Report: Identifying Patterns and Trends in Campus Placement Data using Machine Learning

1. Abstract:

Here we have proposed a solution to the aspect of the education landscape in India: campus placements in engineering institutions which are titled **"Identifying Patterns and Trends in Campus Placement Data using Machine Learning"**, which focuses on machine learning techniques to analyze and understand the factors influencing campus placement outcomes for engineering students in India. We proposed a solution that aims to provide valuable insights to the students, which will help in campus placement, and student's academic and professional growth by securing roles in esteemed organizations before degree completion. This project will include data collection, preprocessing, and identification of features and suitable machine-learning algorithms. We are performing a thorough analysis of a self-prepared dataset that will include a wide range of features.

2. Introduction:

2.1. Overview

Every year approximately 30-40 lakh students enroll themselves in Engineering colleges in India. Each and everyone's dream is to be happy and be placed. Very few are there who follow their passion. But here we will talk about those who want to be placed to get rid of

some problems, many students don’t get placed which then later turns into mental problems, family problems, and many more. To solve this problem, we are here with a solution **"Identifying Patterns and Trends in Campus Placement Data using Machine Learning"** which will predict the chances of a student getting a placement in a company or any organization. It will help them improve themselves as they will be able to know about their status and chances of placement. They will understand themselves and what they will need to improve.

2.2. Purpose:

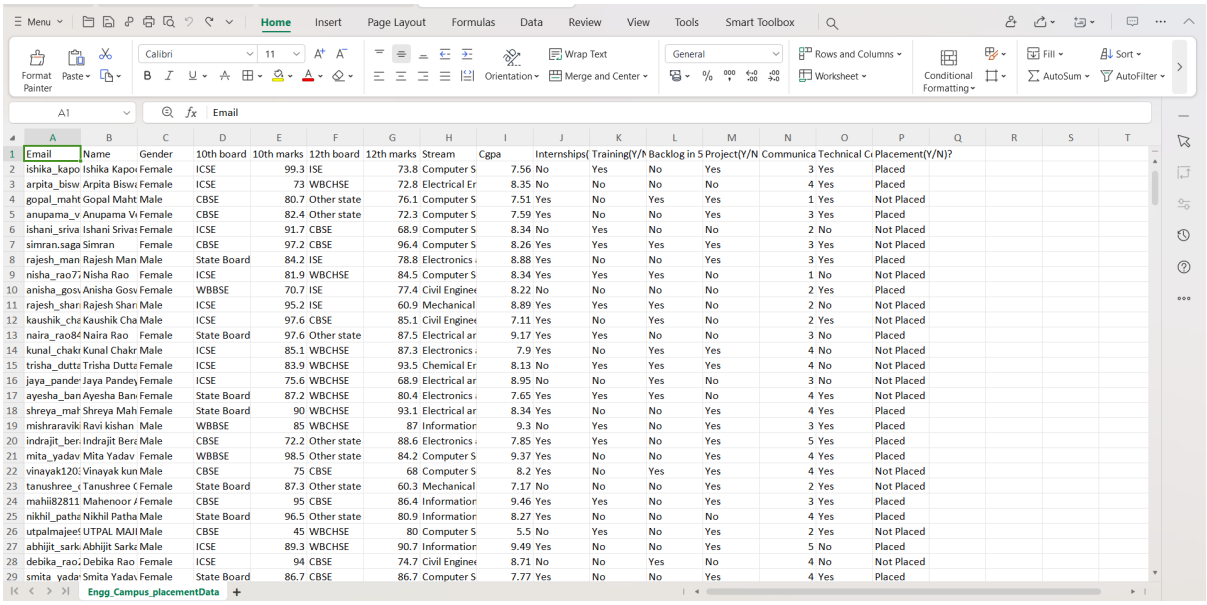
The goal of my project is to enhance the academic and professional growth of students. By accurately forecasting their chances of placement, students will be empowered to take proactive steps toward improvement, thus boosting their overall preparation and confidence. This purpose-driven initiative aims to alleviate common challenges faced by students who seek placement opportunities, including the potential impact on mental well-being and familial dynamics due to unfulfilled career aspirations.

Furthermore, the project's overarching purpose extends to contributing positively to the overall landscape of higher education in India. It seeks to align students' aspirations with actionable insights, fostering a healthier and more productive academic environment. Through rigorous data collection, meticulous preprocessing, and the application of suitable **machine-learning algorithms**, the project intends to equip students with the knowledge they need to navigate the complex landscape of campus placements successfully.

3. Methodology:

3.1. Data Collection:

A comprehensive dataset is collected, containing information about Engineering students through survey Google forms and also collected from my College Training and Placement cell which includes ***10 and 10+2 academic performance, college academic performance, internship experiences, training experiences, any projects or courses completed, and their communication level.*** The dataset comprises both successful and unsuccessful placement cases to ensure a balanced representation. It's a totally new dataset which you will not find in any online websites.



Email	Name	Gender	10th board	10th marks	12th board	12th marks	Stream	Cgpa	Internships(	Training(Y/N	Backlog in 5	Project(Y/N	Communica	Technical C	Placement(Y/N)?
ishika_kapo	Ishika Kapo	Female	ICSE	99.3	ISE	73.8	Computer S	7.56	No	Yes	No	Yes	3	Yes	Placed
arpita_biswa	Arpita Biswa	Female	ICSE	73	WBCHSE	72.8	Electrical Er	8.35	No	No	No	No	4	Yes	Placed
gopal_maht	Gopal Maht	Male	CBSE	80.7	Other state	76.1	Computer S	7.51	Yes	No	No	Yes	1	Yes	Not Placed
anupama_v	Anupama Va	Female	CBSE	82.4	Other state	72.3	Computer S	7.59	Yes	No	No	Yes	3	Yes	Placed
ishani_sriva	Ishani Sriva	Female	ICSE	91.7	CBSE	68.9	Computer S	8.34	No	Yes	No	No	2	No	Not Placed
simran.saga	Simran	Female	CBSE	97.2	CBSE	96.4	Computer S	8.26	Yes	Yes	Yes	Yes	3	Yes	Not Placed
rajesh_man	Rajesh Man	Male	State Board	84.2	ISE	78.8	Electronics	8.88	Yes	No	No	Yes	3	Yes	Placed
nisha_rao77	Nisha Rao	Female	ICSE	81.9	WBCHSE	84.5	Computer S	8.34	Yes	Yes	Yes	No	1	No	Not Placed
anisha_gosv	Anisha Gosv	Female	WBBSE	70.7	ISE	77.4	Civil Enginee	8.22	No	No	No	No	2	Yes	Placed
rajesh_shar	Rajesh Shar	Male	ICSE	95.2	ISE	60.9	Mechanical	8.89	Yes	Yes	Yes	No	2	No	Not Placed
kaushik_cha	Kaushik Cha	Male	ICSE	97.6	CBSE	85.1	Civil Enginee	7.11	Yes	No	Yes	No	2	Yes	Not Placed
naira_rao84	Naira Rao	Female	State Board	97.6	Other state	87.5	Electrical ar	9.17	Yes	Yes	No	No	3	No	Placed
kunal_chakr	Kunal Chakr	Male	ICSE	85.1	WBCHSE	87.3	Electronics	7.9	Yes	No	Yes	Yes	4	No	Not Placed
trisha_dutt	Trisha Dutt	Female	ICSE	83.9	WBCHSE	93.5	Chemical Er	8.13	No	Yes	Yes	Yes	4	No	Not Placed
jaya_pande	Jaya Pande	Female	ICSE	75.6	WBCHSE	68.9	Electrical ar	8.95	No	No	Yes	No	3	No	Not Placed
ayesha_ban	Ayesha Ban	Female	State Board	87.2	WBCHSE	80.4	Electronics	7.65	Yes	Yes	Yes	No	4	Yes	Not Placed
shreya_mal	Shreya Mah	Female	State Board	90	WBCHSE	93.1	Electrical ar	8.34	Yes	No	No	Yes	4	Yes	Placed
mishraravik	Ravi kishan	Male	WBBSE	85	WBCHSE	87	Information	9.3	No	Yes	No	Yes	3	Yes	Placed
indrajit_ber	Indrajit Ber	Male	CBSE	72.2	Other state	88.6	Electronics	7.85	Yes	Yes	No	Yes	5	Yes	Placed
mita_yadav	Mita Yadav	Female	WBBSE	98.5	Other state	84.2	Computer S	9.37	Yes	No	No	Yes	4	Yes	Placed
vinayak120	Vinayak kun	Male	CBSE	75	CBSE	68	Computer S	8.2	Yes	No	Yes	Yes	4	Yes	Not Placed
tanushree_(	Tanushree (	Female	State Board	87.3	Other state	60.3	Mechanical	7.17	No	No	No	Yes	2	Yes	Not Placed
mahii82811	Mahenoor A	Female	CBSE	95	CBSE	86.4	Information	9.46	Yes	No	No	Yes	3	Yes	Placed
nikhil_patha	Nikhil Patha	Male	State Board	96.5	Other state	80.9	Information	8.27	Yes	No	No	No	4	Yes	Placed
utpalmajee	UTPAL MAJI	Male	CBSE	45	WBCHSE	80	Computer S	5.5	No	Yes	No	Yes	2	Yes	Not Placed
abhijit_sark	Abhijit Sark	Male	ICSE	89.3	WBCHSE	90.7	Information	9.49	Yes	No	No	Yes	5	No	Placed
debika_rao	Debika Rao	Female	ICSE	94	CBSE	74.7	Civil Enginee	8.71	No	No	Yes	No	4	No	Not Placed
smita_yada	Smita Yada	Female	State Board	86.7	CBSE	86.7	Computer S	7.77	Yes	No	No	Yes	4	Yes	Placed

3.2. Data Preprocessing:

The collected dataset undergoes preprocessing to ensure its quality and suitability for analysis. This involves cleaning to ensure the data's

quality and accuracy, preventing biased or incorrect model training, handling missing values, and normalizing features will help analyze the features (columns) in the dataset and select the most relevant ones for placement prediction irrelevant or redundant features can lead to overfitting or increased computational costs, removing outliers, and organizing the raw data to prepare it for analysis and model training, transforming categorical features into numerical representations using techniques like **one-hot encoding or label encoding.** Using **MinMaxScaler** to normalize or standardize numerical features to ensure they're on a similar scale, Data splitting divides the dataset into training and testing sets. The training set is used to train the model and the test set evaluates the model's performance, handling imbalanced data If the dataset has an imbalance between the number of placed and non-placed students, consider techniques like oversampling or under sampling to balance the classes, dealing with text data then at last saving pre-processed in a suitable format (e.g., CSV) for future use during model training and evaluation.

3.3. Feature Extraction:

Domain knowledge and Exploratory Data Analysis help in identifying the most relevant features affecting placement outcomes such as *academic performance*, *technical skills*, *communication skills*, *training and internship experiences*, *and any history of backlog*.

3.4. Machine Learning Algorithms:

Several machine learning algorithms are employed to build the placement prediction model. Classification algorithms like Logistic

Regression, Naive Bayes, Adaboost Classifer, Support Vector Classifier, and Gradient Boosting Classifier are experimented with to find the best-performing model.

3.5. Model Evaluation:

The model's performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Cross-validation techniques are applied to ensure the model's robustness and generalization capability.

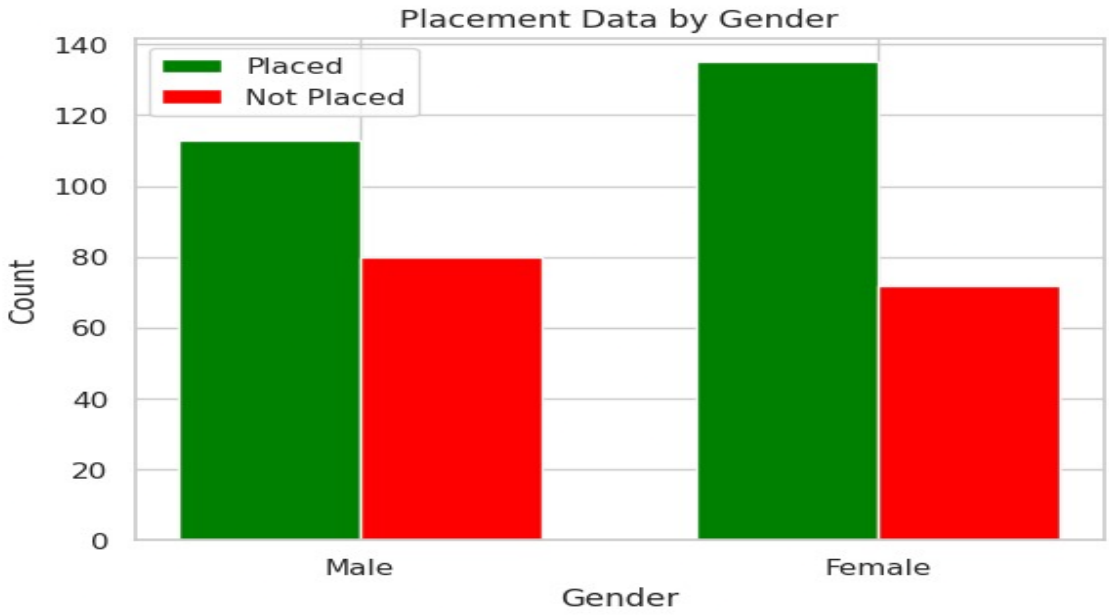
Logistic Regression Classifier

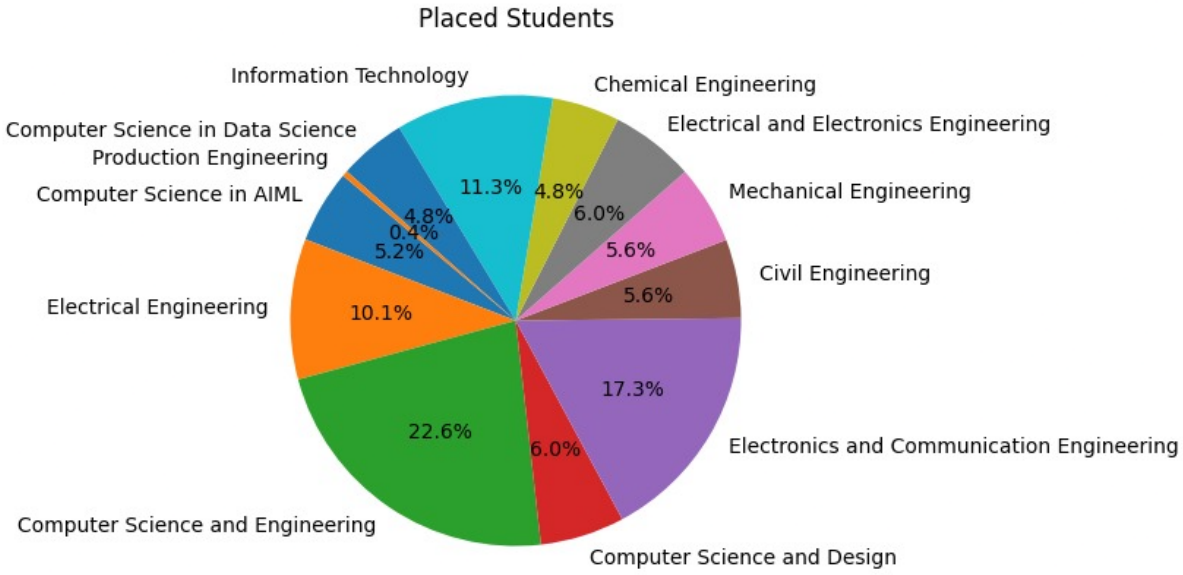
In [18]:
```
from sklearn.linear_model import LogisticRegression

clf = LogisticRegression(random_state=0, solver='lbfgs', max_iter=1000).fit(X_train, y_train)
# printing the acc
y_pred2=clf.predict(X_test)
print(classification_report(y_test,y_pred2))
```

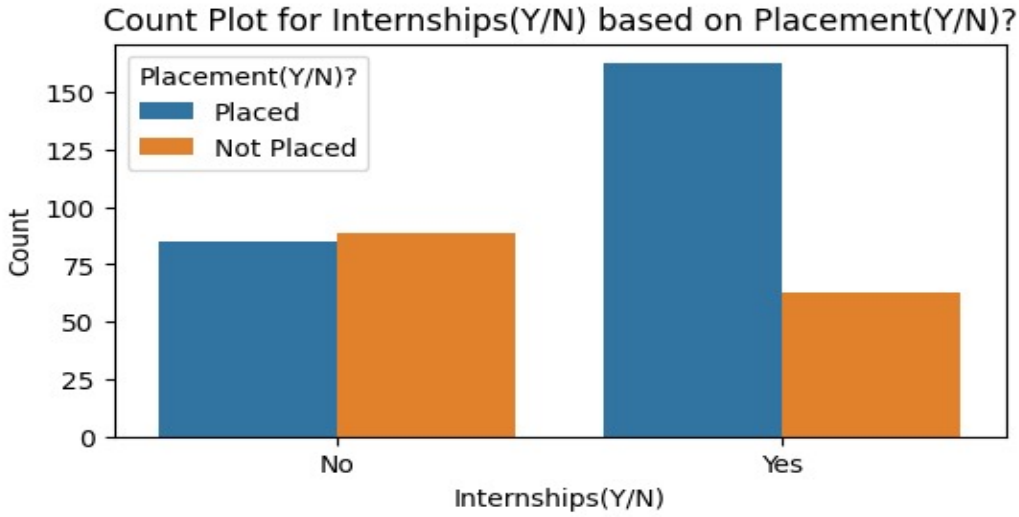
```
              precision    recall  f1-score   support

           0       0.95      0.58      0.72        33
           1       0.77      0.98      0.86        47

    accuracy                           0.81        80
   macro avg       0.86      0.78      0.79        80
weighted avg       0.84      0.81      0.80        80
```

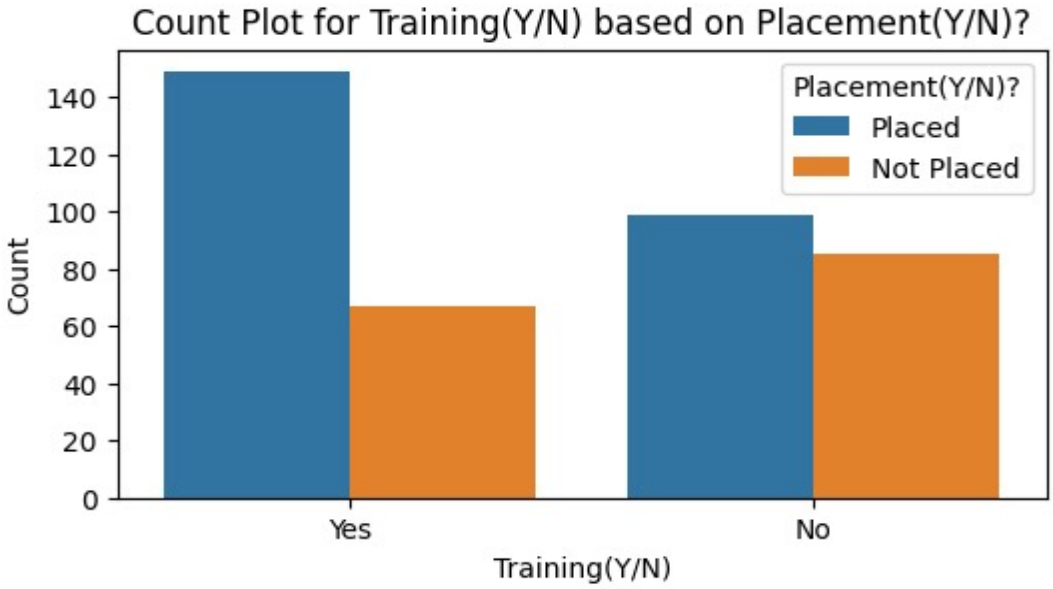
4. Results and Analysis:

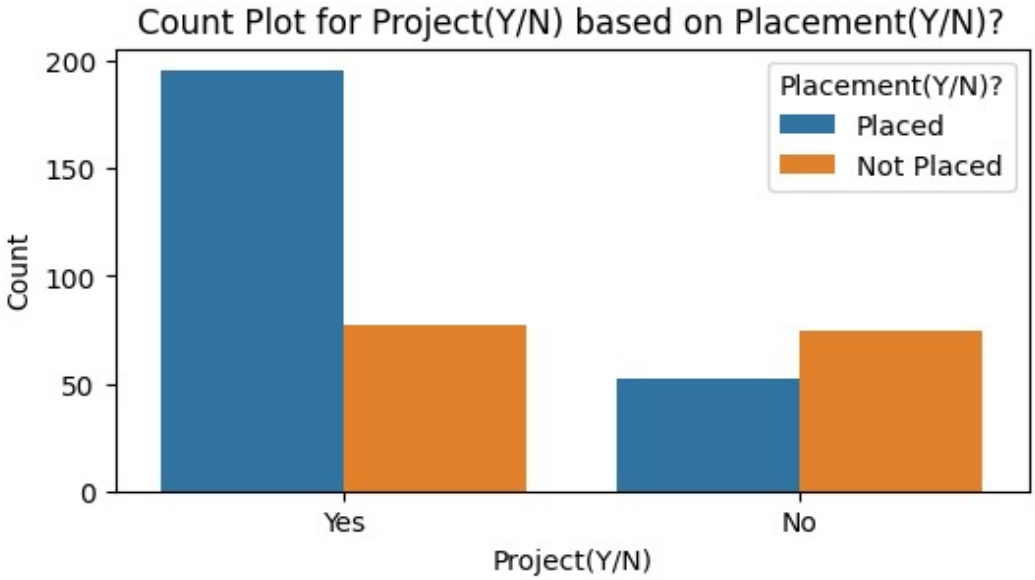
Upon training and evaluating the placement prediction model, we gain insights into the factors that significantly influence placement outcomes. Through feature importance analysis, we identify the relative contribution of each feature to the prediction. This information is valuable for students, educators, and institutions to focus on enhancing those factors.


Placement Data by Gender
Placed
Not Placed
Count
140
120
100
80
60
40
20
0
Male
Female
Gender

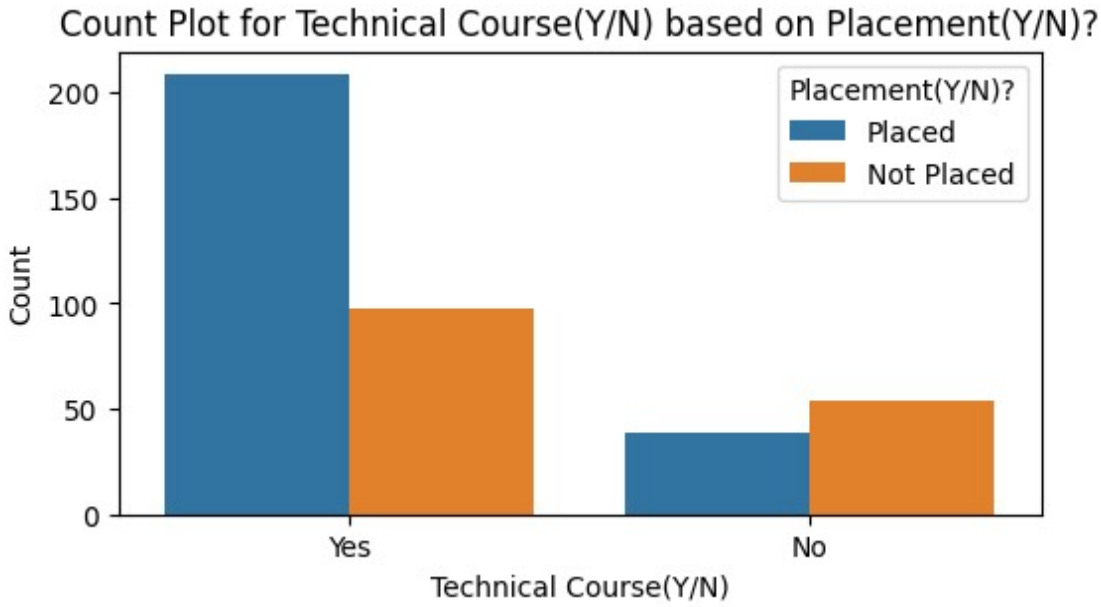

Placed Students
Information Technology
Chemical Engineering
Electrical and Electronics Engineering
Computer Science in Data Science
Production Engineering
Computer Science in AIML
Mechanical Engineering
Civil Engineering
Electrical Engineering
Electronics and Communication Engineering
Computer Science and Engineering
Computer Science and Design
11.3%
4.8%
6.0%
4.8%
0.4%
5.2%
5.6%
5.6%
10.1%
17.3%
22.6%
6.0%


Count Plot for Internships(Y/N) based on Placement(Y/N)?
Placement(Y/N)?
Placed
Not Placed
Count
150
125
100
75
50
25
0
No
Yes
Internships(Y/N)


Count Plot for Training(Y/N) based on Placement(Y/N)?
Placement(Y/N)?
Placed
Not Placed
Count
0
20
40
60
80
100
120
140
Yes
No
Training(Y/N)

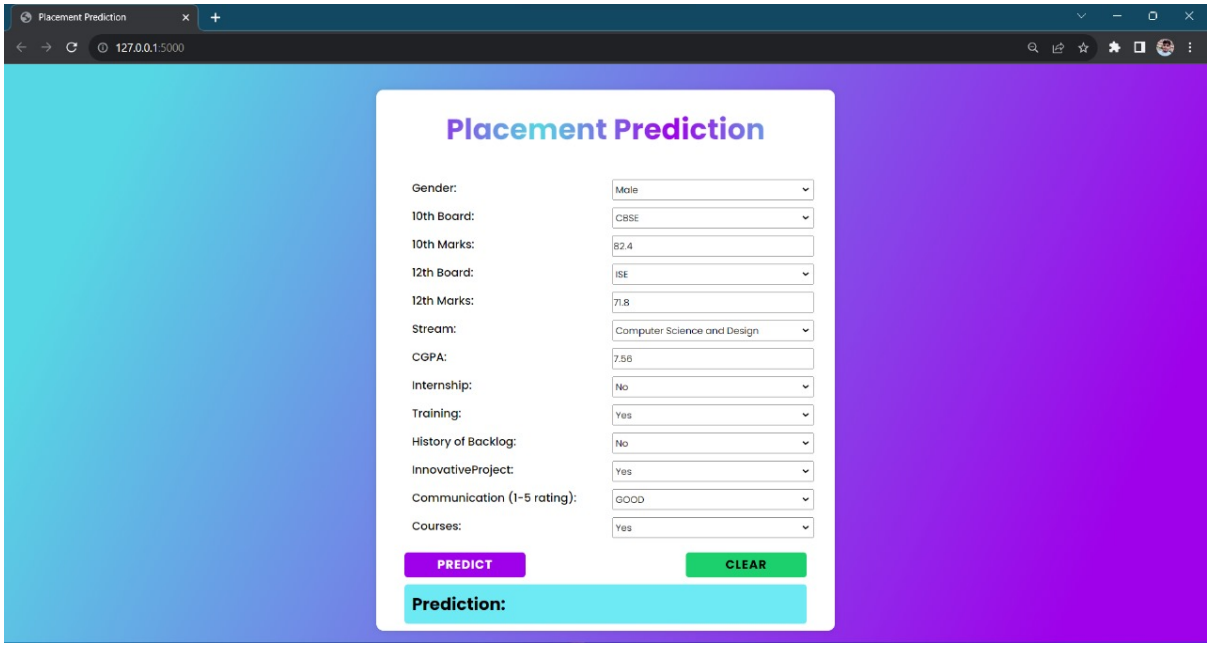

Count Plot for Project(Y/N) based on Placement(Y/N)?
Placement(Y/N)?
Placed
Not Placed
Count
0
50
100
150
200
Yes
No
Project(Y/N)

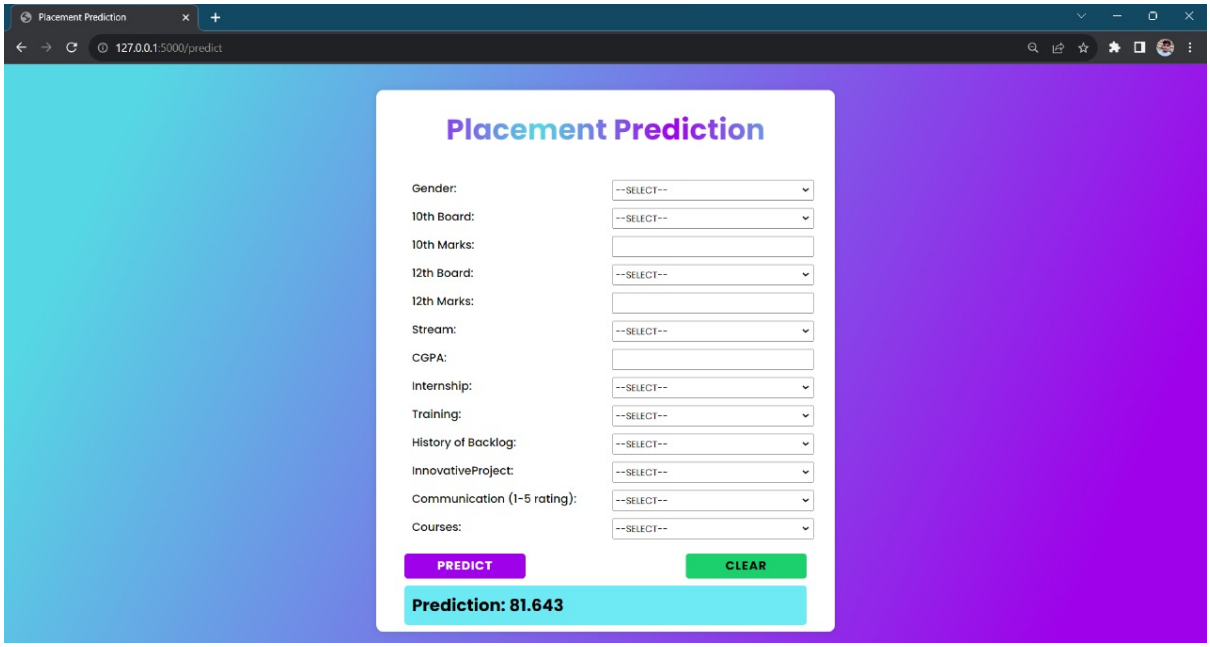
Heat


Count Plot for Technical Course(Y/N) based on Placement(Y/N)?
Placement(Y/N)?
Placed
Not Placed
Count
0
50
100
150
200
Yes
No
Technical Course(Y/N)

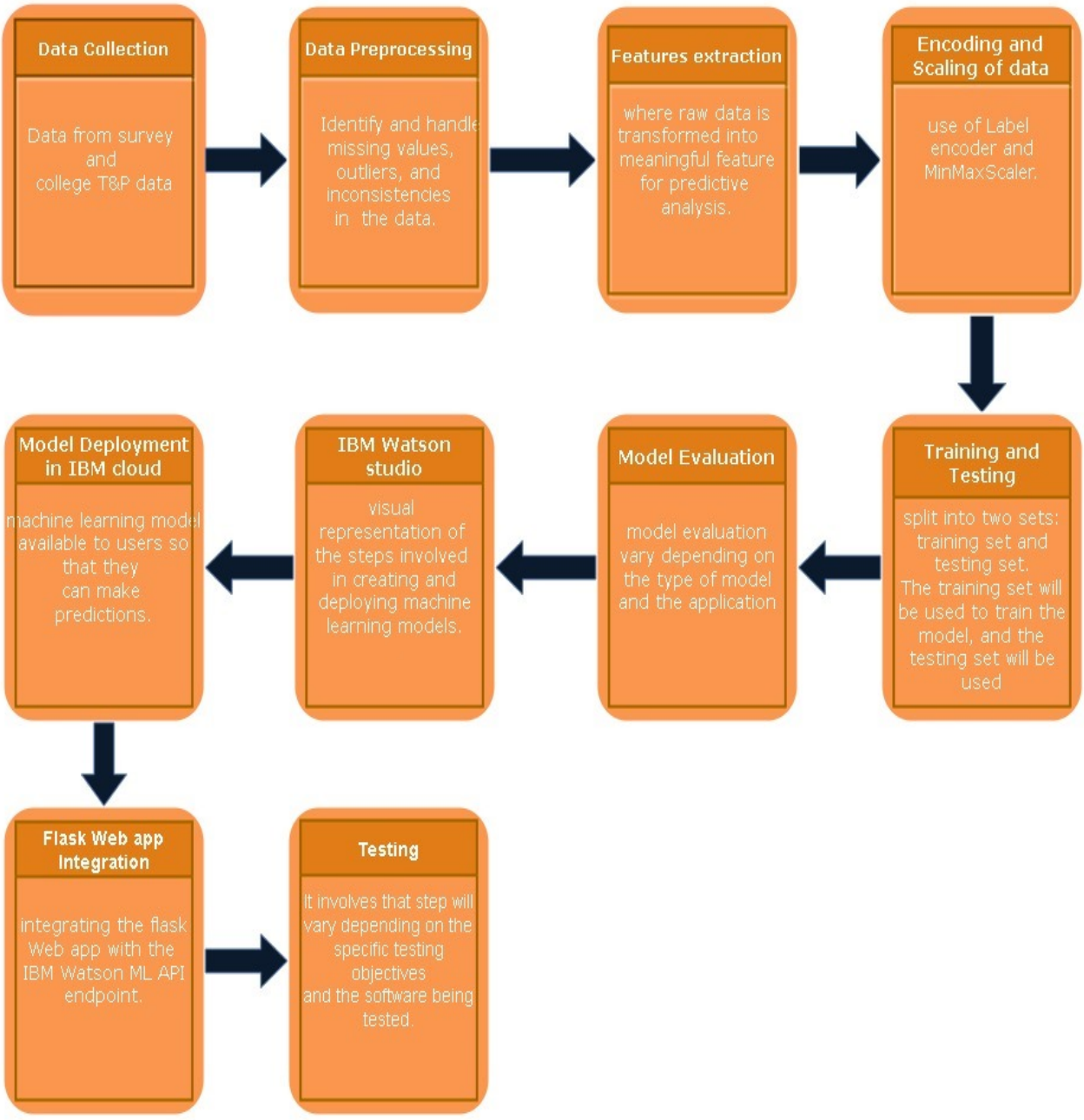
Team StrawHat

Web Implementation of ML model:



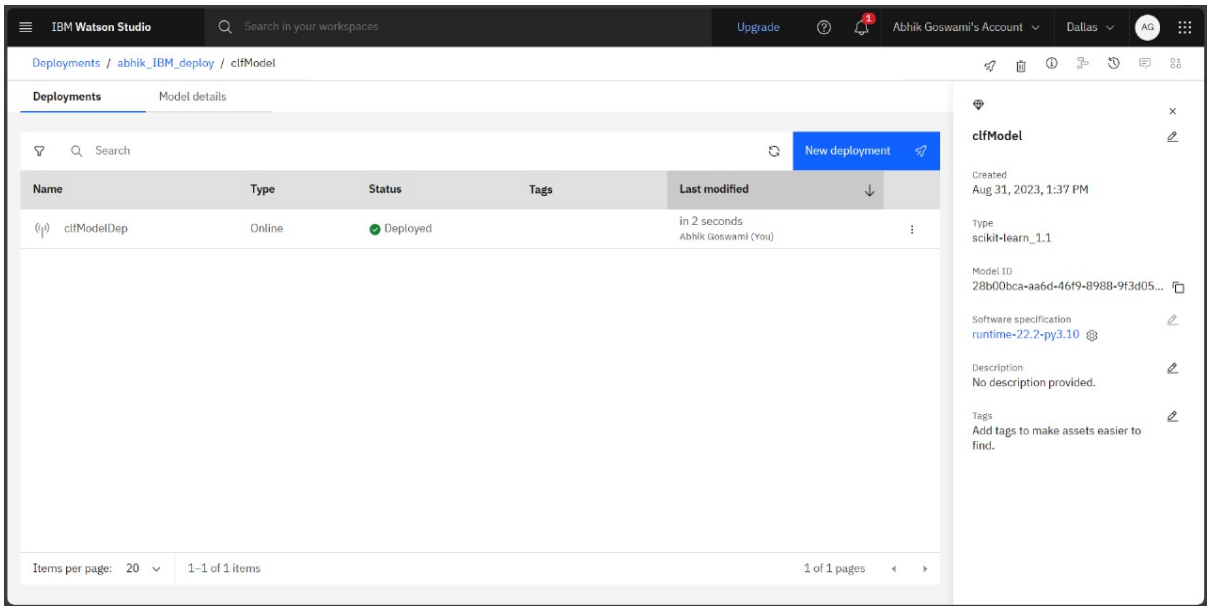


Processes involved in this Project:



Architectural Flow of Identifying the patterns and trends in engineering campus placement data using machine learning

IBM cloud Deployment of ML model using IBM watson:



5. Future Enhancements:

The project lays the foundation for future enhancements and applications, including: - Integration of real-time data to keep the model up-to-date. - Incorporating natural language processing (NLP) techniques for analyzing student resumes and interview performance. - Collaboration with institutions and organizations to provide personalized recommendations for students. - Expanding the scope to other disciplines and industries for broader applicability.

6. Conclusion:

The proposed solution of utilizing machine learning to analyze and predict campus placement outcomes has the potential to greatly benefit engineering students in India. By understanding the underlying patterns and trends, students can make informed decisions to improve their academic performance, acquire relevant skills, and increase their chances of successful placements. This project contributes to the alignment of educational institutions with industry demands, leading to improved placement rates and students' professional growth.