# *Identifying Patterns and Trends in Campus Placement Data using Machine Learning*

### Contents:

### 1.Introduction

### 1.1 Overview

### 1.2 Purpose

### 2.Literature Survey

### 2.1 Existing problem

### 2.2 Proposed solution

### 3.Theoritical Analysis

### 3.1 Block diagram

### 3.2 Hardware/Software designing

### 4. Experimental Investigations

### 5. Flowchart

### 6. Result

### 7. Advantages & Disadvantages

### 8. Applications

### 9. Conclusion

### 10. Future scope

### 11. Bibilography

**Introduction:**

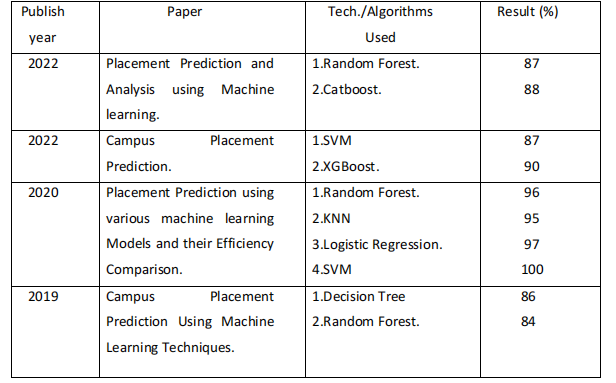
**Overview:**

The placements of students are considered to be the main objective of each institution. The basic success of college is measured by the placements of students. It is important for the student as well institutions to an early idea of the current state of skillset of students for future placement. Educational institutions are always working for the placement of student by introducing new courses and skills. They will get to know that on which parameters they have to focus more. Sometimes it is difficult for the student to prepared for the placement with all the skills. This project will help student to focus on the skillset only required for the placement. The analytical study of the skills of student, both technical and soft skills give idea about the student will be placed or not. This model will give idea about the skill to be prepared for the placement. The proposed model predicts whether the student will place or not. It uses technical and soft skills. In this project the previous year student’s data is used to predict the chance of student placement. This classification model is based on the decision tree classification algorithm. This classification model classifies the student into the placed or not placed category. This classification model will help the student to check their progress easily from time to time. Parameters used for the prediction are the academic score, internship done or not, total number of backlogs etc. The proposed model is also compared with the other classification model with their accuracy.

**Purpose:**

To study the machine learning algorithms and implementation of same for the prediction of the student placement details. The model built using random forest algorithm able to predict the placement of student based on previous year student data. Help students to improve their skillset from time to time. Compare Built model’s accuracy with already existing models for placement prediction to understand efficiency of the algorithm used.

**Literature Survey:**



**Proposed Solution:**

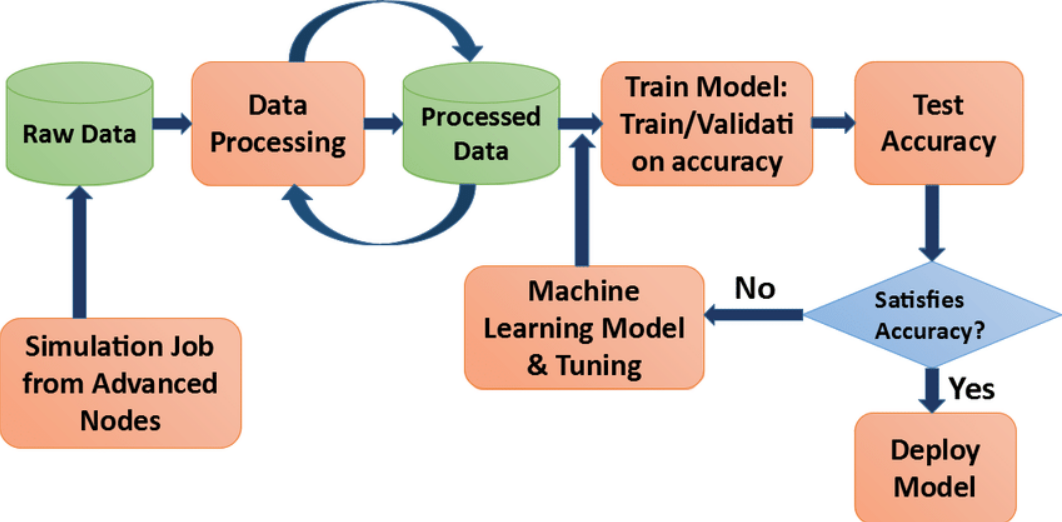
1. In order to anticipate placement, information is gathered from a variety of sources, including academic transcripts, resumes, and prior job experience.
2. After that, this data is cleaned and pre-processed to remove any discrepancies or mistakes.
3. After being cleaned, the data is divided into two categories: training data and testing data.
4. The machine learning algorithm is trained using the training data, and its effectiveness is assessed using the testing data. The system is taught using a variety of methods, including neural networks, decision trees, random forest and regression analysis.

A statistical method for determining the relationship between two or more variables is regression analysis. Regression analysis is used in placement prediction to determine the link between numerous variables, including academic achievement, skill set, prior job experience, and the likelihood of being hired by a firm.

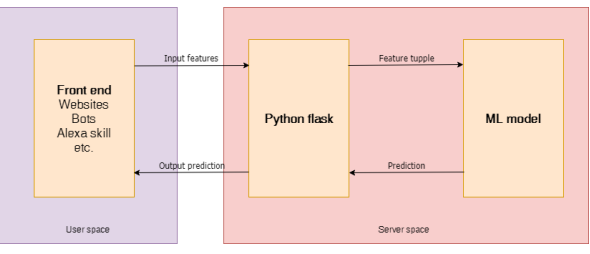
A sort of machine learning algorithm known as a random forest models decisions and potential outcomes using a tree-like structure. Random Forest are employed in the placement prediction scenario to simulate the hiring process decision-making of businesses.

**Theoritical Analysis:**

**3.1 Block diagram:**



**3.2 Hardware / Software designing:**

Detail Architecture

Frontend gives the input features from the user to the server. Frontend is designed using Simple HTML and CSS. HTML defines the meaning and structure of web content. HTML makes the structure of the website. HTML is often accompanied by CSS. CSS is used for styling the web pages. CSS describes how elements should be rendered on screen, on paper, in speech, or on other media. HTML and CSS together form the visual representation of the web page that is visible to the user. The input request will be given to the trained ML model which is deployed on Python’s Flask framework. Flask is a lightweight web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. Making RESTful APIs by Flask is easy.

CSS – Cascading Style Sheets.

HTML- Hyper Text Markup Language.

ML- Machine Learning.

REST- Representational State Transfer.

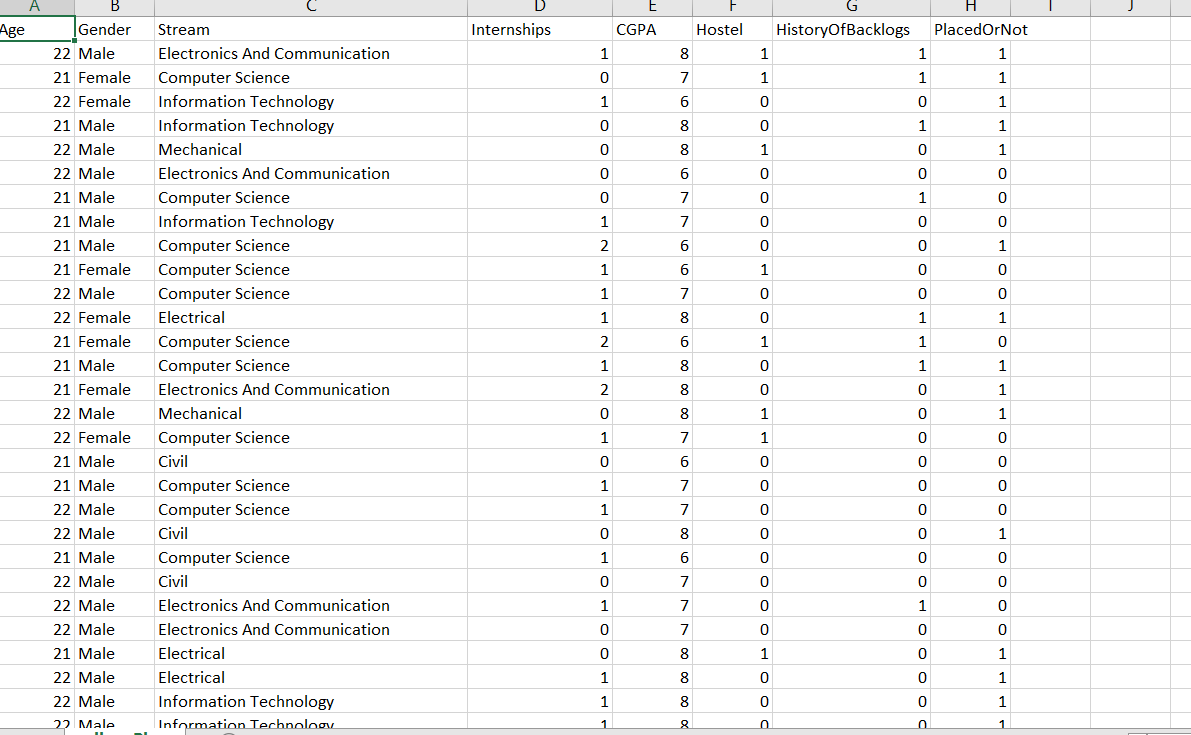
API- Application Programming Interface.

**Experimental Investigations:**

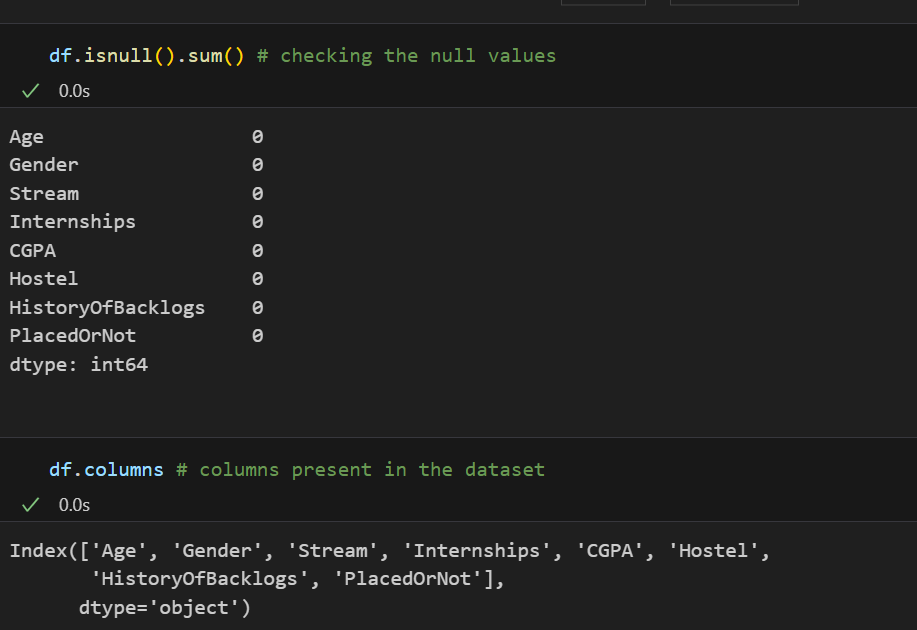
1. **Dataset Description:**

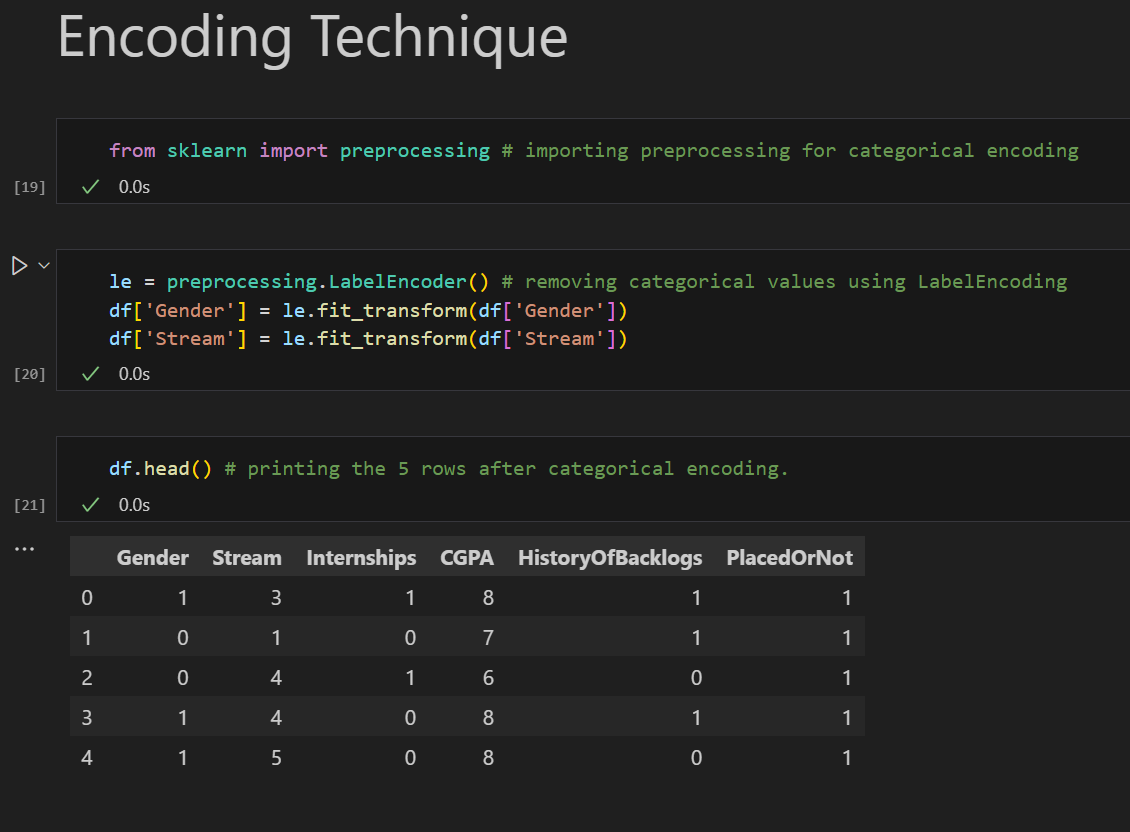
Engineering Placement Prediction Dataset was downloaded from Kaggle. The dataset consists of 2967 datapoints. Dataset has following attributes: Age, Gender,

Stream (i.e., I.T., CSE, Mechanical, Civil etc.), number of internships done, CGPA, whether lives in hostel or not, Total backlogs (Active as well as Passive), Placed or not. Placed or not attribute has just two values 0 or 1 where 0 indicating that candidate is not placed while 1 indicating that the candidate is placed.

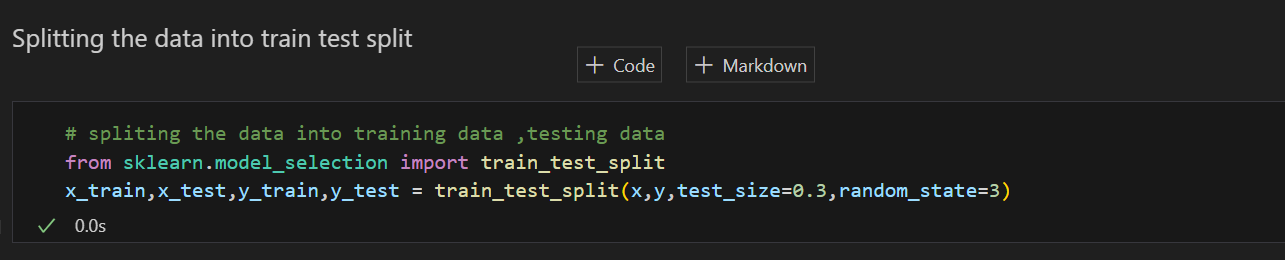


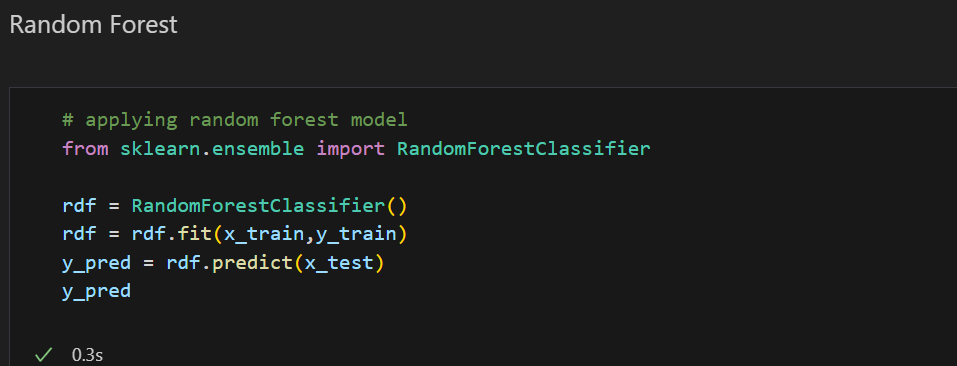
1. **Detail Phases:**
2. Data preprocessing: It includes finding null values in dataset, choosing important features for training and label encoding the dataset. In some cases, the dataset contains missing values. A common plan to handle the matter is to require a mean of all the values of the same column and have it to replace the missing data.

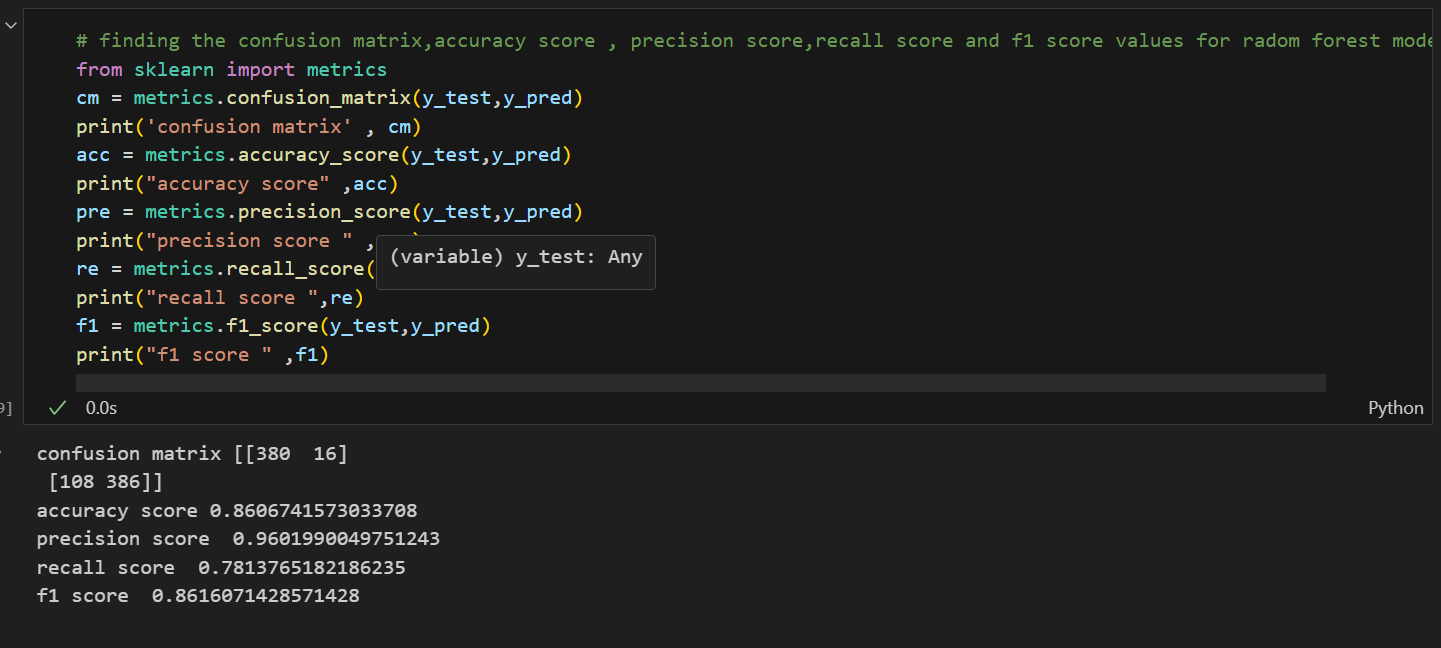




1. Training and testing dataset and checking its performance: Now the next step is to split our dataset into two. Training set and a Test set. We will train our machine learning models on our training set, i.e., our machine learning models will try to understand any correlations in our training set and then we will test the models on our test set to examine how accurately it will predict.







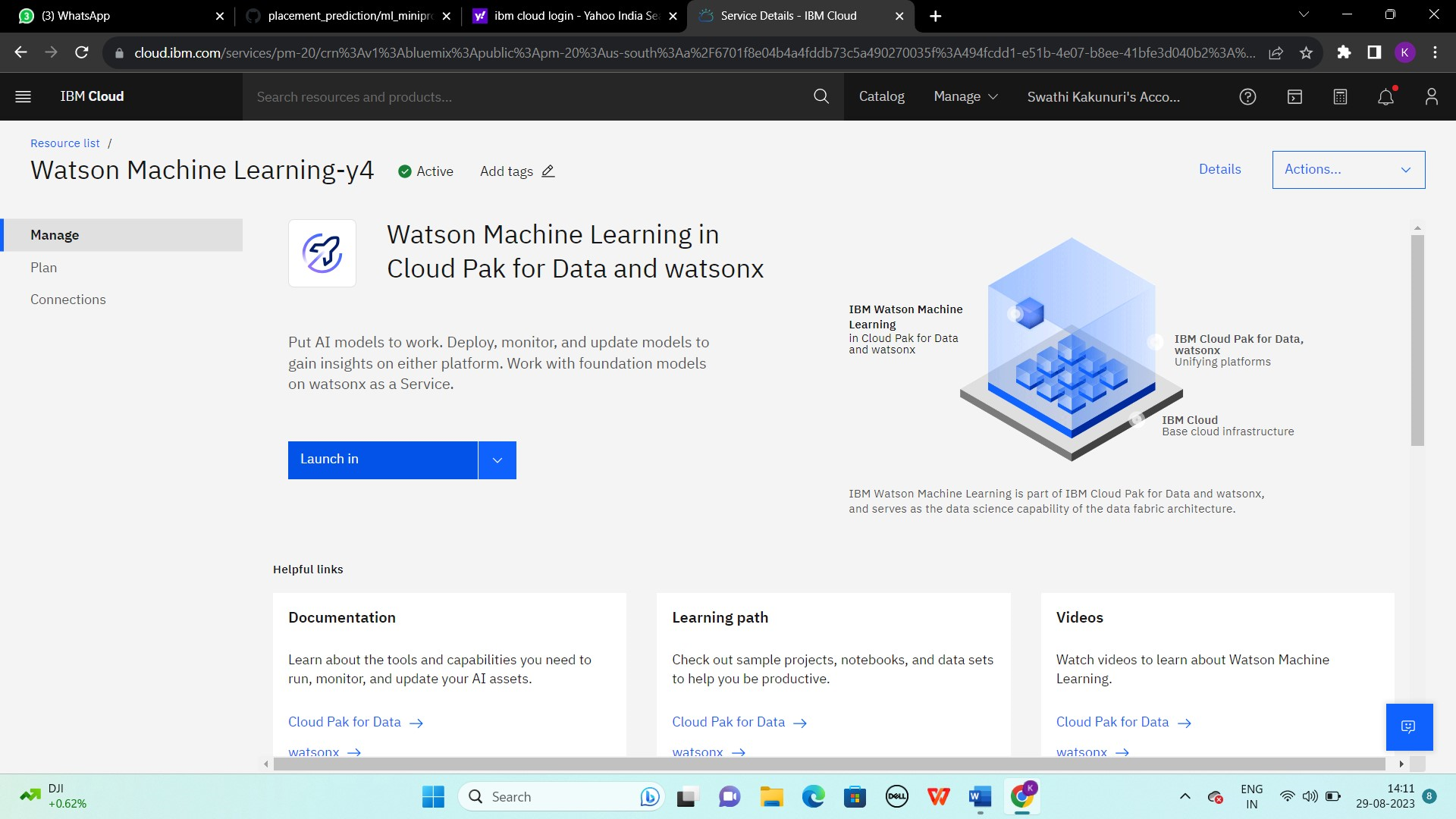
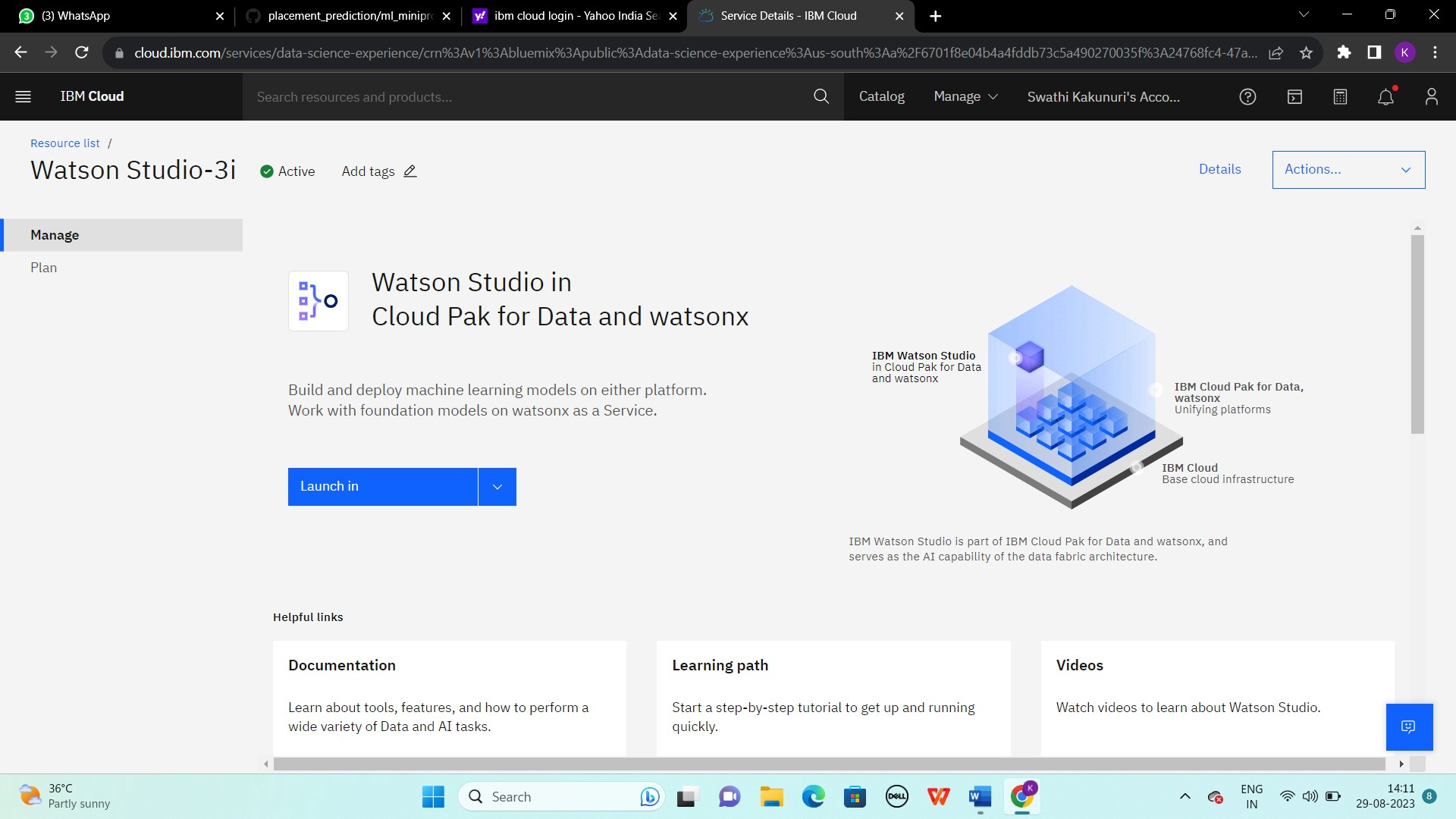
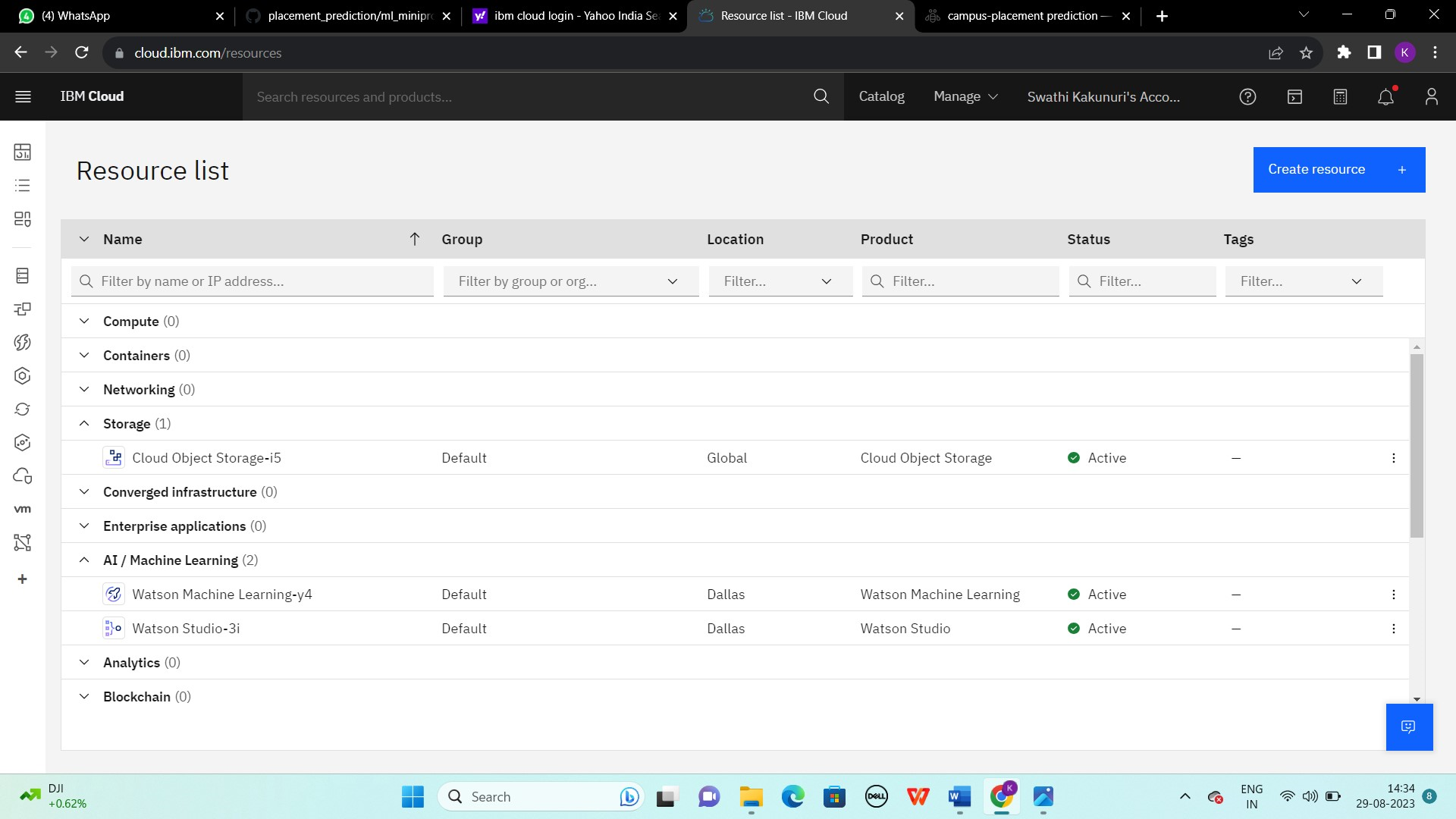
3)Deploying ML model on Flask: ML model which we have prepared will be deployed on Flask.

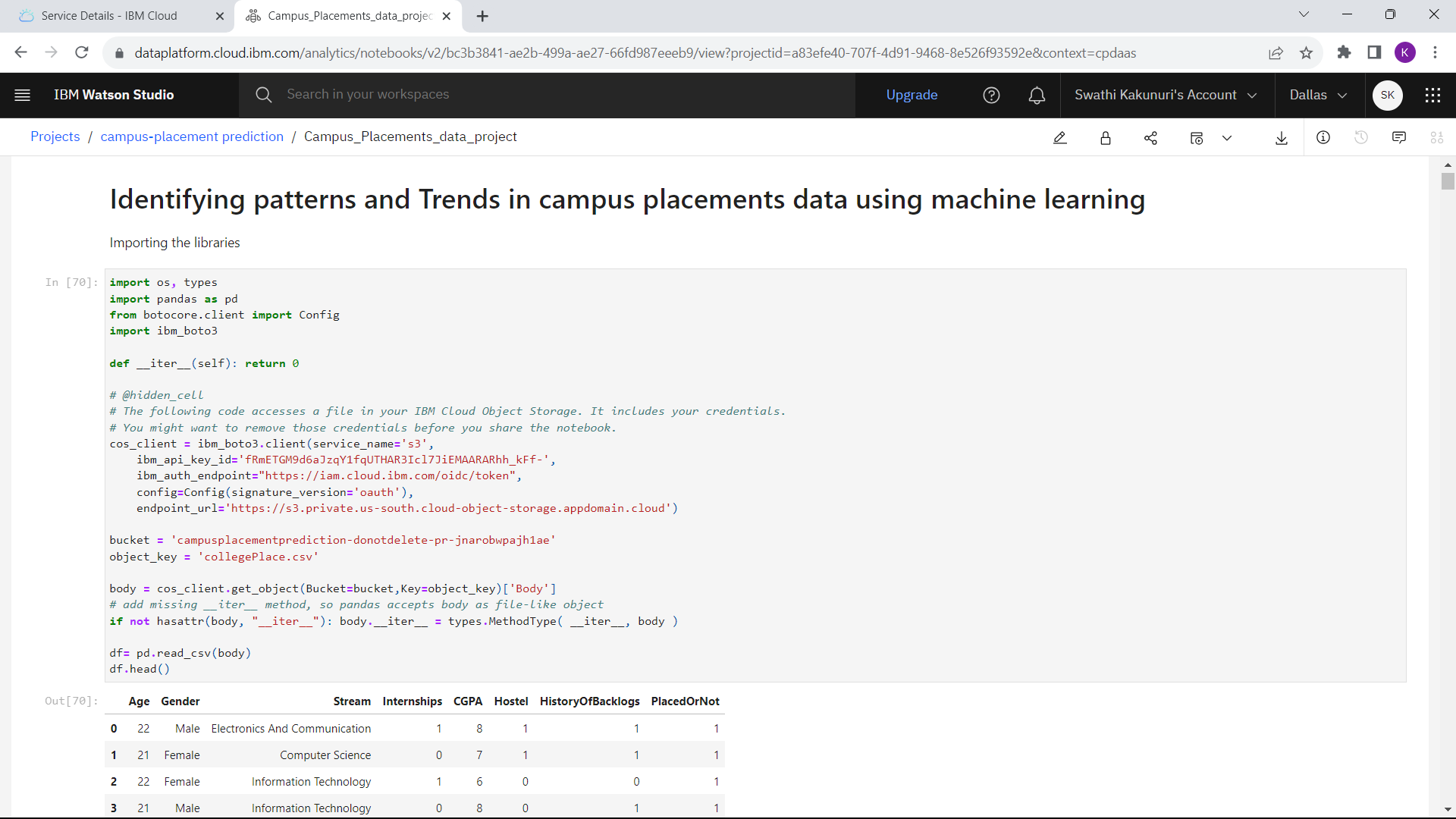
Flask is a Python-based micro framework used for developing small-scale websites. Flask is used to handle API requests.



**Deployment into the Cloud:**

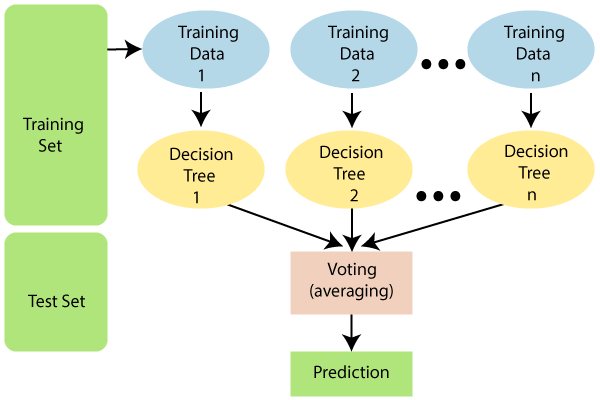
**Tools used in the cloud :**

* **Watson Machine Learning**
* **Watson Studio**
* **Cloud Object Storage**
* 



**Algorithms:**

Random Forest algorithm can be used to train the model on Placement prediction dataset. Random Forest is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.



**Flow Chart:**

**1. Data Collection and Preprocessing:**

Gather historical data on student profiles, academic records, skills, internships, and placement outcomes.

Clean the data by handling missing values and outliers.

**2. Feature Selection and Engineering:**

Select relevant features such as gender, academic scores, skills, and internships that could influence placement outcomes.

Engineer new features like a "total\_score" by combining academic scores and skill assessments.

**3. Data Splitting:**

Split the dataset into training (80%) and testing (20%) subsets.

**4. Random Forest Model:**

Choose the Random Forest algorithm as your baseline model due to its interpretability and ability to handle non-linear relationships.

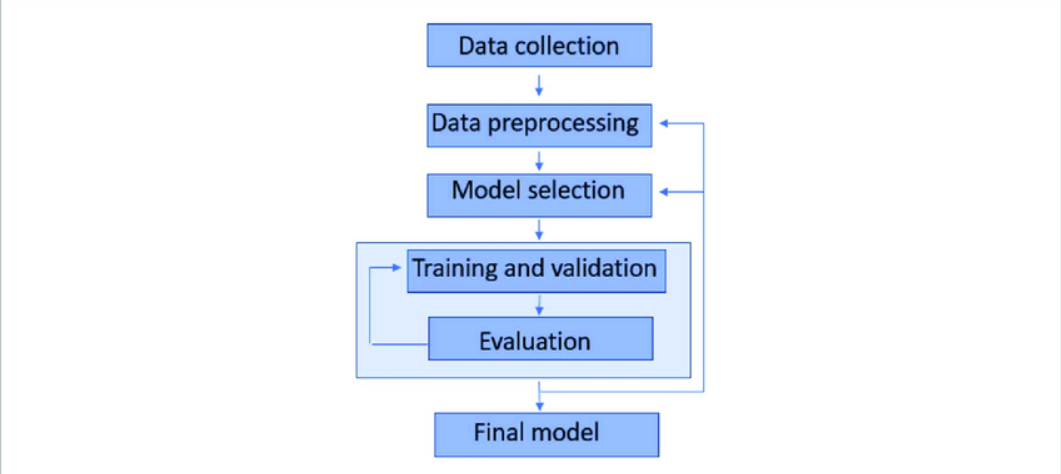
**5. Model Training:**

Train the Random Forest model on the training data.

**6. Model Evaluation:**

Evaluate the model's performance on the testing data using metrics like accuracy, precision and confusion matrix.

Interpret the random forest structure to understand the feature importance and decision-making process.



**Result:**

**Phase-wise Results:**

**Phase 1:** The dataset didn’t contain any null or missing values. Also, the attributes ‘Hostel’ and ‘Age’ didn’t were not important for training the dataset, hence these attributes were deleted from the dataset. Label encoding was applied on attributes ‘Stream’ and ‘Gender’ to convert values in these attributes to it.

**Phase 2**: The dataset was divided into training and testing dataset in the 70:20 proportion. 70% in training and 30% in testing. The model showed 85.7% accuracy, 96.6% precision, 76.9% recall and 85.7% F1 score.

**Phase 3:** Flask was installed and build the frontend part of this project was prepared as well as API requests were prepared. At the end, website is ready.

**Explanation with example:**

**Test 1:**

The website works based on the information provided by the user. This input will be given to trained model to predict the output. For example, user provides these inputs: Gender is male, Stream is CSE, Previously done 1 internship, CGPA is 8 and no backlog. The trained model will predict based on input and it will be shown to user as :

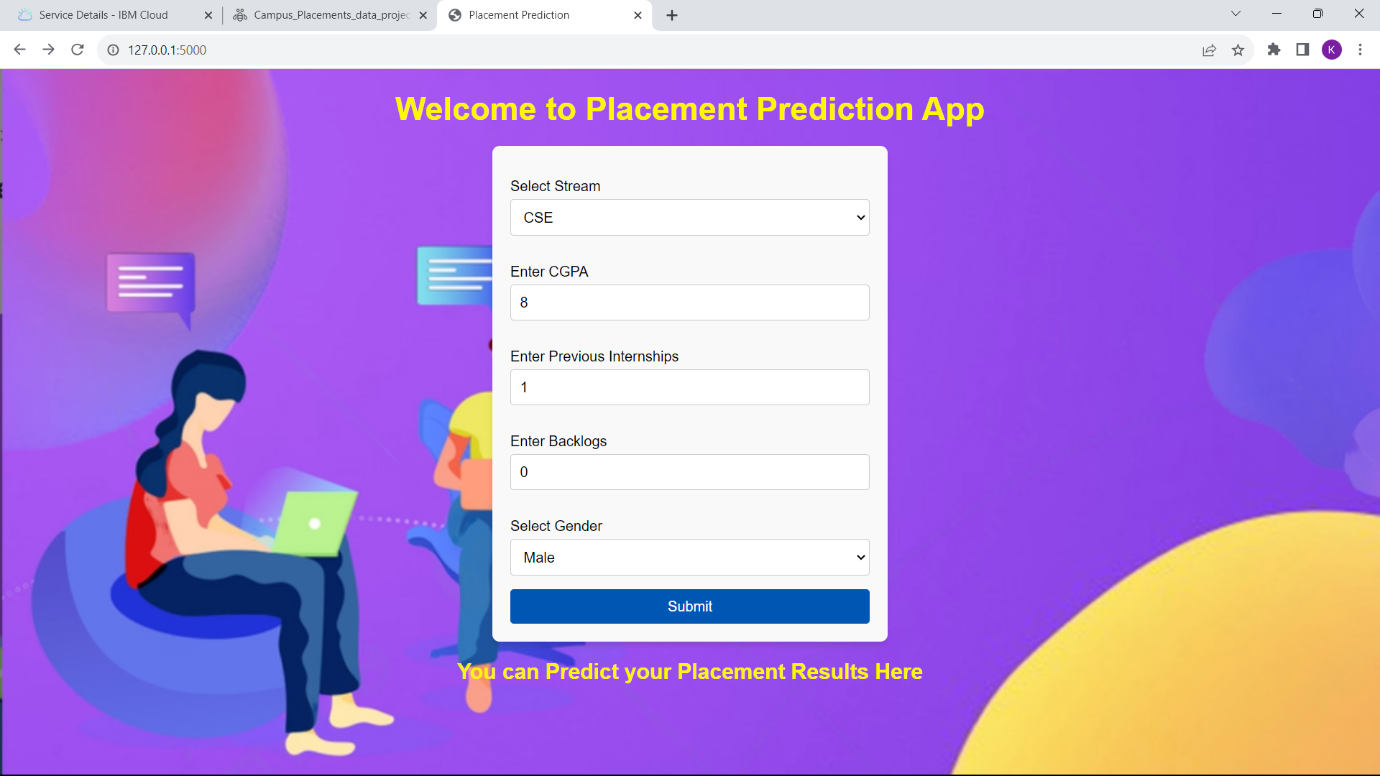
Gender : Male

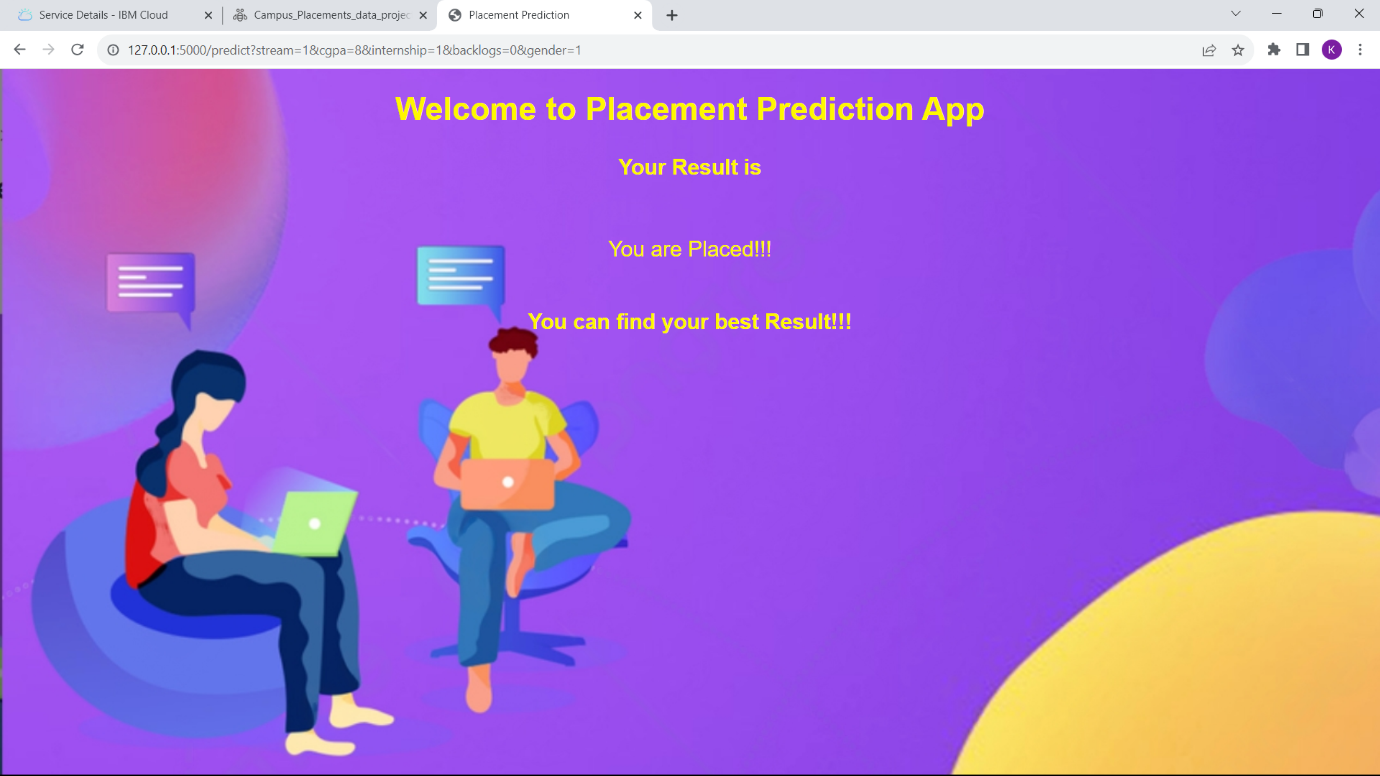
Stream :CSE

CGPA: 8

Internships : 1

Backlogs : 0





**7. Advantages and Disadvantages :**

**Advantages:**

**1**.**Data-Driven Decisions**: Machine learning enables universities and companies to make data-driven decisions regarding student placements, potentially leading to better outcomes for both students and employers.

**2.Efficiency:** Automated placement prediction can significantly reduce the time and effort required to review and process applications, allowing institutions to manage larger volumes of data efficiently.

**3.** **Feature Importance:** Machine learning models can identify which factors contribute most to successful placements, helping students focus on areas that matter the most.

**Disadvantages:**

**1**.**Data Quality:** Accurate placement prediction relies on high-quality data. Inaccurate or incomplete data can lead to erroneous predictions.

**2.Complexity:** Developing and fine-tuning accurate machine learning models can be complex and time-consuming. It requires expertise in data science and machine learning.

**8.Applications:**

**1. Educational Institutions:**

* **Optimized Career Guidance**: Educational institutions can use the model to provide personalized career guidance to students based on their academic performance, skills, and interests.
* **Resource Allocation**: Universities can allocate resources more efficiently by predicting which departments or programs are likely to have higher placement rates.
* **Curriculum Enhancement**: Using insights from the model, institutions can adapt their curriculum to align with industry demands, increasing the chances of student placements.

**2. Students:**

* Career Planning: Students can use the model's insights to plan their career path more effectively by focusing on areas that are likely to lead to better placements.
* Skill Enhancement: The model's feature importance analysis can help students prioritize skill development based on what employers value the most.
* Confidence Boost: Students can gain confidence from knowing they are making informed decisions based on data-driven predictions.

**9.Conclusion:**

The campus placement activity is incredibly a lot of vital as institution point of view as well as student point of view. In this regard to improve the student’s performance, a work has been analyzed and predicted using the classification algorithms Decision Tree and then this model is deployed on Flask. Making website of this model will help to increase the reach of this model to as many students as possible.

**10. Future Scope :**

More instances can be added to the dataset to get more accurate results. Dataset can be trained by different machine learning algorithms to check which performs the best and deploy the website on that model for better results.

**11.Bibilography**

**Books :**

1. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
2. "Introduction to Machine Learning with Python" by Andreas C. Müller and Sarah Guido.

**Research Papers:**

For academic research, you can search databases like IEEE Xplore, ACM Digital Library, and Google Scholar for papers related to placement prediction, education analytics, and machine learning in education.