**DATA ANALYSIS AND PREDICTION FOR CAMPUS PLACEMENTS**

Aditi Indoori 1601187370011 , Sathvika Patwari 1601187370162

1,2 B.E., Department of Information Technology, CBIT, Hyderabad

Under the guidance of Smt.T.Prathima, Assistant Professor, Department of IT, CBIT, Hyderabad

**ABSTRACT**

**Background**

One of the main goals of the educational institution is the placement of students. An institution's reputation and annual enrollment rely inevitably on their students' placements. That is why all the institutions work hard to develop their organization in its entirety by strengthening the placements department. Any help in this specific field would have a positive effect on the campus placements. The students as well as the institution will benefit from this. The institution's fundamental progress is assessed by the students' campus placement. Analysing the proportion of campus placements is one of the primal factors that students consider when selecting and applying to colleges. In this context, the strategy is therefore to analyze and forecast the importance of campus placements, which not only help strengthen an institution’s reputation but also optimize placements for students.

**Design**

The objective of this project is to analyse previous year's student's data and study the trends, patterns from the analysis to know the features that are influencing a student’s placement. Data pertaining to the study was collected from the institution for which the Data Visualisation is done, and also suitable data pre-processing methods like cleaning of data set (removing noise and inconsistent data), Integration of data (integrating multiple sources), Selection of data, Transformation of data (transforming into the appropriate form), Mining the data (extracting relevant data) are applied. This project also aims at developing an interactive dashboard with charts surfaced in different formats to offer concise ways to express the change in the placements in the past 11 years. Interesting plots allow a visual inspection of the situation of the institution regarding the placements. The goal is to track and display information on what percent of students got placed for different branches, on what percent of students got placed for different companies and many more. This research provides a comprehensive understanding of campus placements and compares its impact on different areas of the syllabus with the help of Data Visualization and it will also help derive a better solution in the future. The other main objective is to develop a predictor which can help in predicting the placement chance of the students. The Machine Learning model is proposed with algorithms to predict the same. But the problem was to find a suitable classification algorithm that could do the job with maximum accuracy for our data set. Different algorithms have different accuracies depending on the type of problem it has to solve and the data set it has to work with. So, we decided to select five algorithms, namely KNN (K-Nearest Neighbor), SVM (Support Vector Machine), Decision Tree, Random Forest and Naive Bayes and to compare the accuracy levels of each of these algorithms, with respect to our problem and data set. The result of this test would help us in determining which algorithm to use while implementing our predictor in the placement management system. For this, we trained each of the algorithms with the data set that we acquired and tested it against some test data to find the accuracy of the algorithms. For each algorithm, we can easily obtain the True Positive, True Negative, False Positive and False Negative. With these four values, it was a matter of finding the accuracy using the accuracy equation. Then the algorithm with high accuracy is chosen to create the predictor which is then deployed to a web application using Flask.

**Software Requirements**

Software requirements required for this project are Jupyter NoteBook, Flask, Python Libraries like plotly, numpy, pandas, seaborn, matplotlib, sklearn, Dash etc.

**KeyWords:-** Campus Placements, Data Pre-Processing, Data Visualization, Machine Learning Model, KNN (K- Nearest Neighbor), SVM (Support Vector Machine), Decision Tree, Random Forest, Flask

1. **INTRODUCTION**

**Objective**

The main objective of this project is to analyse Placements Data pertaining to the College and mine trends and patterns to understand better the factors influencing the Placements of students. Data relating to the Companies offering Placements is studied as well, in order to draw various inferences such as - Salary Trends, Branch Wise Recruitments etc. The project also aims at developing a prediction model to determine Placement of Students, which enables students to gain insights into what aspects are important for College Placements and how they can improve their chance of being placed. Predicting the probability of students getting placed helps in uplifting their skills before the recruitment process starts. Performing such a comprehensive analysis would help both recruiters as well as students during College Placements and related activities.

**Scope**

The main purpose of this project is to deliver a coherent solution to students being faced with Placement related activities. It provides them a way of easily checking their chance of being Placed just by entering a few parameters and it also shows the previous trends as well as identifies key roles and factors that enable a student to successfully navigate the Placements process. This saves them a lot of time which they would otherwise have to spend on analysing the data manually and painstakingly which is a cumbersome task. The analysis not only helps the students, but also the potential Recruiters for the Campus Placements - surveying the previous recruitment trends gives them a clear picture of the standard of the College. It also helps the College identify and highlight key areas that need attention for improving the Placements.

**Motivation**

Placements are undoubtedly one of the most important parts of any college student’s life. It essentially determines their career path and any guidance during this phase would be invaluable to the Students. It is extremely important that students are steered into the right direction and invigorated so that they aren’t too overwhelmed with the Placement proceedings. Keeping in mind the state of students during this phase motivated us to begin this project and help contribute towards their betterment.

**Overview**

Data Analysis & Prediction of Campus Placements is a project which deals with analysing the campus placements of the past 10 years and predicting the placement chance of students - whether one can be placed or not. This is developed using various Python modules and libraries with Jupyter NoteBook as the platform for working. The project includes Data Analysis, Data Visualization, Training, building a model and testing. The user can enter details like Gender, CGPA, No. of Backlogs, Branch and the model predicts whether he/she can be placed or not. The model is deployed to a Flask Application.

1. **LITERATURE SURVEY**

In [1], a model based on different algorithms like Naive Bayes, Decision Tree, Random Forest and KNN (K- Nearest Neighbor) is proposed, to know which one has high accuracy. This model will predict whether the student gets placed or not in campus recruitment. For this the data considers the academic history of students like overall percentage, skills, CRT training (Campus Recruitment Training) etc. After all the comparisons they concluded that Random Forest and KNN (K- Nearest Neighbor) showed high accuracy for their dataset.

In [2], a model based on the best algorithm from three Weka-based classification algorithms is proposed. They were Multilayer Perceptron, Naive Bayes Classification, and Decision Tree. The best algorithm based on the data is Naive Bayes Classification with an accuracy of 86.15% .This approach found an enhanced evaluation method for predicting the placement for students. This model can be used to assess the relationship between a student's academic success and their placement in the campus selection process.

In [3], the aim is to use linear regression techniques to construct a model that predicts the success of Engineering students. Based on the data collected, the model's indicator or independent variables are the number of hours spent on the internet doing various activities. The estimation of end-of-semester examination grades, i.e. C.G.P.A., is the performance or dependent variable (Cumulative Grade Points).

In [4], the decision tree algorithm was used to predict the placement of students. They used Decision Tree (DT) algorithms like C4.5, ID3, and CHAID, which were created with the help of the Data Mining Rapid Miner software tool. The three algorithms are validated, and their accuracy is determined. Based on the collected placement results, the best algorithm from C4.5, ID3, and CHAID is ID3, which has a 95.33 percent accuracy.

In [5], a model to classify the performance of the placement of students is built. The error rate for classifying validation data, result prediction classification tree, and validating placement prediction classification tree was found to be 38.46 percent and 45.38 percent, respectively.

In [6], a study on Student Placement Prediction was made using Machine learning models which included Naive Bayes Classifier and K-Nearest Neighbors [KNN] algorithm. The study highlighted the efficiency of the algorithms which used the data of the previously passed students and aimed to predict the placement probabilities of the current students.It 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. Here two different machine learning classification algorithms, namely Naive Bayes Classifier and KNearest Neighbors [KNN] algorithm are used.

In [7], a model using linear regression is proposed where they have used the R-squared method to calculate the standard error rate. Linear Regression, a supervised machine learning method, can help with this by contributing its features. It is an effective method helping in predicting future trends of student placement based on advance placement practice test marks. The results will help students better understand their weak areas and where they should focus their efforts. Working on these areas will let students achieve a higher number of placements in an institution.

In [8], master learning strategies are used to predict student placement by using the dataset. The predictive data set parameters include Quantitative scores, Logical Reasoning scores, verbal scores, programming scores, CGPA,number of hackathons participated, number of certificates and current backlog number. The placement prediction is done by machine learning using Logical Regression, Random Forest, KNN (K- Nearest Neighbor) , SVM (Support Vector Machine). The accuracies of each algorithm were Logical Regression 97.59% , Random Forest 96.38%, KNN (K- Nearest Neighbor) 95.18%, SVM (Support Vector Machine) 100%.

In [9], a model is developed using Random Forest and Decision Tree algorithms. This model would predict whether a student will be able to get a job on campus or not. The data takes into account the academic history of students, such as average percentage, backlogs, and credits. The algorithms are run on the students' previous year's results. The decision tree's accuracy after review is 84 percent, while the Random Forest's accuracy is 86 percent. As a consequence of the above study and estimation, it is preferable to use the Random Forest algorithm to predict placement outcomes.

1. **METHODOLOGY**

The Placement Prediction Model is proposed with the intention of making the important Placement proceedings a smooth sail for College Students. It also has the potential to help the Company Recruiters by allowing them to look into important statistics such as year wise and branch wise trends that will help them make informed decisions with respect to their recruitment process.

The proposed system consists of 4 main stages - Data Analysis, Data Visualisation, Building a Prediction Model, Deploying the Dashboard and Prediction Model to a Web Application.

**Data Analysis and Data Visualisation**

Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

To perform Data Analysis, the first step is to obtain the Datasets. The Datasets used in this project have been procured from the Placements Department of our college.

**Description of the Datasets**

The Datasets contain data pertaining to the last 11 years, that is from 2010-2021 in a Branch Wise and Class Wise format. Additionally, Company Wise and Branch Wise data has also been obtained for the last ten years.

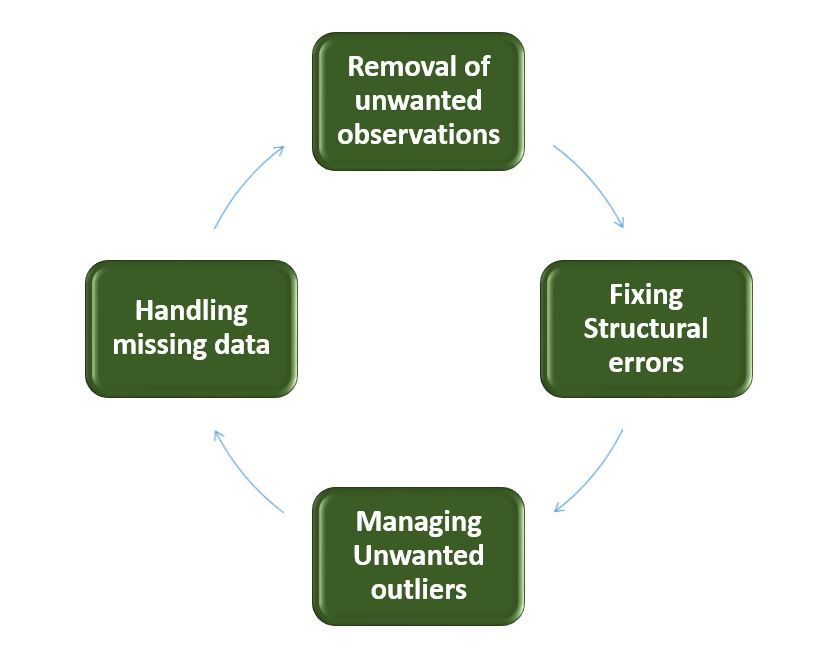
**Students datasets structure**



The **students datasets** are stored in a hierarchical manner. Root directory consists of folders in a year wise manner. Each year folder consists of the .csv files for the sections of each branch. In total there are 153 .csv files. Each dataset consists of data that gives deep insights into the students' placements. Columns such as CGPA, Gender, Branch, No. of placements are present.

The **companies datasets** consists of 11 .csv files, year wise from 2010 to 2021. Each file has data such as the name of the company, salary per annum (in lakhs), the number of placements for a particular company in each branch as well as the grand total of the number of placements.

The Datasets undergo extensive preprocessing for which two Python libraries Pandas and NumPy are used.



**Fig 3.1:** Data Preprocessing cycle

Once the Datasets are cleaned, exploratory data analysis is done in order to extract insightful information.

This information is then displayed in the form of various colourful and interactive charts and graphs. To help with the Data Visualization process, a plethora of python libraries are available. This project employs the use of libraries Matplotlib (Pyplot), Seaborn and Plotly (Plotly Express).



**Fig 3.2:** Flow of Data Analysis and Visualisation

**Building a Prediction Model**

**Selection of Algorithm**

To build a perfect Model with high accuracy, the first step is to analyse what all algorithms are suitable for the given Dataset. For our Data the algorithms used are classification algorithms

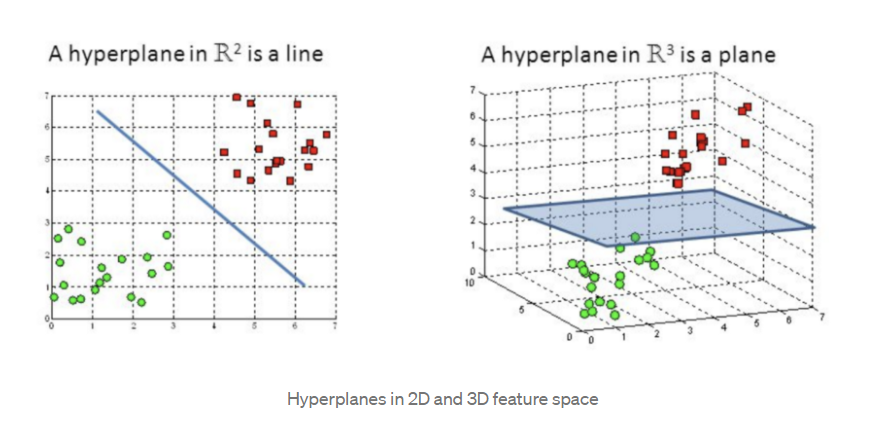
1. SVM (Support Vector Machine)
2. KNN (K- Nearest Neighbor)
3. Decision Tree
4. Random Forest.
5. Naive Bayes

**SVM (Support Vector Machine)**

The Support Vector Machine, or SVM, is a common Supervised Learning algorithm that can be used to solve both classification and regression problems. However, it is mainly used in Machine Learning for Classification problems.

The SVM algorithm's aim is to find the best line or decision boundary for categorizing n-dimensional space into classes so that new data points can be conveniently placed in the correct category in the future. A hyperplane is the name for the best judgment boundary.

The extreme points/vectors that help create the hyperplane are chosen by SVM. Support vectors are the extreme cases, which is why the algorithm is called Support Vector Machine.



**Fig 3.3:** Hyperplanes in 2-Dimensional and 3-Dimensional feature space

**KNN (K- Nearest Neighbor)**

K-Nearest Neighbors is one of Machine Learning's most simple yet essential classification algorithms. Pattern recognition, data mining, and intrusion detection are only a few of the applications it considers in the supervised learning domain.

The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning algorithm that can be used to solve classification and regression problems.

The KNN algorithm assumes that items that are identical are close together. To put it another way, related things are close together. KNN combines the concept of similarity (also known as distance, proximity, or closeness) with some math we might have learned in childhood— and depending on the problem we're trying to solve, one approach could be preferable. The straight-line distance (also known as the Euclidean distance) is a common and well-known choice.

It's commonly used in real-world situations because it's non-parametric, which means it doesn't make any assumptions about data distribution (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

The following algorithm can be used to illustrate how K-NN works:

Step 1: Decide on the number of neighbors (K).

Step 2: Determine the Euclidean distance between K neighbors.

Step 3: Using the measured Euclidean distance, find the K closest neighbors.

Step 4: Count the number of data points in each group among these k neighbors.

Step 5: Assign the new data points to the group with the greatest number of neighbors.

Step 6: We've completed our model.

**Decision Tree**

In machine learning, classification is a two-step method that includes both learning and prediction. The model is based on training data given during the learning stage. The model is used to predict the answer for given data in the prediction stage. The Decision Tree is one of the most straightforward and widely used classification algorithms.

The supervised learning algorithms family includes the Decision Tree algorithm. The decision tree algorithm, like other supervised learning algorithms, can be used to solve regression and classification problems.

The aim of using a Decision Tree is to build a training model that can be used to predict the class or value of the target variable by learning simple decision rules derived from prior data (training data).

When we predict a class label for a record, we start at the root of a tree. Comparison of root attribute values and record attribute. We follow the branch that is the same and jump to the next node on the basis of the comparison.

The decision to make strategic splits has a significant impact on a tree's accuracy. The decision criteria for classification and regression trees are different.

To determine whether to divide a node into two or more sub-nodes, decision trees employ a variety of algorithms. The homogeneity of the resulting sub-nodes improves with the construction of sub-nodes. To put it another way, the purity of the node improves as the goal variable increases. The decision tree divides the nodes into sub-nodes based on all available variables, then chooses the split that produces the most homogeneous sub-nodes.

The following algorithm can be used to understand the process:

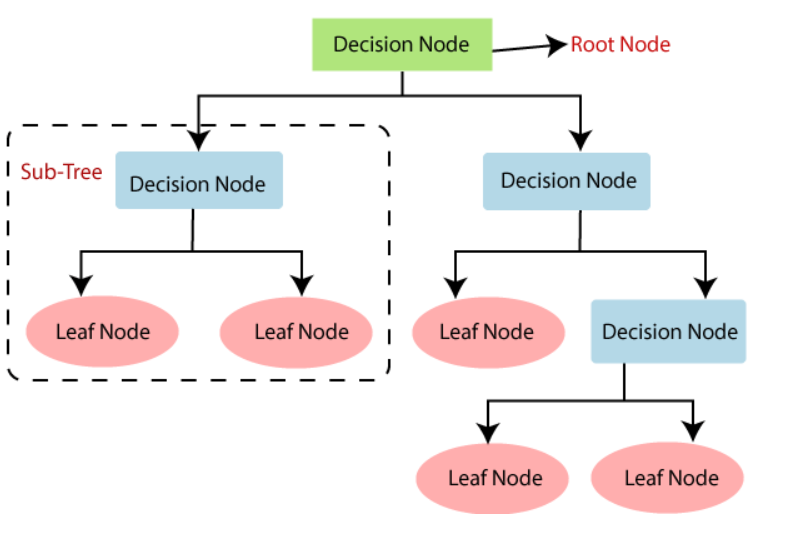
Step 1: Start with the root node, which contains the entire dataset, says S.

Step 2: Using the Attribute Selection Measure, find the best attribute in the dataset (ASM).

Step 3: Subdivide the S into subsets that contain the best attribute's possible values.

Step 4: Create the node of the decision tree that contains the best attribute.

Phase 5: Generate new decision trees in a recursive manner using the subsets of the dataset generated in step 3. Continue this method until you can no longer identify the nodes any further and refer to the final node as a leaf node.



**Fig 3.4:** Working Process of Decision Tree

**Random Forest**

Random Forest is a well-known supervised learning algorithm. In machine learning, it can be applied to both classification and regression problems. It is based on the idea of ensemble learning, which is a method of combining multiple classifiers to solve a complex problem and improve the model's accuracy.

According to the term, a Random Forest is “a classifier that combines multiple decision trees on different subsets of a dataset and averages their outputs to increase the dataset's predictive accuracy.”

The Random Forest algorithm operates in two phases: first, it creates a random forest by combining N decision trees, and then it makes predictions for each tree generated in the first step.

The following steps and diagram can be used to illustrate the working process:

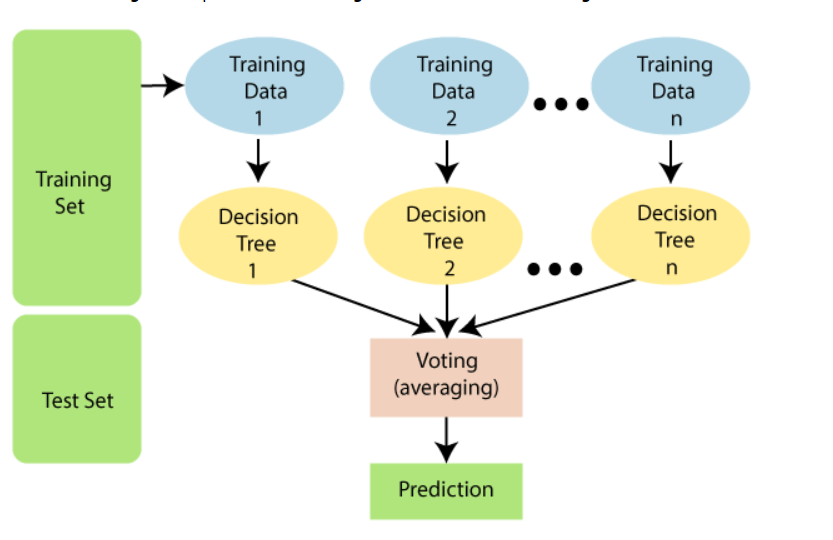
Step 1: Pick K data points at random from the training collection.

Step 2: Create decision trees for the data points you've chosen (Subsets).

Step 3: Decide on the number N for the decision trees you want to build.

Step 4: Repetition of Steps 1 and 2.

Step 5: Find the predictions of each decision tree for new data points, and allocate the new data points to the group with the most votes.



**Fig 3.5:** Working process of Random Forest

**Naive Bayes**

Naive Bayes is a supervised learning algorithm which helps models to make quick predictions. According to the term Naive Bayes it assumes that occurance of one feature in a class is independent of other features (so is called Naive) and uses the Bayes Therom as its main principle.

**Bayes’ Theorm**: Bayes’ Theorem describes the probability of an event occurring given the probability of another event that has already occured.

P(A|B)= [P(B|A)P(A)]/P(B)

where A and B are events and P(B)!=0

* P(A|B) is a conditional probability: the likelihood of event A occurring given that B is true.
* P(B|A) is also a conditional probability: the likelihood of event B occurring given that A is true.
* P(A) and P(B) are the probabilities of observing A and B independently of each other, it is known as the marginal probability.

The following steps can be used to illustrate the working process:

Step 1: Prior Probability calculation for the given labels.

Step 2: Calculate conditional probability with each feature for each class.

Step 3: Multiply same class conditional probability.

Step 4: Multiply prior probability with step 3 probability.

Step 5: See which class has the higher probability(higher probability class belongs to the given input step).

Once the algorithms are selected, the next step is to check what features are required for the prediction.

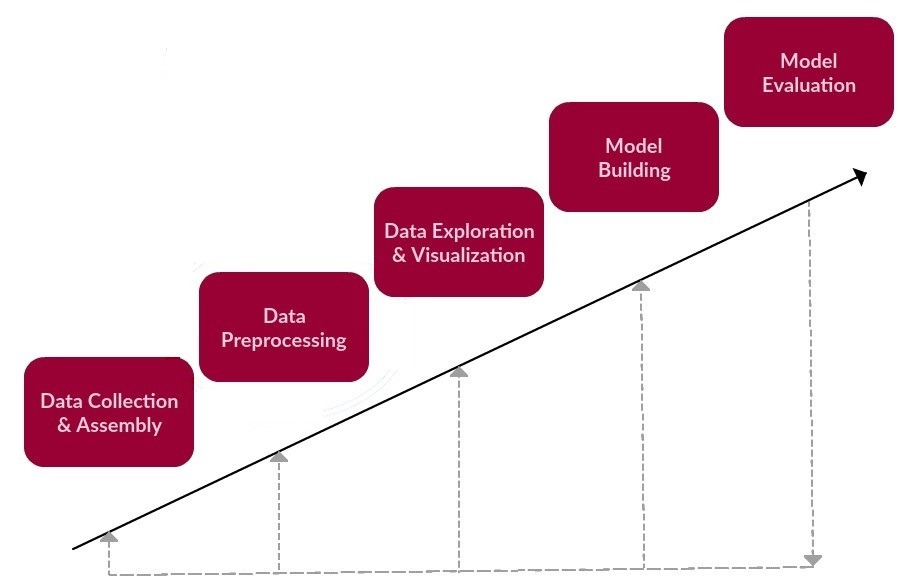
**Working of Machine Learning Algorithm**

After thorough observation we found that Naive Bayes Algorithm is the best for our Dataset. The features used in our model are Gender, CGPA, Number of Backlogs, Branch. The data is converted to categorical type. It is then encoded using the .cat.codes method. The features are stored in variable X and the response vector is stored in variable y. The models are developed and are trained with 70% of the Dataset and tested with 30% of the Dataset. After importing the model from the sklearn module of python, an instance of the model is created and fitting of the training data is done. We were able to predict the chance of a student being placed. For this, we used the predict and predict\_proba methods to predict one's placement and the probability of his chance, respectively. predict\_proba is a function that returns the chance of being placed as well as the chance of not being placed in the form of an array.

**Deploying to a Web Application**

Once a prediction model is built, the next step typically entails deployment. The model is deployed using Flask to integrate it into an existing production environment which can take in any input and return an output which can be used by any student to know whether they have a chance of being placed or not. The data analysis and visualisation is presented in the form of a dashboard which is built using a python library called Dash. This dashboard is integrated to the existing flask application containing the predictor. Once the flask application is developed and all bugs are resolved, it is deployed to a Heroku web link.

**Workflow**

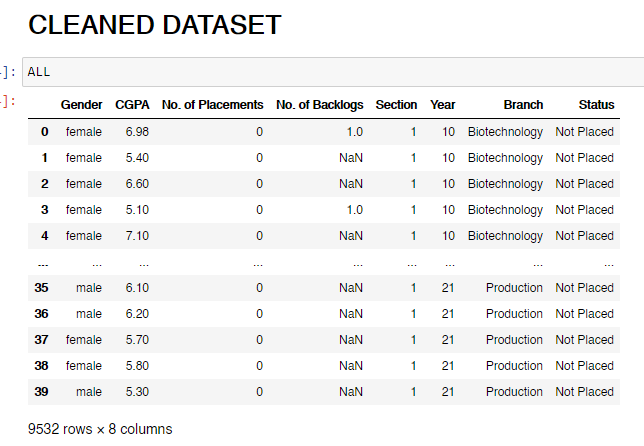
****

**Fig 3.6:** Workflow of the project

1. **RESULTS**

For the Students Data, firstly, unnecessary columns are dropped from the dataset and the columns are renamed to a particular format. The Data Types are extremely important and in this regard, the entire dataset is formatted to achieve homogeneity. Next, the missing values are filled in. Take for example the CGPA column. For this, a “Replace” function has been used - the null values are replaced with the average value of that column. This essentially fills in the average CGPA of the class for a student with a missing CGPA.

Once suitable cleaning has been done, the datasets are merged year wise and branch wise.



**Fig 4.1:** Cleaned Dataset for students data

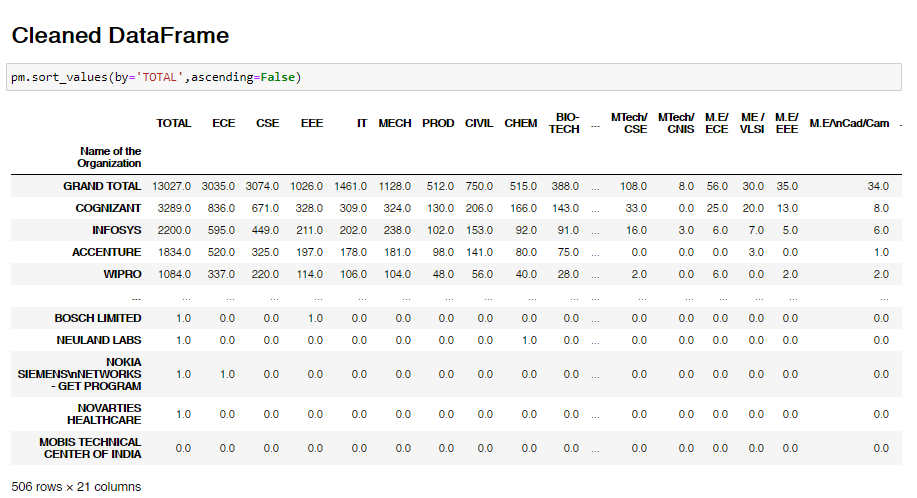
As shown in Fig 4.1, the cleaned dataset contains columns -

1. Gender
2. CGPA
3. No. of Placements
4. No. of Backlogs
5. Section
6. Year
7. Branch
8. Status

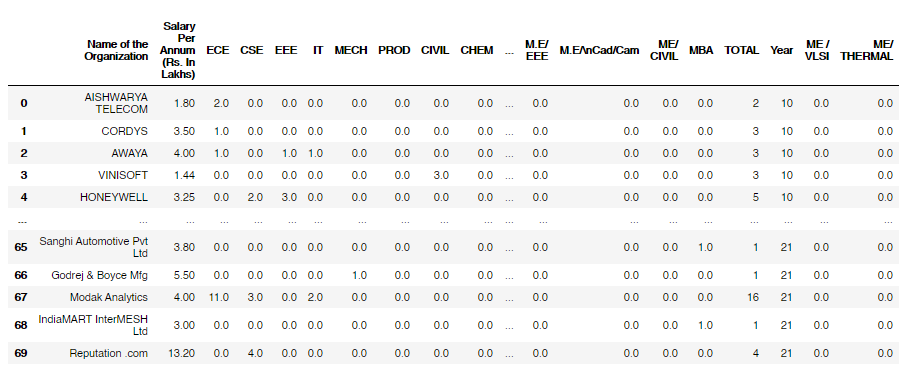
This cumulative dataset provides deep insights into how the students from various branches have fared academically over the last 11 years. By analysing this data various trends can be mined.

For the Company Data, similar preprocessing strategies are applied - unnecessary columns are dropped from the dataset and the columns are renamed to a particular format. The strings are all converted to the same format to achieve homogeneity. Next, the missing values are filled in.

Once suitable cleaning has been done, the datasets are merged / concatenated year wise.

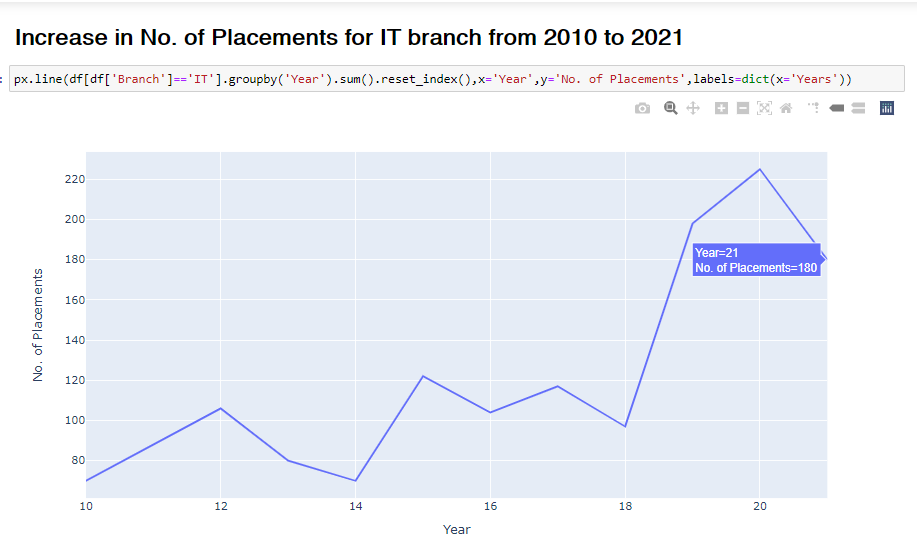


**Fig 4.2:** Cleaned & Merged Dataset for Companies data (Without Salary Column)

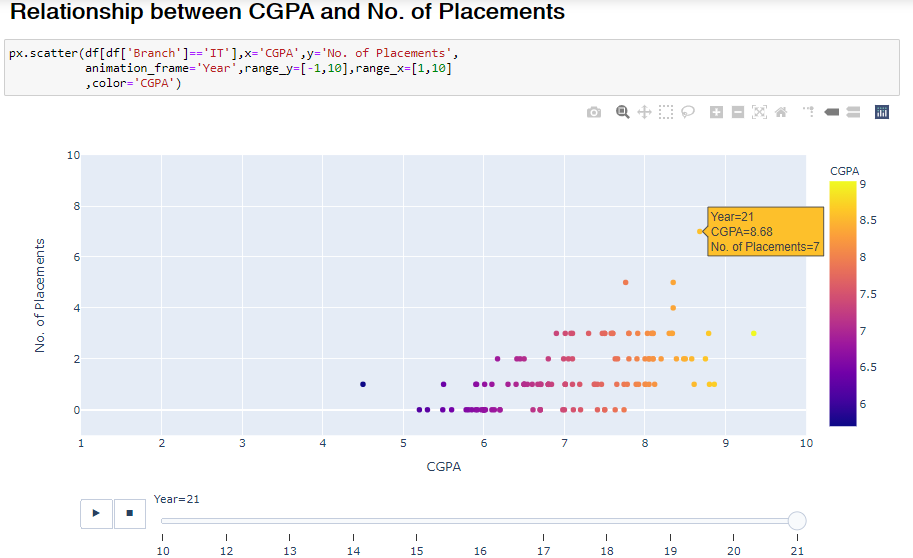


**Fig 4.3:** Cleaned & Concatenated Dataset for Companies data (With Salary Column)

From the Students Dataset, a lot of useful patterns and trends can be extracted. Investigating if a relationship exists between various parameters helps profoundly with making an approach towards building the Prediction Model.

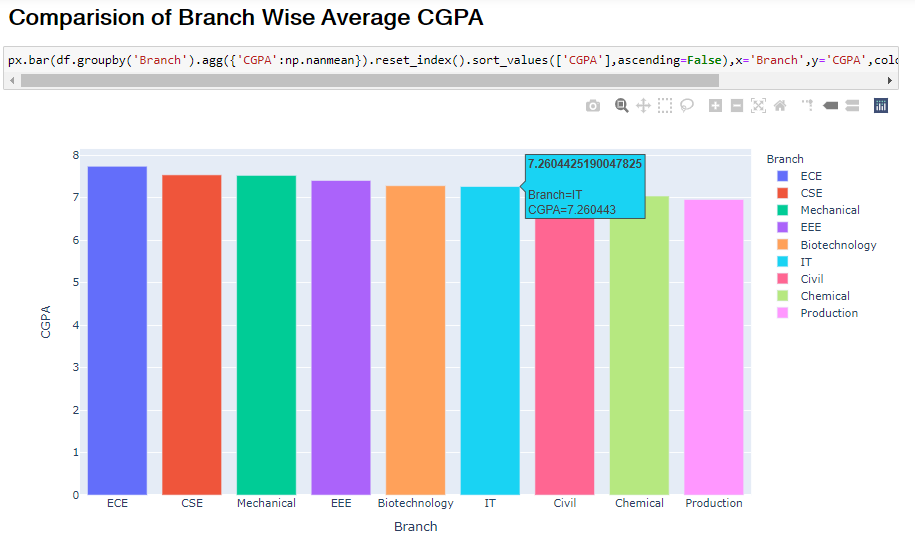
****

**Fig 4.4:** A Line plot depicting the increase in No. of Placements over the years for IT Branch

****

**Fig 4.5:** Relationship between CGPA and No. of Placements

From Fig 4.5, it is clear that a linear relationship exists between CGPA and No. of Placements. The higher the CGPA of a student, the more No. of Placements they receive. To have a better chance of being placed, the best bet would be to improve the CGPA.

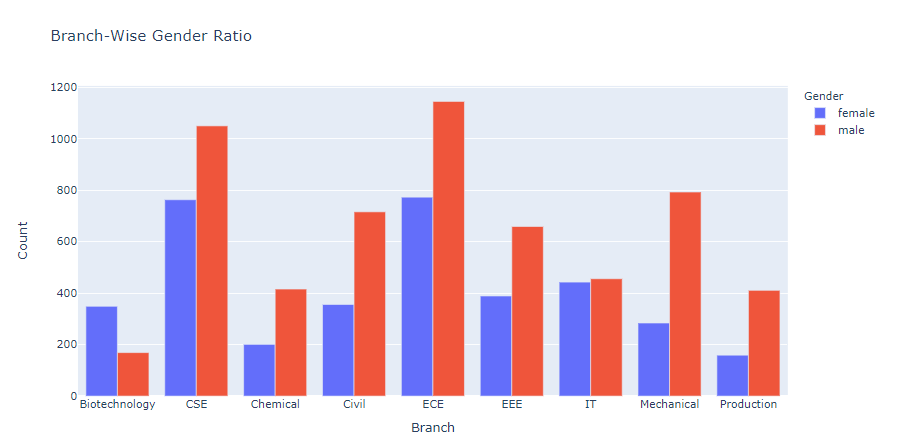
****

**Fig 4.6:** Branch Wise Average CGPA

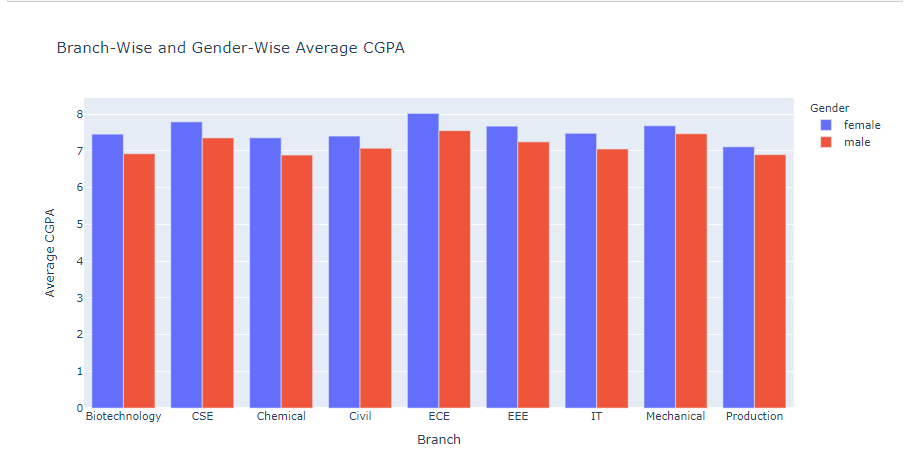
****

**Fig 4.7:** Branch Wise No. Of Placements

From Fig 4.6 and Fig 4.7, it is further evident that CGPA has a huge impact on the No. of Placements. ECE and CSE are the top branches in terms of CGPA as well as No. of Placements. However, there are numerous other parameters that come into play when considering the chance of being placed. Although IT stands at the sixth place in terms of Avg. CGPA when compared to other branches, it ranks third when it comes to No. of Placements. This indicates the importance of Branch when considering Placements.

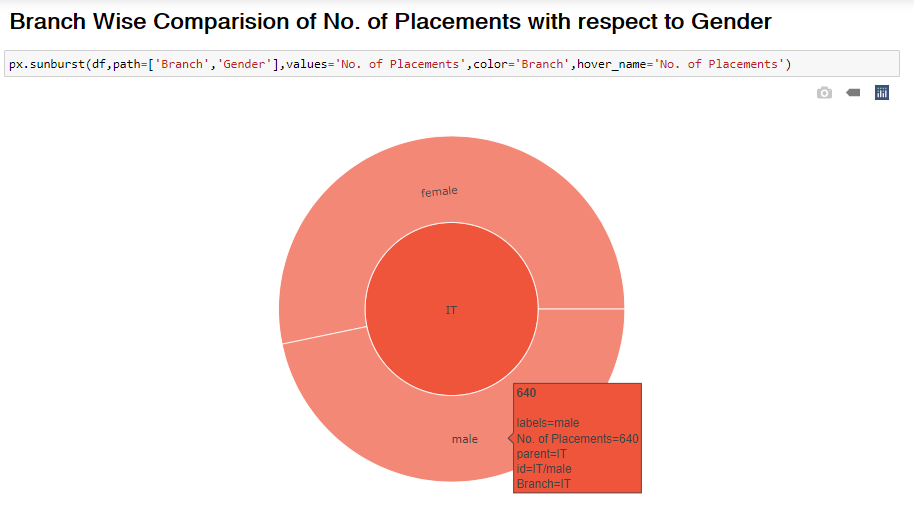
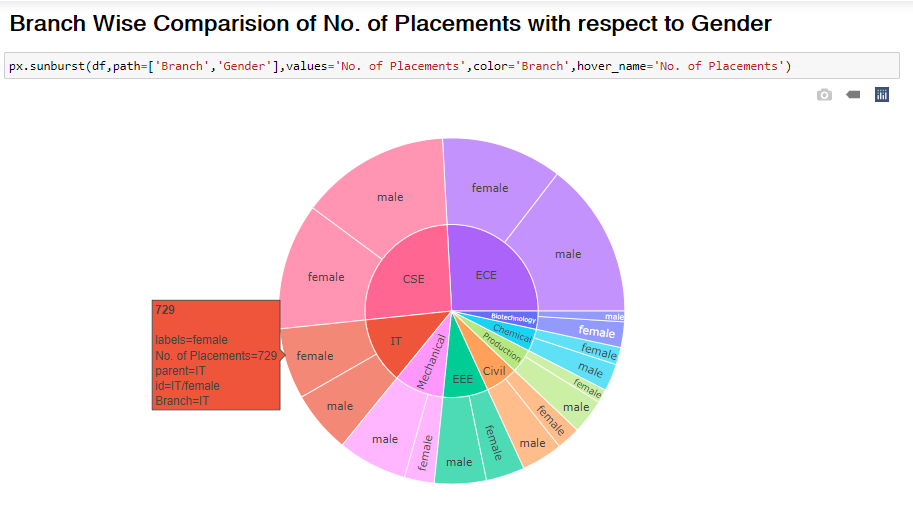


**Fig 4.8:** Branch Wise Gender count



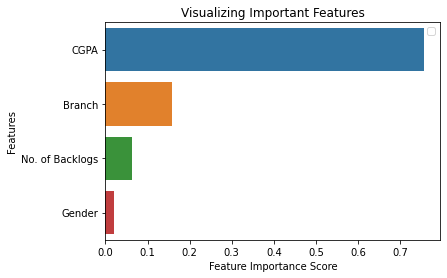
**Fig 4.9:** Branch Wise and Gender Wise effect on Avg. CGPA

From Fig 4.8 and Fig 4.9, it is conspicuous that the count of Male students is usually higher than Female students. However, Female students in general tend to have a slightly higher CGPA than Male students on average. But, does Gender have any effect on Placements? Taking a look into the effect of gender on No. of Placements can help answer this question.



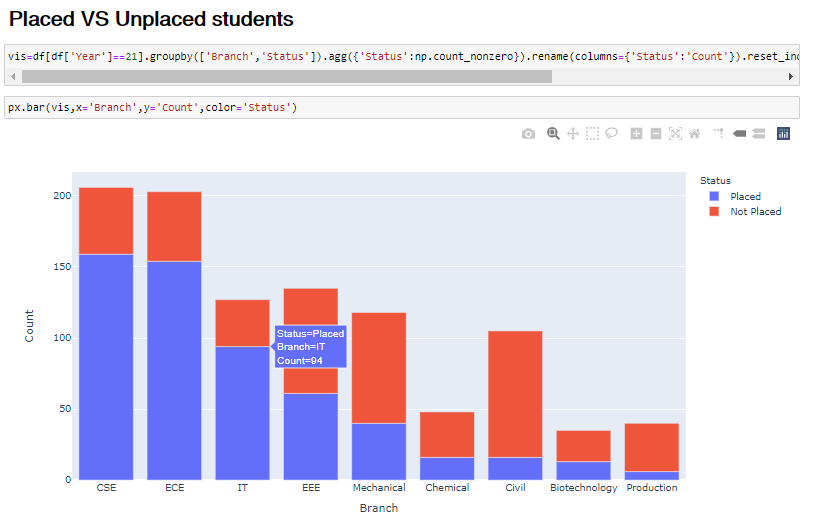
**Fig 4.10:** Branch Wise and Gender Wise effect on Placements

In Fig 4.8 and Fig 4.9 and Fig. 4.10, the relationship between Gender, CGPA and No. of placements has been studied. In some branches such as Biotechnology and IT, Female Students tend to have a better chance of being placed, while the Male Students dominate in other branches like CSE and ECE. Despite having a lower average CGPA than Female Students, the higher Placement figures of Male Students can be attributed to the fact that their count is a lot higher when compared to Female Students. Hence the effect of Gender on Placements may be disregarded when compared to other important factors such as CGPA and Branch.

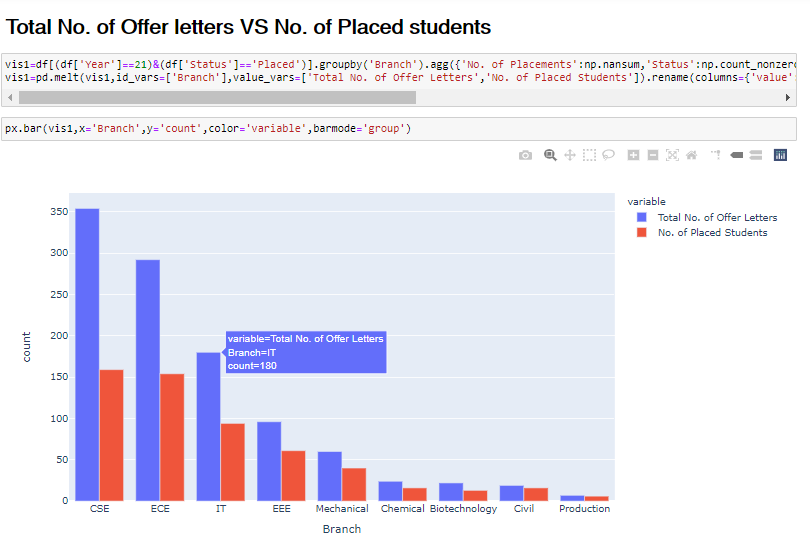


**Fig 4.11:** Visualising important features

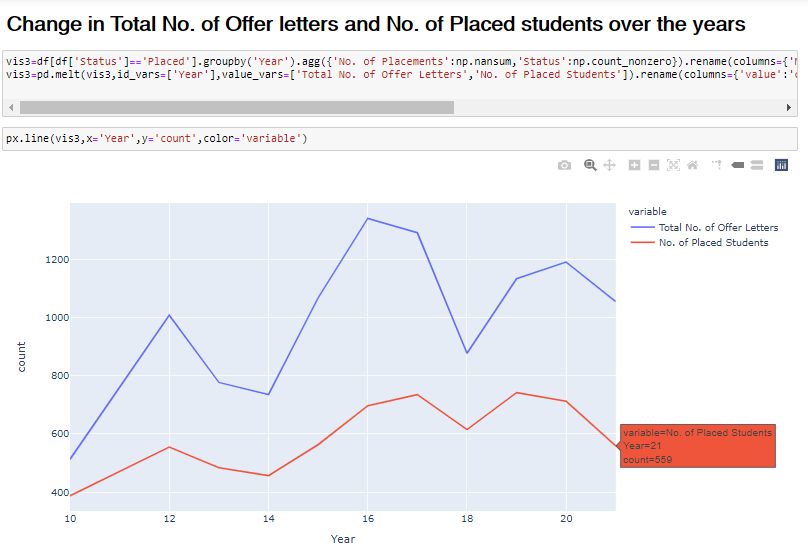
Random Forest Classifier is used to predict the importance of each parameter in the dataset with respect to Placements. It is apparent that CGPA has the most effect on Placements, followed by Branch, No. of Backlogs and Gender.



**Fig 4.12:** Branch-Wise Placed and Unplaced students count for the year 2021

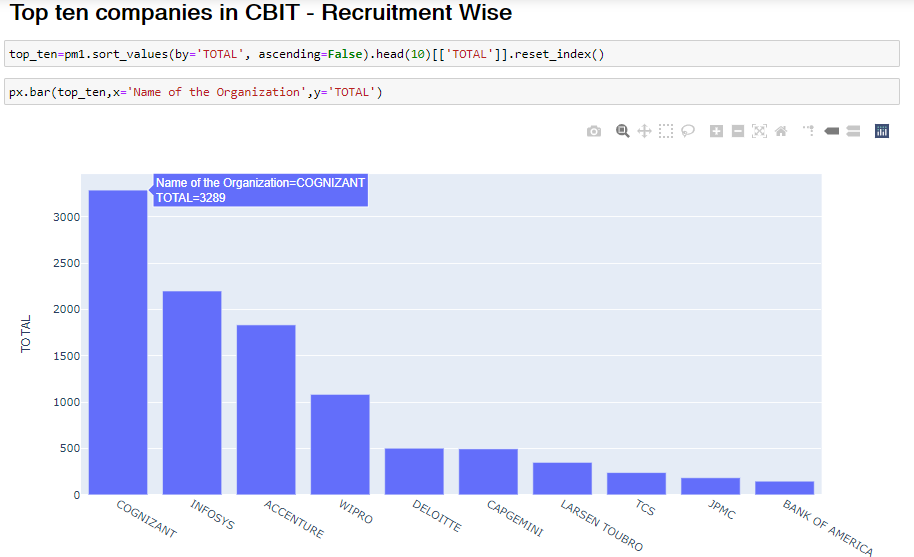


**Fig 4.13:** Comparison of No. of offer letters and No. of placed students Branch-Wise for the year 2021

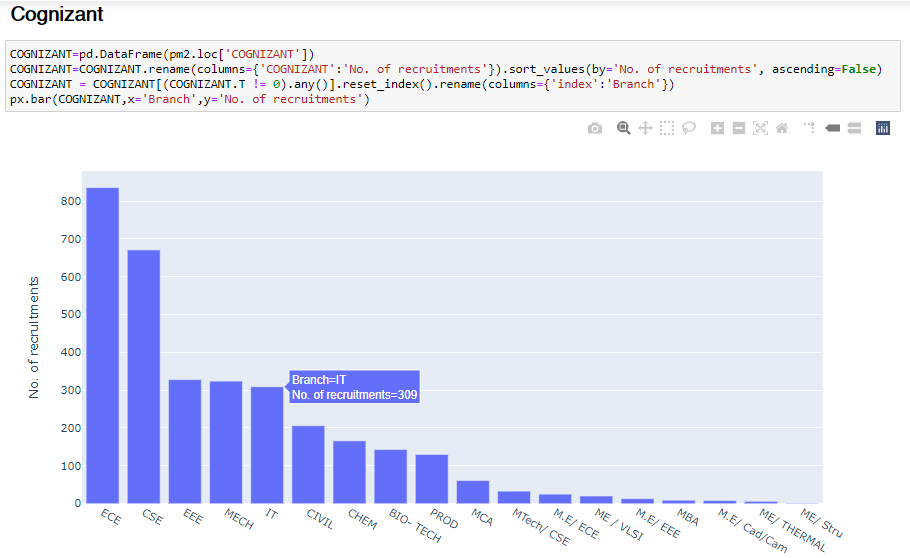


**Fig 4.14:** Change in Total No. of Offer letters and No. of Placed students over the years

From 4.12, 4.13 & 4.14, clearly the number of offer letters is always more than the number of placed students, implying that competent students may be placed in more than one company, although some students may not be placed in any company at all. In order to be placed, students must strive to perform their best, as only the ones with good potential are offered the jobs.

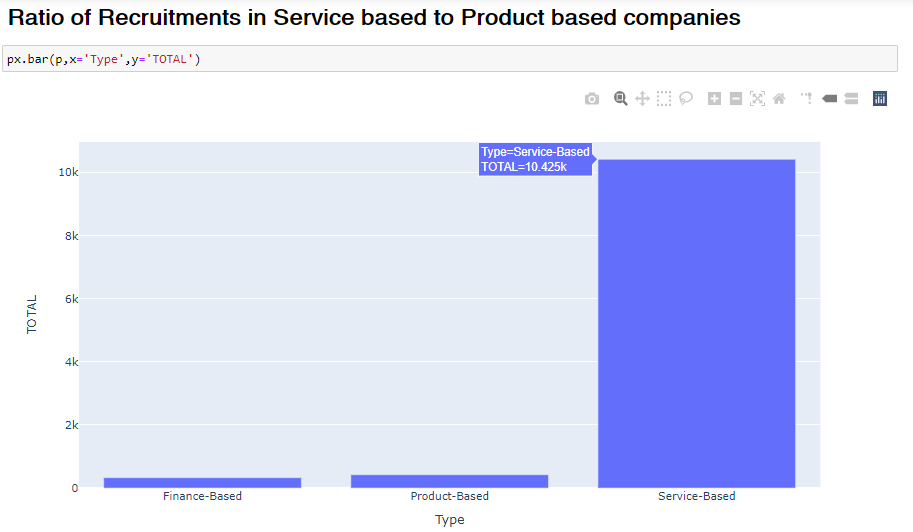


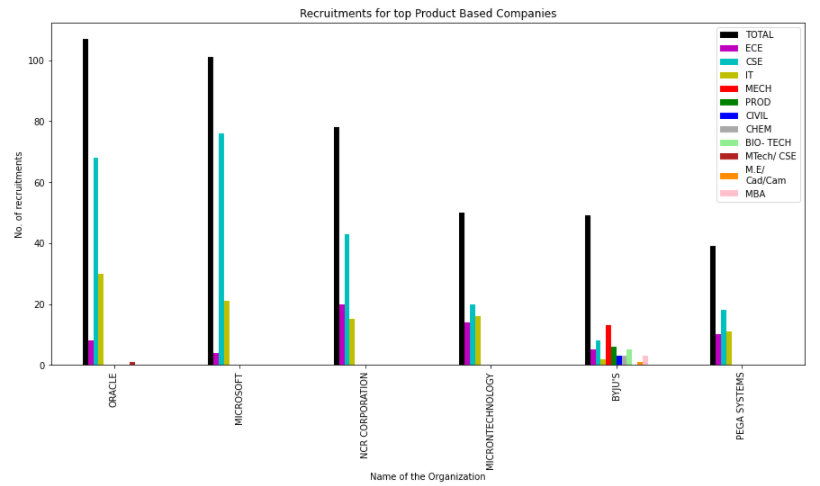
**Fig 4.15:** Top 10 companies according to No. of Recruitments

****

**Fig 4.16:** Branch Wise Recruitments count for Cognizant

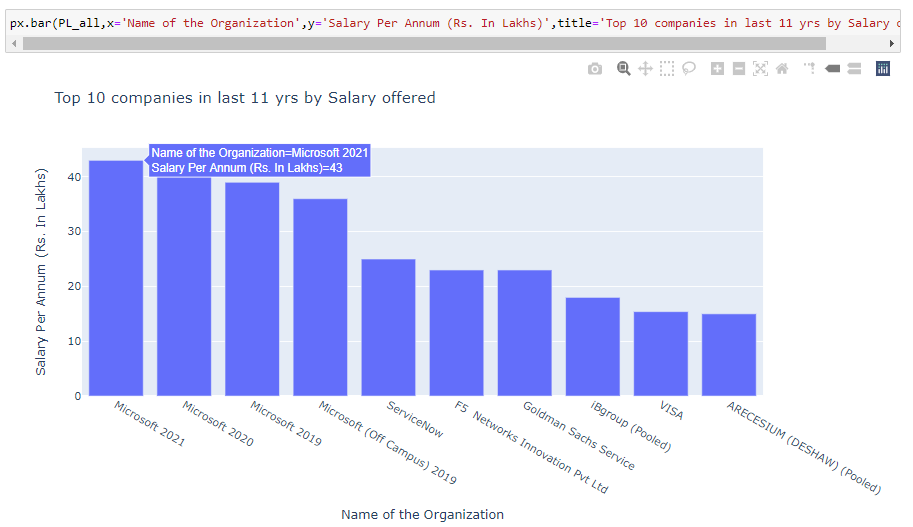
Cognizant is the company with the highest number of recruitments as displayed in Fig 4.15. The top branches that Cognizant recruits from are ECE, CSE, EEE and Mechanical, while IT ranks fifth with 309 placements.

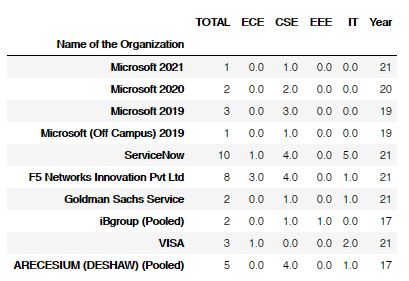


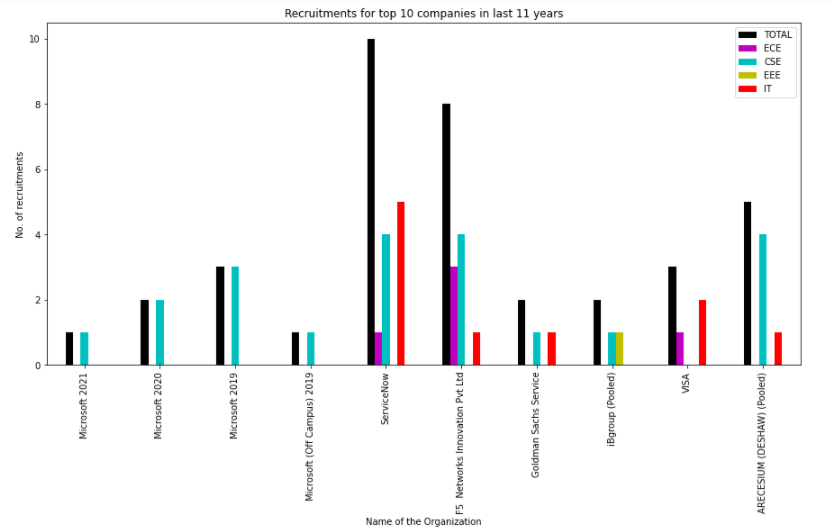


**Fig 4.17:** Recruitment trend for top Product/Finance Based Companies

From Fig 4.17, clearly the Service based companies have majority recruitments, while the Product based companies mostly recruit from the circuit branches with CSE having majority of the recruitments.

**Fig 4.18:** Top 10 Companies by Package offered



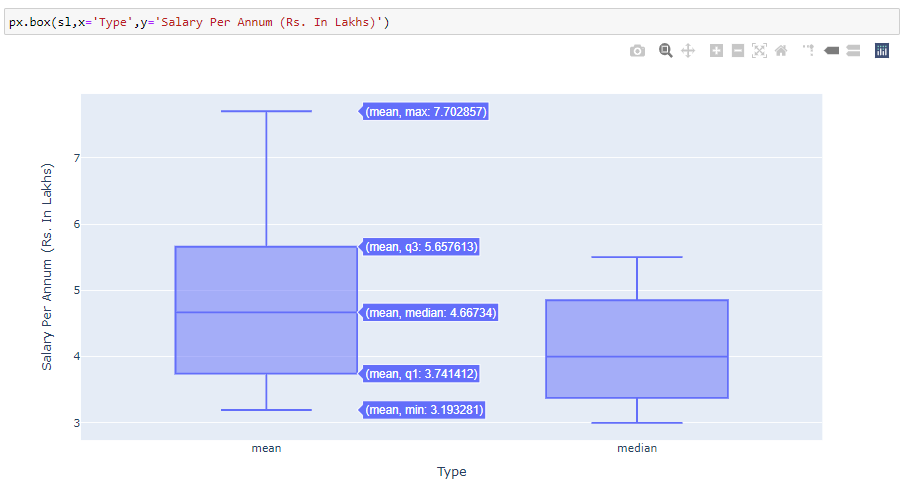


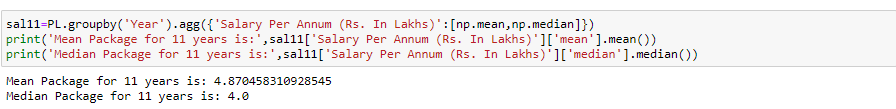
**Fig 4.19:** Branch Wise recruitments for top 10 Highest Paying Companies

From Fig 4.18 and Fig 4.19, it can be inferred that Microsoft takes the spot for the company with the highest package, offering a whopping Rs. 43 LPA. In the history of the last 11 years, Microsoft has only recruited from the CSE branch for such high packages.

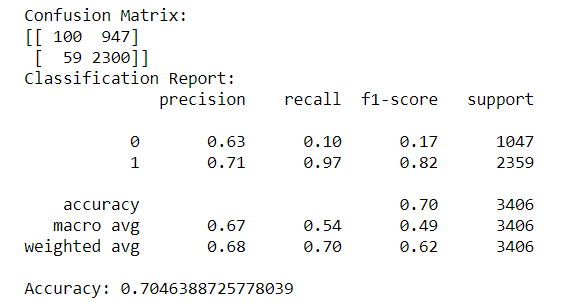
ServiceNow ranks fifth in terms of package offered, but has the most number of recruitments for the top paying companies.

****

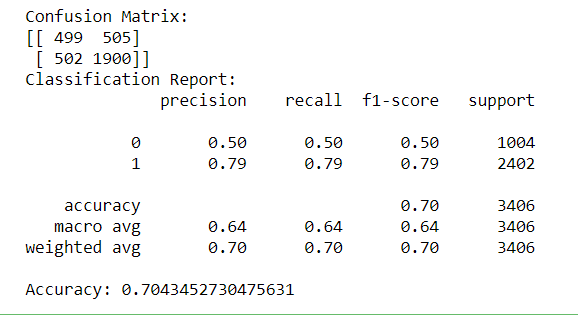
****

****

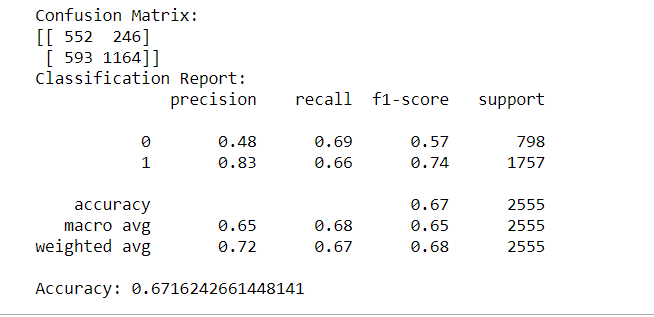
**Fig 4.20:** Average & Median Package Trends



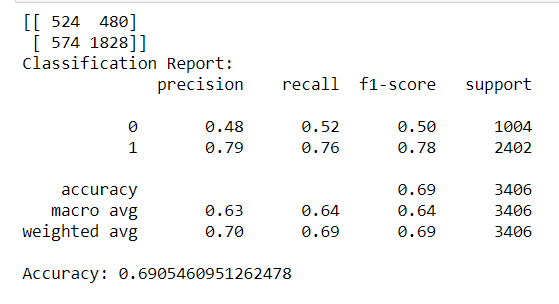
**Fig 4.21:** Confusion Matrix and Classification Report of SVM



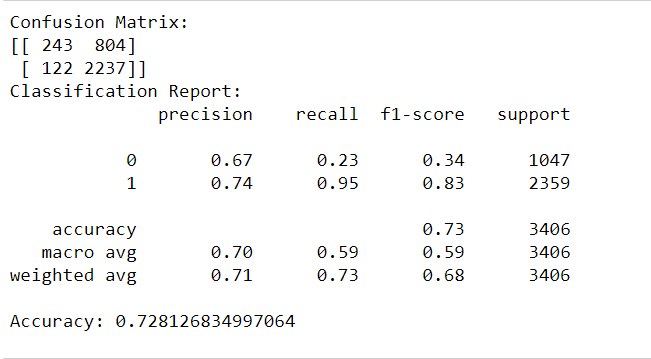
**Fig 4.22:** Confusion Matrix and Classification Report of Random Forest



**Fig 4.23:** Confusion Matrix and Classification Report of KNN

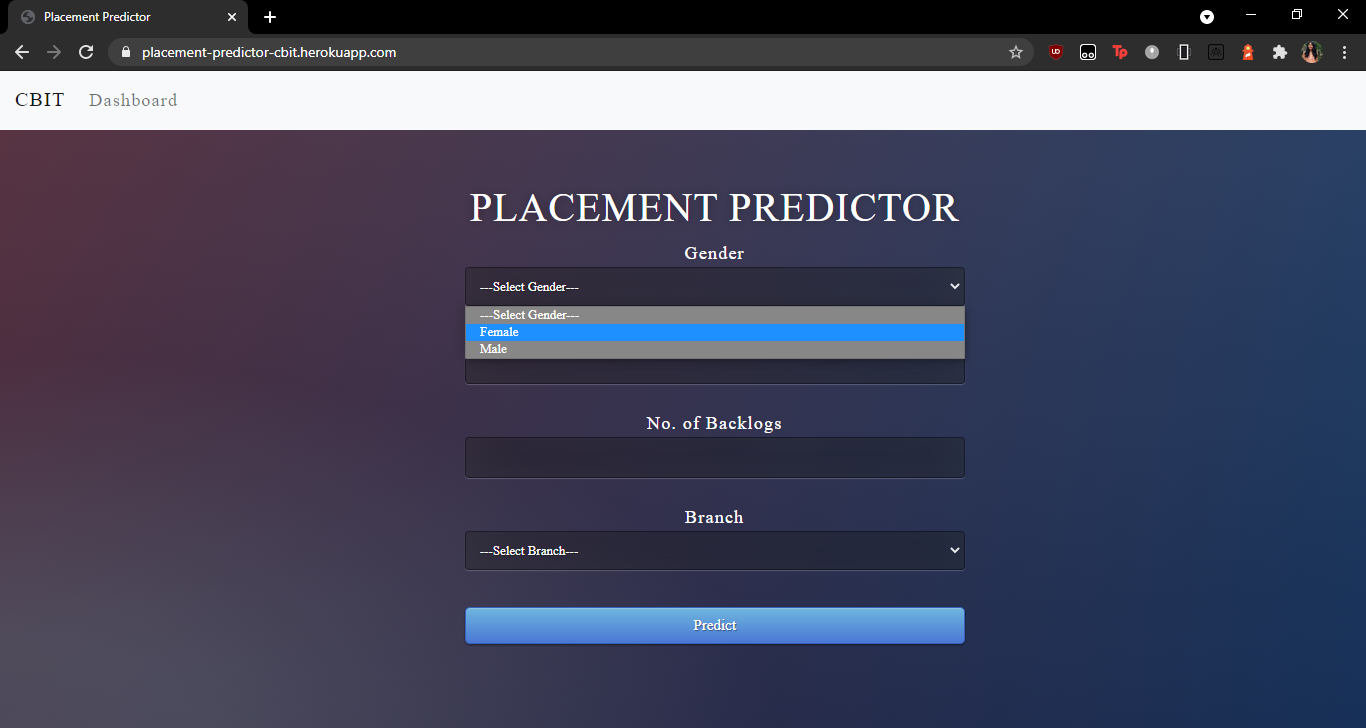


**Fig 4.24:** Confusion Matrix and Classification Report of Decision Tree

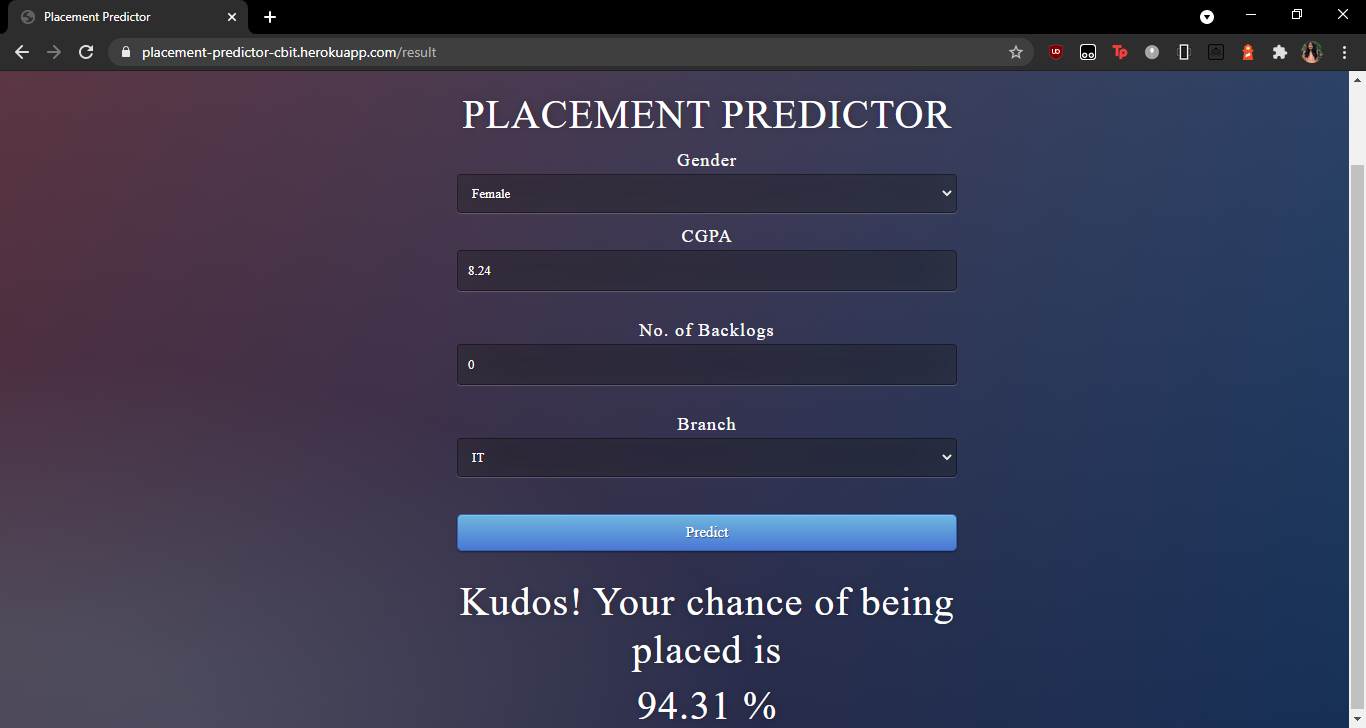


**Fig 4.25:** Confusion Matrix and Classification Report of Naive Bayes

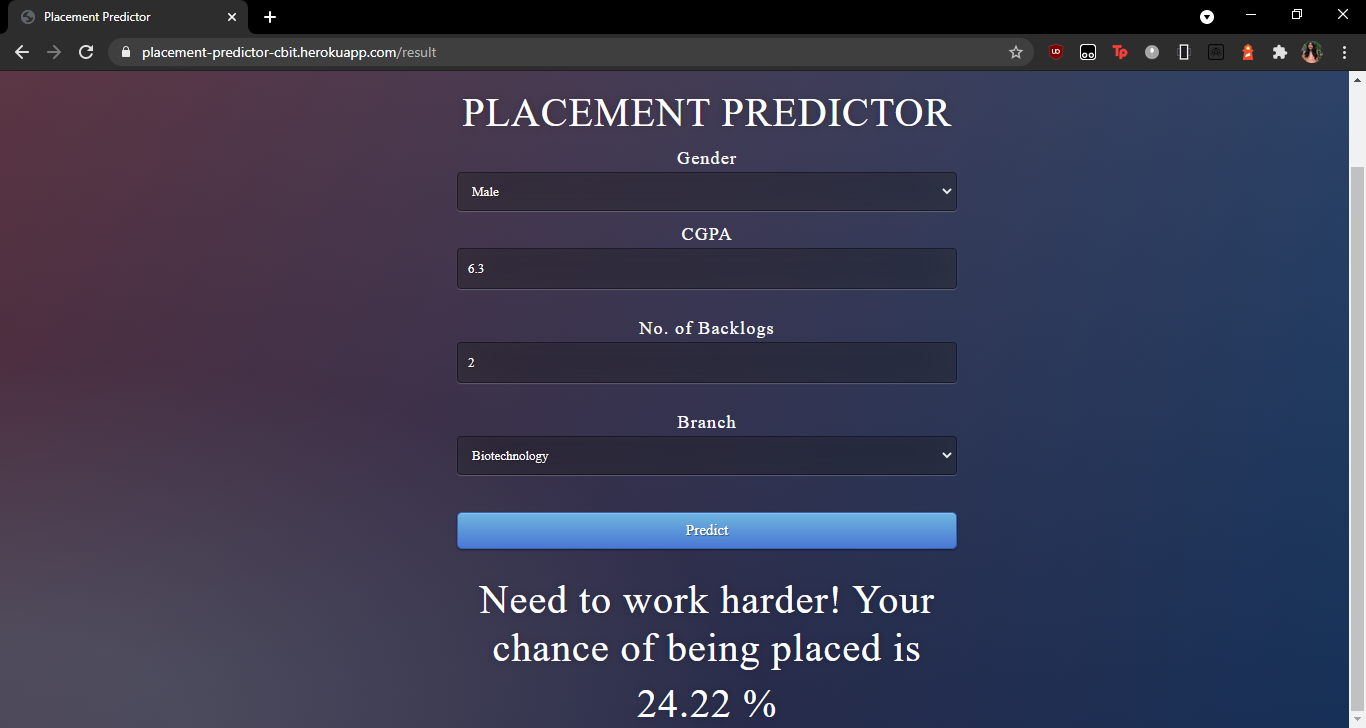
From Fig 4.21, Fig 4.22, Fig 4.23, Fig 4.24, Fig 4.25 We can state that the best and most suitable algorithm for the given Dataset is Naive Bayes which has an accuracy of 72.81%.



**Fig 4.26:** Placement Predictor

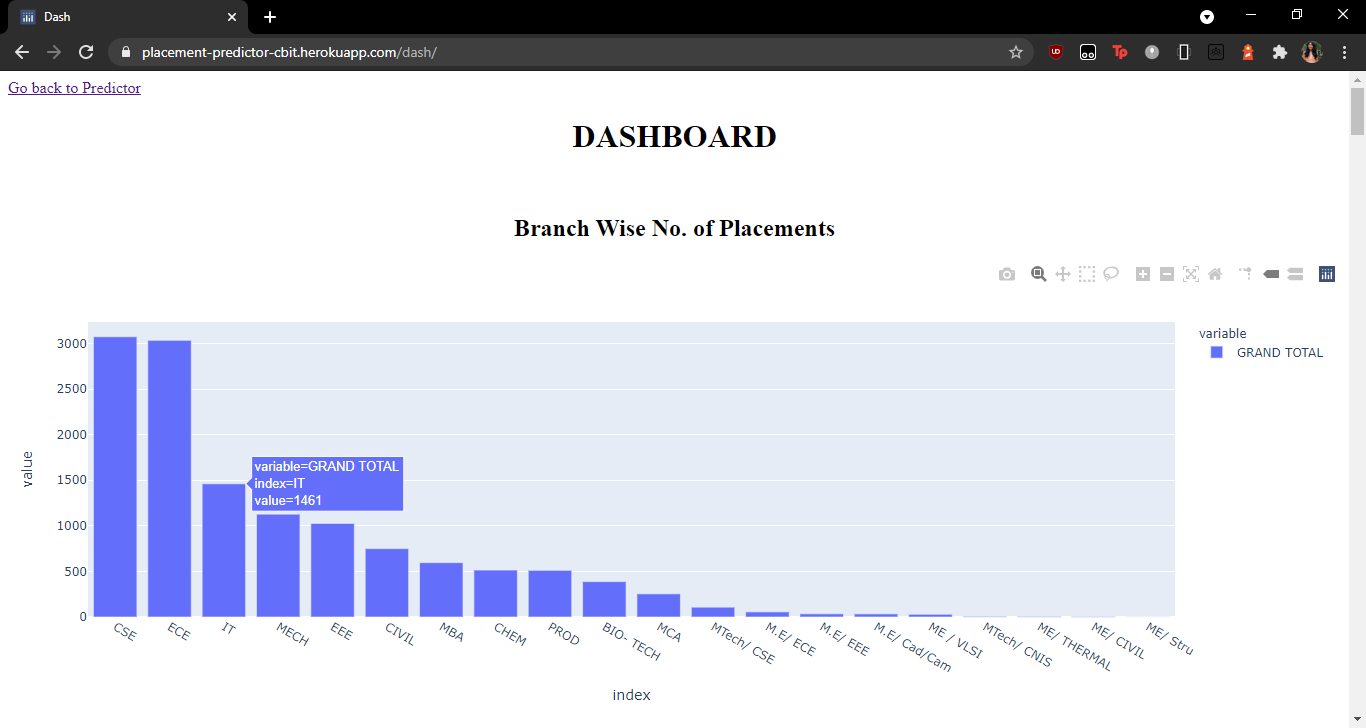


**Fig 4.27:** Placement Predictor displaying a Message if a Student can be Placed

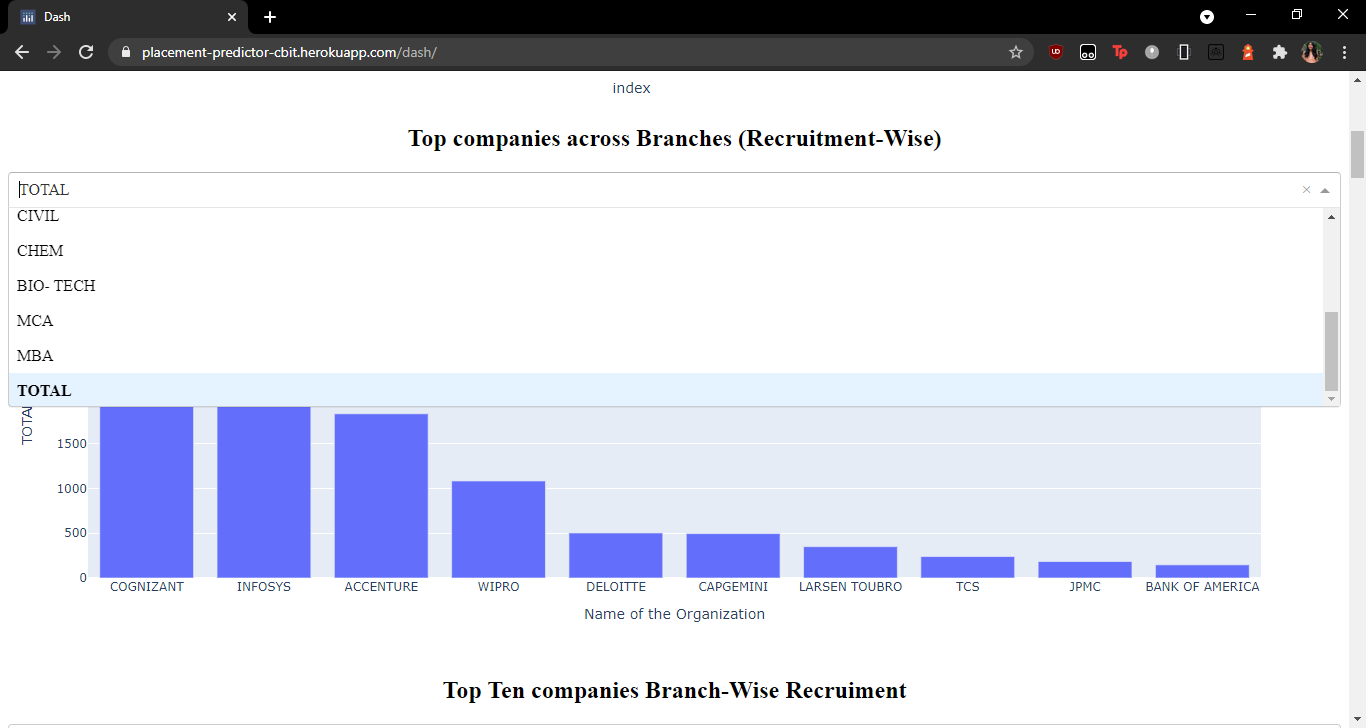


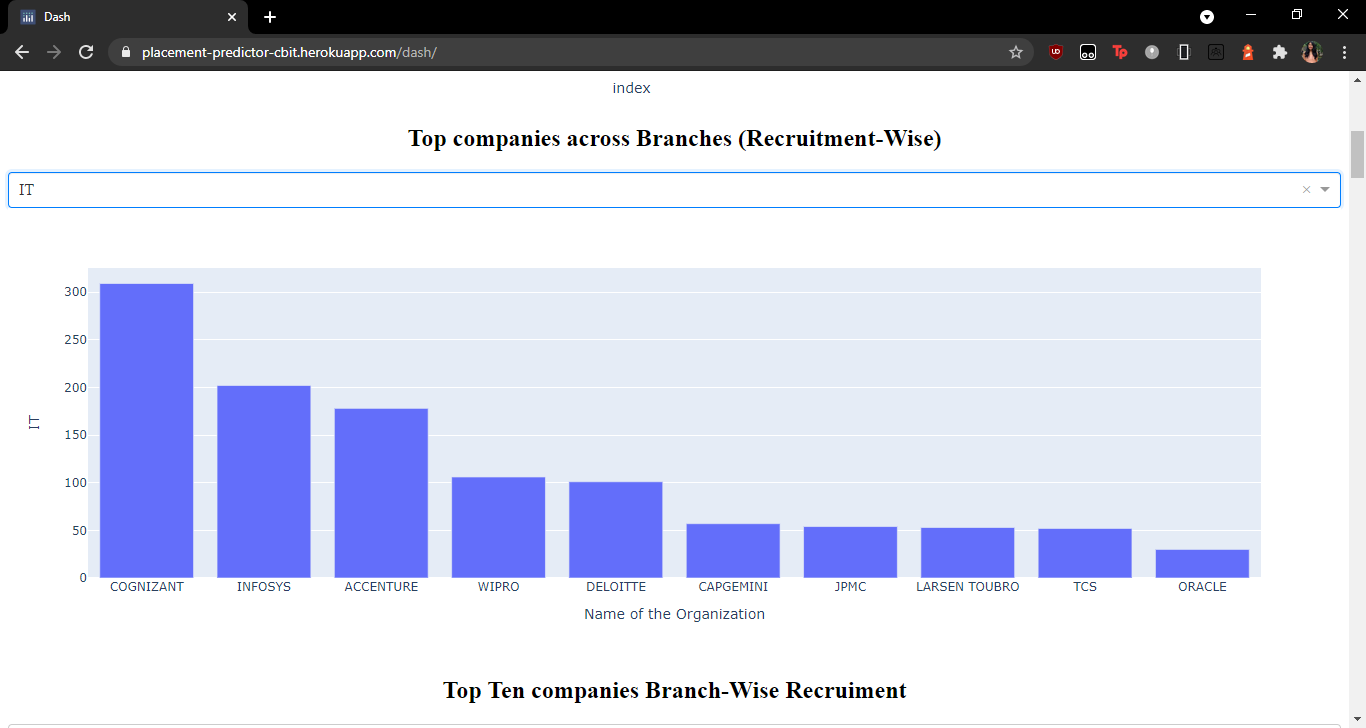
**Fig 4.28:** Placement Predictor displaying a message if a student can’t be placed

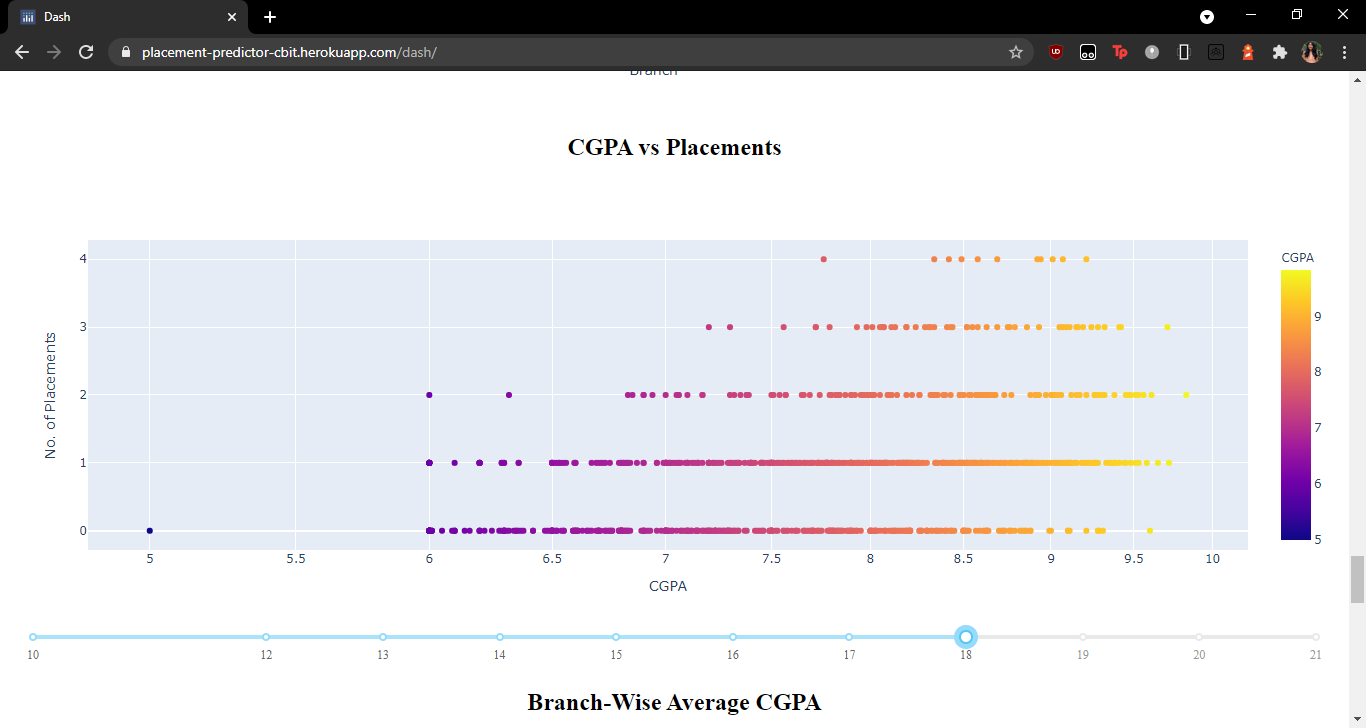
In Fig 4.26, Fig 4.27, Fig 4.28, the best model (here it is Naive Bayes with an accuracy of 72.81%) is deployed to a flask application which can take in any input and return an outputthat can be used by any student to know whether he/she can be placed or not.



**Fig 4.29:** Dashboard for the visualization and analysis of placement trends







**Fig 4.30:** Dynamic inputs

Fig 4.29 and 4.30, show a dynamic dashboard created using a python library Dash. It offers the users a way to visualize and analyse the placement trends. It is completely dynamic and the graphs change according to user inputs such as year, branch etc.

1. **CONCLUSIONS**

Placements are undoubtedly one of the most important aspects of a student’s college life. Students tend to come under extreme pressure around the Placements season. Placements are what pave the path for a student’s future career and being cognizant of the fact that one’s entire career would be dependent on the events that unfold over just a few months can be quite overwhelming. Students face the incessant compulsion to perform well, and this invariably takes a toll on their confidence levels and drives them over the edge. To overcome such extreme levels of stress, they need to be motivated and guided well. Boosting one’s morale can have an overall positive impact on how they perform in the Placements. One way this can be done is by analysing the previous placement trends and assuring the students of their caliber. This inspires them to be diligent and do their best. It also serves as a means of constructive criticism that helps students realise how they can better themselves and build the necessary skills required for their ideal placement. This will be of great help to both the students, as well as the institution. Placements have become a benchmark to compare the efficiency of Institutions. It’s a known fact that placements are what essentially determine and impact the number of students willing to join an Institution. Hence, with regard to this, a solution has been offered in the form of a Placement Analyser and Predictor.

To improve the student’s performance, previous data has been analyzed, visualised and future prediction using a Machine Learning Model has been done. Models are developed using the classification algorithms SVM (Support Vector Machine), KNN (K- Nearest Neighbor), Decision Tree Random forest and Naive Bayes algorithm to validate the approaches. The algorithms are applied on the data set and attributes used to build the model.

|  |  |
| --- | --- |
| **Algorithms** | **Accuracy (%)** |
| SVM (Support Vector Machine) | 70.46 |
| KNN (K- Nearest Neighbor) | 67 |
| Decision Tree | 69 |
| Random forest | 70.43 |
| Naive Bayes algorithm | 72.8 |

**Fig 5.1** Table showing accuracies of proposed Algorithms

Hence, from table Fig 5.1,we can conclude that Naive Bayes has the highest accuracy and is the best algorithm for our Dataset.

The accuracies of all the algorithms used in this project ranges from 65-75, this is because the data was very noisy, there was a lot of missing data. Even after preprocessing of the datasets, the accuracy of the model was compromised due to insufficient data for some of the features.

Out of these Models, the best one is deployed to a Flask Web Application which takes in inputs such as CGPA, Gender, Branch and No. of Backlogs and returns an outputthat determines whether a student will be placed or not.

1. **FUTURE WORK**

The scope of this project isn’t just confined to the existing state of art. There are many more tweaks that can be incorporated to make it a lot more beneficial on a whole.

The Prediction Model is one of the most important parts of the project and it can be further improved to get better accuracy. It can also be enhanced to take in many more parameters, excluding the standard parameters, such as a student’s dream company and predicting their chance of getting placed in that company along with feedback on how they can increase that chance. Predictions on how the average package of the institution can be boosted also will certainly prove to be extremely helpful to the Institution.

1. **REFERENCES**
2. Ankita Mahalle, Divisha Samrit, Nishigandha Wagh, Rucheeka Gothe “campus placement predicting using data science”,ijetemr ,[Volumn 7](http://ijetemr.org/category/volumn-7/) , June 2020
3. Ajay Kumar Pal, Saurabh Pal "Classification Model of Prediction for Placement of Students", I. J. Modern Education and Computer Science, Published Online in MECS(2013).
4. P.Natesan , N.Krishnamoorthy, S.Ponni,R.R. Rajalaxmi, “Regression Model for Predicting Engineering Students Academic Performance”, researchgate, April 2019.
5. T. Jeevalatha,N. Ananthi,D. Saravana Kumar, “Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms”, International Journal of Computer Applications (0975 – 8887),Volume 108 – No 15, December 2014
6. Neelam Naik and SeemaPurohit, “Prediction of Final Result and Placement of Students using Classification Algorithm”, Interna-tional Journal of Computer Applications (0975–8887), Volume56–No.12, October 2012.
7. Shreyas Harinath, Aksha Prasad, Suma H S, Suraksha A and Tojo Mathew, Student placement prediction using machine learning, (IRJET), e-ISSN: 2395-0056, pISSN: 2395-0072, Volume: 06 Issue: 04 | Apr 2019.
8. G.Gautami, Monam Hayat,”Prediction of Student Placement using Machine Learning Algorithms”,IJCRT | Volume 8, Issue 5, May 2020
9. Joel James , Daibin Raju , Jeebu Abraham Aniyankunju,Irene Treesa Jose ,“Placement Prediction using Various Machine Learning Models and their Efficiency Comparison”,International Journal of Innovative Science and Research Technology,Volume 5(May 2020).
10. Pothuganti Manvitha and Neelam Swaroopa, Campus placement prediction using supervised machine learning techniques, International Journal of Applied Engineering Research ISSN Volume 14 (2019).