

Final Project

Project Overview:

In this project, teams of **minimum of 3 and maximum of 5** will collaborate to explore, analyze, and model a dataset using SVM to either predict an outcome (classification) or estimate a value (regression).

Tasks:

1. Dataset:

Find a dataset containing more than 10,000 records and at least 20 features.

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 Cleaning:

This should cover issues like missing values, outliers, and inaccurate data entries.

4. Dimensionality Reduction:

Implement dimensionality reduction technique(s) covered in this course to reduce the number of features while retaining helpful information.

5. SVM Model Development:

Build an SVM model, 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. Dimensionality reduction technique(s) used.
 - iv. SVM training process including parameter tuning and model evaluation.