# *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 primary objective of educational institutions is often seen as securing successful placements for their students. A college's effectiveness is often gauged by the achievements of its students in finding placements. This holds significance for both students and institutions in assessing students' current skillsets to anticipate their potential for future employability. Institutions continuously strive to facilitate student placements by introducing new courses and skill-enhancement programs. This endeavor is crucial for institutions to determine their areas of focus. However, students might find it challenging to prepare comprehensively for placements, given the diverse skill requirements.

This project aims to address this issue by enabling students to concentrate solely on the skillsets essential for successful placements. By conducting an analytical evaluation of students' technical and soft skills, insights can be gained into their potential placement outcomes. The proposed model assists in identifying specific skills crucial for placement readiness. By utilizing a dataset comprising information from previous students, the model predicts the likelihood of a student's placement. The model employs a decision tree classification algorithm for this purpose. It categorizes students as either "placed" or "not placed." This classification model proves beneficial for students to conveniently track their progress over time.

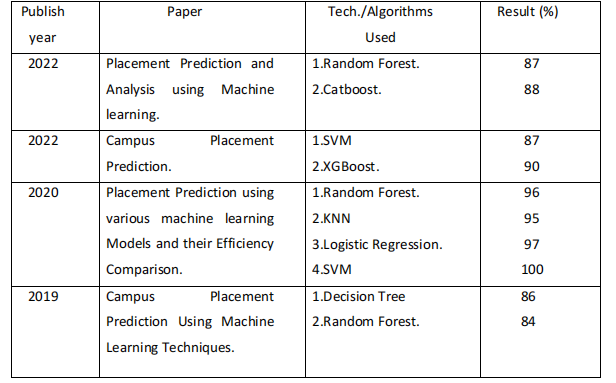
The predictive factors include academic scores, internship experience, and the total number of backlogs. The model's effectiveness is benchmarked against alternative classification models based on their respective accuracies.

**Purpose:**

This project involves a comprehensive exploration of machine learning algorithms and their practical implementation to forecast student placement outcomes. The focal point is the development of a predictive model utilizing the his HistGradientBoostingClassifier algorithm, which effectively anticipates student placement results using historical data from previous years. This model not only provides insights into potential placements but also serves as a valuable tool for students to continually enhance their skill profiles.

Additionally, a critical aspect of this endeavor is to assess the model's accuracy in comparison with pre-existing predictive models for placement outcomes. By doing so, the project aims to ascertain the efficacy of the chosen algorithm. This comparative analysis aids in understanding the algorithm's proficiency in making accurate predictions. Through this process, the project contributes to refining the understanding of machine learning techniques for placement prediction and their real-world applicability.

**Literature Survey:**



**Proposed Solution:**

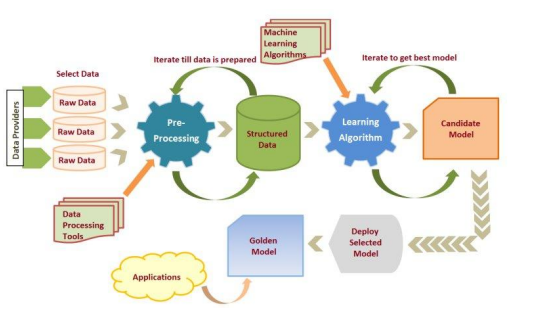
1. To predict placements, data is collected from diverse sources such as academic transcripts, resumes, and previous work experiences.
2. Following this, the data undergoes a process of thorough cleansing and preprocessing to rectify any inconsistencies or errors.
3. Once cleaned, the data is categorized into two sets: training and testing data, pivotal for model development.
4. Employing a range of techniques including neural networks, decision trees, random forests, and regression analysis, the machine learning algorithm is trained using the training data. Its proficiency is subsequently gauged through testing data.

Regression analysis, a statistical tool for establishing relationships among multiple variables, is pivotal in placement prediction. It enables the understanding of connections between factors like academic performance, skill repertoire, prior job engagements, and the prospects of securing a position with an organization.

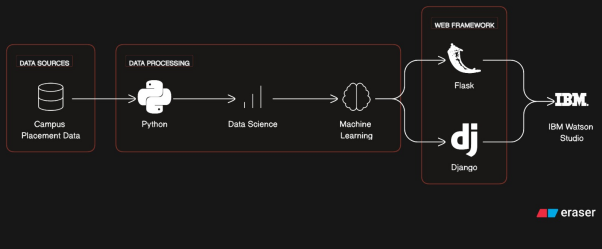
In the context of placement prediction, a specific machine learning algorithm, termed a random forest, employs a tree-like framework to model decisions and potential outcomes. The application of random forests simulates the intricate decision-making involved in a business's hiring process.

**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.

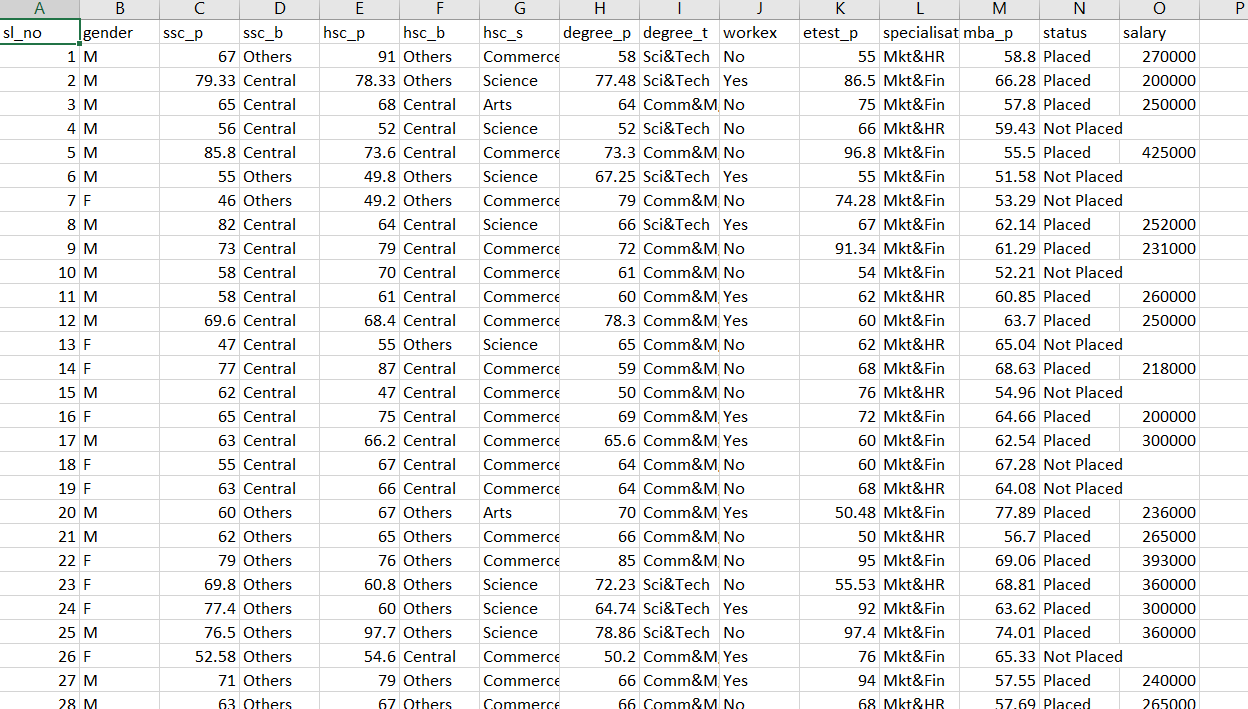
REST- Representational State Transfer.

API- Application Programming Interface.

**Experimental Investigations:**

1. **Dataset Description:**

Engineering Placement Prediction Dataset was downloaded from Kaggle. The dataset consists of 215datapoints.Dataset has following attributes: si\_no, ssc\_p,ssc\_b, hsc\_p, hsc\_b, hsc\_s, degree\_p, degree\_t, workex, etest\_p,specialization,mba\_p, Placed or not attribute has just two values placed or Not placed where 0 indicating that candidate is not placed while 1 indicating that the candidate is placed.



**2)Detail Phases:**

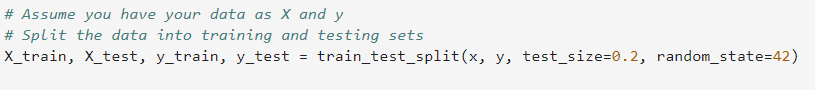
1. Data preprocessing: It includes finding null values in dataset, choosing important features for training and label encoding the dataset. In some cases, the dataset contains missing values. A common plan to handle the matter is to require a mean of all the values of the same column and have it to replace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

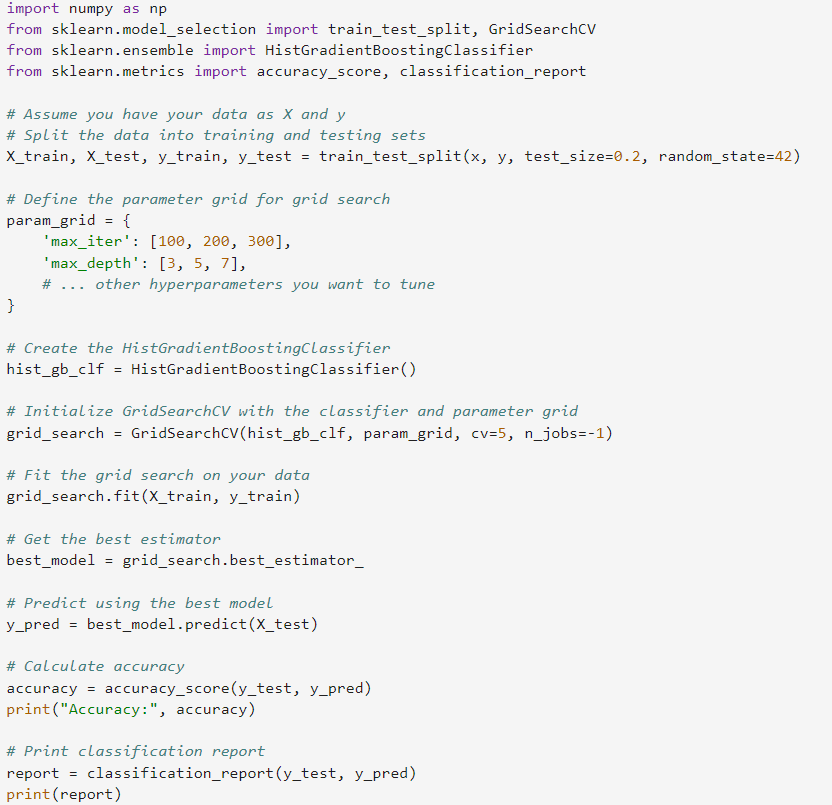


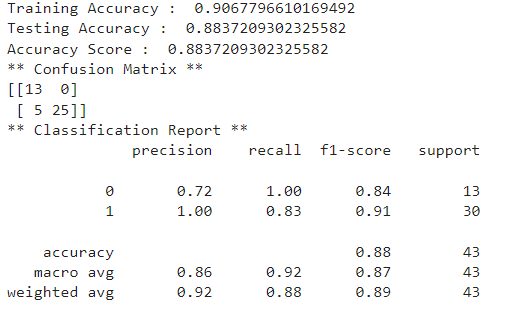
**ENCODING TECHNIQUE**

****

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.

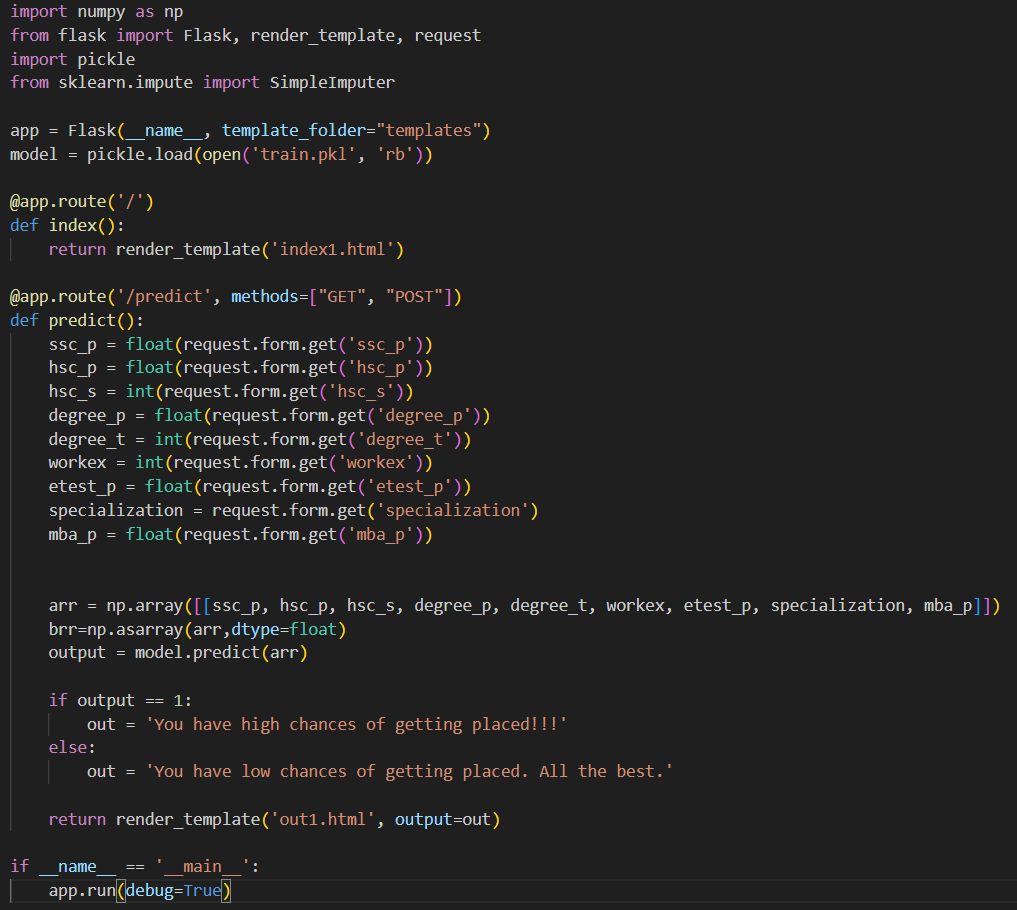






3)Deploying ML model on Flask: ML model which we have prepared will be deployed on Flask.

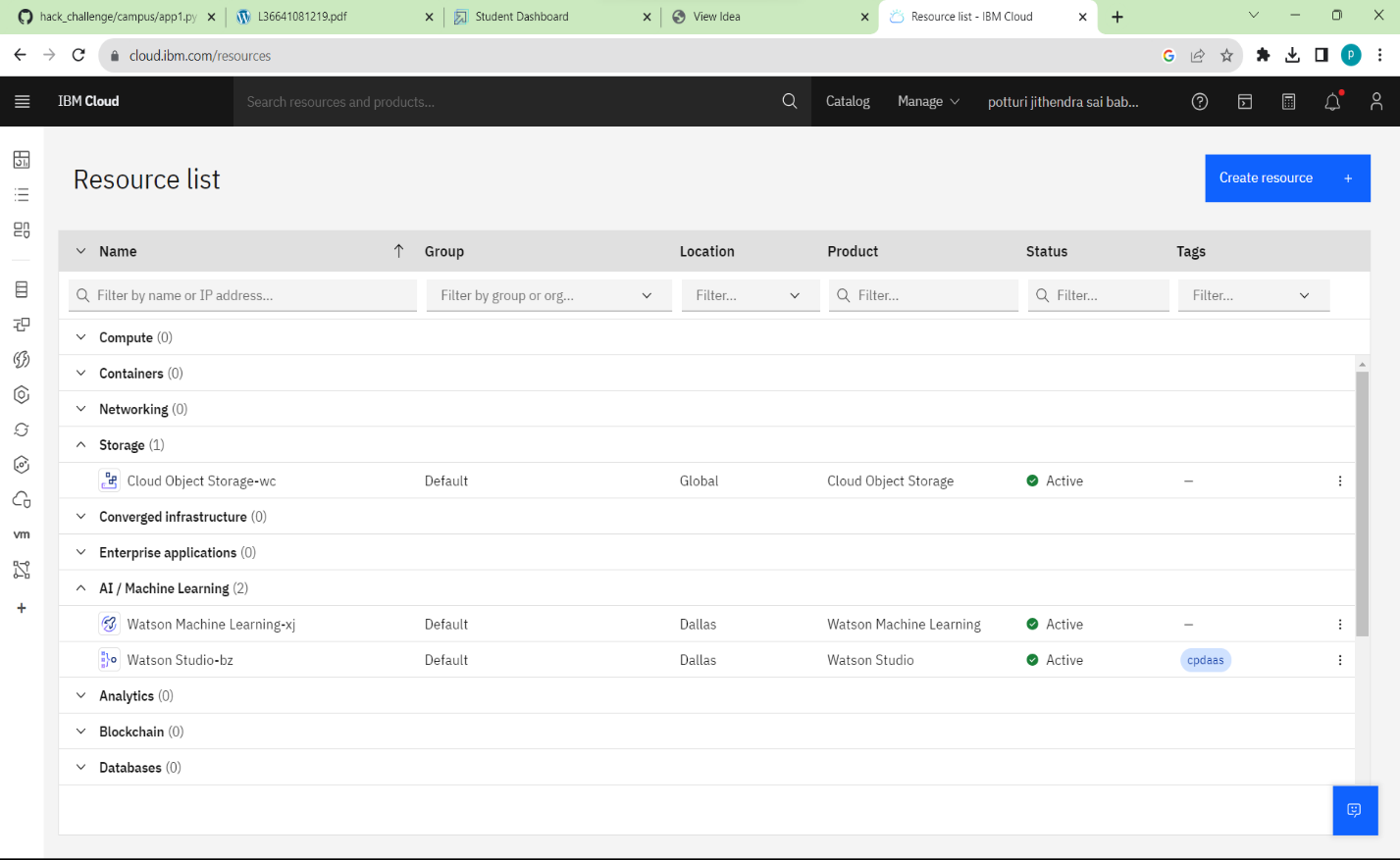
Flask is a Python-based micro framework used for developing small-scale websites. Flask is used to handle API requests.

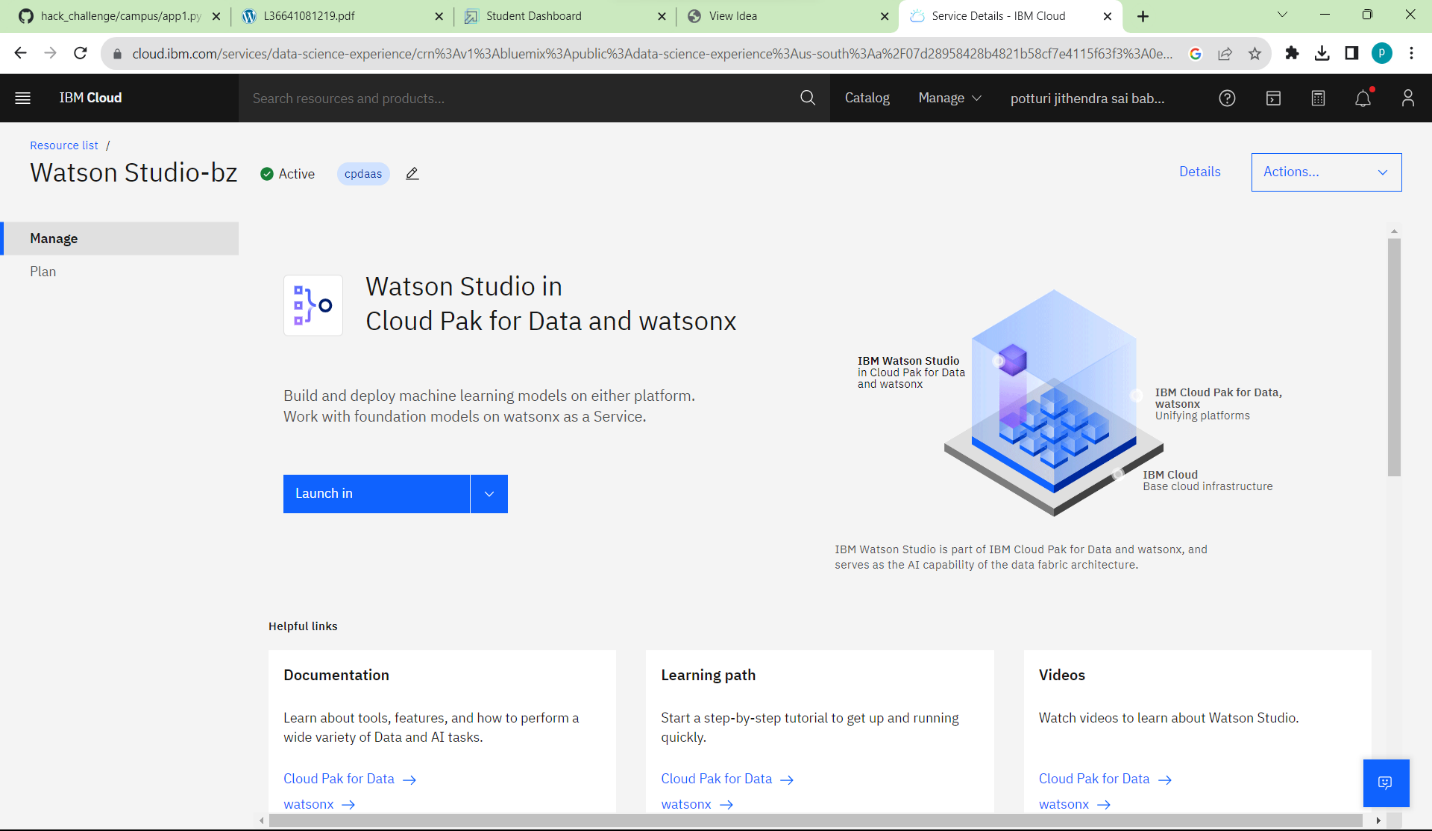


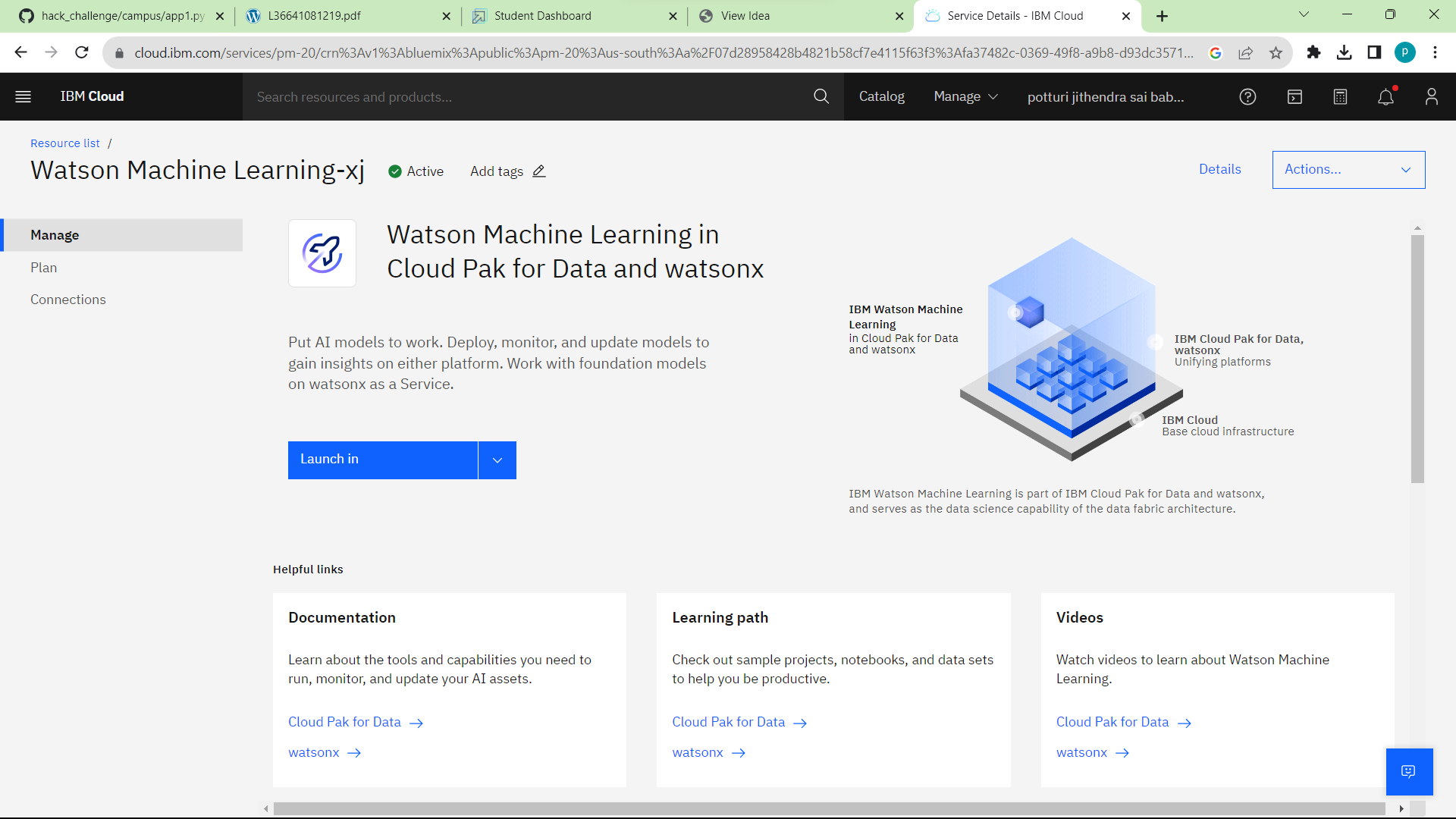
**Deployment into the Cloud:**

**Tools used in the cloud :**

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

****

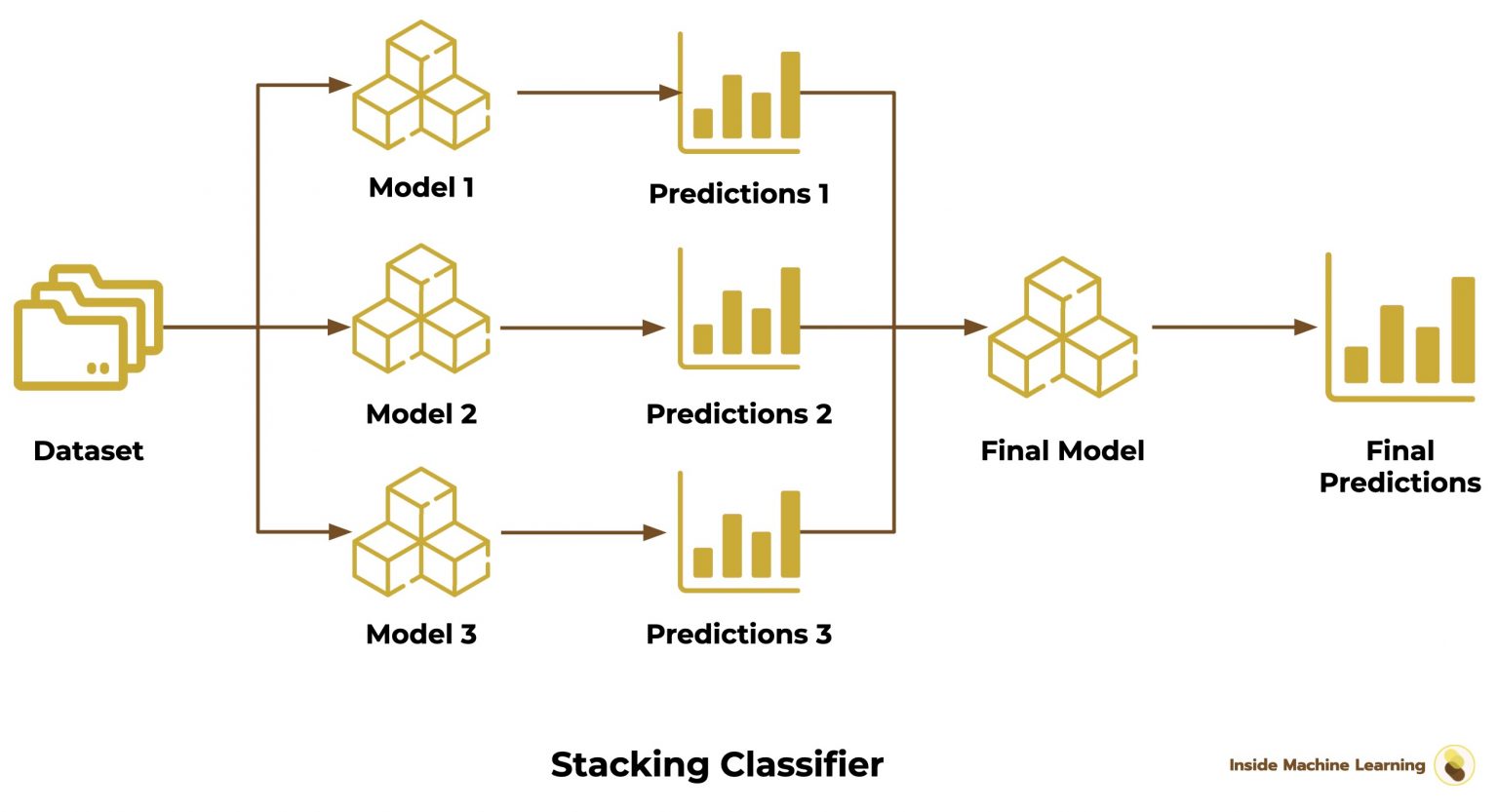
****

****

****

**Algorithms:**

HistGradientBoostingClassifier is a machine learning classifier available in scikit-learn library. It's designed for high-performance gradient boosting on large datasets. It uses histogram-based techniques to accelerate the training process by grouping data into bins, reducing memory usage and computation time. This classifier is particularly suitable for datasets with a large number of samples or features. It's based on the gradient boosting framework, which combines multiple weak learners (typically decision trees) to create a strong predictive model.



**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. HistGradientBoostingClassifier:**

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

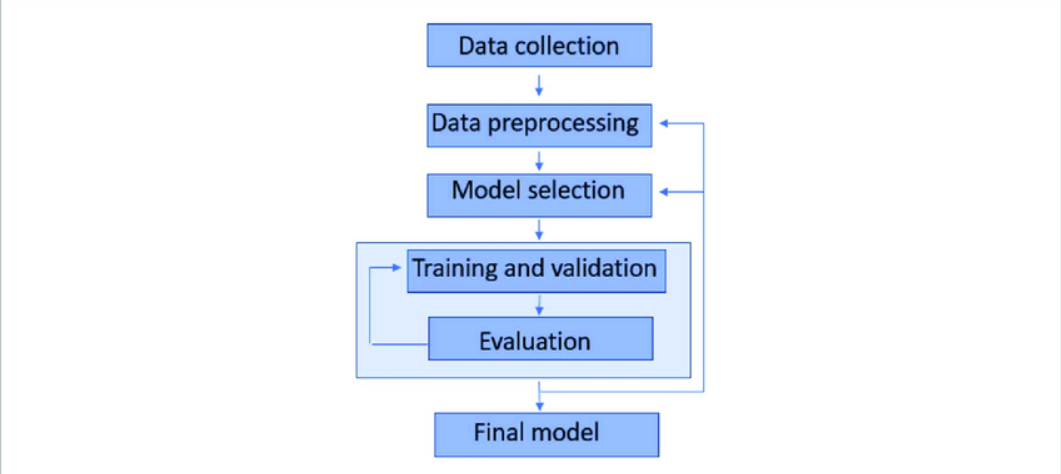
**5. Model Training:**

Train the Random HistGradientboosting 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 ‘ssc\_b’,’hsc\_b’, ‘salary’and ‘sino’ didn’t were not important for training the dataset, hence these attributes were deleted from the dataset. Label encoding was applied on attributes to convert values 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.

ssc\_percentage:80.00

hsc\_percentage:78.00

stream:science

degree percentage:70.00

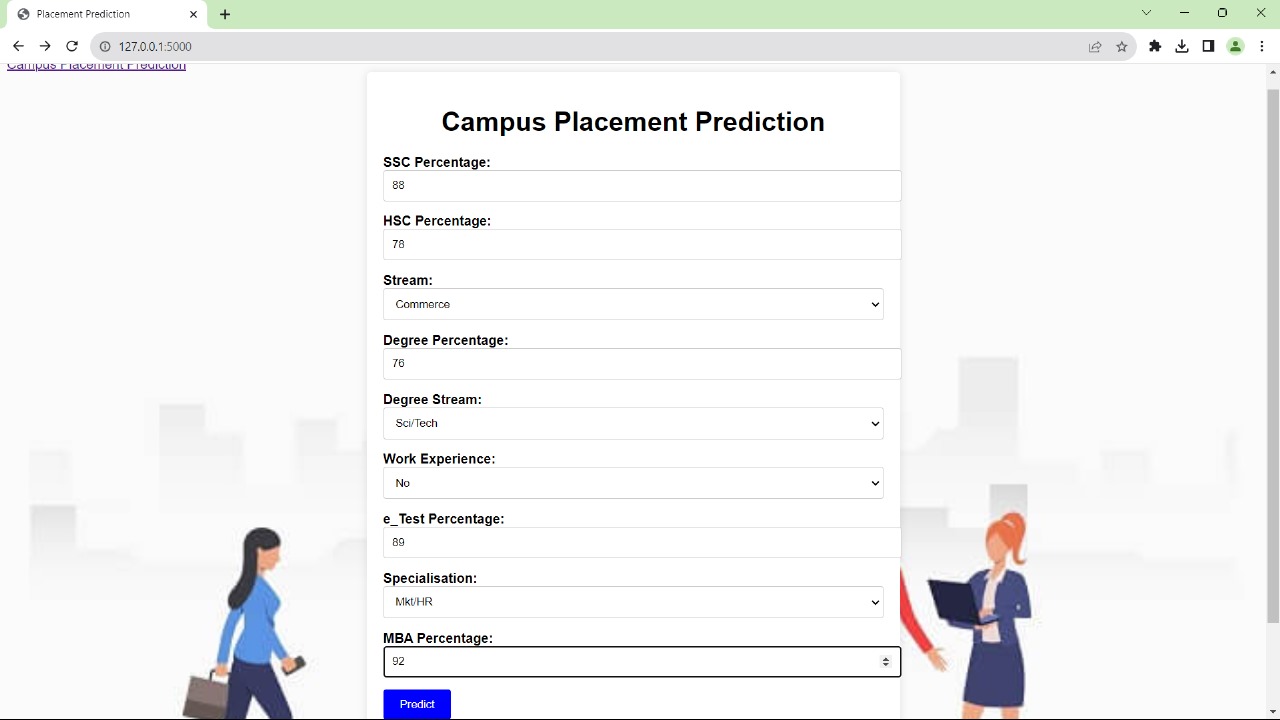
degree stream:comm/mgmt

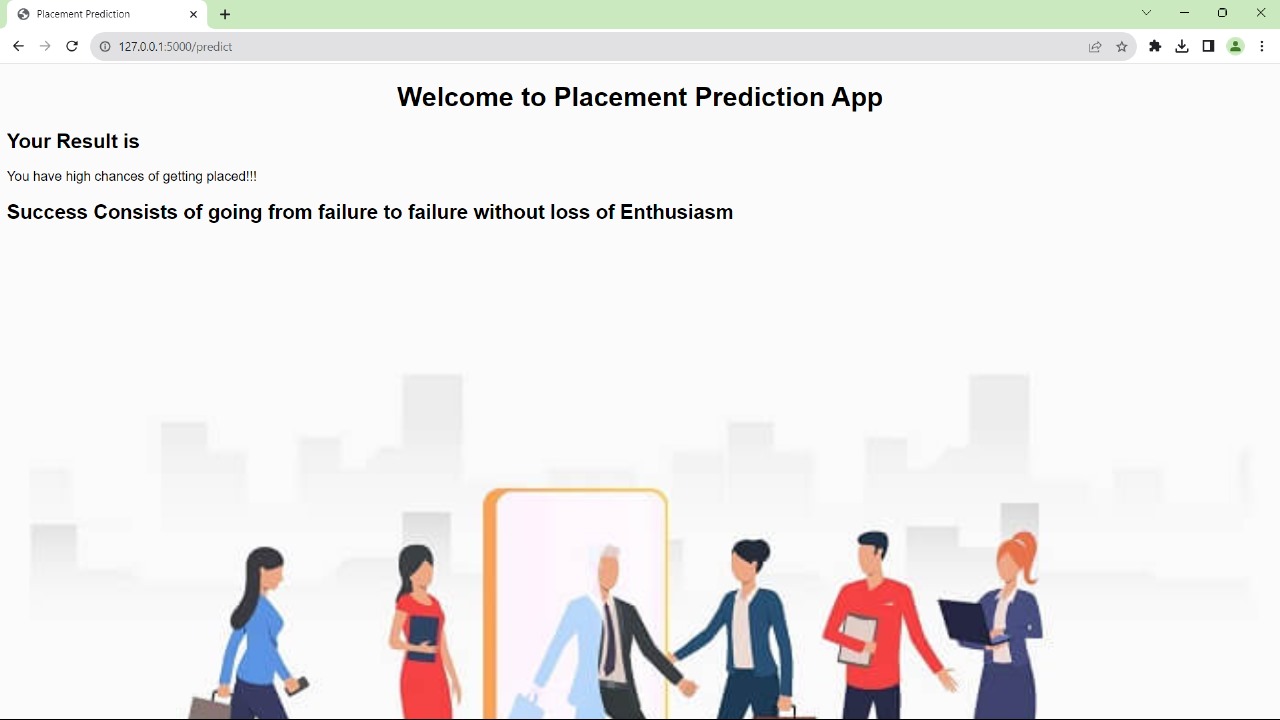
work experience:no

e\_test percentage:70.00

specialization:mkt/hr

mba\_percentage:92.00



****

**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 employer 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.

**General Description**

  Campus placement data comprises information about students, their academic performance, skills, internships, and their eventual placement outcomes. The objective is to extract valuable insights from this data to understand factors influencing placement success and develop strategies for improving the placement process. by identifying patterns and trends in campus placement data using machine learning techniques

 campus recruitment during the placement season of a Business School in India. We work closely with university career services centers and attend career fairs to connect with college students and recent graduates in person. By analyzing various factors, such as work experience, exam percentages, and more, we identify the key determinants of candidate success. The data also includes recruitment status and remuneration details. Through data-driven insights and personalized approaches, we optimize the hiring process for both students and employers, ensuring the right fit for internships and entry-level positions. Our goal is to enhance placement outcomes and faster long-term career growth and organizational success.

**Novelty / Uniqueness:**

Using machine learning to identify patterns and trends in campus placement data is a novel and innovative approach. It leverages advanced algorithms to uncover hidden relationships and correlations in the dataset. This unique analysis allows educational institutions to make data-driven decisions and develop targeted strategies to improve placement outcomes. Each campus's data is distinct, and machine learning models can be customized accordingly. The insights gained from this analysis empower students with personalized career guidance and support services. Overall, the use of machine learning optimizes the placement process, fastering better industry-academics partnerships and enhancing students' employability for successful career paths.

**Business / Social Impact:**

The campus placement business model involves universities and colleges collaborating with companies to offer job opportunities to their students. It may also include supplementary services like career counseling, resume review, and interview preparation, which can be provided to students for an extra charge. This model benefits students by facilitating job placements and enhancing their employability. For universities, it strengthens industry ties and boosts the institution's reputation. Additionally, companies gain access to a pool of potential candidates, streamlining their recruitment process. The model fosters a symbiotic relationship between academia and industry, creating a positive impact on the overall job market and contributing to economic growth.

**Technology Architecture:**

Python, Data science, Machine Learning, Web Framework like flask or Django, IBM Watson Studio

**Scope of the Work:**

 The scope of work for the campus placement business model involves university and company partnerships, connecting students with job opportunities. It includes additional services like career counseling, resume review, and interview preparation for students. The model aims to enhance students' employability, strengthen industry ties, and streamline company recruitment. Continuous evaluation and adaptation are essential to optimize the placement process and ensure successful career outcomes for students.