ID:	Name:

Green University of Bangladesh

Department of Computer Science and Engineering

Lab Final Examination, summer 2025

Course Code: CSE 412 | Course Title: Machine Learning Lab

Full Marks: 30 | Time: 1 Hour 20 Minutes

Employee Performance and Salary Analysis

The HR analytics team at DataCorp aims to analyse employee data to predict performance, segment employees into groups based on work-related metrics, and build predictive models to identify high-performing employees. The analysis will help in strategic HR decision-making, including promotions, salary adjustments, and remote work policies.

Objective:

1. Build predictive models to forecast employee **performance scores.**
2. Cluster employees into groups with similar work-related attributes for HR insights.
3. Develop an advanced deep learning model for salary prediction.

Tasks:

1. Linear Regression (Salary Prediction) [3+4+3]

I. **Data Preparation (Dataset)**
   - ○ Identify and handle missing or outlier values in Age, Salary, and Years_at_Company.
   - ○ Explore relationships between Years_at_Company, Number_of_Projects, and Salary using visualizations.

II. **Model Building:**
   - ○ Split the dataset into training (80%) and testing (20%) sets.
   - ○ Build a linear regression model to predict Salary based on relevant features.

III. **Model Evaluation:**
   - ○ Evaluate the model using Mean Absolute Error (MAE) and R^2 score.
   - ○ Provide insights on how features influence salary.

2. Naïve Bayes Classification (Employee Performance Classification) [2+2+3+3]

I. **Target Attribute:** `Performance_Score` will be converted into categories for classification. Since it is originally a numerical value (0–100), it should be discretized into **High Performance** (e.g., score ≥ 80) and **Low Performance** (score

< 50). The rest of the employees are **Medium Performance.**

II. **Data Preparation:** Encode categorical variables. Normalize or scale the numerical feature.

III. **Model Building:** Train a Naïve Bayes classifier to predict employee performance categories based on work-related and demographic attributes.

IV. **Model Evaluation:** Assess the classifier using accuracy, precision, recall, and F1-score. Provide discussion on the strengths and weaknesses of Naïve Bayes in this context compared to neural networks.

3. Neural Networks (Performance Classification) [5+5]

I. Model Building:
   ○ Build a feedforward neural network using Keras/TensorFlow to classify high vs. low performance.
   ○ Train for 10 epochs and track accuracy.

II. Model Evaluation:
   ○ Evaluate the model using accuracy, precision, recall, and F1-score.
   ○ Compare performance with the **Naive Bayes** model's predictive capability on the performance score.

Data Description

The dataset synthetic_employee_data.csv includes the following columns:

- Employee_ID: Unique identifier for each employee.
- Employee_Name: Name of the employee.
- Age: Age of the employee.
- Salary: Annual salary in USD.
- Years_at_Company: Number of years the employee has been with the company.
- Number_of_Projects: Number of projects the employee is involved in.
- Performance_Score: Performance score ranging from 0 to 100.
- Is_Manager: 1 if the employee is a manager, 0 otherwise.
- Works_Remotely: 1 if the employee works remotely, 0 otherwise.
- Department: Department in which the employee works.
- Education_Level: Highest level of education attained.
- Location: Work location of the employee.
- Hire_Date: Date the employee was hired.