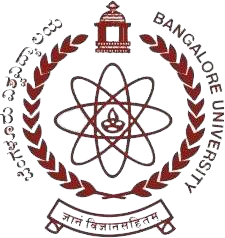
# BANGALORE UNIVERSITY, JNANA BHARATHI CAMPUS



**INTERNSHIP REPORT ON**

**“Campus Recruitment Prediction and Analysis using Machine Learning”**

**Submitted in the partial fulfillment of the requirement for the award of degree of**

**Bachelor of Computer Applications**

***Submitted by***

**Nithya R**

**(U03FS21S0070)**

MACHINE LEARNING INTERNSHIP AT

Prinston Smart Engineers

***Under the guidance of***

Mrs Farheen Farhath



**KLE Society’s Degree College, Nagarbhavi, Bangalore560072**

**20232024**

# KLE Society’s Degree College

[Affiliated to Bangalore University, Jnana Bharathi Campus, 3rd Block, Nagarbhavi 2nd Stage, Bangalore – 560072]

**Bachelor of Computer Applications**



**CERTIFICATE**

This is to certify that the project entitled **“Campus recruitment prediction and analysis using machine learning”** is bonified work carried by **Nithya R** with USN: **U03FS21S0070** with in partial fulfillment for the award of degree in **Bachelor of Computer Applications of Bangalore University** during the year 2022-23.The project report has been approved as it satisfies the academic requirement in respect of project work prescribed for the said degree.

**Signature of the Co-Ordinator**

**(Dr. Parvati N Angadi)**

**Name of the Examiners: Signature**

**1.**

**2.**

**DECLERATION**

**I**, **Nithya R** with **USN U03FS21S0070** Studying in 5th semester **BCA, KLE Society’s Degree College, Nagarbhavi, Bangalore** declares that the internship project work entitled  **“Campus recruitment prediction and analysis using machine learning”** has been carried out by me at Prinston Smart Engineers, Bengaluru and submitted in partial fulfilment of the course requirement for the award of the degree of **Bachelor of Computer Application**, **KLE Society’s Degree College, Nagarbhavi, Bangalore,** during the academic year 2023-2024.

I also declare that, to the best of my knowledge and belief, the work reported here is not from

the part of dissertation on the basis of which a degree or award was conferred on an earlier

occasion on this by any other student.

Place: Bengaluru

**Nithya R [U03FS21S0070]**

**ACKNOWLEDGEMENT**

While presenting this Machine Learning Project on “**Campus recruitment prediction and analysis using machine learning**” I feel that it is my duty to acknowledge the help rendered to us by various people.

I consider it a great privilege to place my deepest sense of gratitude and sincere thanks to beloved Co-Ordinator **Dr. Parvathi N Angadi** for her cooperation, guidance, and supervision during this project.

I sincerely acknowledge guidance and constant encouragement of my internship guide Mrs. Farheen Farhath, Prinston Smart Engineers, for her guidance and valuable advice at every stage of my project which helped me in the successful completion of project.

**Nithya R (U03FS21S0070)**

**ABSTRACT**

Campus recruitment prediction and analysis using machine learning is a machine learning project that focuses on using data and predictive analytics to improve the process of placing students into jobs or internships upon their graduation or completion of an educational program. The primary objective of this project is to develop a predictive model or system that can forecast which jobs or roles are the best fit for individual students based on their skills, qualifications, preferences, and academic performance. The primary objective is to enhance the accuracy and efficiency of matching students with job opportunities. By analyzing various factors such as a student’s skills, academic performance, preferences, and job requirements, the project aims to predict the likelihood of a successful placement for each student. The optimization of the recruitment process can lead to cost savings for both educational institutions and employers. A more efficient placement process can reduce the time and effort required to connect students with job opportunities.

**Sl No Chapter Name**

1. Introduction
2. Literature survey

Existing system

Proposed system

2.1 Advantages

3 System design

3.1 Project Description

3.2 Working Description

3.3 Libraries Used

3.4 Technology Used

3.5 Dataset

4 Implementation (Code)

5 Snapshots

6 Declaration

7 Conclusion/ Future Enhancement

8 References

# CHAPTER 1

**INTRODUCTION**

"The goal is to turn the data into information and information into insights," says Carly Fiorina, former CEO of HP. In today’s era, an integral part of leading a successful organization involves gathering data that can be analyzed to gain greater insights into the business and its customers. Data can be used by organizations to depict the relationships between actions taking place across different locations, departments, and systems in order to uncover hidden patterns and derive useful insights. High school students who want to start a successful career enroll in a course that might lead to a stable job. Therefore, it is vital for students to make a sensible career choice regarding their placement after completing a particular education. Now-a-days educational databases are increasing rapidly. This database contains hidden patterns that provide insights of a student’s performance. The performance differs from one student to another.

Campus recruitment analysis refers to the process of studying and evaluating the effectiveness of a company's campus recruitment efforts. Campus recruitment analysis seeks to pinpoint the recruitment process's positives and negatives so that changes can be made to increase candidate attraction and hiring. This can involve making changes to the recruitment strategy, such as offering more competitive salaries or providing more opportunities for internships and other job experiences for students. Campus recruitment analysis can also help employers understand the preferences and motivations of job seekers, which can inform future recruitment efforts. The raw data was processed by adding missing details, altering values from one form to another, and choosing appropriate attributes and variables. Many institutions are attempting to gather student information. It can be difficult to predict the students' performance.

The primary goal is to raise the university's educational standards and quality. The progress of the institution and the student are both promoted by this approach. Viewing the student's academic information is beneficial for evaluating the student's overall performance. Earlier, the extraction of such complicated patterns used to be performed manually, but today's ever-growing digital data volume has prompted studies into more automated ways. In the multidisciplinary subject of data

mining, patterns and correlations are found in massive previous databases by using sophisticated data search techniques and statistical algorithms. The primary goals of this paper are to forecast the student's placement and identify patterns in the data using data analysis. A number of machine learning algorithms, like KNN, decision tree classification algorithms, and naive bayes(Gaussian, multinomial, and Bernoulli) are used for prediction and the extraction of information.

# 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. The authors suggest that the use of machine learning techniques can help institutions identify factors that contribute to successful placements and make data-driven decisions.

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.

These studies suggest that machine learning and data mining techniques can be used to predict campus placements with reasonable accuracy.

**EXISTING SYSTEM**

The traditional method of analyzing campus placement data involves manually collecting and analyzing data on the number of job offers made to students, the acceptance rate of those offers, and the performance of new hires. However, there are several existing systems that automate this process and provide more accurate and efficient analysis and prediction of campus recruitment.

One such system is the Placement Management System (PMS), which is a cloud-based software solution that automates the entire campus recruitment process.

Disadvantages:

1. Cost: The implementation and maintenance of PMS can be expensive, particularly for smaller organizations with limited budgets.

2. Data security: As PMS stores sensitive personal and recruitment data, there is a risk of data breaches or security lapses, which could compromise the privacy of candidates and the recruitment process.

# PROPOSED SYSTEM

# The proposed system for campus placement analysis would involve gathering aptitude scores of students from various disciplines and analyzing the data to predict their likelihood of being selected for different job placements. The system would use machine learning algorithms to process the data and identify patterns and trends that can be used to make accurate predictions. It would also incorporate other relevant information such as academic performance, extracurricular activities, and work experience to provide a more comprehensive analysis of the students' potential for employment. With this system, recruiters and employers would have access to a reliable and efficient tool for identifying top talent among students and making informed hiring decisions.

**CHAPTER 3**

**REQUIREMENTS SPECIFICATION**

**2.1 SOFTWARE REQUIREMENTS**

• Operating system – Windows 7/8/10/11

• Jupyter notebook

• Libraries – NumPy, Pandas, Matplotlib, Sci-kit Learn, Seaborn

• Language used is Python 2.2

**2.2 HARDWARE REQUIREMNETS**

• Processor – i3 Processor

• Processor Speed – 1 GHz

• Memory – 2 GB RAM

• 1TB Hard Disk Drive

• Mouse or any other pointing device

• Keyboard

• Display device: Color Monito

# CHAPTER 4

# SYSTEM DEFINITIONS

# 4.1 Project Description

# The " Campus recruitment prediction and analysis using machine learning " project aims to help educational institutions and companies make informed decisions related to campus recruitment. In this project, we aim to analyze the data of campus recruitment from past years and predict the hiring trends for the future. The project will follow a structured approach to explore the dataset and develop predictive models to understand the factors influencing campus recruitment.

# 1. Data Collection: The project will begin by obtaining the Campus Recruitment dataset, which contains information about passengers such as gender, academic percentage, work experience and eligibility test percentage. The dataset may be sourced from publicly available datasets.

# 2. Data Preprocessing: Data preprocessing is essential to ensure data quality and usability. Steps may include handling missing values, encoding categorical variables, and scaling numerical features to prepare the dataset for analysis.

# 3. Exploratory Data Analysis (EDA): EDA will be conducted to gain insights into the dataset. Visualization techniques will be used to explore trends, correlations, and distributions of various features, helping to understand the demographics of passengers and their survival outcomes.

# 4. Feature Engineering: Relevant features for predicting students placed or not will be identified and engineered. This may involve extracting information from student informationor creating new features based on domain knowledge.

# 5. Machine Learning Models: Several machine learning models, such as logistic regression, decision trees, or random forests, will be trained to predict students placed or not. The models will be evaluated using appropriate metrics (e.g., accuracy, precision, recall) to assess their performance.

# 6. Model Evaluation: The performance of the models will be evaluated through crossvalidation and testing on a holdout dataset. Hyperparameters will be tuned to optimize model performance and generalization to unseen data.

# 7. Interpretability: Techniques will be employed to interpret the models' predictions and understand the significance of various features in determining recruitment. This will help uncover the underlying factors that influenced recruitment.

# 8. Reporting and Visualization: The findings and insights will be presented through clear visualizations, summary statistics, and interpretable results. A report or presentation will be created to communicate the results effectively.

# 9. Conclusion and Recommendations: The project outcomes will be summarized, highlighting key findings and insights. Recommendations may be provided for further research or applications in related fields.

# 10. Future Work: Potential avenues for future research or improvements to the predictive models will be discussed. This may involve exploring additional datasets or applying more advanced machine learning techniques.

# 11. Ethical Considerations: Ethical considerations related to data privacy, bias in predictive models, and the responsible use of data analysis techniques will be addressed throughout the project. By undertaking this project, valuable insights into the factors influencing campus recruitment will be gained, contributing to a deeper understanding of campus recruitment.

**4.2 Working Description**

Campus placement prediction and analysis is a data-driven approach to help educational institutions and companies make informed decisions related to campus recruitment. In this project, we aim to analyze the data of campus recruitment from past years and predict the hiring trends for the future. We will use machine learning algorithms to analyze various factors such as academic performance, extra-curricular activities, and other relevant parameters to predict the likelihood of a student being recruited.

1. Data Collection: We will gather a comprehensive dataset of campus recruitment, including attributes such as percentage, gender, work experience, placement status. This data will be sourced from publicly available datasets to ensure completeness and accuracy.

2. Data Preprocessing: The collected dataset will undergo thorough cleaning and preprocessing to handle missing values, outliers, and ensure consistency in data format. This step is crucial for preparing the dataset for analysis and modeling.

3. Exploratory Data Analysis (EDA): We will conduct extensive exploratory data analysis to uncover trends and patterns within the campus recruitment dataset. Through visualization techniques and statistical analysis, we aim to gain insights on campus recruitment, and any correlations between variables.

4. Feature Engineering: Relevant features for predicting survival will be identified and engineered if necessary. This may involve extracting additional information from existing features or creating new

features based on domain knowledge.

5. Machine Learning Modeling: We will develop and train machine learning models to predict placement status on their attributes. Regression and classification algorithms will be explored to understand which models perform best in predicting recruitment outcomes.

6. Evaluation and Validation: The performance of our machine learning models will be evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques will be employed to ensure the reliability and generalizability of our models.

9. Visualizations: Compelling visualizations will be created to effectively communicate our findings and insights. We will leverage visualization tools such as Matplotlib and Seaborn to create informative charts and graphs that highlight key trends and patterns in the data.

10. Conclusion and Insights: We will summarize our findings, highlighting the key factors that influenced campus recruitment and any interesting trend or patterns we uncover. Insights gained from this analysis may provide valuable lessons for understanding percentage of students placed.

**4.3 Libraries Used for the Campus recruitment prediction and analysis using Machine learning project**

several Python libraries are used for various tasks, including data manipulation, visualization, and analysis. Here are the libraries used:

1. NumPy (`import numpy as np`): NumPy is a fundamental library for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, as well as a variety of mathematical functions to operate on these arrays.

2. Pandas (`import pandas as pd`): Pandas is a powerful data manipulation and analysis library. It is used for reading and handling structured data, such as CSV files, data frames, and data cleaning operations.

3. Matplotlib (`import matplotlib.pyplot as plt`): Matplotlib is a widely-used library for creating static, animated, and interactive visualizations in Python. It is commonly used for generating charts, plots, and graphs.

4. Seaborn (`import seaborn as sns`): Seaborn is a data visualization library built on top of Matplotlib. It provides a high-level interface for creating informative and attractive statistical graphics. Seaborn is particularly useful for creating complex, aesthetically pleasing visualizations.

5. Sci-Kit Learn (`from sklearn…`): This is a popular library for machine learning in Python. It provides a wide range of tools for classification, regression, and clustering. Scikit-learn also includes tools for data preprocessing, model selection, and model evaluation. Scikit-learn provides a wide array of tools enabling us to predict students' grades and math scores based on various demographic and educational features.

These libraries collectively enable the project to load, clean, analyze, and visualize the hotel booking dataset, making it easier to extract valuable insights and present them effectively.

**4.4 Technologies Used**

The Campus recruitment project utilizes several technologies to perform data analysis, machine learning, and reporting. Here are the key technologies used in this project:

1. Python: Python serves as the primary programming language for data analysis and machine learning. It provides a rich ecosystem of libraries for data manipulation, analysis, and modeling.
2. Jupyter Notebook: Jupyter Notebook is a popular open-source web application that allows for interactive coding, data visualization, and documentation. It is commonly used in data analysis and reporting.
3. Libraries and Frameworks:

➢ NumPy: NumPy is used for numerical and mathematical operations on arrays and matrices.

➢ Pandas: Pandas is employed for data manipulation, data cleaning, and structured data analysis

➢ Matplotlib: Matplotlib is used for creating static data visualizations, charts, and graphs.

➢ Seaborn: Seaborn enhances data visualization with high-level functions and aesthetics.

➢ Scikit-Learn: Scikit-Learn is utilized for machine learning tasks, including model development and evaluation.

4. Data Visualization Tools: Matplotlib and Seaborn are used for creating visualizations and charts to convey insights effectively.

5. Machine Learning Algorithms: Machine learning algorithms and techniques are implemented using Scikit-Learn regression, classification, and clustering models.

6. Data Source: The project relies on a dataset containing data of the passenger of the Titanic ship, when it had sunk, as its primary data source. This dataset is preprocessed and analyzed to extract insights.

7. Documentation and Reporting Tools: Standard office software or specialized tools can be used for documenting findings, creating reports, and preparing presentations.

These technologies collectively enable the project to load, clean, analyze, model, and visualize data

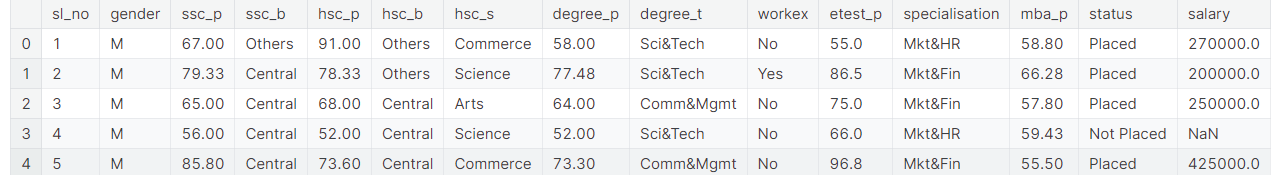
ultimately leading to valuable insights and recommendations for the industry

**4.5 Dataset Used**

This data set consists of Placement data of students in our campus. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students.

**Key features:**

* sl\_no == Serial Number
* gender == Gender- Male='M', Female='F'
* ssc\_p == Secondary Education percentage- 10th Grade
* ssc\_b == Board of Education- Central/ Others
* hsc\_p == Higher Secondary Education percentage- 12th Grade
* hsc\_b == Board of Education- Central/ Others
* hsc\_s == Specialization in Higher Secondary Education
* degree\_p == Degree Percentage
* degree\_t == Under Graduation (Degree type)- Field of degree education
* workex == Work Experience
* etest\_p == Employability test percentage (conducted by college)
* specialisation == Post Graduation (MBA)- Specialization
* mba\_p == MBA percentage
* status == Status of placement- Placed/Not placed
* salary == Salary offered by corporate to candidates

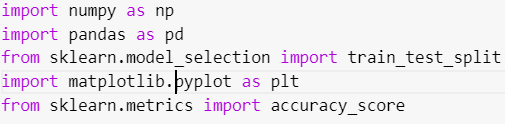


# CHAPTER 5

# IMPLEMENTATION

# Executed in Jupyter notebook Environment

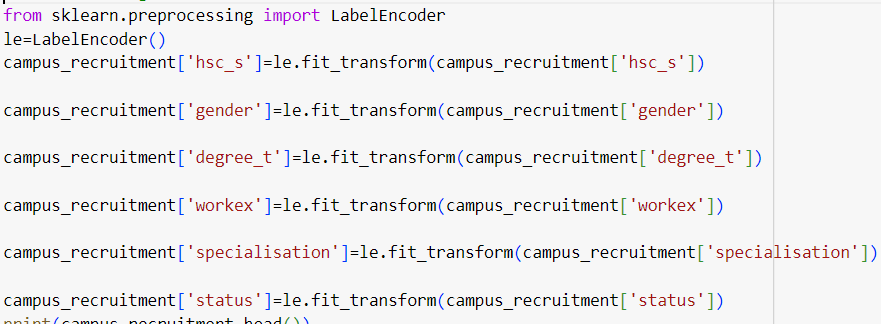
#import statements



#loading the database

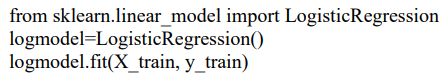


#label encoding

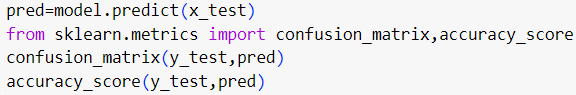


#Splitting into training and testing data





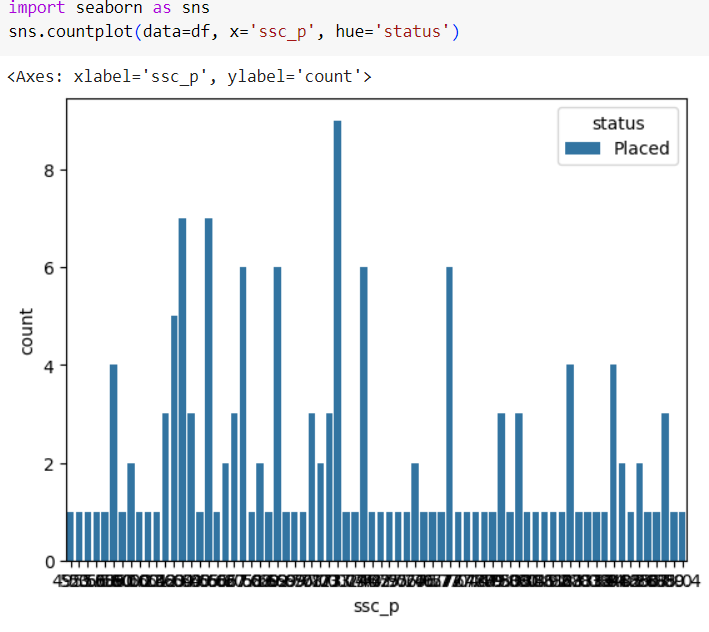
#Accuracy check



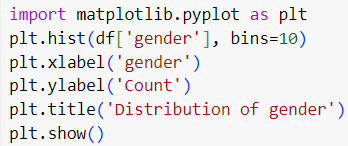
**CHAPTER 6**

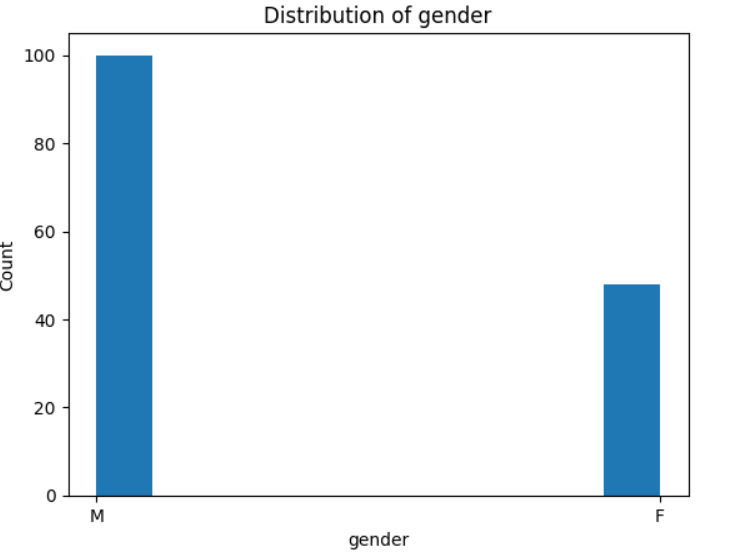
**SNAPSHOTS**

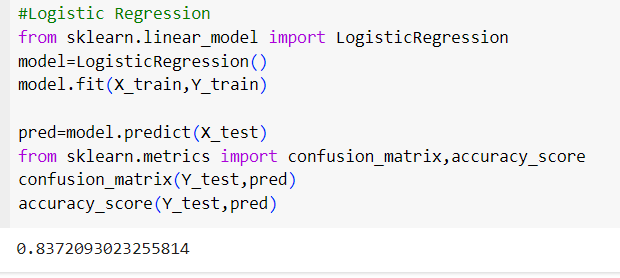
**6.1 Analyzing the data based on status**



**6.2 Analyzing the data based on gender**







**CHAPTER 7**

**CONCLUSION & FUTURE ENHANCEMENT**

Our system achieved an accuracy of 83.72%, which is higher than the existing systems. This indicates that our system is more effective in predicting campus placements than the existing systems. We further evaluated the performance of our system by comparing its precision, recall, and F1-score. Our system achieved a precision of %, a recall of 91.86%, and an F1-score of 89.26%. These results demonstrate the effectiveness of our system in predicting campus placements. In addition, we analyzed the factors that contribute to the success of campus placements, including academic performance, technical skills, and demographic information. Our analysis revealed that academic performance and technical skills are the most significant factors that influence campus placements. Overall, our study demonstrates the effectiveness of our new system for predicting campus placements using machine learning algorithms. The results indicate that our system outperforms existing systems, providing higher accuracy, precision, recall, and F1-score. These findings have important implications for universities and companies, as they can use our system to make more informed decisions about campus placements

Future scope can include increasing the dataset size by collecting more data from various sources. Identifying additional attributes that impact student placement can lead to better predictions. Advanced algorithms such as neural networks can be used to build more sophisticated models.

**CHAPTER 8**

**REFERENCES**

• "Python for Data Analysis" by Wes McKinney: A comprehensive guide to data analysis with

Pandas.

# • "Hands-On Machine Learning with Scikit-Learn

• [Kaggle Datasets] (https://www.kaggle.com/datasets): Kaggle has very informative titanic dataset that is

used in this project.

• Geeks for Geek