# Mid-Term Project Guidelines

Course Title: Machine Learning & Data Science

Instructor: Gulshan Yasmeen

Institution: NAVTAC AI Training Program

Project Type:

Submission Deadline: October 19, 2025

## 1. Project Overview

The Mid-Term Project is designed to give students hands-on experience in applying the complete data science and machine learning pipeline. Each student is required to choose a project idea of their own interest and implement all the key steps we have covered in class — from data collection and preprocessing to model building and evaluation.

The project should demonstrate your understanding of the core concepts we have studied so far, including:  
- **Data mining and exploration  
- Data preprocessing and cleaning  
- Data visualization and feature understanding  
- Model selection and application of machine learning algorithms  
- Evaluation and performance reporting  
- Documentation and presentation of your results**

## 2. Objectives

1. Apply theoretical concepts learned in class to a practical, real-world problem.
2. Demonstrate end-to-end understanding of the data analysis and model-building process.
3. Develop skills in data wrangling, visualization, and predictive modeling.
4. Improve research, documentation, and presentation abilities.

## 3. Project Requirements

Each project must include the following sections:

### A. Project Proposal

* Title: Clearly state the name of your project.
* Problem Statement: Describe the real-world problem you are trying to solve or explore.
* Objectives: What do you aim to achieve through your analysis or model?
* Dataset Description: Mention the dataset you are using — its source, size, and key features.

### B. Data Mining and Exploration

* Collect or identify a dataset related to your project idea.
* Perform initial data exploration to understand structure, size, missing values, and feature distributions.
* Use descriptive statistics and visualizations (histograms, boxplots, scatterplots, etc.) to summarize findings.

### C. Data Preprocessing

* Handle missing values, duplicates, and outliers.
* Encode categorical variables and normalize or scale numerical features where necessary.
* Split the dataset into training and testing sets for modeling.

### D. Data Visualization

* Create clear and meaningful visualizations to represent patterns and relationships in data.
* Use libraries such as Matplotlib, Seaborn, or Plotly to produce high-quality plots.

### E. Model Development

* Apply at least one supervised and one unsupervised machine learning algorithm.
* Explain the reason for choosing each algorithm.
* Train and test your models and compare performances.

### F. Model Evaluation

* Evaluate model performance using metrics (Accuracy, Precision, Recall, F1-score, AUC-ROC, etc.).
* For clustering, use metrics such as inertia or silhouette score.
* Interpret the results and discuss model implications.

### G. Conclusion and Recommendations

* Summarize key insights and findings.
* Discuss limitations and suggest future improvements.

### H. Documentation

* Submit a comprehensive project report (Word or PDF).
* Include cover page, table of contents, screenshots of code, outputs, and references.

## 4. Deliverables

* Project Report (Word or PDF): Complete documentation as described above.
* Python Notebook (.ipynb): Containing the implementation of your project.
* Presentation: A short summary of your project to present in class.

## 5. Example Project Ideas

* Predicting house prices based on location and features
* Sentiment analysis of Twitter data
* Customer segmentation for a retail store
* Disease prediction using medical data
* Credit card fraud detection
* Student performance analysis and prediction
* Air quality or weather pattern analysis

## 6. Submission Instructions

Submit all files on your Github profiles and add the instructor as collaborator.

Name your files in the following format: StudentName\_ProjectTitle.ipynb and StudentName\_ProjectReport.docx