CHAPTER 1

INTRODUCTION

1.1 Overview

In the modern educational landscape, campus placement plays a pivotal role in shaping the careers of students. The Campus Placement Analysis project aims to leverage the power of data analysis and machine learning to gain valuable insights from the vast amount of data generated during the campus placement process. This data includes student profiles, academic achievements, skills, internships, and eventual placement outcomes. By delving into this data, institutions can make informed decisions, enhance the placement process, and better prepare students for successful careers.

By integrating machine learning algorithms with web technologies, the Campus Placement Analysis project aims to revolutionise how institutions approach placement processes. This holistic approach empowers institutions to make informed decisions, enhances student employability, and establishes strong connections between academia and industry.

1.2 Purpose

The main purpose of the Campus Placement Analyzer is to provide colleges and universities with a comprehensive tool that leverages data analysis and machine learning to optimize the campus placement process. This tool aims to offer valuable insights, personalized recommendations, and predictions for students' placement outcomes. Analyzing student data and historical placement trends, helps institutions enhance their placement strategies, improve student employability, and foster stronger connections with potential employers.

Data Collection and Input:

Colleges provide an Excel sheet containing detailed information about students, including academic records, skills, internships, and any other relevant details.

Data Processing and Analysis:

The system takes the Excel sheet and processes the data to create a structured dataset. Feature engineering is performed to extract key attributes like academic scores, skills proficiency, number of internships done, and more.

Stats and Insights Dashboard:

The system generates a comprehensive dashboard with detailed statistics about upcoming placements. Colleges can view placement stats across all branches and the entire campus. Insights include placement success rates and average salary offers.

Personalized Recommendations:

Based on individual student profiles and skill sets, the system provides personalized recommendations.

Salary Prediction:

Utilizing machine learning algorithms, the system predicts potential salary ranges for each student based on their data and skills.

Visualization and Reporting:

The dashboard includes visualizations such as graphs, and charts to present data trends and patterns. Colleges can use these insights to make informed decisions about refining placement strategies.

Student Engagement:

Students can access the system to view their personalized recommendations and predicted salary ranges.

CHAPTER 2

LITERATURE SURVEY

Campus placement analysis and prediction is an area of research that has gained considerable attention in recent years, particularly with the increasing demand for skilled and employable graduates. Here are some key findings from recent literature on this topic:

1) Machine Learning Techniques for Campus Placement Prediction: A Review of Literature (2021) by S. Suresh and S. Sengottuvelan: This review paper examines the use of machine learning techniques in predicting campus placements. The study finds that machine learning models like Random Forest, SVM, and Decision Trees have shown promising results in predicting campus placements.

2) "Exploring campus placement practices in India: A qualitative study" by S. Kumar, R. K. Mishra, and S. Shankar (2021): This qualitative study explores the experiences and perspectives of stakeholders in the campus placement process, such as students, recruiters, and placement coordinators. The authors identify several challenges and opportunities for improving campus placement practices, such as better communication and collaboration between universities and industry partners.

3) Predicting Campus Placement Using Data Mining Techniques: A Review (2020) by R. Shankar, S. P. Gupta, and S. K. Singh: This paper provides a comprehensive review of various data mining techniques used for predicting campus placements. The authors highlight the importance of feature selection and data preprocessing techniques in improving the accuracy of the prediction models. The study finds that Naïve Bayes, K-Nearest Neighbor, and Decision Trees are some of the commonly used techniques for campus placement prediction.

4) Analysis and Prediction of Campus Placement Using Decision Tree Algorithm (2018) by S. A. Deshmukh, P. M. Mokadam, and A. V. Kulkarni: This study uses the Decision Tree algorithm to predict campus placements. The authors find that academic performance, communication skills, and technical skills are the most important factors affecting campus placements. The study also suggests that Decision Tree is an effective algorithm for predicting campus placements as it provides an easy-to-understand visual representation of the decision-making process.

2.1 Existing problem

Data Quality and Consistency: Many institutions struggle with inconsistent and incomplete data collection, which can hinder accurate analysis. Variability in how student data is recorded and the lack of standardized formats can lead to challenges in data processing and interpretation.

Lack of Personalization: Traditional placement approaches often lack personalization. Students have diverse skill sets and aspirations, but generic placement strategies may not effectively cater to individual needs. This can result in suboptimal placement outcomes and unfulfilled potential.

Dynamic Industry Landscape: The job market's fast-paced nature poses challenges in accurately predicting future placement trends. Economic fluctuations, technological advancements, and shifting industry demands can make it difficult to create models that consistently predict salary ranges.

Skill Gap Identification: Identifying the precise skills that students need to develop for successful placements is challenging. Often, there need to be more clarity between industry requirements and the integration of relevant skills into academic curricula.

2.2 Proposed Solution

Inconsistent Data Handling: Implemented a data preprocessing module that standardizes and cleans incoming data, converting it into a consistent format. This ensures that regardless of how data is provided by colleges, it is transformed into a structured dataset for analysis.

Personalization Gap: Develop a recommendation engine that considers individual student profiles, skills, and preferences. By leveraging machine learning algorithms, the system can suggest specific skills, that align with each student's strengths and aspirations.

Skill Gap Bridging: Skills Mapping and Development: Feature a "Skill Gap Analysis" chart that visualises the alignment between student skills and industry requirements. It will offer clear and insightful data visualisations to colleges and universities, aiding them in making informed decisions about placements.

CHAPTER 3

METHODOLOGY

3.1 Datasets

1. Database: The dataset was collected from our college T&P (training and placement) cell. The dataset consists of various attributes like the gender, ssc percentage, inter percentage, ssc board, inter board, b tech percentage, branch etc., with a total 731 records

2. Data Preprocessing: Data preprocessing involves transforming raw data into well-formed data sets so that data mining algorithms can be applied. Raw data often lacks consistency in formatting and is insufficient. Handling missing values In the dataset, missing values are present in the status column due to the fact that these figures apply to the students who were not placed during any placement drive. The missing values in the are therefore considered to be zero and replaced by 0 using Python's fillna(0,inplace=True) method. Handling categorical values Categorical values cannot be dealt directly, so mapping is done for attributes having categorical data. Here, status attribute values ‘placed’ is replaced by 1 and ‘not placed’ is replaced by 0.

3. Feature Selection: Feature selection in campus recruitment analysis involves identifying and selecting the most relevant and significant variables or features that have a strong impact on predicting the outcome of the recruitment process. This process involves filtering out the unnecessary or irrelevant features that may lead to overfitting and poor performance of the model.

4. Build training and testing sets: The data is divided into training and testing data. The data used to train an algorithm or machine learning model to predict a result is known as training data. Data created or selected to fulfill the prerequisites for execution and the input required to run one or more test cases. Various proportions were created and used for the review, including 80:20, 70:30, and 60:40. An 80:20 ratio indicates that 80% of the information is considered preparation and the remaining 20% is considered testing.

3.2 Machine learning Algorithms

● **Naive Bayes Classification:** Naive Bayes is a probabilistic algorithm that is commonly used for classification tasks. There are three types of Naive Bayes classifiers: Gaussian, Multinomial, and Bernoulli.

a) Gaussian Naive Bayes classifier is suitable for continuous data where the features have a normal distribution. In the context of campus placement analysis, this classifier can be used to predict the likelihood of a student getting placed based on their academic performance, such as their GPA or percentage.

b) Multinomial Naive Bayes is suitable for discrete data where the features represent counts, such as word frequencies in a text document. this classifier can be used to predict the likelihood of a student getting placed based on their skills and experience, such as the number of internships and work experience they have completed or based on skills they posses.

c) Bernoulli Naive Bayes is similar to Multinomial Naive Bayes, but it assumes that the features are binary (i.e., they are either present or absent). this classifier can be used to predict the likelihood of a student getting placed based on their job preferences, such as whether they are willing to relocate or not.

● **Decision tree**: The algorithm uses a tree-like model to represent decisions and their possible consequences, and it can be used to predict the outcome of a decision based on a set of input variables. Once the tree is constructed, it can be used to predict the placement success of new students based on their input variables.

● **KNN:** KNN can be used to predict the likelihood of a student being placed in a particular company based on various factors such as academic performance, profile, and job preferences. The algorithm works by finding the K nearest data points to the new data point (i.e., the student's profile), based on a distance metric, and then predicting the outcome variable (i.e., placement) based on the majority class of the K nearest neighbours.

● **Logistic Regression**: it is a statistical technique used to analyze the relationship between a categorical dependent variable and one or more independent variables. This model can also be used to identify which independent variables are most strongly correlated with placement success. This information can be used to identify areas where students can improve their chances of getting placed, such as by focusing on improving their academic performance or gaining more work experience.

CHAPTER 4

THEORITICAL ANALYSIS

4.1 Block Diagram

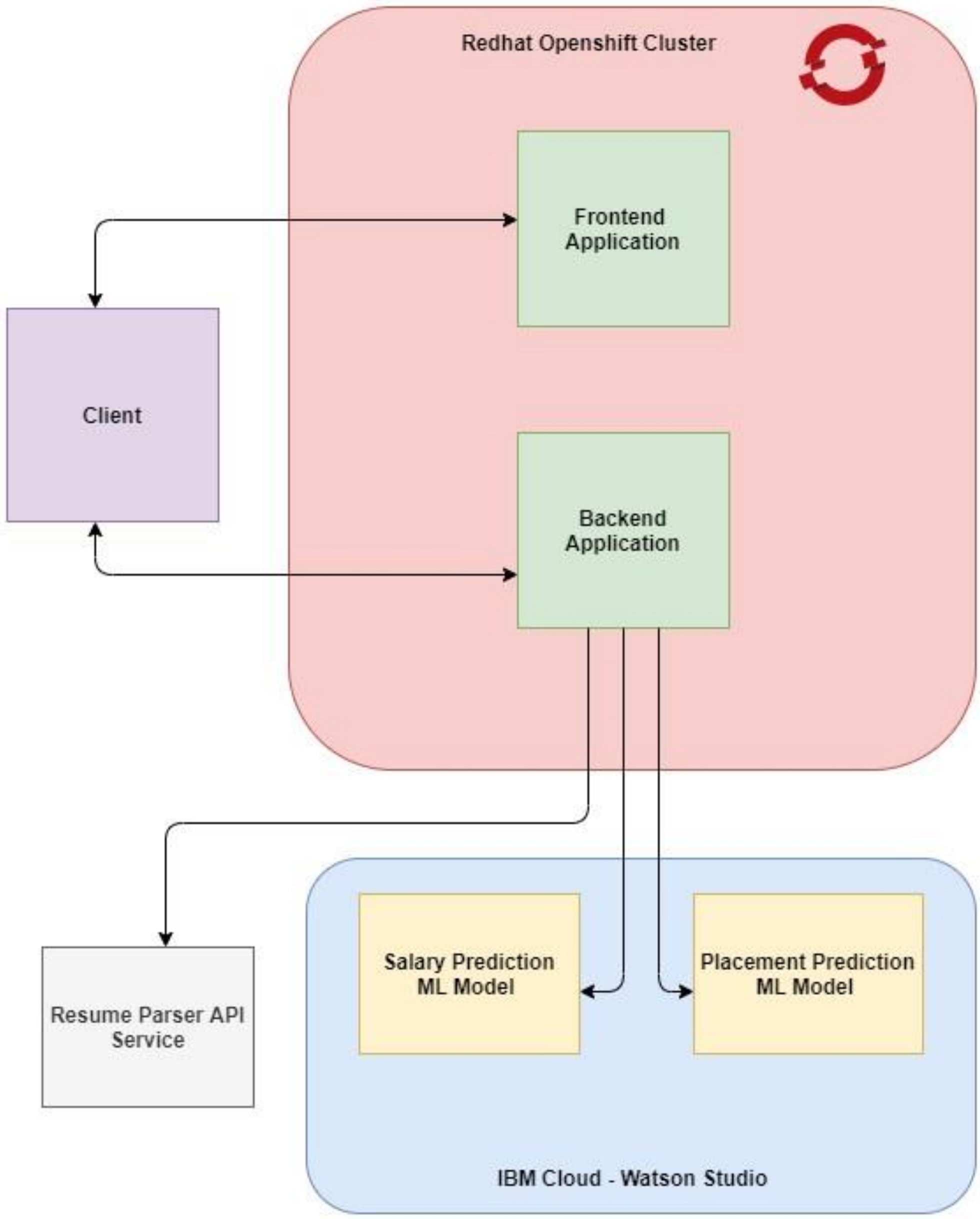

Fig 4.1: Resume Analysis and Prediction System

4.2 Hardware / Software Designing Requirements

1. **Hardware Requirements**

   Web Server - Redhat Openshift

   Load Balancers (if applicable)

2. **Software Requirements**

   Operating System – Linux

   Server-Side Scripting Language – Python

   Frontend Technologies (HTML, CSS, JavaScript)

   Framework for Web Development (Flask and React)

   Postman to test backend APIs

   Microsoft Excel to store college student's data.

3. **Data Science Tools**

   IBM Watson Studio to perform ML techniques on Jupyter Notebook

   Numpy, Pandas, and Matplotlib for Data Analytics

   Scikit-Learn to train Machine Learning models.

4. **Deployment Tools**

   Redhat Openshift to host the application

   IBM Watson Machine Learning to deploy ML models

   Git for version control and collaboration

   Docker Hub to upload docker images to the docker repository

5. **API services**

   Affinda resume parser API to parse the resume and enhance user accessibility

   Affinda Skill suggestions API to recommend skills based on existing skills

CHAPTER 5

EXPERIMENTAL INVESTIGATIONS

5.1 Experimental Investigation

Tier-wise Salary Distribution Investigation:

Investigated the relationship between student tiers (Tier 1, Tier 2, Tier 3) and placement outcomes to understand how higher tiers correlate with higher salary packages.

Role of CGPA and Technical Skills Investigation:

Explored the significance of CGPA and technical skills in influencing student salaries, emphasizing the importance of these factors in the placement process.

Resume Scoring ATS and Student-Oriented Applications Investigation:

Identified the absence of applications catering to student needs beyond resume scoring. Researched potential avenues to develop an application that provides personalized insights, including salary expectations and skill recommendations.

API Exploration Investigation:

Investigated available APIs to enrich the project's functionality, with a focus on integrating external data sources to enhance insights and predictions.

External Factors and Recession Trends Investigation:

Explored the impact of external factors, such as the country's GDP and economic trends, on placement outcomes. Analyzed how placement success varies during different economic scenarios, including recessions.

User-Friendly Data Visualization Investigation:

Explored options for creating user-friendly charts and visualizations that effectively communicate insights to users. Emphasized the importance of clear and intuitive data representation.

5.2 Flowchart

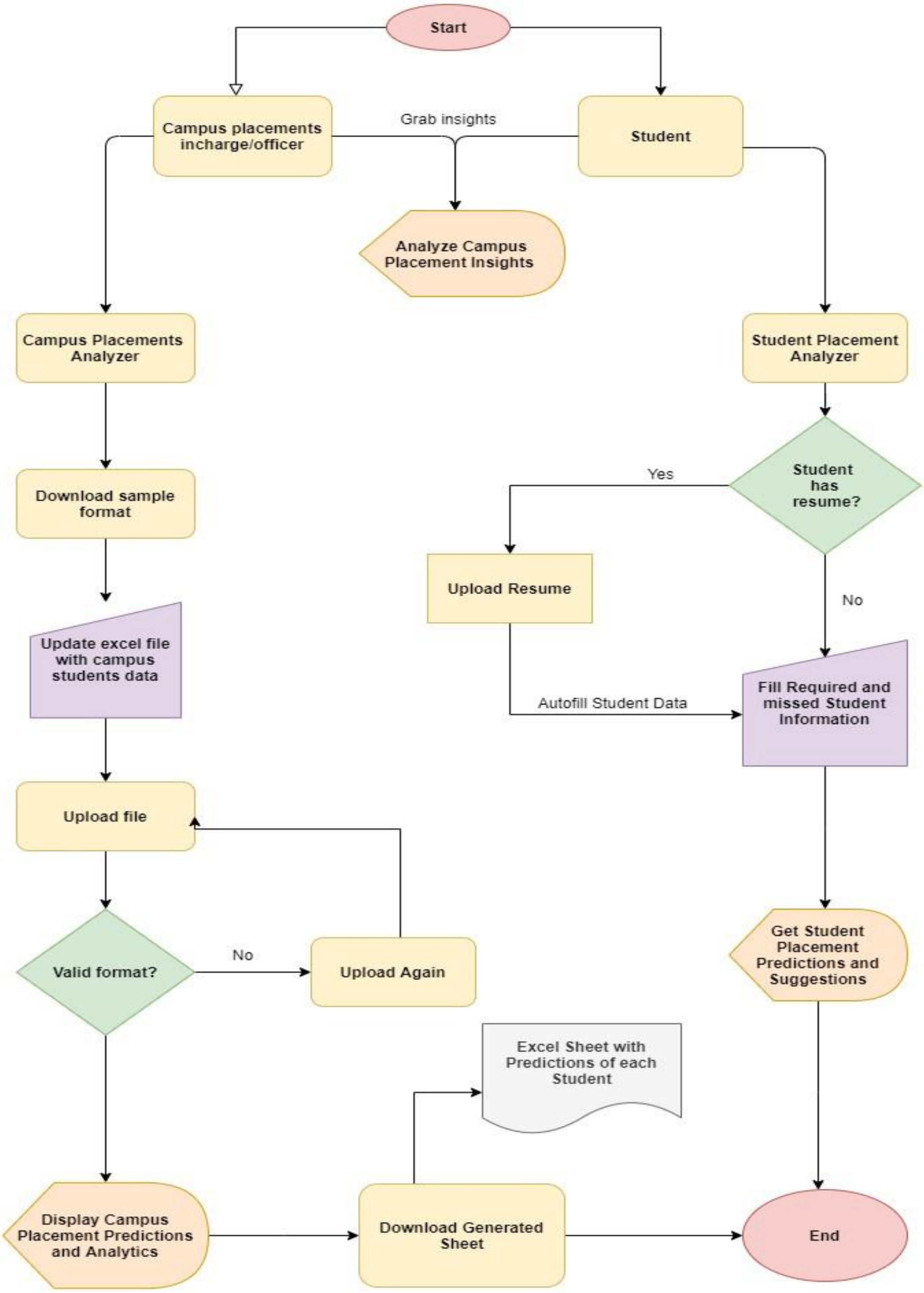

Fig 5.1: Flowchart

5.3 Advantages

The Campus Placement Analysis project offers several advantages to educational institutions, students, and recruiters, revolutionising the way campus placements are managed:

Informed Decision-Making:

The project empowers institutions with data-driven insights, enabling them to make informed decisions about curriculum enhancements, skill development programs, and industry partnerships based on placement trends and industry demands.

Enhanced Student Employability:

By providing personalized recommendations, skill gap analysis, and predicted salary ranges, the project helps students prepare more effectively for placements, improving their chances of securing desired job roles. Optimized Placement Strategies: Institutions can refine their placement strategies by analyzing historical data, improving placement success rates, and catering to industry-specific requirements.

Improved Industry-Academia Collaboration:

With real-time job market trend analysis, colleges can align their offerings with industry needs, training them on those skills, and ensuring that graduates are job-ready.

Visual Data Insights:

The project offers intuitive visualizations that simplify complex data, making it easier for stakeholders to understand placement trends, strengths, and areas for improvement.

Personalized Student Support:

The system's recommendations and insights enable institutions to provide targeted support and counselling to students, helping them align their skills and goals with industry requirements.

Student Motivation:

Visualizing potential salary ranges and success rates motivates students to actively participate in skill enhancement programs and engage more enthusiastically in the placement process.

5.4 Disadvantages

Changing Trends: Employment trends and job requirements can change rapidly. Models trained on historical data might not accurately predict future placement outcomes if the job market evolves significantly.

Lack of Contextual Information: While machine learning can provide insights, it might lack the rich contextual information so that human experts can provide. This can limit the depth of understanding regarding certain placement factors.

Unpredictable External Factors: Economic shifts, changes in industries, and unforeseen events (e.g., pandemics) can significantly impact placement outcomes, rendering historical data less relevant.

5.5 Applications

The Campus Placement Analysis project holds significant real-world applications

Educational Institutions:

Colleges and Universities: Educational institutions can use the insights to refine their curriculum, offer skill development programs, and strengthen industry collaborations to enhance student employability.

Students Individual Career Planning:

Students can utilise personalized recommendations and predicted salary ranges to make informed decisions about their career paths and skill development priorities.

Skill Enhancement:

The skill gap analysis guides students on areas that need improvement, allowing them to focus on building competencies relevant to their desired roles.

Educational Policy Makers:

Curriculum Design: Educational policymakers can leverage the project's data to inform curriculum design and update educational programs based on industry demands.

Recruiters and Companies: Recruiters can use historical placement trends and insights to identify colleges producing graduates with relevant skills for their industry, streamlining their talent acquisition process.

Data-Driven Hiring: Access to student profiles and skill data helps recruiters make data-driven hiring decisions, aligning candidate profiles with their organizational needs.

Industry Experts and Consultants:

Consultation Services: Industry experts can offer consultation services to institutions, helping them interpret the data and make strategic decisions for better placement outcomes.

Job Portals and Recruitment Platforms: Job portals can enhance their matching algorithms by incorporating the skill and placement data to suggest the most relevant job opportunities to candidates.

CHAPTER 6

SNAPSHOTS

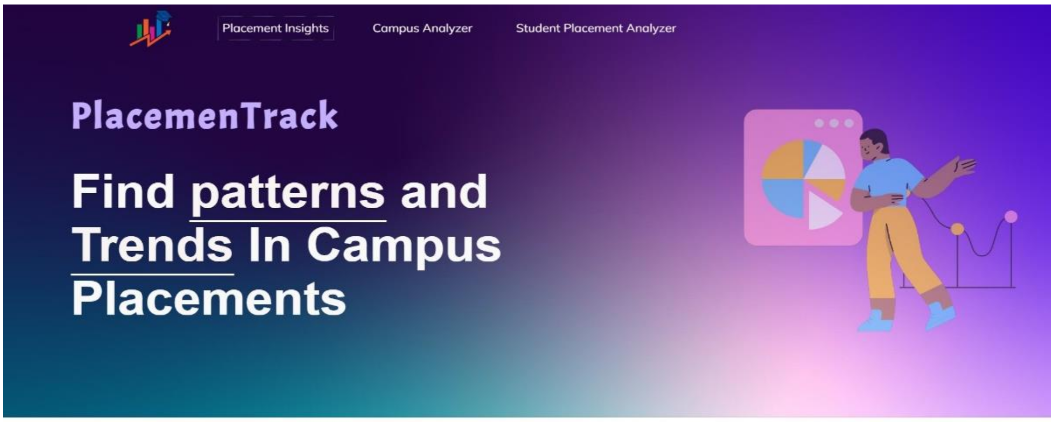

Fig 6.1: Website view

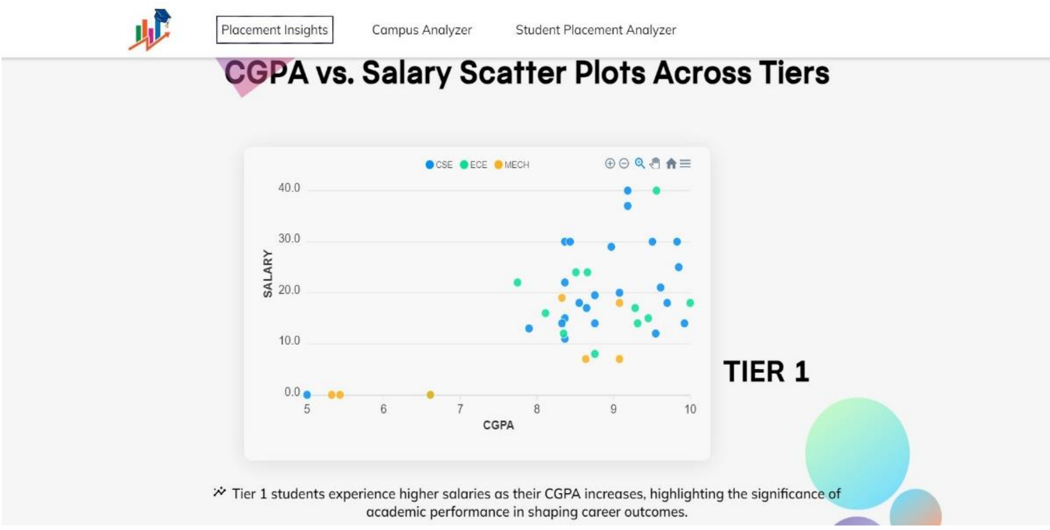

Fig 6.2: CGPA vs Salary plots

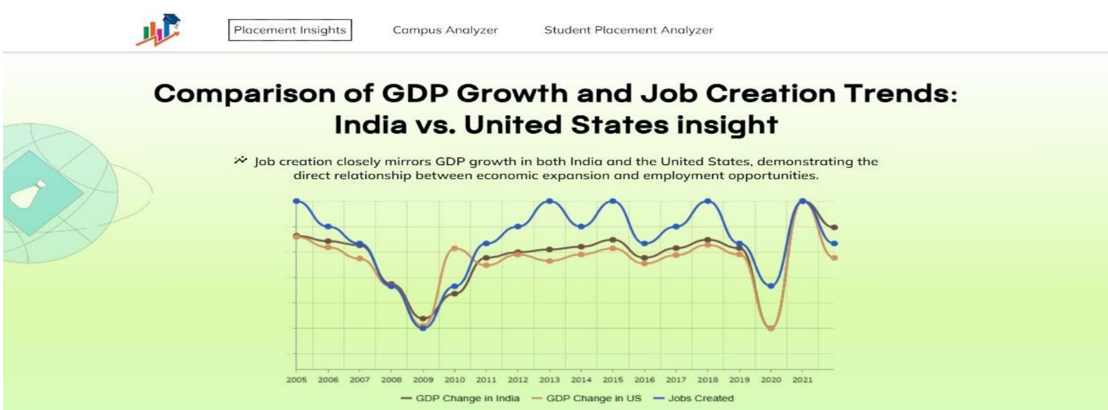

Fig 6.3: Comparison graph

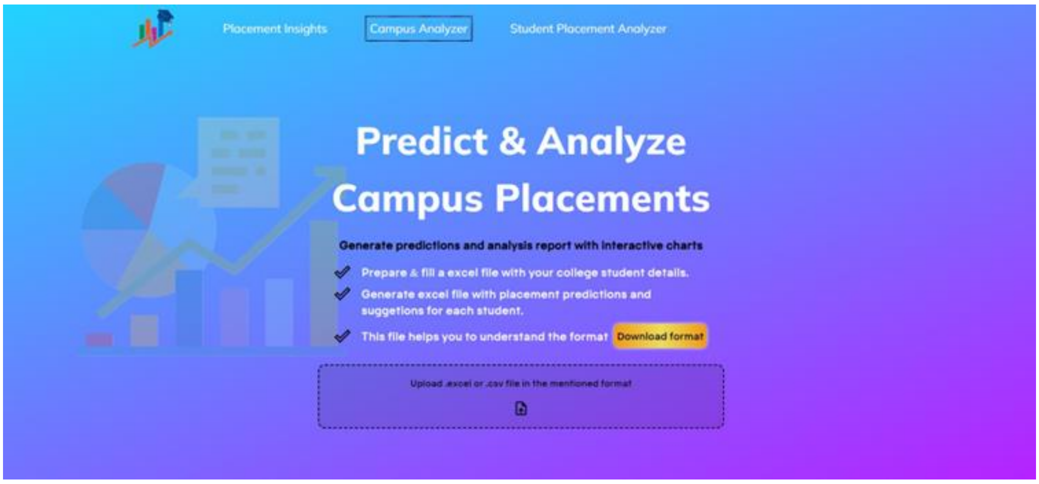

Fig 6.4: Campus Analyzer

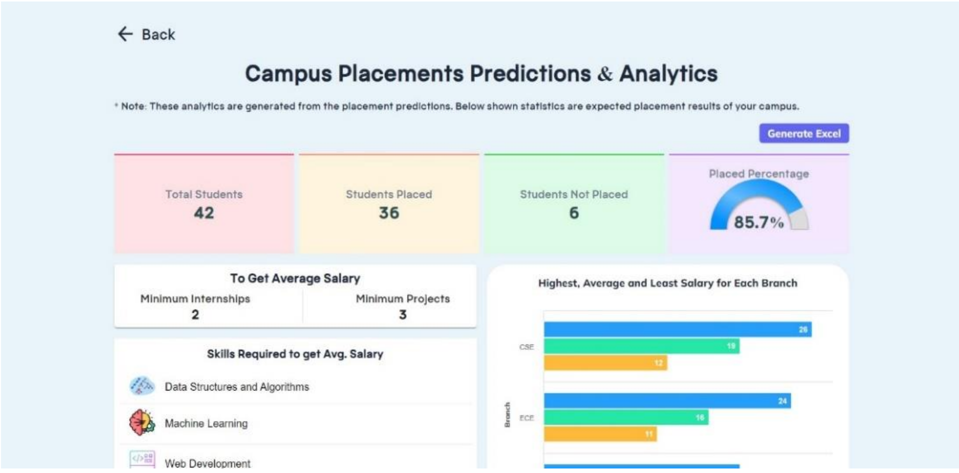

Fig 6.5: Prediction graph

Fig 6.6: Graphs

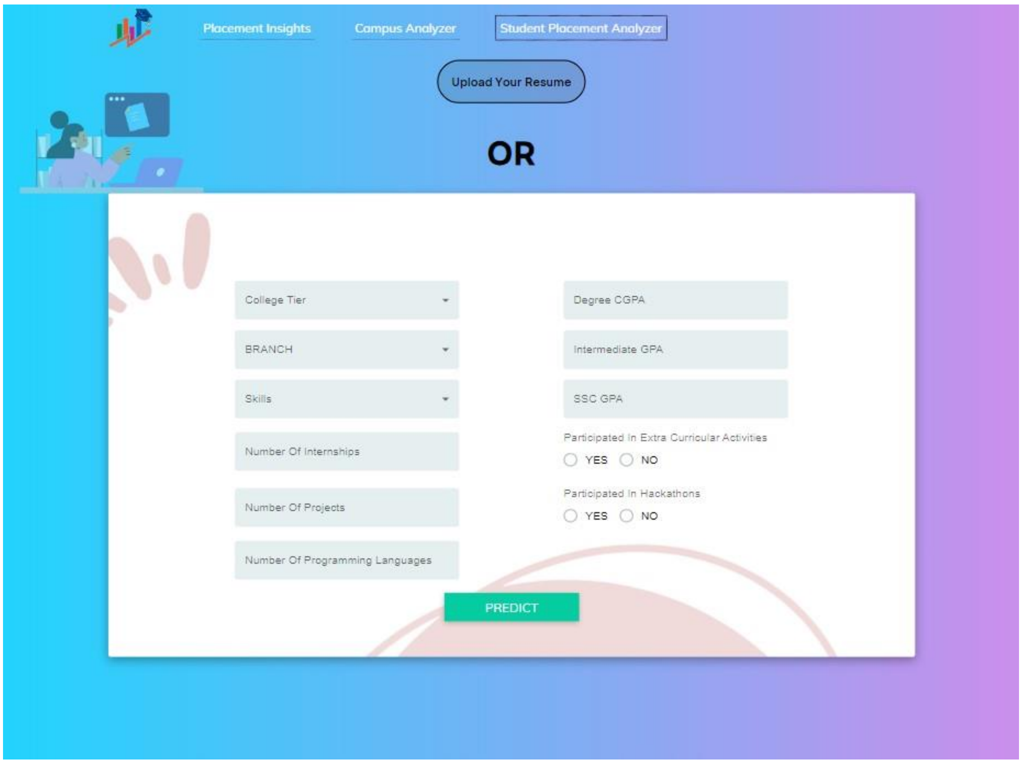

Fig 6.7: Student placement analyzer

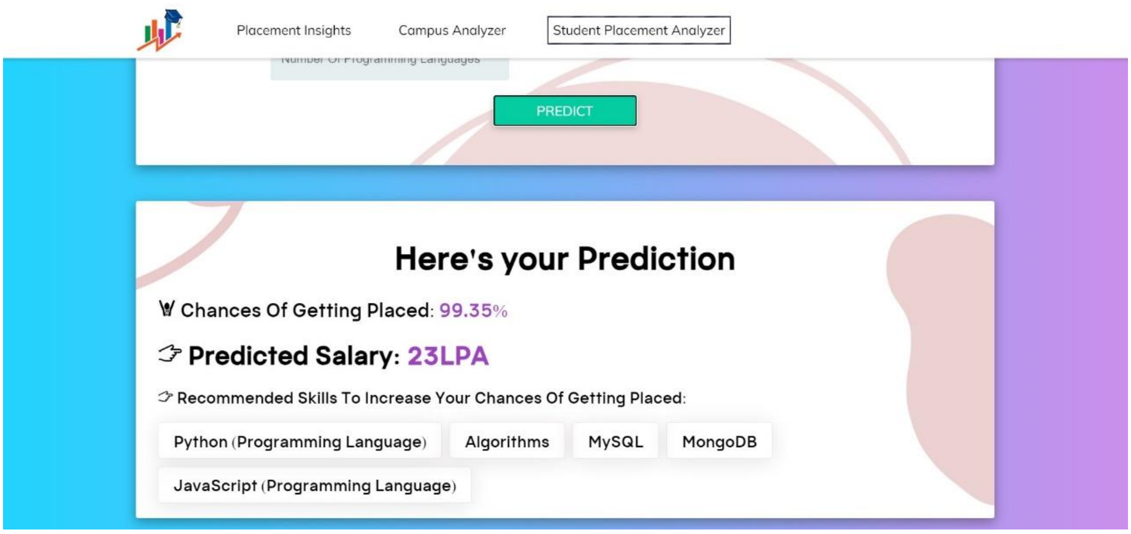

Fig 6.8: Predicted output