also provide links to visit the various PES department websites for more information.

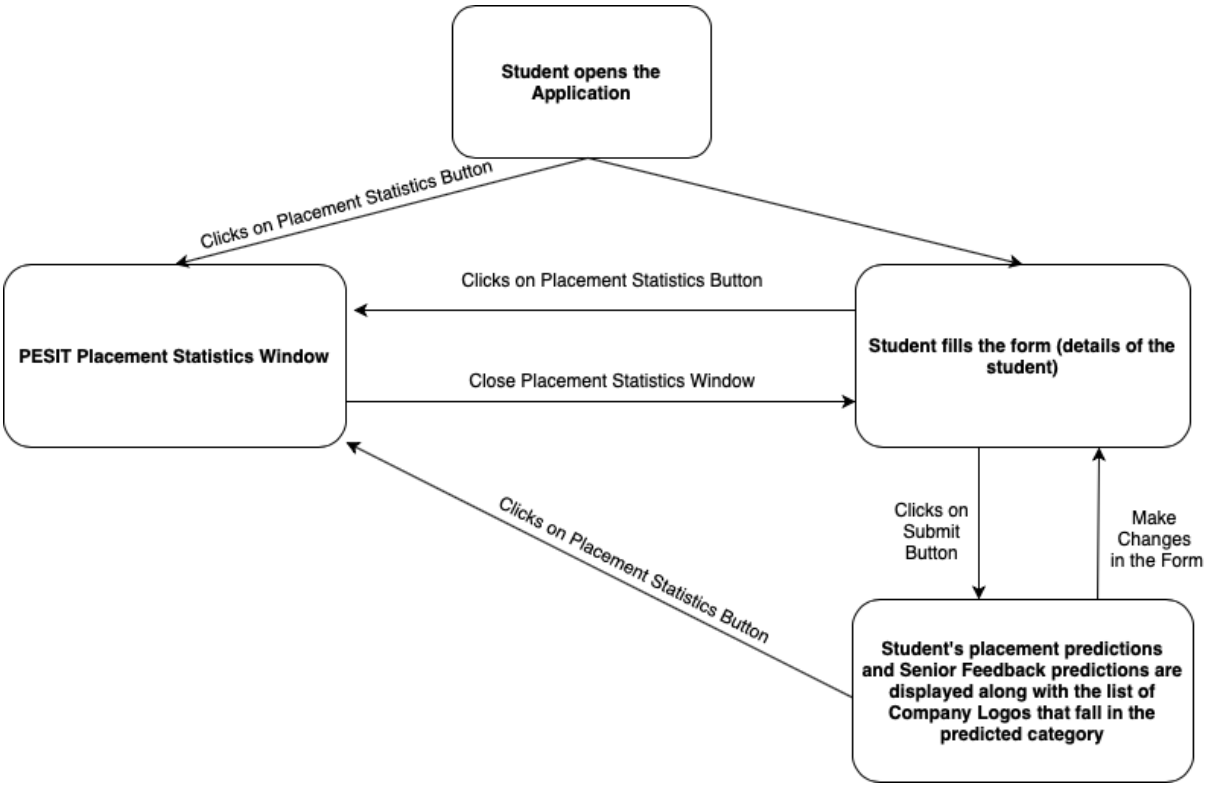

Figure 1.1: Application Block Diagram

Chapter 2

Literature Survey

2.1 Related Work

Liya Claire Joy and Asha Raj in [1] presented a study of the various models that can be used as a placement chance predictor for pre-final year students. With the emergence of data mining techniques, many predictor models were introduced by analyzing the previous year student's performance dataset.

Algorithms Considered: Logistic Regression, Fuzzy Approach, Decision Tree Algorithm, Random Forest Algorithm, classification and clustering techniques, Sum of Difference method, and Job Competency Modeling

Conclusion: Their study was conducted based on various placement prediction models and was depicted that the student dataset containing academic and placement attributes are a potential source for predicting the future placement opportunities.

Animesh Giri et al in [2] presented a method (Composer System) to augment student performance prediction system. They used KNN Classifiers, Decision Tree, Logistic Regression, Naive Bayes.

K Nearest Neighbour - A data point is classified by a majority vote of its neighbors, with the data point being assigned to the class most common among its k nearest neighbors.

Decision Tree - Supervised learning algorithm that works by splitting the sample into two or more homogeneous sets.

Logistic Regression - Predict the probability of an dichotomous outcome.

Native Bayes - Presence of a particular feature in a class is independent to the presence of any other feature.

Attributes Used: High school scores, secondary school scores, programming, general aptitude, engineering semester and competitive exams grades, engineering branch, age, year of qualification of degrees.

Conclusion: Their work of the Composer System to reduce the chances of classification based on a algorithm, reduced the bias and variation caused by any of the poor base algorithm.

S. K. Thangavel et al in [3] presented a recommendation system that predicts the students to have the one of following placement status, Dream Company, Core Company, Mass Recruiters, Not Eligible and Not Interested in. They used Decision Tree, Naive Bayes, and Naive Bayes Classification

Attributes Used: Department, Gender, 12th Board, 12th Marks, Location, No of Standing Arrears, Arrear History, CGPA

Conclusion: Results identify that Decision Tree classifier has 0.01 seconds of running time 84.42% accuracy while Naive Bayes has 0 second running time and 44.29% accuracy and Logistic Regression had a 0.19 second running time and This proves the efficiency of the methodology utilized in the system.

Animesh Giri et al in [4] proposed a system that predicts the chance of a student getting placed by applying the K Nearest Neighbor (KNN) model. They used KNN, Logistic Regression, Support Vector Machine.

K-nearest neighbours classification - Is an supervied learning algorithm. The classification according to this algorithm is based on the distance between the training and testing data.

Logistic Regression - Is a statistical method for analysing data with many independent variables that determine an outcome. The outcome of this method is a dichotomous variable.

Support Vector Machine - SVM can be used for binary classification.

Attributes used: Gender,10th percentage,12th percentage,B.E.Aggregate,Backlogs, Communication Skill,Analytical Skills,Teamwork, and Technical Skill.

Conclusion: The result that was obtained from this model KNN was compared against the logistic regression result that gave an accuracy of around 75% and SVM with an accuracy of around 77.38%.

Ashok M V and Apoorva in [5] proposed a model. This was compared with other classification algorithms such as decision tree, naive bayes, and neural network

Decision Tree - This algorithm gave an Accuracy of 0.84, TPR of 0.95, Precision of 0.74.

Naive Bayes - This algorithm gave an Accuracy of 0.87, TPR of 0.93, Precision of 0.78.

Neural Network - This algorithm gave an Accuracy of 0.83, TPR of 0.88, Precision of 0.69.

Proposed Model - This algorithm gave an Accuracy of 0.92, TPR of 0.96, Precision of 0.87.

Attributes used: Year, Reg-No, Branch, Percent, Skills, Effective Score, and Placed.

Conclusion: In terms of precision, accuracy, and true positive rate their proposed model proved to be the best predicting model for solving the placement prediction problem.

Animesh Giri et al. in [6] presented an innovative recruitment system using social networking websites such as Twitter and LinkedIn along with code repository website such as GitHub and competitive coding platforms like SPOJ.

They have used K-means Clustering Algorithm along with few methodology used in this model that is Proficiency Competency Analysis, Personality Quotient Calculation, Feedback Mechanism for Job Aspirants.

K-Means Clustering Algorithm - K-Means is an unsupervised machine learning algorithm to cluster candidates.

Attributes Used: No of programming challenges solved, difficulty level of programming language, programming language, the cluster assigned to, and professional quotient.

Conclusion: The personal and professional background of a student is accounted for. K-means is used as a clustering algorithm to cluster candidates based on their competency.

Mangasuli Sheetal B, Prof. Savita Bakare in [7] presented a method to present campus placement. They used Fuzzy Logic and K Nearest Neighbors Classification.

Fuzzy Logic - Used to build the basis as rules for inferences which include fuzzy sets. The fuzzy logic system is defined as the mapping of the non-liner input data set to the output of scalar data and is used as whole interval of real number between zero and one. The fuzzy logic system is defined as mapping of the non-liner input data to the output of scalar data and is used in the interval of real number between zero and one.

K Nearest Neighbor Classification - KNN is an algorithm used for classification and prediction in Data mining.

Attributes Used: SSLC and PUC Marks, Practical Marks, UG Aggregate, Participation in Co-curricular activities, Seminar Skills, Communication Skills, Lead-

ership Quality and Interaction.

Conclusion: Fuzzy Logic resulted in an Accuracy of 92.67% and an Execution Time of 450 ms while KNN resulted in an Accuracy of 97.33% and an Execution Time of 13458 ms.

K. Pruthi and P. Bhatia, in [8] aimed to predict the placement of students by applying decision tree algorithms, Naive Bayes using WEKA.

Multi-way Decision Tree is created which stores the placement statistics in an optimised manner using Information gain. The Decision Tree stored using Adjacency List of storing Hierarchical data in a relational system.

Attributes Used: 27 Attribute - CGPA, Marks in Operating System, Computer System Architecture, Discrete Mathematics, Principles of Programming Languages, Algorithms Design, Database Systems, Software Engineering, Data structures, Object Oriented Programming,Computer Networks, System Analysis and Design.

Conclusion: Results suggest the Decision Tree had an accuracy of 62.1% and Naive Bayes had an accuracy of 45.78%.

Ramanathan L et al. in [9] presented a method to predict the performance of student placements. They have used Sum of Difference method for placement prediction.

Sum of Differences - Using similarity measure with mathematical method which is called sum of difference. It is used to analyze the performance of students using the dataset attributes and to find a method is more accurate result for prediction.

Attributes Used: Gender, Category, Communication skill, Academic gap, Grade in BTech exam, Grade in 10th exam,Grade in 12th exam, Extra technical course, Number of arrear faced.

Conclusion: The work of this paper says that on using SOD(sum of difference),this predicts whether a student will get placed or he will not get placed.

A. S. Sharma et al in [10] presents the development of placement predictor system (PPS). This system uses the logistic regression algorithm to predict placements of students.

Logistic Regression - Placement predictor, is binary. It has yes (student placed) or no (student not placed). Hence, as the placement predcitor is binary in nature, Logistic regression is used. Gradient descent algorithm is used to learn the parameters which calculate the chance a student being placed.

Attributes Used: 10th Percentage, 12th Percentage, Subject Marks, Sex, Residency (Rural/Urban)

Conclusion: Results suggest that the training and testing accuracy of the algorithm was 98.93% and 83.3% respectively.

Ajay Kumar Pal et al in [11] presented a method to predict the performance about placement. They used the Multilayer Perception, C4.5 Tree, and Naive Bayesian Classification,

Naïve Bayesian Classification - suited when high input dimensionality. It is simple, over fitting protected, and Complex resulting classifier can be reliably determined from less amount of data.

Multi layer Perceptron - The network has a set of sensory elements that includes the input layer, one or more hidden layers of processing elements, and output layer.

C4.5 Tree - C4.5 has features such as categorization of continuous attributes, pruning of decision trees,handling missing values, and rule derivation.

Attributes Used: MCA Results, Communication Skill, Graduation Background, Sex, Seminar Performance, Lab Work.

Conclusion: Results suggest that among the algorithm tested, Naive Bayes classifier has the potential to predict placements with an accuracy of 85.6%, the highest compared to the others. Multi Layer Perceptron resulted in an Accuracy of 80.00% and C4.5 Tree resulted in an Accuracy of 75.38%.

S. Elayidom et al in [12] attempt to help students to make appropriate career decisions. Using Information Gain theory a Multi-way Decision Tree is created which helps admission seekers to choose a branch with high industrial placement. Once the information is entered, the system will predict which branch is Excellent, Good, Average or Poor for the student.

Multi-way Decision Tree is created which stores the placement statistics in an optimised manner using Information gain.

Attributes Used: Entrance Rank, Reservation(OBC/ SC/ST/GEN), Sector (Rural/Urban), and Gender(Male/Female)

Conclusion: Results suggest the Multi-way Decision Tree had an accuracy of 82.40%.

Chapter 3

Hardware and Software Requirements Specification

3.1 Hardware Requirements

3.1.1 Processor

x86 compatible processor

3.1.2 RAM

512 MB or greater

3.1.3 Hard Disk

20 GB or greater

3.1.4 Keyboard

104 keys standard

3.2 Software Requirements

3.2.1 Front End Programming Languages

Python - Tkinter

3.2.2 Backend Programming Languages

Python

Chapter 4

System Design

4.1 Data Preprocessing

The original placement dataset consisted of 23 columns as mentioned in the Table 4.1.

The first step was to remove attributes that would not have impact on the target attribute. Hence, the attributes Serial No, USN, Student Name, DOB, Father Name, Year of 10th and 12th Pass were removed from the original dataset. As not all students write the CET or COMEDK exam, we have removed these two columns from the dataset. The student's mode of admission (Admission Entrance) have values CET, COMEDK, and Management.

The 10th, 12th marks, percentage in each semester, and aggregate of all semesters were given as continuous numerical values. We discretized these values into classes mentioned in the Table 4.2.

The survey we conducted for the PESIT alumni was divided into two datasets, Survey Dataset and Senior Feedback Dataset. As all questions had to be answered in the survey, there were no missing values present in the data. Variable encoding was done to both dataset as the machine learning model required binary and multi class encoding.

4.1.1 Placement Dataset

After data pre-processing, the final placement dataset used consists of 12 independent variables and 2 dependent variables. The Firm CTC and Firm Type are the two dependent variables used in this dataset. Table 4.4 describes the placement dataset.

Attribute	Data Type
Serial No	Continuous
USN	Continuous
Student Name	String
Branch	Discrete
Company Placement	Discrete
Company Tier	Discrete
Gender	Discrete Boolean
10th % and 12th %	Continuous
Year of 10th and 12th Pass	Discrete
12th College Name	String
Individual Semester % till 6th Sem	Continuous
Aggregate % till 6th Sem	Continuous
DOB	Date/Time
Father Name	String
Admission Entrance	Discrete
CET Rank	Continuous
COMED-K Rank	Continuous

Table 4.1: Original Placement Dataset Details

Original Values	Converted Values
Equal or Above 80	First Class with Distinction (FCD)
”60-79”	First Class (FC)
”40-59”	Second Class (SC)
Below 40	Fail (F)

Table 4.2: Discretization of Marks

Firm Name	Firm Type	Firm Domain
HCL	MNC	Information Technology
Amazon	MNC	E-Commerce
Pharmeasy	Private	E-Commerce
Cognizant	MNC	Information Technology
Goldman Sachs	MNC	Investment Banking

Table 4.3: Firm Classification

Placement Dataset
Every Semester Percentage Class (FCD, FC, SC, F)
Aggregate Percentage Class (FCD, FC, SC, F)
Engineering Stream (CSE, MECH, ISE, ECE)
Gender (M or F)
10th Standard Marks
12th Standard Marks
Admission (CET/ COMEDK/ Management)
Firm CTC (T1, T2, T3) **”Target”**
Firm Type (MNC,Private) **”Target”**

Table 4.4: Placement Dataset

Survey Dataset
Firm CTC (T1, T2, T3)
Firm Type (MNC,Startup,Govt. Jobs)
Employee Satisfaction (1-5, 5: Most Satisfaction) **"Target"**
Career Growth (1-5, 5: High Potential) **"Target"**

Table 4.5: Survey Dataset

Senior Feedback Dataset
Employee Satisfaction (1-5, 5: Most Satisfaction)
Career Growth (1-5, 5: High Potential)
Feedback from Seniors (Higher Studies or Placement) **"Target"**

Table 4.6: Senior Feedback Dataset

4.1.2 Survey Dataset

After data pre-processing, the survey dataset used consists of 2 independent variables and 2 dependent variables. The Employee Satisfaction and Career Growth are the two dependent variables used in this dataset. Table 4.5 describes the placement dataset.

4.1.3 Senior Feedback Dataset

After data pre-processing, the senior feedback dataset used consists of 2 independent variables and 1 dependent variables. The Feedback (Higher Studies or Directly Work) is the dependent variables used in this dataset. Table 4.6 describes the placement dataset.

4.2 Methodology

4.2.1 K-Nearest Neighbors

paragraph The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning algorithm that can be used to solve classification problem. This algorithm is considered as an example of lazy learning. This algorithm classifies data points/records based on the points that are most similar to it. The model is trained with labeled data. The value of K is determined by the user. The predicted data point is classified based on the classification of its k nearest neighbours. The input consists of the k closest training examples.

4.2.2 Random Forest

Random forest is an ensemble machine learning prediction method by taking the aggregate the outcomes of individual decision trees. Random forest can predict the chances of a student to be placed in campus placements. From the model the system can display the name of companies a student have chances to be placed based on their academic performance. Random forests are presented as an analytical foundation for educational data mining tasks. The Random Forest (RF) algorithm has proven useful in many fields due to its efficiency and accuracy in making predictions with large data sets. The accuracy of Random Forest can be improved by tuning parameters such as number of attributes and trees.

4.2.3 Artificial Neural Networks

Artificial neural networks (ANN) use different layers for processing. An ANN architecture ideally consists of 3 layers, an input layer, hidden layer, and output layer. The input layer receives various types of information consisting of the data the neural network aims to process. From the input unit, the data goes through one or more hidden layers. The hidden layers goal is to transform the input into a format process-able by the output unit. Artificial Neural Network is an iterative approach,it solves problem through iteration. We have used ANN to predict the company tier and company type of the student. Since our dataset has twelve input attributes we have considered twelve nodes in the input layer. Ideally the number of nodes in the hidden layer should be half of the number of input layer nodes hence we have considered six nodes in the hidden layer. The output layer consisted of two or three nodes when predicting the company type and company tier respectively.

4.2.4 Decision Trees

Decision Trees is a supervised learning algorithm. Decision tree algorithm solves a problem by using tree representation, consisting of edges and nodes. It breaks data into smaller subsets while incrementally developing a decision tree. The final result is a tree with leaf nodes which consist of the predicted class. Nodes are identified as leaf nodes as such nodes that have 0 outgoing edges and they represents a decision (class label). A root node is the topmost decision node in a tree corresponding to the best predictor. Each internal node of the tree corresponds to an attribute , and each leaf node corresponds to a class label. Decision tree model has been used to predicting the Career Growth, Employee Satisfaction and Preferred Choice after

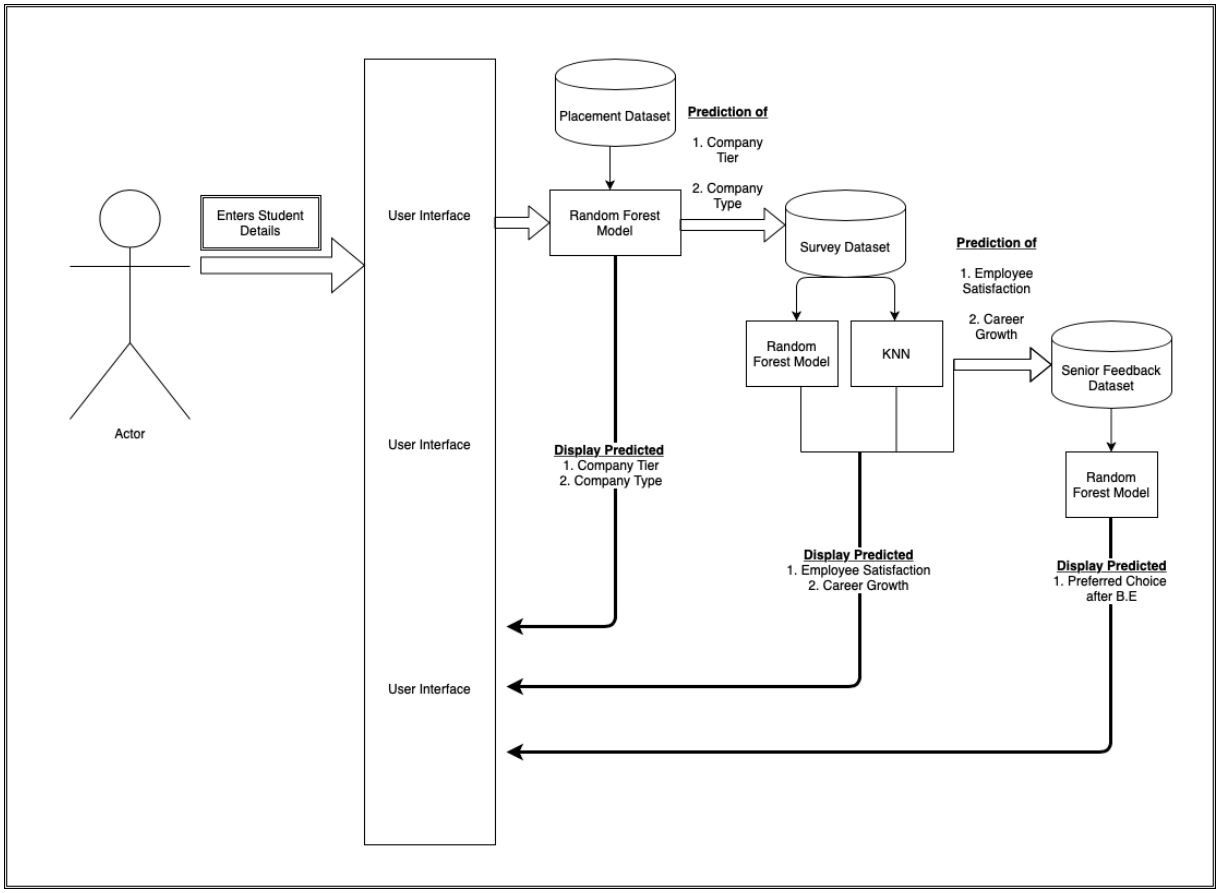

Figure 4.1: Methodology

B.E. based on the Survey and Seniors Feedback.

System Methodology

The student/end-user will enter their details in the User Interface and press submit. Their details will be the input variables to the Random Forest Model trained by the placement dataset. This Random Forest Model will predict the Company Tier and Company Type based on the user's details. The predicted Company Tier and Type will be sent to another Random Forest Model and KNN model (trained by Survey Dataset). KNN Model predicts the Employee Satisfaction and the Random Forest Model predicts the Career Growth. The predicted employee satisfaction and career growth will be sent to another random forest model (trained by the senior feedback dataset) to predict the Feedback/Preferred Choice after BE. All predicted values will be displayed back on the User Interface. The Methodology is illustrated in Figure 4.1.

Chapter 5

Implementation

The User Interface was written in Python using the TKinter library. The opening window is displayed in Figure 5.1. Once the user fills the form and clicks on the button, the results are displayed.

Figure 5.2 describes a student who is predicted to get placed in a Tier 3 Private Company. The student is predicted to have a Mid Career Growth, Satisfied Employee Satisfaction, and alumni who were placed in these companies preferred to go for higher studies directly after completing their B.E. The bottom right side of the screen displays Tier 3 private companies that come to PES for placements.

Figure 5.3 describes a student who is predicted to get placed in a Tier 2 MNC. The student is predicted to have a High Career Growth, Satisfied Employee Satisfaction, and alumni who were placed in these companies preferred to work directly after completing their B.E. The bottom right side of the screen displays Tier 2 MNCs that come to PES for placements.

Figure 5.4 describes a student who is predicted to get placed in a Tier 1 MNC. The student is predicted to have a High Career Growth, Satisfied Employee Satisfaction, and alumni who were placed in these companies preferred to work directly after completing their B.E. The bottom right side of the screen displays Tier 1 MNCs that come to PES for placements.

Figure 5.5 describes a student who is predicted to get placed in a Tier 3 MNC. The student is predicted to have a High Career Growth, Satisfied Employee Satisfaction, and alumni who were placed in these companies preferred to work directly after completing their B.E. The bottom right side of the screen displays Tier 3 MNCs that come to PESIT for placements.

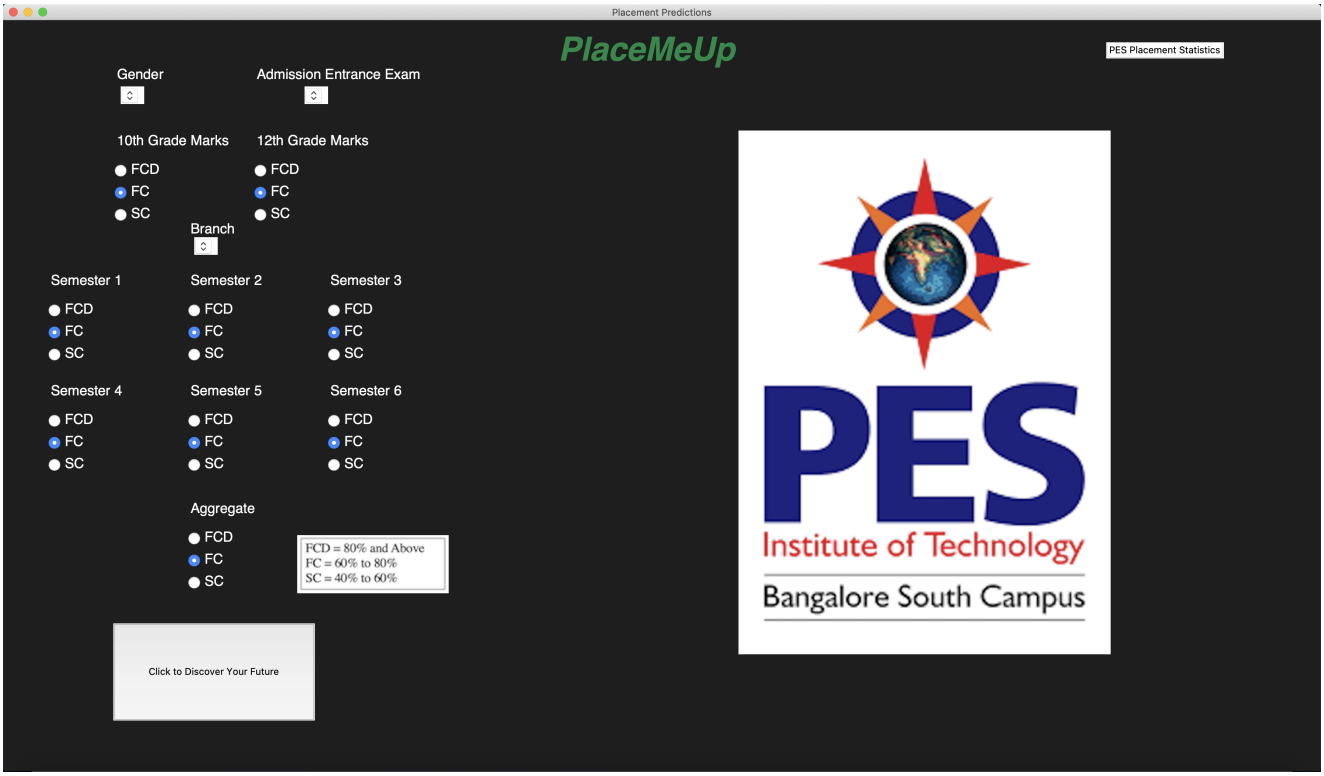

Figure 5.1: Front Page

Figure 5.6 describes a student who is predicted to get placed in a Tier 2 Private Company. The student is predicted to have a Mid Career Growth, Satisfied Employee Satisfaction, and alumni who were placed in these companies preferred to pursue higher studoes after completing their B.E. The bottom right side of the screen displays Tier 2 Private Companies that come to PESIT for placements.

Figure 5.7 displays the new window that appears if the user clicks on the 'PES Placement Statistics' button on the front page. This window contains exploratory data analysis of PESIT Placement Data over the past 6 years. This window also consists of button with the various engineering branches offered in PESIT. Each button, when clicked, takes the user to the PESIT department's homepage, shown in Figures 5.8 - 5.11.

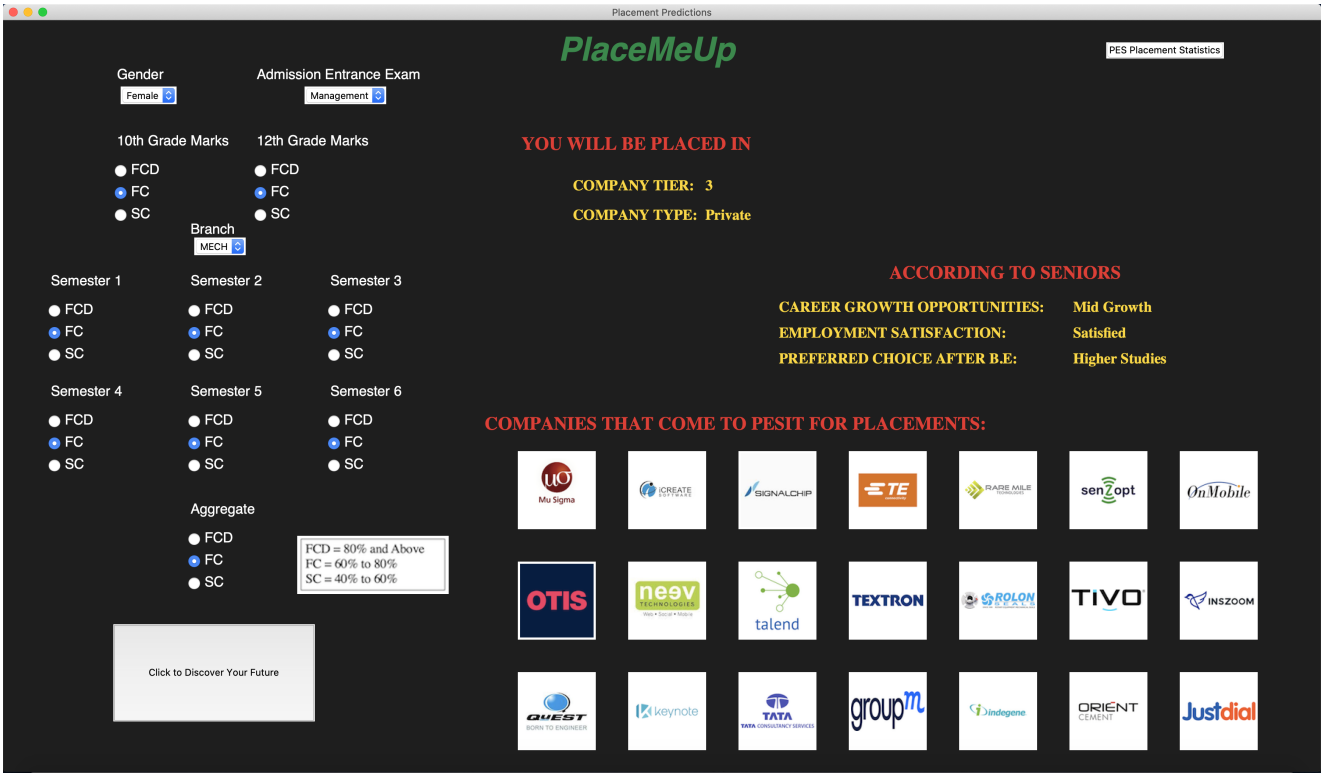

Figure 5.2: Result 1

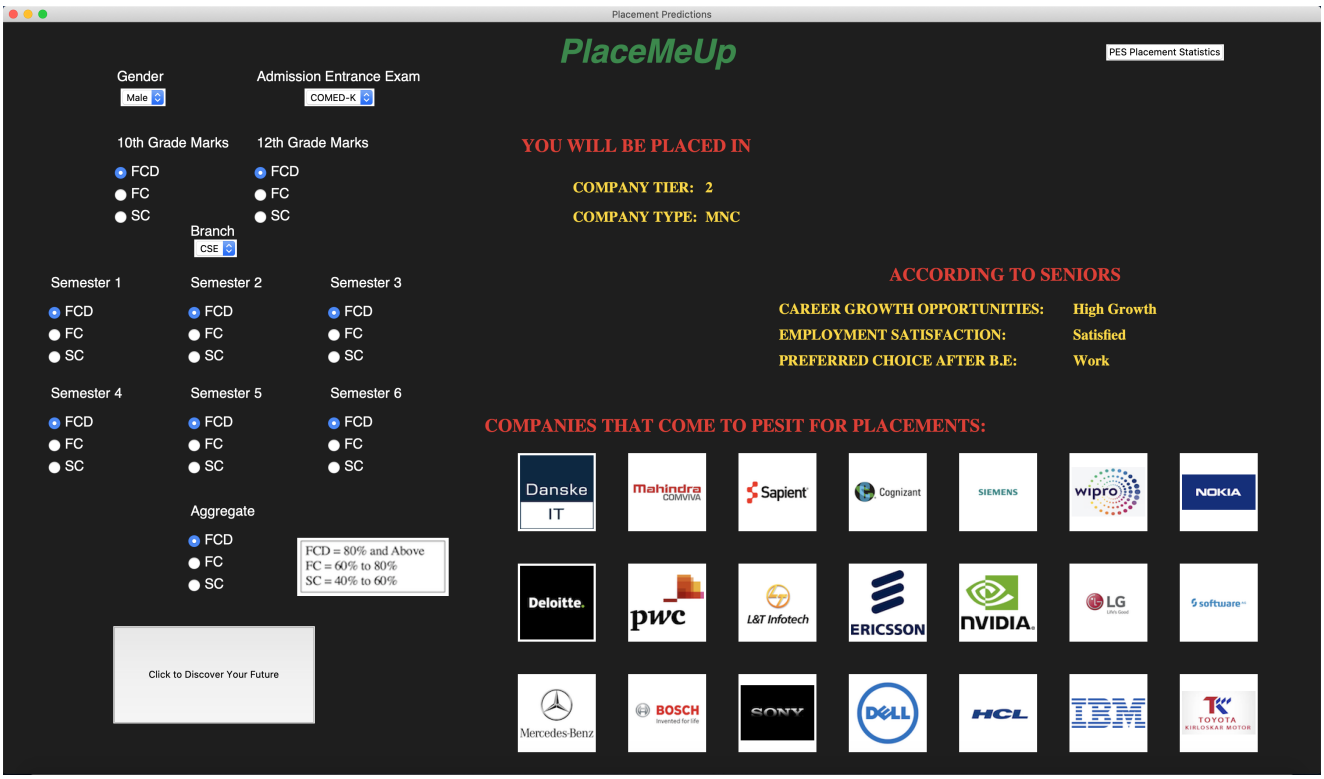

Figure 5.3: Result 2

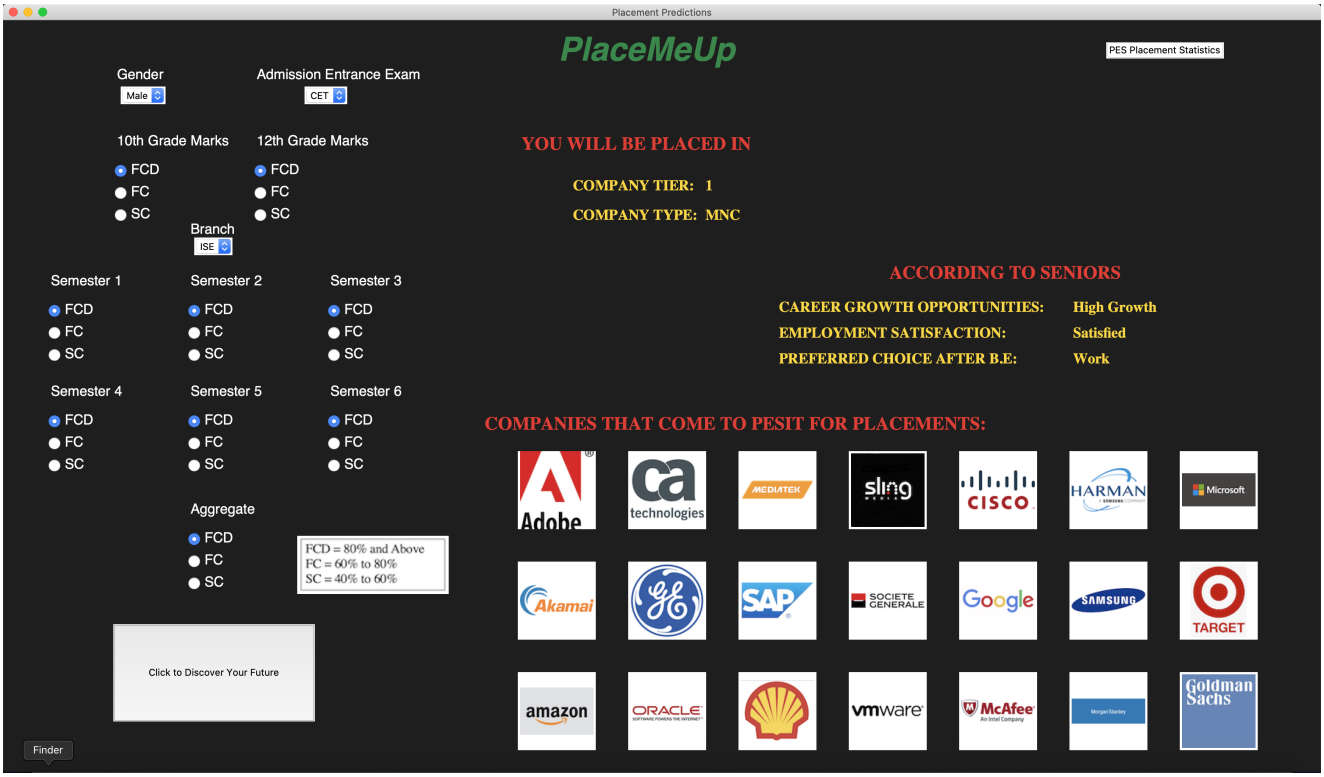

Figure 5.4: Result 3

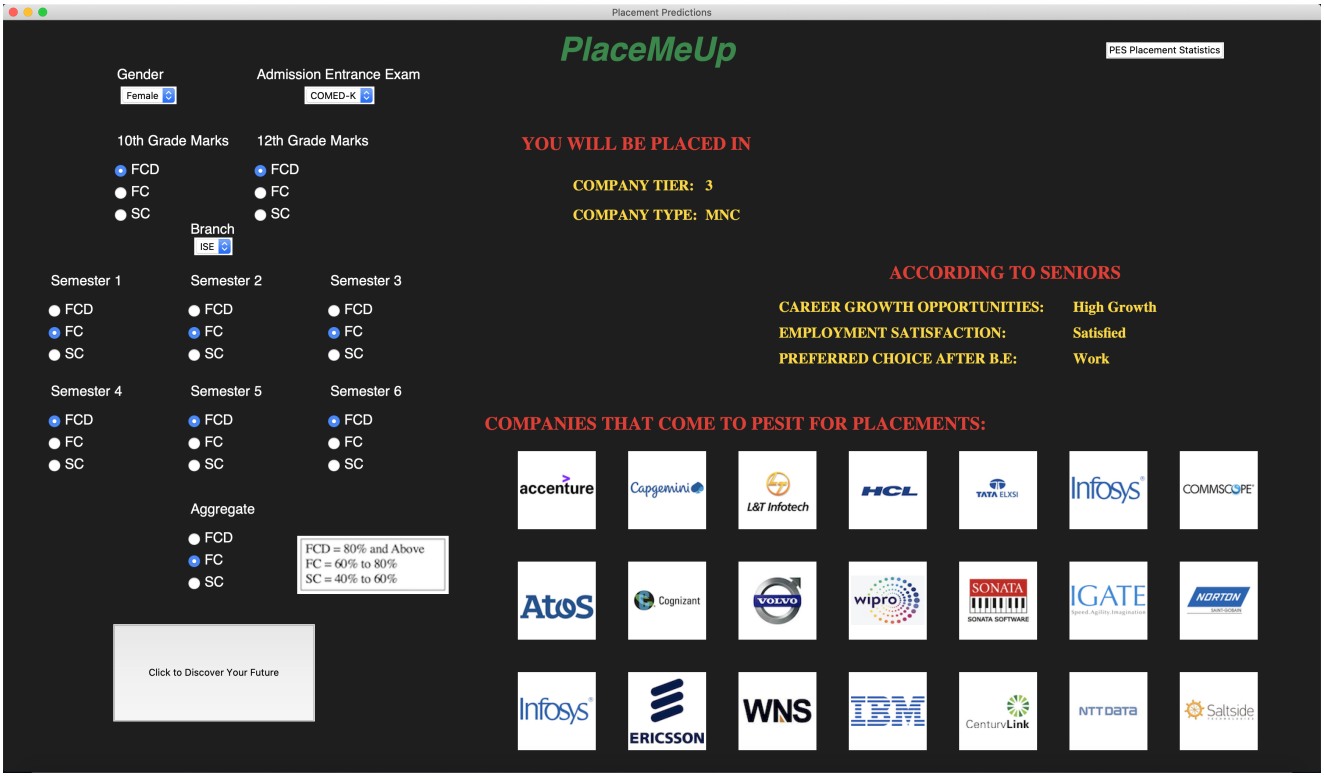

Figure 5.5: Result 4

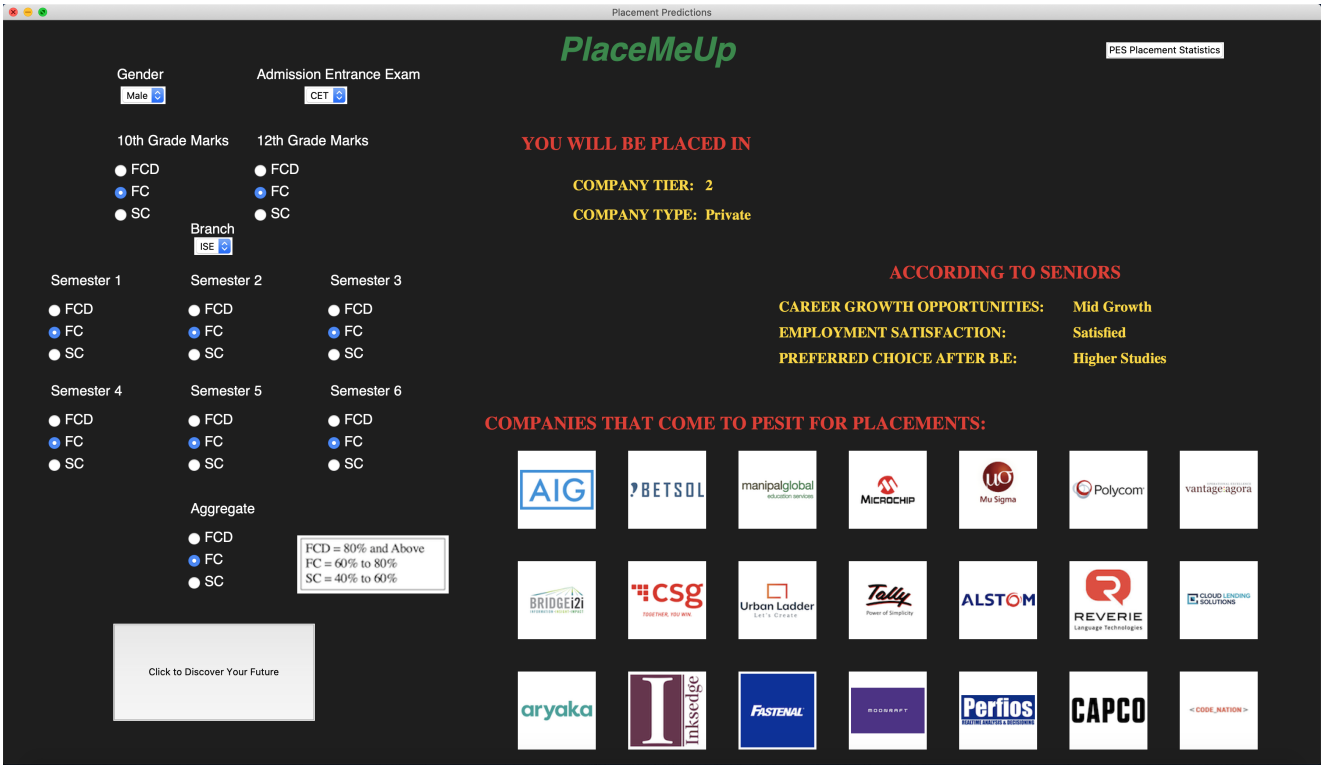

Figure 5.6: Result 5

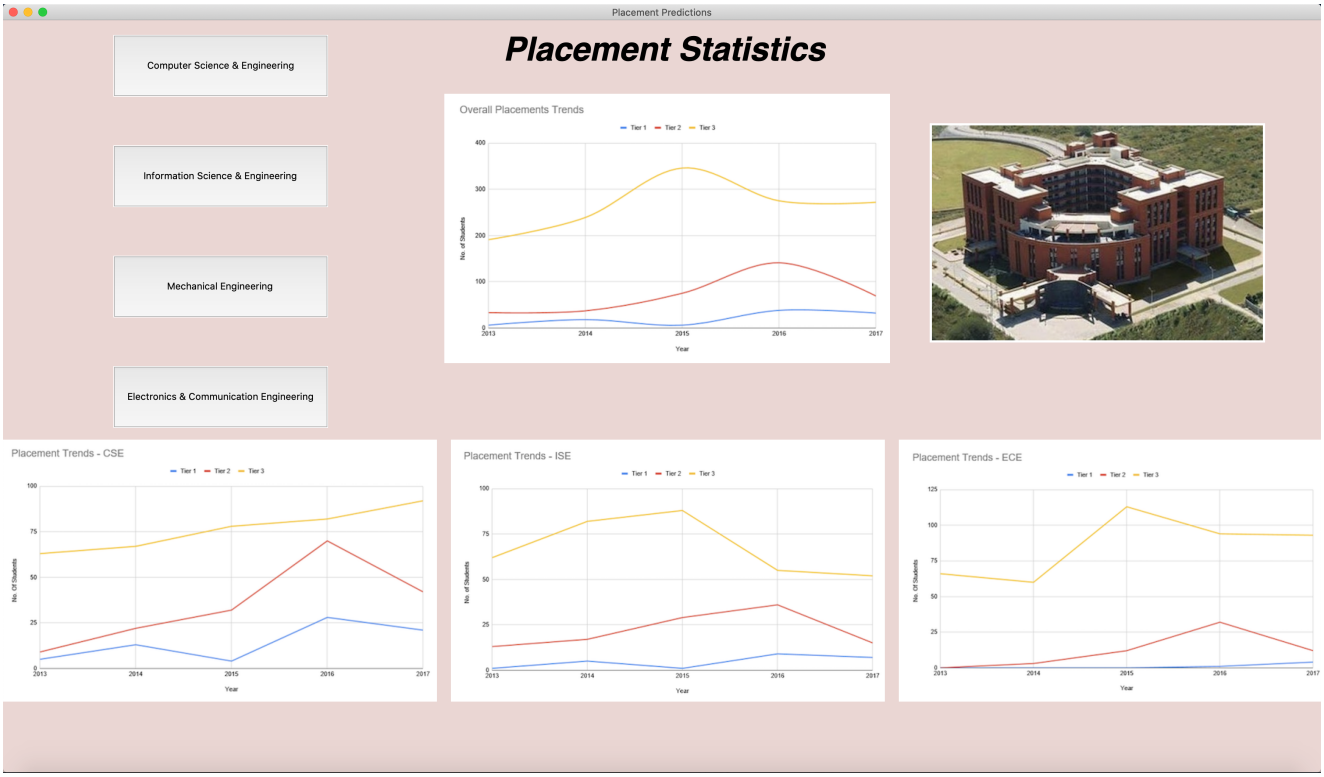

Figure 5.7: Placement Statistics

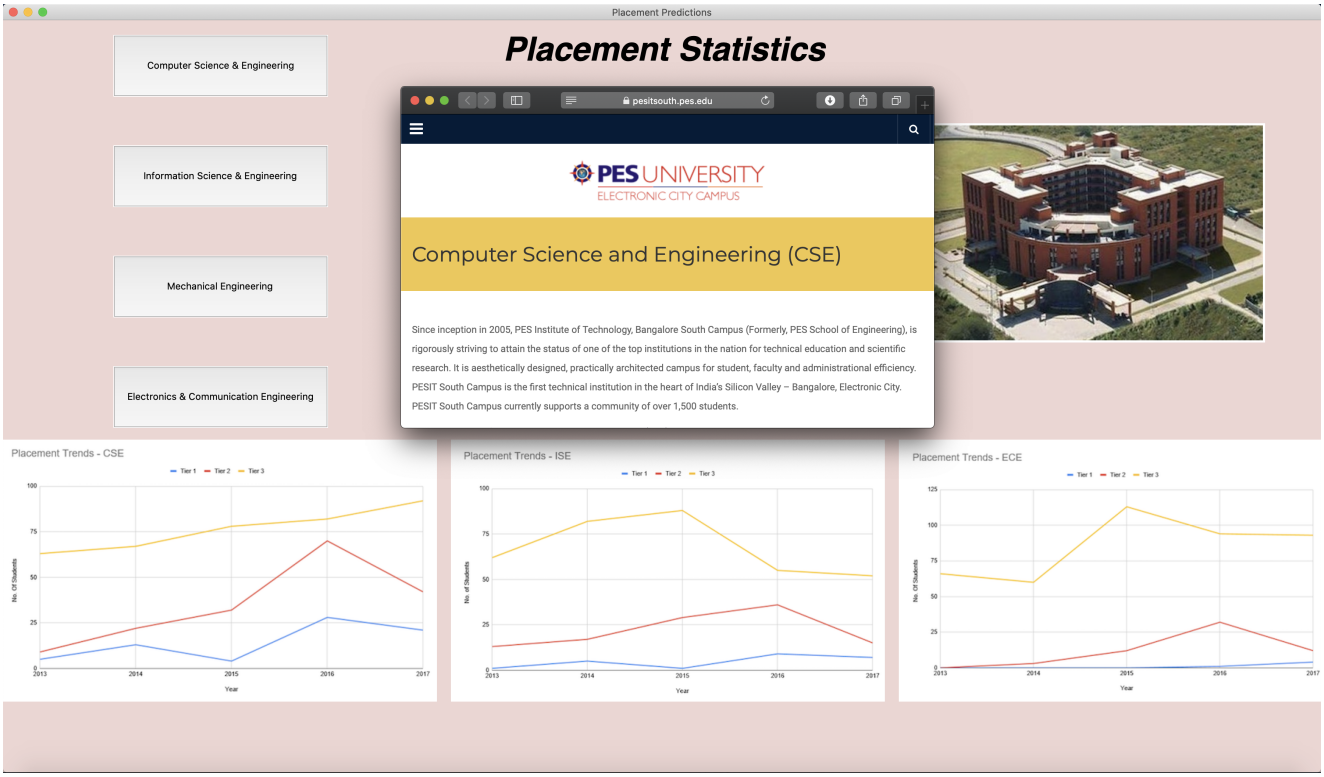

Figure 5.8: Button Click: CSE

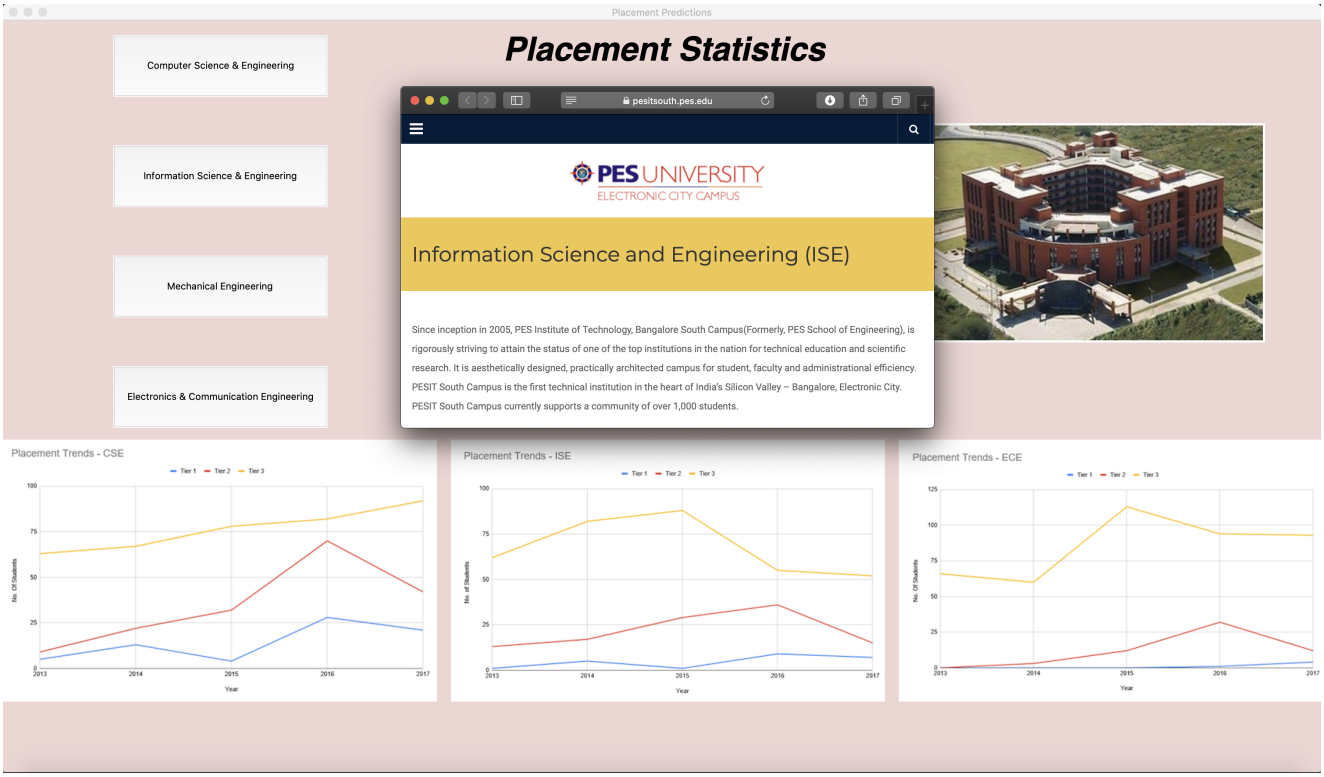

Figure 5.9: Button Click: ISE

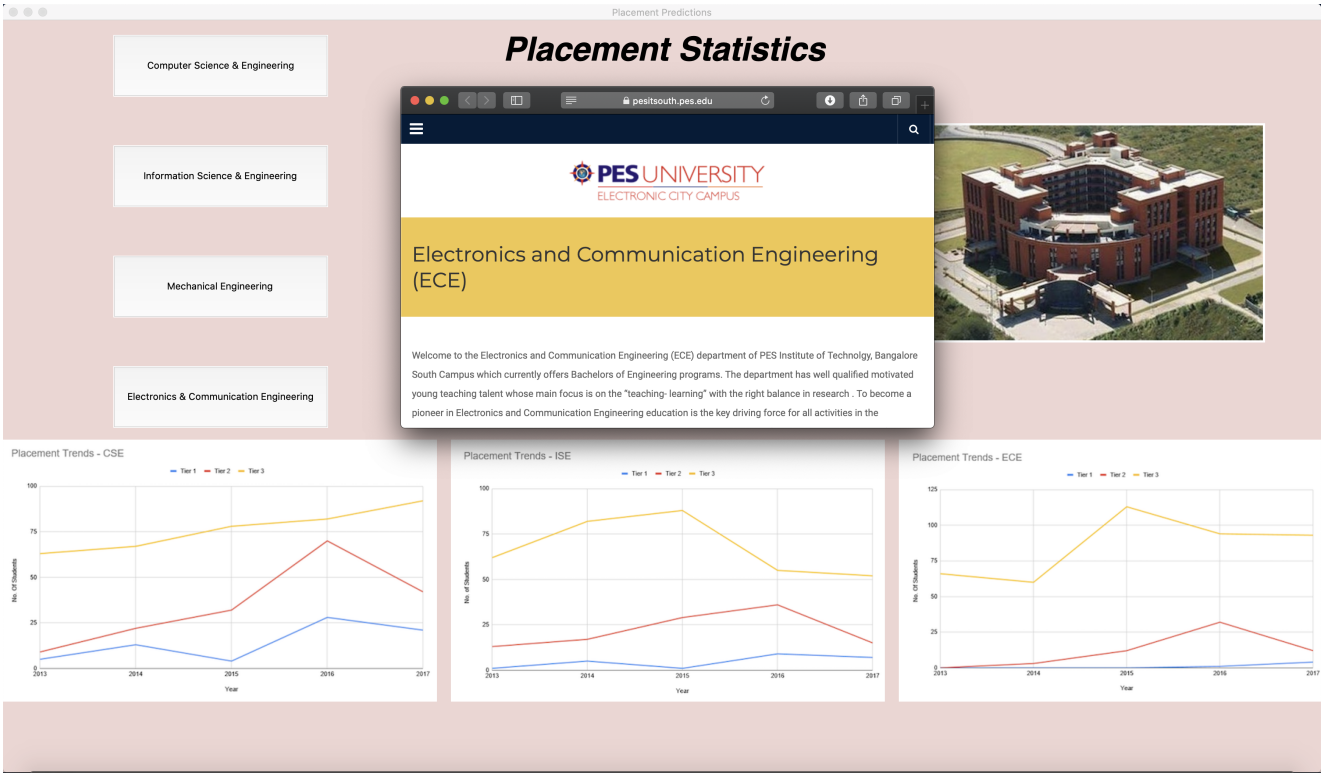

Figure 5.10: Button Click: ECE

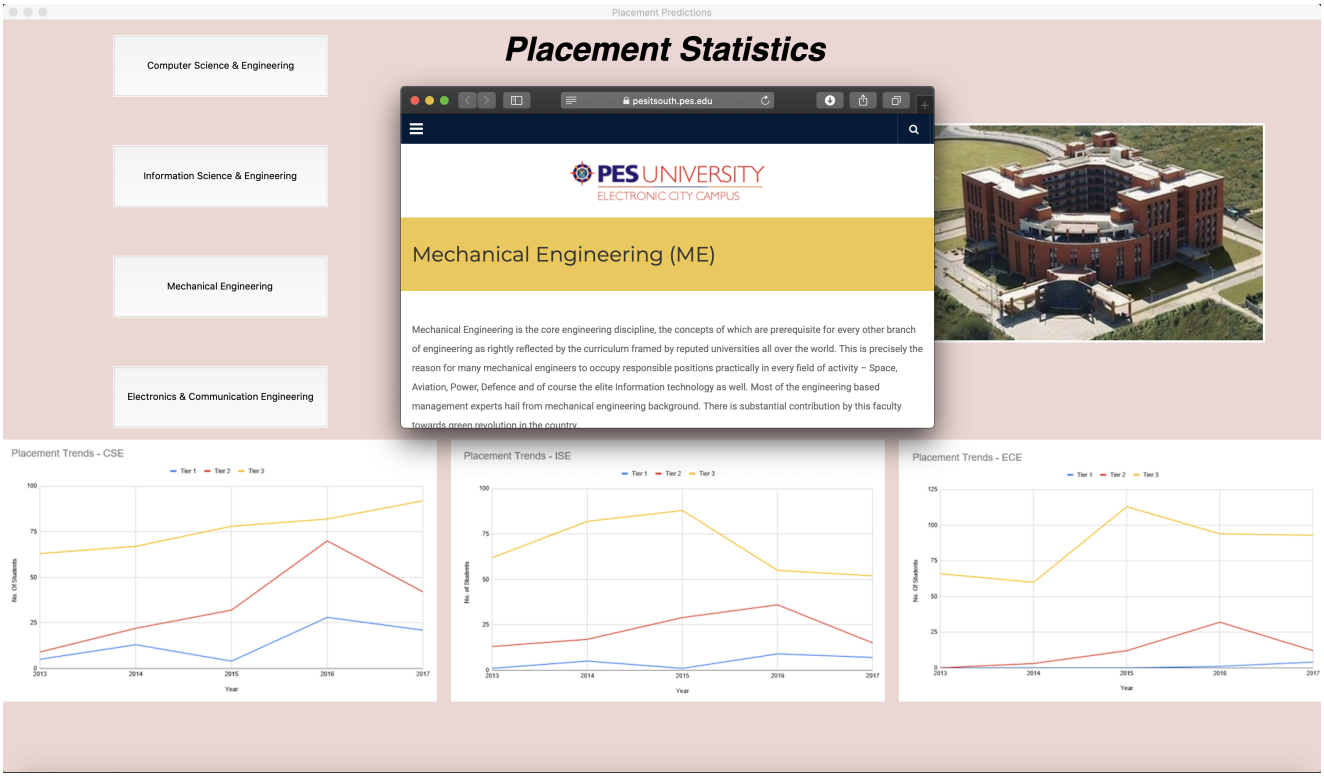

Figure 5.11: Button Click: MECH

Chapter 6

Result Analysis

The aim of this project is to understand the trends in the PESIT placement and predict the placement for final year undergraduate students. The records in the datasets are divided into training and test sets.After data pre-processing, machine learning algorithms namely Random Forest,KNN,ANN and Decision Tree were applied. This section shows the classification report and accuracy of these models, coded in Python.

6.1 Prediction of Company Tier

Prediction of Company Tier:

Algorithm	Accuracy	Classification Report
ANN	57.18%	precision recall f1-score 0 0.36 0.36 0.36 1 0.43 0.38 0.40 2 0.69 0.74 0.71 micro avg 0.57 0.57 0.57 macro avg 0.49 0.49 0.49 weighted avg 0.56 0.57 0.57 samples avg 0.57 0.57 0.57
Random Forest	61.43%	precision recall f1-score 1 0.53 0.42 0.47 2 0.41 0.32 0.36 3 0.70 0.83 0.76 accuracy 0.61 macro avg 0.55 0.52 0.53 weighted avg 0.59 0.61 0.60

Figure 6.1: Company Tier Prediction

Firm CTC is broken down into 3 Tiers - Tier 1 (Above Rs. 8 Lakhs per annum), Tier 2(Rs. 4 Lakhs - 8 Lakhs per annum) and Tier 3 (Below Rs. 4 Lakhs per annum). Comparison of the above two algorithm in Figure 6.1 shows that Random Forest has

higher accuracy, with an accuracy level of 61.43% hence Random Forest is used for Prediction of Company Tier.

6.2 Prediction of Company Type

Prediction of Company Type:

Algorithm	Accuracy	Classification Report
ANN	79.94%	precision recall f1-score 1 0.84 0.94 0.89 2 0.20 0.07 0.11 accuracy 0.80 macro avg 0.52 0.51 0.50 weighted avg 0.73 0.80 0.76
Random Forest	81.60%	precision recall f1-score 0 0.85 0.95 0.90 1 0.29 0.11 0.16 accuracy 0.82 macro avg 0.57 0.53 0.53 weighted avg 0.76 0.82 0.78

Figure 6.2: Company Type Prediction

Comparison of the above two algorithm in Figure 6.2 shows that Random Forest has higher accuracy, with an accuracy level of 81.60% hence Random Forest is used for Prediction of Company Type.

6.3 Prediction of Career Growth

Comparison of the above three algorithm in Figure 6.3 shows that Random Forest has the highest accuracy, with an accuracy level of 64.70% hence Random Forest is used for Prediction of Career Growth.

6.4 Prediction of Employee Satisfaction

Comparison of the above three algorithm in Figure 6.4 shows that KNN has the highest accuracy, with an accuracy level of 64.90% hence KNN is used for Prediction of Employee Satisfaction.

Algorithm	Accuracy	Classification Report
KNN	62.83%	precision recall f1-score High Growth 0.63 0.95 0.76 Low Growth 0.00 0.00 0.00 Mid Growth 0.62 0.27 0.37 accuracy 0.63 macro avg 0.42 0.41 0.38 weighted avg 0.53 0.63 0.54
Random Forest	64.70%	precision recall f1-score 0 0.00 0.00 0.00 1 0.52 0.64 0.57 2 0.71 0.85 0.77 accuracy 0.65 macro avg 0.41 0.49 0.45 weighted avg 0.54 0.65 0.59
Decision Tree	61.94%	precision recall f1-score 0 0.77 0.76 0.76 1 0.00 0.00 0.00 2 0.42 0.67 0.51 accuracy 0.62 macro avg 0.40 0.47 0.43 weighted avg 0.56 0.62 0.58

Figure 6.3: Career Growth Prediction

Algorithm	Accuracy	Classification Report
KNN	64.90%	precision recall f1-score Not Satisfied 0.00 0.00 0.00 Satisfied 0.52 0.64 0.57 Very Satisfied 0.71 0.85 0.77 accuracy 0.65 macro avg 0.41 0.49 0.45 weighted avg 0.54 0.65 0.59
Random Forest	61.94%	precision recall f1-score 0 0.77 0.76 0.76 1 0.00 0.00 0.00 2 0.42 0.67 0.51 accuracy 0.62 macro avg 0.40 0.47 0.43 weighted avg 0.56 0.62 0.58
Decision Tree	64.70%	precision recall f1-score 0 0.00 0.00 0.00 1 0.52 0.64 0.57 2 0.71 0.85 0.77 accuracy 0.65 macro avg 0.41 0.49 0.45 weighted avg 0.54 0.65 0.59

Figure 6.4: Employee Satisfication Prediction

6.5 Prediction of Preferred Choice after B.E

Algorithm	Accuracy	Classification Report
KNN	85.29%	precision recall f1-score Higher Studies 0.79 0.75 0.77 Work 0.88 0.90 0.89 accuracy 0.85 macro avg 0.83 0.82 0.83 weighted avg 0.85 0.85 0.85
Random Forest	86.47%	precision recall f1-score 0 0.82 0.75 0.78 1 0.88 0.92 0.90 accuracy 0.86 macro avg 0.85 0.83 0.84 weighted avg 0.86 0.86 0.86
Decision Tree	86.12%	precision recall f1-score 0 0.82 0.75 0.78 1 0.88 0.92 0.90 accuracy 0.86 macro avg 0.85 0.83 0.84 weighted avg 0.86 0.86 0.86

Figure 6.5: Preferred Choice after B.E Prediction

Comparison of the below three algorithm in Figure 6.5 shows that Random Forest has the highest accuracy, with an accuracy level of 86.47% hence Random Forest is used for Prediction of Preferred Choice after B.E.

6.6 Final Algorithms

The above algorithms were applied to the placement dataset in order to analyze the best algorithm in terms of accuracy.Below Figure 6.6 shows the conclusion of preferred algorithm used for prediction of placement trends in PESIT based on their accuracy level.

Prediction Of	Algorithms
Company Tier	Random Forest
Company Type	Random Forest
Employee Satisfaction	KNN
Career Growth	Random Forest
Choice after B.E	Random Forest

Figure 6.6: Final Algorithms

6.7 Exploratory Data Analysis

To understand and identify the trends in the PESIT placement dataset,we plotted graphs illustrating the number of students that attend and got placed via college placements over the years. This trend we looked at by classifying firms in terms of CTC along with the Engineering branch.Such graphs would be displayed on our application to give insights to the user about PESIT placement trends. The figures below currently consists of data from 2013-2017.

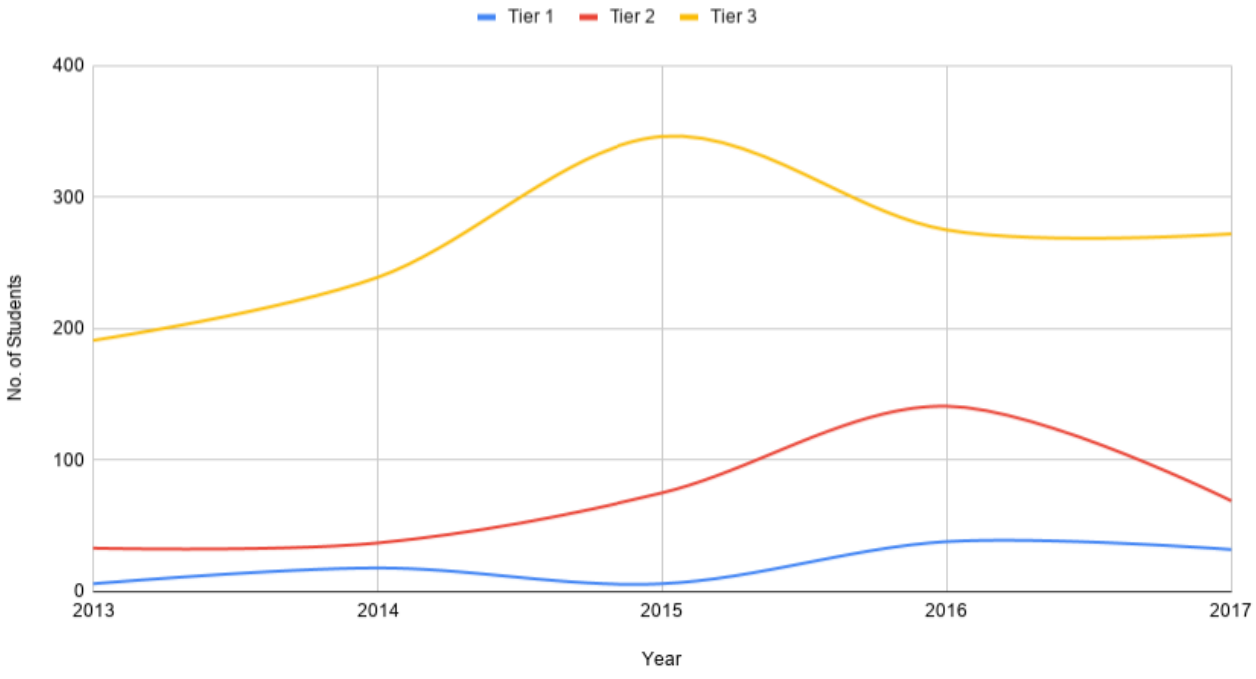

Figure 6.7: Overall Placements

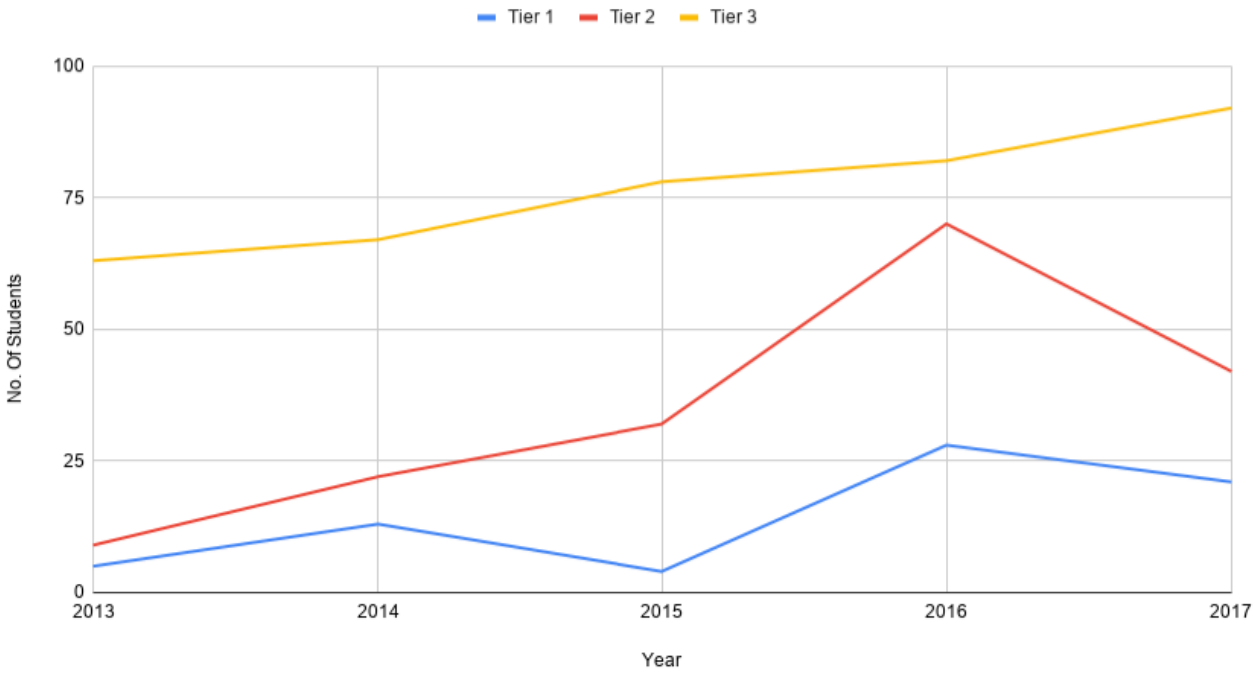

Figure 6.8: CSE Placement Trends

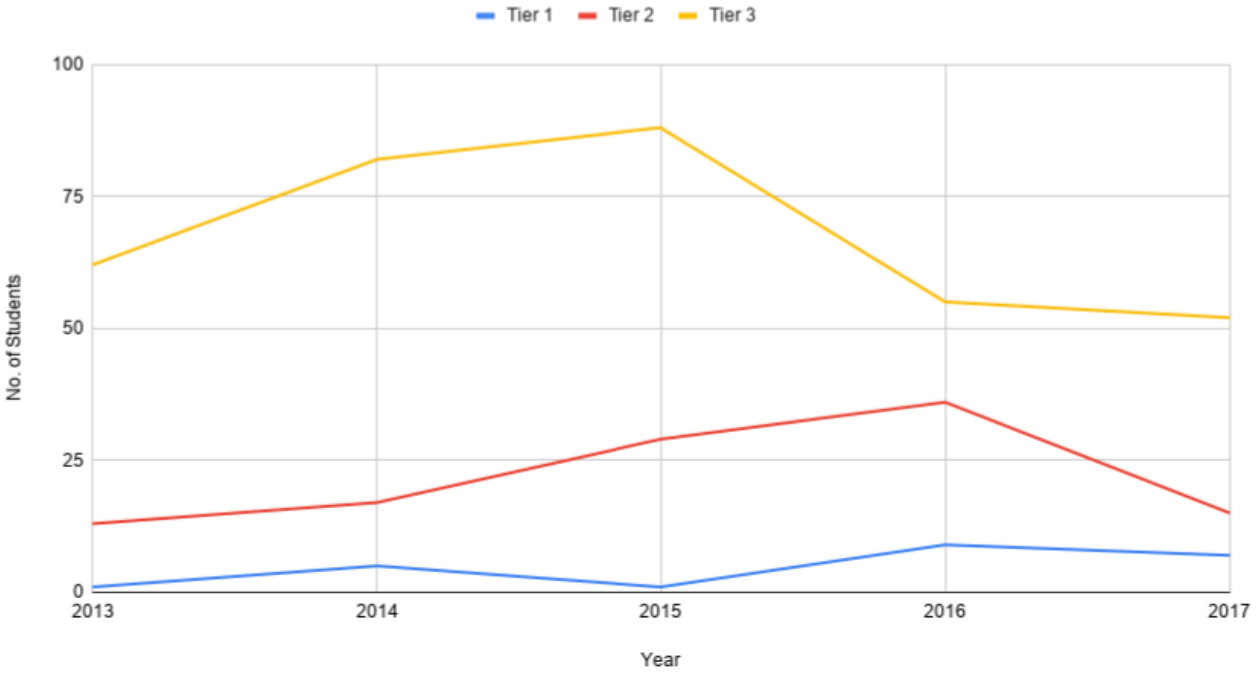

Figure 6.9: ISE Placement Trends

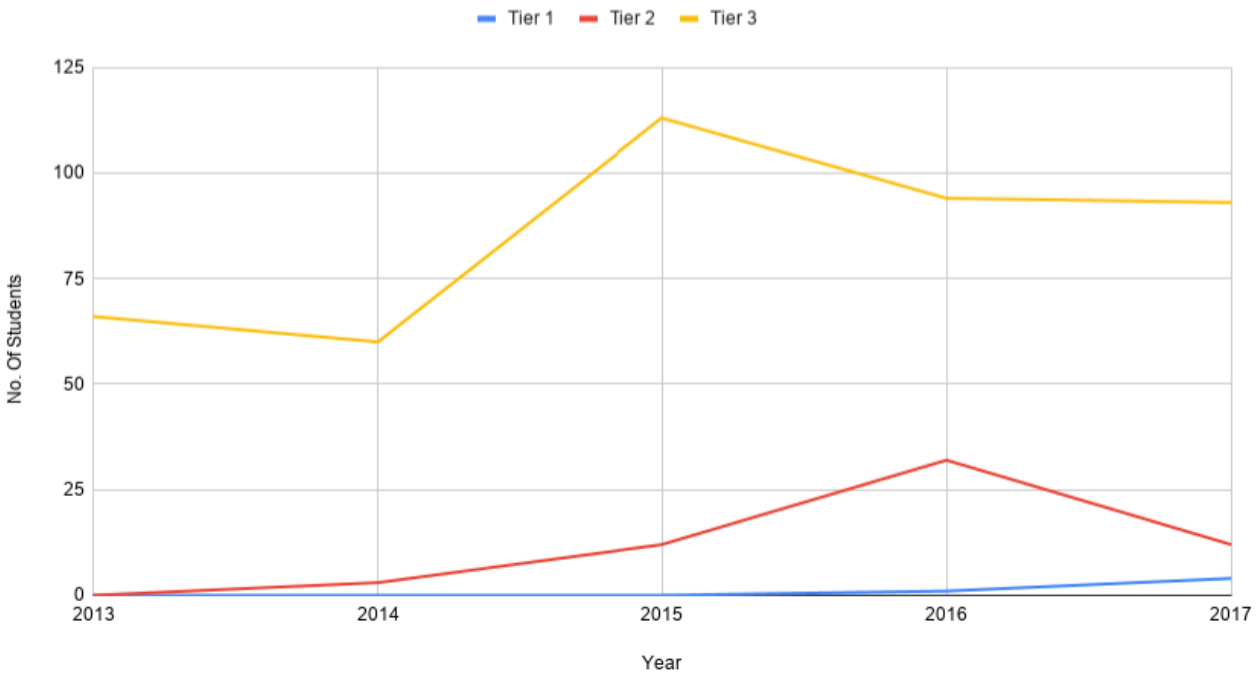

Figure 6.10: ECE Placement Trends

Chapter 7

Conclusion and Future Scope

7.1 Conclusion

The application aims to predict student placements based on their profile and provides alumni advice for the predicted placement. There has been an analytical comparison with various machine learning models to understand and implement the model with the highest accuracy. The comparison included Artificial Neural Networks, Decision Tree, KNN, and Random Forest. Prediction of Company Tier, Company Type, Career Growth, and Choice after B.E is done by the Random Forest Algorithm. Prediction of Employee Satisfaction is done by KNN algorithm. The front end is implemented using Tkinter, a python based user interface. The application uses widgets such as radio buttons, drop down menus, and web-linked buttons. The user has to enter their details such as their 10th and 12th marks, engineering branch, semester marks, gender, and their admission entrance exam. Followed by which, the application will use its machine learning predictive model to predict the Company Type, Company Tier, Employee Satisfaction, Career Growth, and Choice after B.E. A list of company logos that come to PESIT for placements and fulfill the predicted company type and tier are displayed below. An additional feature of the application involves an exploratory data analysis with a goal to understand the placement trends (placement statistics) in PESIT.

7.2 Future Scope

The future scope of the project is to make the application consistent with all the major colleges and universities in Bangalore. By this we will be training the machine learning model using placement data from different colleges around Bangalore to build a generalized application for all university students.

References

[1] *A Review On Student Placement Chance Prediction-2019*; Liya Claire Joy and Asha Raj

[2] *A Composer System Based On Meta-Learning for Student Performance Prediction-2018* ; Animesh Giri

[3] *Student placement analyzer: A recommendation system using machine learning-2017*; Sentkil Kumar Thangavel, P. Divya Bkaratki, Abijitk Sankar

[4] *A Placement Prediction System Using K-Nearest Neighbours Classifier-2016*; Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi, Naini Dubey

[5] *A Data Mining Approach For Predicting Student and Institution Placement Percentage -2016*; Ashok M V and Apporva

[6] *A Theoretical Framework For IT Recruitment based on Machine Learning Techniques Applied over Twitter Conclusion,LinkedIn,SPOJ and GitHub Profiles-2016*; Animesh Giri

[7] *Prediction of Campus Placement Using Data Mining Algorithm-Fuzzy logic and K nearest neighbor-2016*; Mangasuli Sheetal B, Prof. Savita Bakare

[8] *Applying Data Mining Techniques for Placement Chance Prediction-2015*; Karan Pruthi, Partheek Bhatia

[9] *Mining Educational Data for Students' Placement Prediction using Sum of Difference Method-2014*; Ramanathan L, Swanalatha P, D. Ganesh Gopal

[10] *PPS — Placement prediction system using logistic regression-2014*; Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar

[11] *Classification Model of Prediction For Placement Of Students-2013*; Ajay Kumar Pal and Saurabh Pal

[12] *Applying Data Mining Techniques for Placement Chance Prediction-2009*; Sudheep Elayidom, Sumam Mary Idikkula, Joseph Alexander, Anurag Ojha