Project Report

Team #12

WANG Zeyu YANG Xirui Wu Tianxiao

April 4, 2025

Task-1 - Problem definition

Given the dataset with the academic and training performance of students. The goal is to train a model to predict whether each student will be placed (i.e., can find a job) based on the training set.

- Input: CGPA, Internships, Projects, Workshops/Certifications, AptitudeTestScore, SoftSkillsRating, ExtracurricularActivities, PlacementTraining, SSC_Marks, HSC_Marks;
- Ouput: Label for whether will be placed.

Task-1 - Prepare - Data distribution

Task-1 - Prepare - Correlation coefficient

Task-1 - Analyze - Logistic Regression

Given the dataset $(\mathbf{x}_i, y_i)_{i=1}^n$, the logistic regression aims to fit the function

$$f(\mathbf{x}) = \frac{1}{1 + \exp(-(\mathbf{x}^T\mathbf{w} + b))}$$

with the binary cross entropy loss function

$$\text{loss}\left((x_i, y_i)_{i=1}^n\right) = \sum_{i=1}^{n} \left(-y_i \log(f(x_i)) - (1 - y_i) \log(1 - f(x_i))\right).$$

Task-1 - Performance

Model Parameter	Time (Train/Test)	Memory	F1 Score (Train/Test)

Table: Model Performance.

Task-2 - Problem definition

Task-2 - Prepare

Task-2 - Analyze

Task-2 - Performance

Task-3 - Problem definition

Task-3 - Prepare

Task-3 - Analyze

Task-3 - Performance

Reference I