Digital Empowerment Network

Machine Learning

Week 01

Building a Machine Learning Model for Real-world Prediction Task

Mentor : Hussain Shoaib

Task Title:

Building a Machine Learning Model for Real-world Prediction Task

Objective:

To understand and implement the workflow of a machine learning project including data preprocessing, feature selection, model building, training, testing, and evaluation using a real-world dataset.

Step 1: Dataset Selection

Choose one of the following datasets or get approval for a dataset of your choice:

• Heart Disease Prediction Dataset (e.g., from UCI or Kaggle)

• Student Performance Dataset

• Iris Dataset or any other classification dataset

Step 2: Task Instructions

A. Data Understanding & Preprocessing:

• Load and explore the dataset

• Handle missing values and remove duplicates

• Convert categorical variables into numerical format

• Normalize or scale the features

B. Exploratory Data Analysis (EDA):

• Visualize relationships between features

• Use statistical summaries and graphs (scatter plots, bar graphs, heatmaps)

• Identify trends or correlations in the data

C. Feature Selection:

• Select relevant features for model training using correlation or importance analysis

D. Model Building & Evaluation:

• Train at least two ML models (e.g., Decision Tree, SVM, Random Forest, or Logistic Regression)

• Use performance metrics (Accuracy, Precision, Recall, F1-Score, Confusion Matrix)

• Compare models and interpret the results

E. Optional (Bonus):

• Tune hyperparameters using GridSearchCV

• Deploy your model using a simple interface (like Streamlit or Flask)

Deliverables:

• Jupyter Notebook or Python script with well-commented code

• A brief report or presentation explaining your approach and results

• (Optional) Video presentation of your project

Deadline:

Please complete and submit your task within 7 days from the date of assignment.

Last Date 23-July-2025