A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Yumiko Maria Bejarano Azogue |
| *Student Number* | 2024144 |
| *Module Title* | Machine Learning (10 ETCS) |
| *Assessment Title* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date* | 21st April 2024 |
| *Date of Submission* | 21st April 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Abstract

[NOTE: This section is designated for the abstract. Abstracts are not assigned page numbers and should precede the table of contents. If an abstract is unnecessary for your work, please delete this page.]

Attention: All notes must be removed from the document before submission!!

Contents

[Introduction 1](#_Toc158384946)

[Chapter 1 1](#_Toc158384947)

[Chapter 1.1 1](#_Toc158384948)

[Chapter 1.1.1. 1](#_Toc158384949)

[References 2](#_Toc158384950)

[NOTE: The table of contents above has been included for your convenience. To refresh the table, simply click on it, then select 'Update Table' using the mouse. You can choose to update either the page numbers exclusively or the entire table as needed.]

# Student Placement Analysis

## Introduction

In this project, is about to predict whether students will be placed in jobs based on various factors using machine learning techniques. Job placement is crucial for students and educational institutions as it directly impacts the perceived quality of education. The goal is to develop predictive and classification models that can accurately anticipate student job placement.

This report presents an analysis of a dataset containing information about students, including their age, gender, academic performance, internship experiences, hostel accommodation, history of backlogs, and placement status. The dataset comprises 2966 entries, with various numerical and categorical variables. The main objective of this project is to analyze the factors influencing students' placement and to develop predictive models for placement outcomes based on the available data.

## Objectives

The primary objectives of this project are as follows:

* Analyze the demographic distribution of students based on age and gender.
* Investigate the relationship between academic performance (CGPA) and placement status.
* Explore the impact of internship experiences and hostel accommodation on placement outcomes.
* Examine the correlation between different variables in the dataset.
* Develop predictive models, including neural network, decision tree, random forest, logistic regression, and support vector machine, to predict students' placement.

## Data Analysis

The dataset has the following columns:

* Age : Age At The Time Of Final Year
* Gender : Gender Of Candidate
* Stream : Engineering Stream That The Candidate Belongs To
* Internships : Number Of Internships Undertaken During The Course Of Studies, Not Necessarily Related To College Studies Or Stream
* CGPA : CGPA Till 6th Semester
* Hostel : Whether Student Lives In College Accomodation
* HistoryOfBacklogs : Whether Student Ever Had Any Backlogs In Any Subjects
* PlacedOrNot : Target Variable

### Age Distribution by Gender

Analysis indicates that the majority of students fall within the 21-22 age range, with a higher proportion of male students across all age groups.

### Placement Details by Gender

Although the total number of male students surpasses that of female students significantly, the percentage of successfully placed male students is slightly lower than that of females.

## Methods Used

This project employs various methods to analyze the dataset and predict job placement outcomes. These methods include:

Descriptive Statistical Analysis: Analyzing the distribution and summary statistics of different variables in the dataset.

Correlation Analysis: Examining the relationships between different variables to identify potential predictors of job placement.

Machine Learning Techniques: Training and evaluating machine learning models such as Neural Networks, Decision Trees, Random Forests, Logistic Regression, and Support Vector Machines to predict job placement outcomes.

Cross-Validation: Validating the performance of machine learning models using k-fold cross-validation techniques.

## Results Obtained

### Correlation Matrix

CGPA demonstrates a strong positive correlation with placement status, while hostel accommodation and history of backlogs exhibit weaker correlations.

### Predictive Models

The predictive models yield varying accuracy levels in predicting placement status. Support vector machines achieve the highest accuracy, with an 88.05% accuracy rate on the test dataset.

## Conclusion

In conclusion, this study sheds light on the determinants of students' placement outcomes. Academic performance, as gauged by CGPA, emerges as a pivotal predictor of placement success. Additionally, internship experiences and hostel accommodation may also influence placement outcomes. The predictive models developed herein offer valuable tools for educational institutions and recruiters to gauge students' likelihood of placement, thereby facilitating more informed decision-making.

## References