Data-Driven Placement Prediction for College Students Using Machine Learning

# CERTIFICATE

(This section will be filled as per institutional format)

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the Director of MIT Bengaluru, the Head of the Department of Computer Science and Engineering, and my project guide [Guide Name], for their constant support and guidance throughout this project. I am also thankful to the faculty members and technical staff of the department for their valuable assistance and for providing the necessary resources to carry out this work.

# ABSTRACT

This project aims to leverage machine learning and data analytics to predict placement outcomes for college students based on academic, technical, and extracurricular factors. The system analyzes features like CGPA, internships, certifications, project experience, and soft skills to forecast a student's placement probability. By using supervised learning models and visual dashboards, the project offers data-driven insights to help students and institutions optimize career readiness strategies. The end goal is to build a smart, interpretable tool that aids both students and placement cells in understanding employability trends.

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction to the Area

With the rise of data analytics, educational institutions can now leverage machine learning to understand student performance trends, employability potential, and placement outcomes. Predictive analytics helps transform student data into actionable insights.

## 1.2 Present-Day Scenario

Companies increasingly value not just academic scores but also skills, internships, and projects. Institutions seek automated tools that can assess a student's employability based on holistic profiles. This project addresses that need using ML.

## 1.3 Project Schedule

The project is divided into six phases: Planning and Data Collection, Preprocessing, Model Building, Dashboard Design, Optimization, and Final Submission — spread across Jan to June 2025.

## 1.4 Organization of the Report

Chapter 2: Literature Review  
Chapter 3: Project Objectives  
Chapter 4: Methodology  
Chapter 5: Conclusion

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

Placement prediction has been explored using logistic regression, decision trees, and ensemble models with student datasets involving academic and co-curricular data.

## 2.2 Summary of Related Work

Several works used Random Forests and SVMs for prediction accuracy. Some included psychometric and resume features, while others stuck to academic metrics. This project extends this by including internships, technical skills, and certifications.

## 2.3 Outcome of Review

There is a clear gap in user-facing dashboards and interpretable feedback systems. Most research lacks real-time prediction capabilities and holistic profiling, which this project aims to provide.

# CHAPTER 3: OBJECTIVES

The specific goals of the project are:

* To collect and preprocess student academic and co-curricular data.
* To identify the key features that influence placement success.
* To build predictive models that forecast placement likelihood.
* To visualize placement insights using an interactive dashboard.
* To suggest recommendations to students for improving employability.

# CHAPTER 4: METHODOLOGY

## 4.1 Methodology Overview

The approach involves data collection, preprocessing, feature engineering, ML model training, evaluation, dashboard development, and refinement.

## 4.2 Assumptions

* All data collected is anonymized.
* The placement decision is binary (placed/unplaced).
* Soft skill scores are approximated based on survey/feedback.

## 4.3 Tools Used

Python, Scikit-learn, Pandas, Seaborn, Streamlit/Flask  
Power BI/Tableau for visualization  
Jupyter Notebook for experimentation

## 4.4 Major Steps

* Data Collection: From placement cell, simulated/real data
* Preprocessing: Null handling, encoding categorical values, scaling
* Feature Engineering: CGPA, internships, certifications, projects
* Modeling: Logistic Regression, Random Forest, XGBoost
* Evaluation: Accuracy, F1-score, ROC-AUC
* Dashboard: Predictive score + recommendations

# CHAPTER 5: CONCLUSION

This project presents a complete pipeline to predict the placement outcome of a student using data analytics and machine learning. It offers insights into key performance indicators and career readiness factors. The project is designed to support students with actionable feedback and help placement cells with decision-making tools. Future extensions could include salary prediction and career track suggestions.

# REFERENCES

1. [1] Jain, A., et al., “Student Placement Prediction System using ML,” IEEE, 2021.
2. [2] Aggarwal, C., “Data Mining: The Textbook,” Springer, 2015.
3. [3] Kaggle Dataset: Factors Affecting Campus Placement. https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement
4. [4] Scikit-learn Documentation: https://scikit-learn.org
5. [5] Chennuri, S., “Advancing AI in Recruitment,” Texas A&M, 2023.