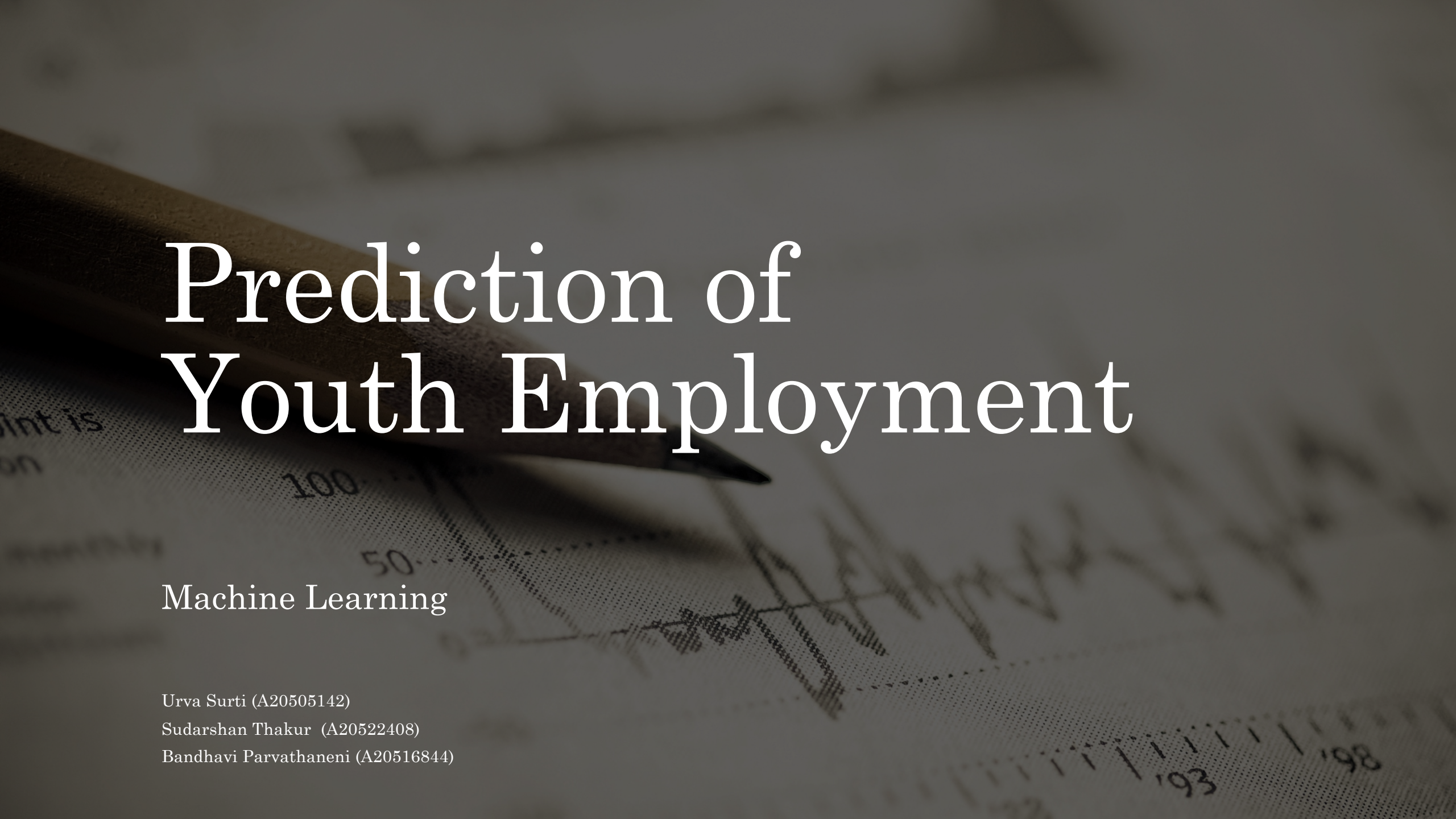


Prediction of Youth Employment

Machine Learning

Urva Surti (A20505142)
Sudarshan Thakur (A20522408)
Bandhavi Parvathaneni (A20516844)

Introduction

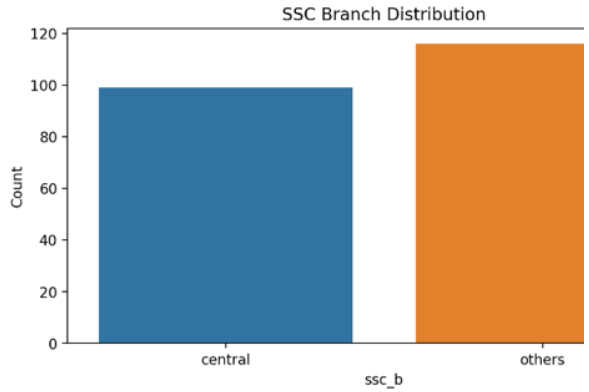
- This project focuses on placement of a college student considering the parameters/features such as Highschool percentage marks, graduation gpa, salary expectation, work experience, stream the student belongs to, the board in which he/she studied etc.
- The project tests various machine learning algorithms like Gaussian Naive Bayes, Random Forest, XGBoost and K Nearest Neighbours. The best performing algorithms were chosen for stacking and creating a new hybrid model. Based on the result the student can decide about how he/she has to go around for the preparation of campus placements.
- With the help of Machine Learning Algorithms and Prediction Modelling, candidates just by providing some data can find out whether they're employable or not and be prepared for the future by enhancing their skills in the domain they wish to be employed into.

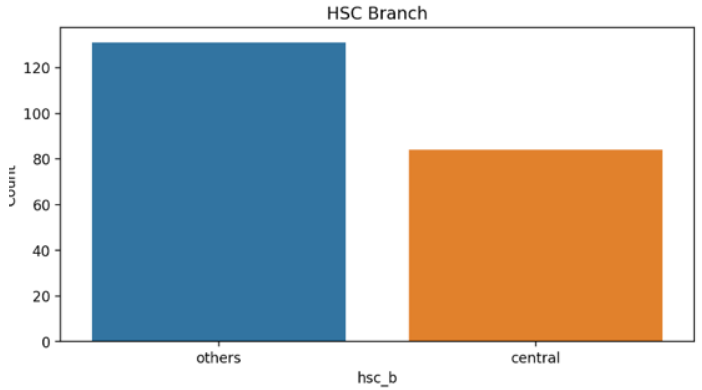
EDA

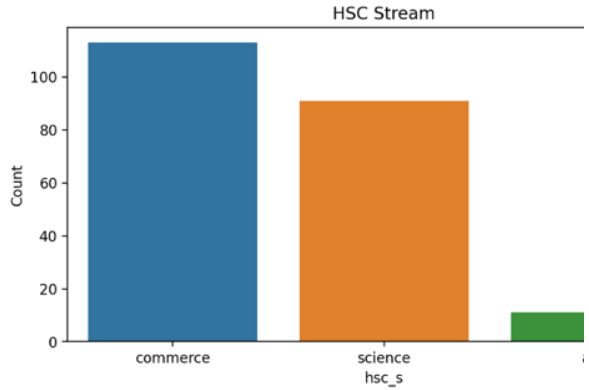
We have used EDA to extract insights from the data. EDA can also be useful in finding different patterns within the data and also help in detecting and dealing with the outliers which are present within the data yet hard to find. The collected data is preprocessed to eliminate/replace the null and missing values from the data. Some inferences from EDA

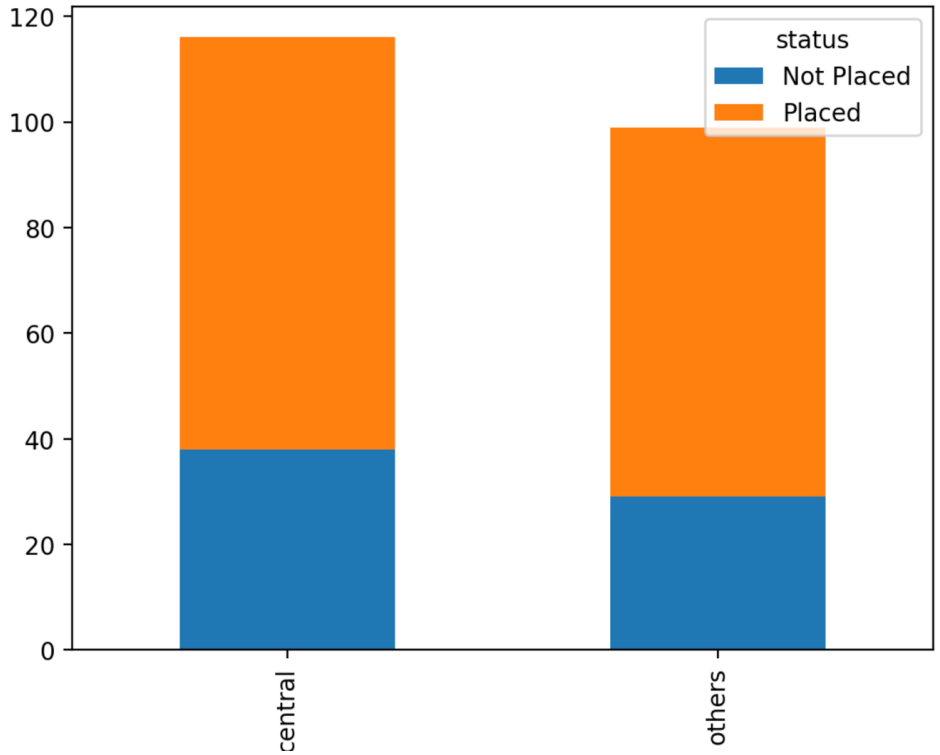
- 1) The students with no work experience have very less placements whereas the students with some work experience are placed
- 2) Most of the placed students belonged to Science and arts.
- 3) Majority of the students have no work experience, whereas very small percentage has some sort of work experience.

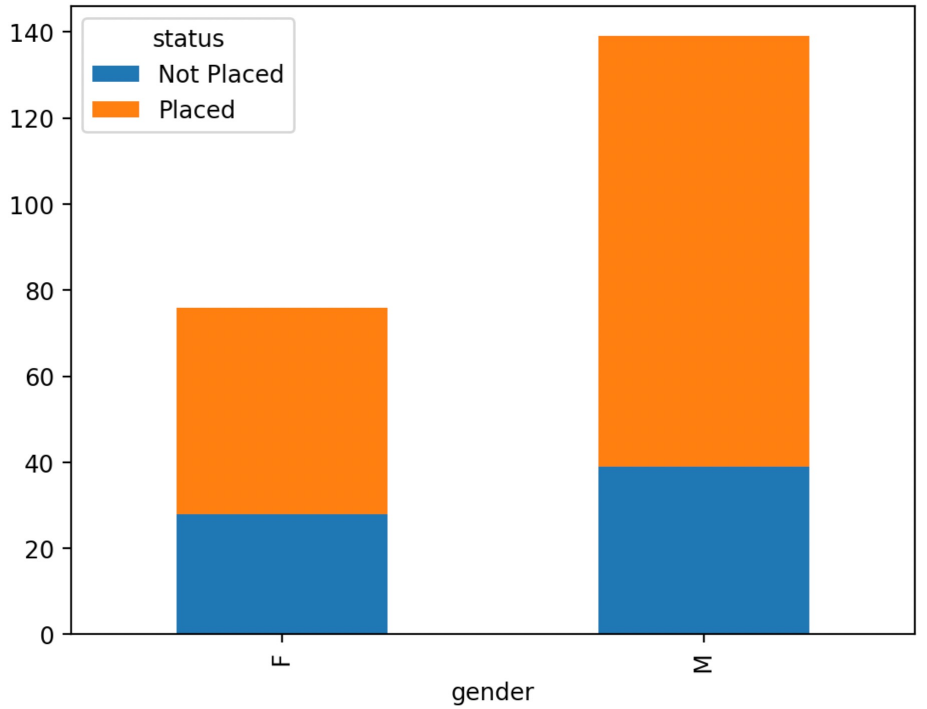
lcome to Exploratory Data Analysis of
ıcement Prediction Project

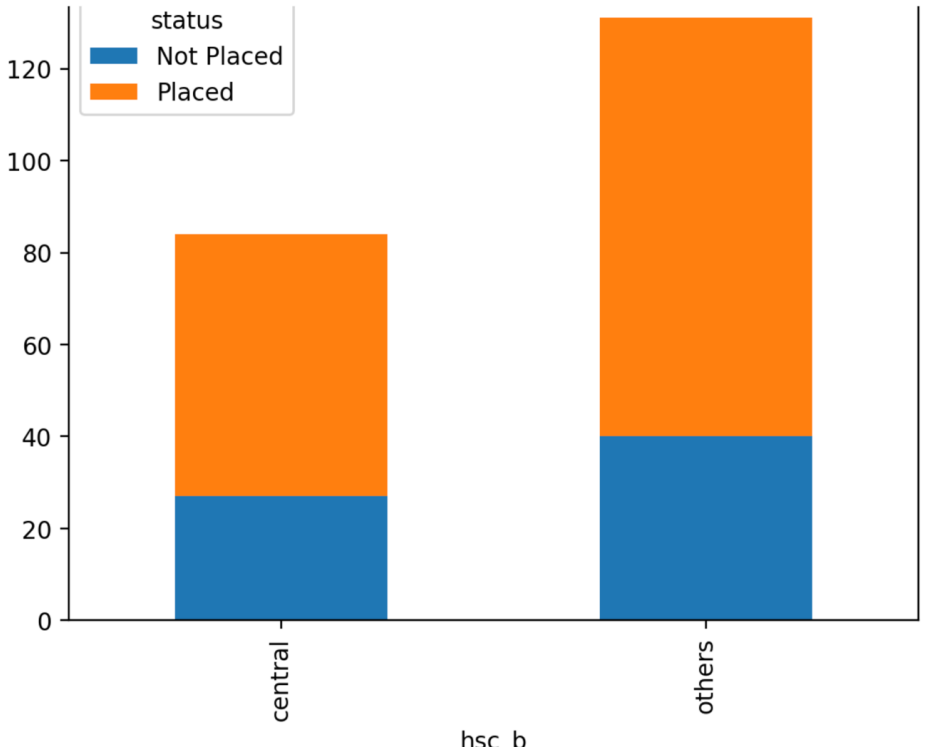


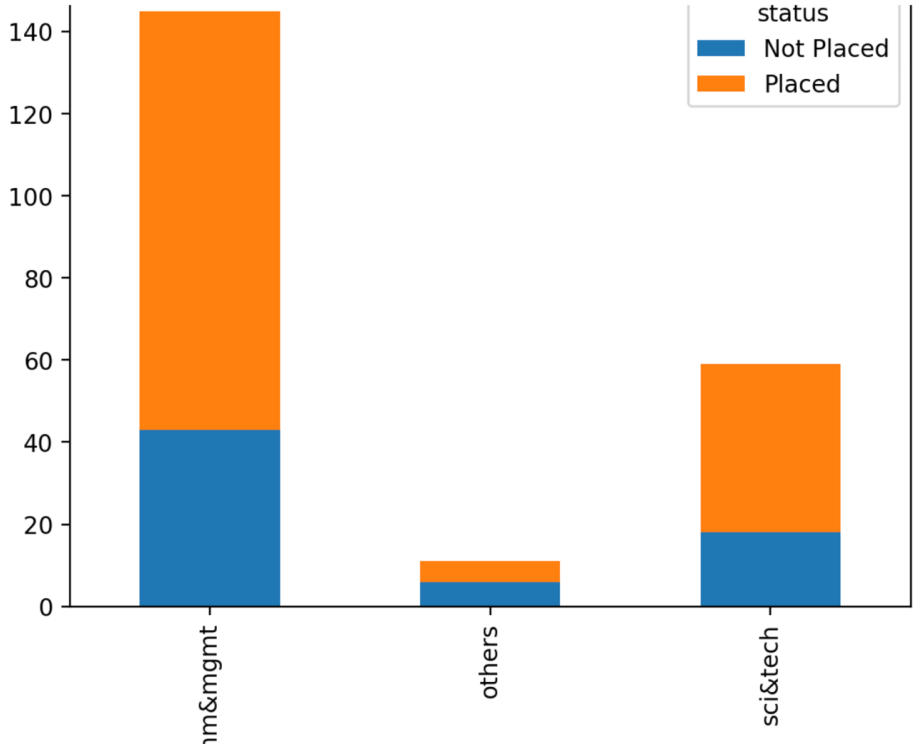


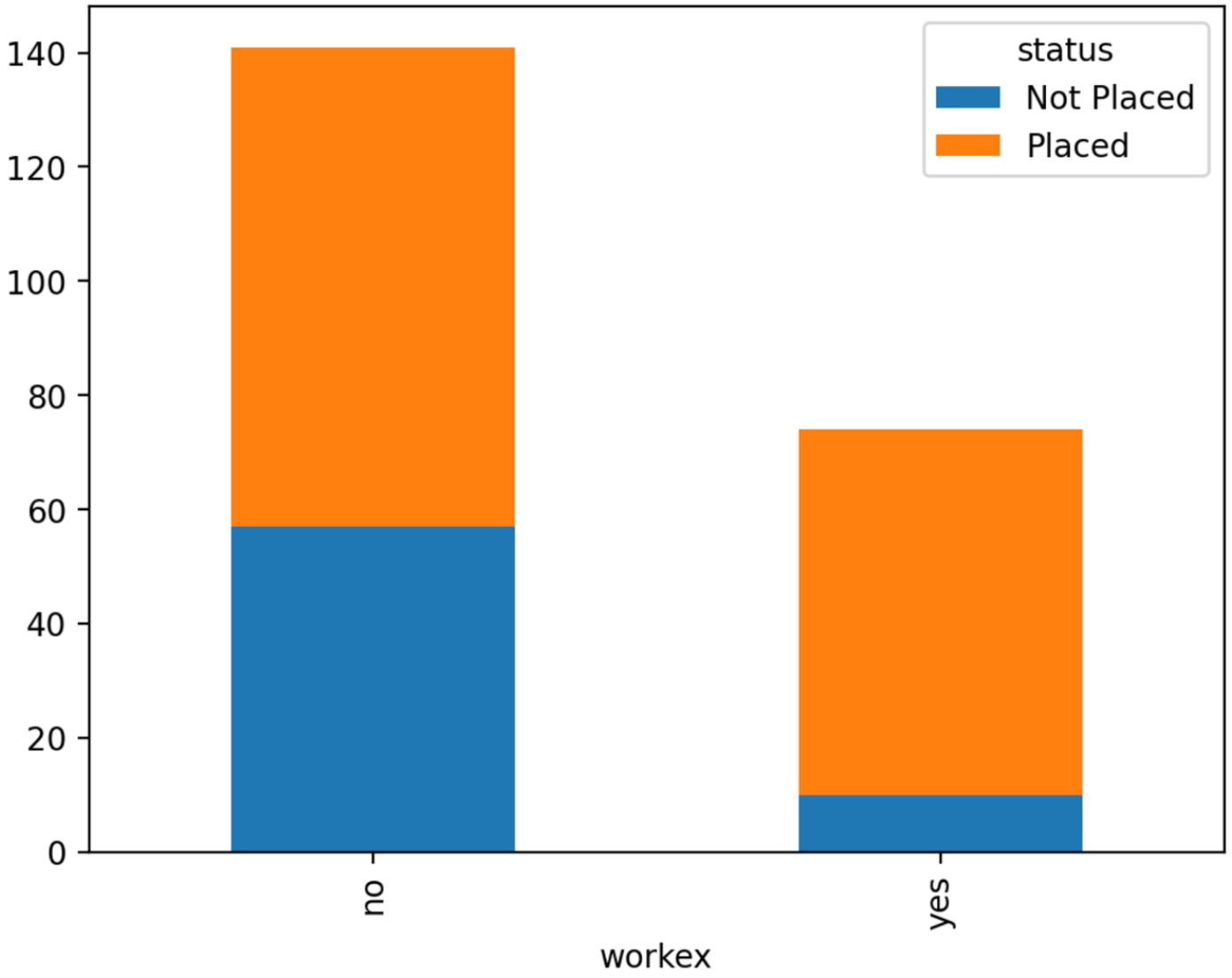



status
Not Placed
Placed
120
100
80
60
40
20
0
central
others

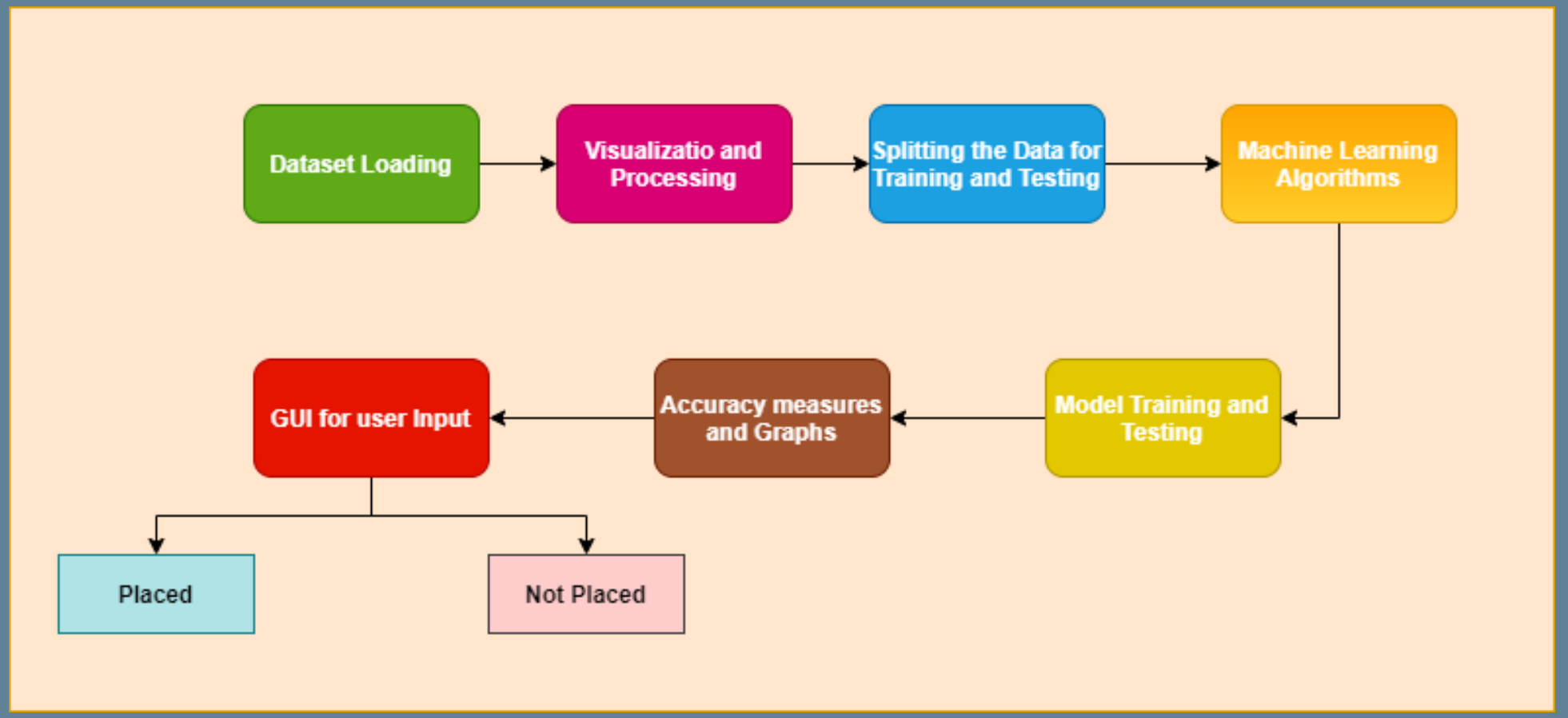

status
Not Placed
Placed
140
120
100
80
60
40
20
0
F
M
gender


status
Not Placed
Placed
120
100
80
60
40
20
0
central
others
hsc_b


status
Not Placed
Placed
140
120
100
80
60
40
20
0
others
sci&tech


status
Not Placed
Placed
140
120
100
80
60
40
20
0
no
yes
workex

System Architecture



DataSet

- The dataset used in the research is from an online source i.e. Kaggle which consists of 15 attributes and 215 student records. The candidates are classified into two categories i.e. placed and not placed

Welcome to Exploratory Data Analysis o Placement Prediction Project

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	s
0	1	M	67.0000	others	91.0000	others	commerce	58.0000	sci&tech	no	55.0000	r
1	2	M	79.3300	central	78.3300	others	science	77.4800	sci&tech	yes	86.5000	r
2	3	M	65.0000	central	68.0000	central	arts	64.0000	comm&mgmt	no	75.0000	r
3	4	M	56.0000	central	52.0000	central	science	52.0000	sci&tech	no	66.0000	r
4	5	M	85.8000	central	73.6000	central	commerce	73.3000	comm&mgmt	no	96.8000	r
5	6	M	55.0000	others	49.8000	others	science	67.2500	sci&tech	yes	55.0000	r
6	7	F	46.0000	others	49.2000	others	commerce	79.0000	comm&mgmt	no	74.2800	r
7	8	M	82.0000	central	64.0000	central	science	66.0000	sci&tech	yes	67.0000	r
8	9	M	73.0000	central	79.0000	central	commerce	72.0000	comm&mgmt	no	91.3400	r
9	10	M	58.0000	central	70.0000	central	commerce	61.0000	comm&mgmt	no	54.0000	r

Methodology

- The parameters/features such as Highschool percentage marks, graduation gpa, salary expectation, work experience, stream the student belongs to, the board in which he/she studied etc are used as an input data to the model.
- The machine learning model inputs the features from the user and gives out a result in the form of placed and unplaced.
- We have tested various machine learning algorithms like Gaussian Naive Bayes, Random Forest, Support Vector Machine and K Nearest Neighbors.
- The best performing algorithms are chosen for stacking and creating a new hybrid model.

Algorithms

We have tried implementing different algorithms and calculating their accuracy and stacking up the algorithms with high number to accuracy to create our model. The algorithms used for model building:

- Gaussian Naïve Bayes
- K-Nearest Neighbours Classifier
- Random Forest Classifier
- XGBoost

The performance evaluation was done for all the above algorithms using metrics like accuracy, confusion matrix, misclassification rate, recall. Out of all the 4 algorithms the XGBoost had the highest accuracy of 89.46%.The lowest accuracy was given by Random forest algorithm which was 83.07%.

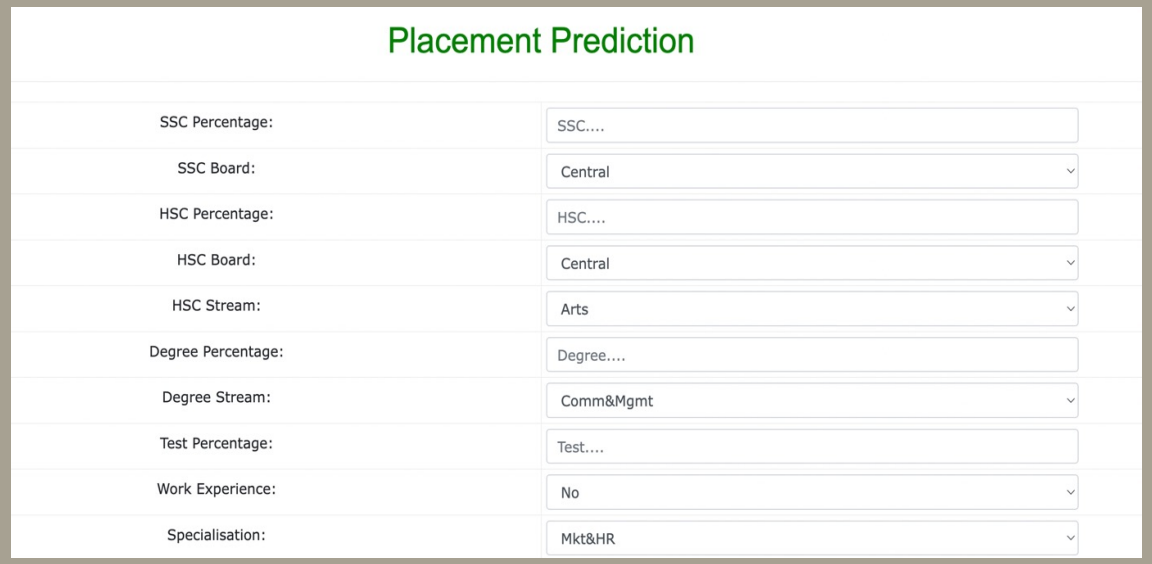
The hybridized algorithm was creating by using the stacking classifier from the MLX tend library and the three algorithms used for stacking were

- XGboost
- KNN
- Gaussian NB

Further this model was tested for performance, and it was observed that the stacked model gave an accuracy of 93.384% which was good enough for deployment.

Previous Work Methods

- Previously various models were deployed using single algorithms such as logistic Regression, Gaussian naïve baye's with accuracy of about 85-90%. We have stacked 3 algorithms and have created a hybrid model for our project.



Screenshots

Conclusion

We have developed a model which can assist students to predict whether they'll be able to get a job or not using various inputs dealing with student's academic record and the field/domain the student comes from. The EDA performed during the making of this project provided our team some important insights within the data itself which will be further used for research purpose. .The hybrid model provides a better performance compared to the work of previous researchers we can say that we have achieved our goal.

Since, Machine Learning is quite a vast field there will be researchers coming up with new algorithms which would provide better results and performance compared to the existing work so this process is an ongoing one

Future Work

- In the future, the researchers must focus more on gathering more diverse features within the data and not just focus on percentage of marks which are more of a qualifying criteria for a particular student to participate in a placement drive for an organization. Having a dataset with more volume can also impact the result in several ways. The data should also focus on students' other achievements and also nonacademic roles/achievements they were a part of which will allow the research to take a new outlook in predicting students’ employment. Hybridization of other algorithms apart from the ones used in this paper can also be done to obtain better results