Department of CSE, Jagannath University Dhaka-1100 Bangladesh., **Phone:** 01929-125856 **E-mail**: nasir.jnu.cse@gmail.com

Machine Learning with Python

Final Examination for Batch - IUBAT I 08

Vanue: IUBAT software lab - 08
Date & Tme: 04/07/2025, 3:30 PM

Time: 150 Minutes
Full Marks: 25

- You are required to complete **all task** step by step.
- Submit your Jupyter Notebook (.ipynb) file at the end of the exam.

Dataset Overview:

The visitor_count_dataset.csv contains various features such as time_segment, DayOfWeek, holiday indicators (is_official_holiday, is_school_holiday, is_bridge_day), Weekday, Month, seasonal flags (Winter, Spring, Summer, Autumn), post-holiday indicators (after_holiday, after_bridge, after_school_holiday), and weather-related attributes (precip_quantity, temp_avg, wind_speed_10m, humidity_rel_shelter_avg, pressure, sun_duration). This dataset is structured for a machine learning regression problem where the goal is to predict the VisitorCount, which serves as the target column. Please download the dataset from google classroom.

Task 1: Data Exploration and Visualization (5 Marks)

a) Load the dataset and display the first 5 rows.
b) Display summary statistics of the dataset.
c) Check for missing values.
d) Plot the distribution of any 2 numerical and 2 categorical features.
e) Plot the correlation matrix (heatmap) among numerical features.

Task 2: Data Preprocessing and feature engineering (10 Marks)

a) Handle missing values and outlier appropriately (imputation or removal).
b) Show best 5 features using multiple feature engineering technique
c) Encode categorical variables properly using Label Encoding or One Hot Encoding.
d) Normalize or standardize numerical features where necessary.
e) Split the dataset into training and testing sets (80%-20%).

Task 3: Model Building and Evaluation (6 Marks)

a) Train at least 2 regression models (e.g., linear Regression, decision Tree, Random Forest, XGBoost, etc.).
b) Evaluate the models using MSE, and R2 score.
c) Plot the result comparison for both models.
e) Based on your results, explain which model is more appropriate for predicting **VisitorCount** and why.

.

Task 5: Conclusion and Recommendations (4 Marks)

a) Write a short conclusion summarizing your findings.

b) Provide at least 2 recommendations that could help with more precise prediction based on your analysis.