

Final Project

Project Overview:

In this project, teams of **minimum of 2 and maximum of 5** will collaborate to explore, analyze, and model a dataset using a Neural Network to either predict an outcome (classification) or estimate a value (regression).

Tasks:

1. Dataset:

Find a dataset containing at least 10,000 records.

2. Exploratory Data Analysis (EDA):

Conduct a thorough EDA to uncover patterns, anomalies, trends, and relationships within the data. Visualizations should be used to help understand the distribution of data and the relationships between features.

3. Data Preprocessing:

This should cover issues like missing values, outliers, and inaccurate data entries. It should also cover the pipeline in which the data flows (scaling, imputation, etc.).

4. Network Development:

Build a Neural Network using your preferred framework, focusing on either classification or regression. The model should be robust, and its parameters should be fine-tuned to get optimal performance. Evaluate the model using appropriate metrics.

Deliverables:

1. Python code.
2. Presentation that includes:
 - i. An overview of the dataset, explaining the types and nature of features.
 - ii. Insights and visualizations from the EDA.
 - iii. Data preprocessing pipeline.
 - iv. Network structure and training process including parameter tuning and model evaluation.

Discussion:

The project discussion is scheduled for Saturday, May 11th, at 2:30 PM in Hall 415.