**NED UNIVERSITY OF ENGINEERING & TECHNOLOGY**

Centre for Multidisciplinary Postgraduate Programmes (CMPP) – NED Academy

Postgraduate Diploma in Data Science & AI

Final Examination Fall-I-2025

Course: **PGD / DSAI-562** **Machine Learning**

Time: 3 Hours Max. Marks: 60

Instructions:

1. Attempt all questions
2. Marks for each question are given.
3. You are required to abide by all rules and regulations set for the examination by the NED Academy.
4. Total time of examination including uploading is 3 hours. No extra time will be provided after the time is over.

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Question** | **Marks** |
| 1. | Download the following data set  <https://drive.google.com/file/d/1we2ZCzvLemI-fLWCFuWVSHKHs63zlH-f/view?usp=sharing>  **Loan Status Prediction**  Predict the loan to be **approved or to be rejected** for an applicant.   * In this Loan Status Prediction dataset, we have the data of applicants who previously applied for the loan based on the property which is a Property Loan. * The bank will decide whether to give a loan to the applicant based on some factors such as Applicant Income, Loan Amount, previous Credit History, Co-applicant Income, etc... * Our goal is to build a Machine Learning Model to predict the loan to be approved or to be rejected for an applicant. * Try to achieve accuracy more than 85% * Perform Data pre-processing if required. * Perform EDA and show visualization among different features * Select a model for your data and train your model on 80%-20% * Plot confusion matrix * plot classification report | 15 |
| 2. | (a) What is Over fitting, and How Can You Avoid It?  (b) What do you understand by the term Bias Variance trade off?  (c) When Will You Use Classification over Regression?  (e) What is Cross-Validation?  (f) What do you understand by the F1 score? Calculate F1from the following confusion matrix. | 15 |
| 3. | **Fraud Detection using Pipeline Architecture**  <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>   1. Perform **EDA** (check imbalance between fraud and non-fraud cases). 2. Build a **Pipeline** that includes preprocessing + classification. 3. Compare **Logistic Regression, Decision Tree, and Random Forest**. 4. Apply **hyper parameter tuning** and handle **class imbalance** (class weights / resampling) | 15 |
| 4. | **Medical Insurance Cost Prediction**  <https://www.kaggle.com/datasets/mirichoi0218/insurance>   1. Perform EDA (impact of age, BMI, smoking on charges). 2. Perform preprocessing (encoding categorical features, scaling numerical features). 3. Train and compare with different Regressors. 4. Use GridSearchCV to tune hyperparameters and evaluate using R² and RMSE. | 15 |