



DSA 1080B FS 2024 Group Project

Project Title: Predicting House Prices Using Regression Analysis

Week 1: Project Introduction and Data Collection

Tasks:

Explain the problem: "Predict the price of houses based on various features such as size, location, number of rooms, etc."

Discuss the importance of regression analysis in predictive modeling.

Data Collection:

Provide students with a dataset from Kaggle or an open dataset (e.g., House Prices dataset).

Tools: Python, Jupyter Notebook, pandas.

Week 2: Data Exploration and Cleaning

Tasks:

exploratory data analysis (EDA) techniques:

Descriptive statistics (mean, median, etc.)

Identifying missing values, outliers, and data types.

Clean the data:

Fill or drop missing values.

Handle outliers.

Normalize or scale numerical data if necessary.

Tools: pandas, matplotlib, seaborn for data visualization.

Week 3: Feature Engineering and Selection

Objective: Enhance the dataset by creating new features and selecting the most relevant ones.

Tasks:

Explain what feature engineering is and why it's important.

Create new features (e.g., price per square foot, interaction features).

Feature selection using correlation analysis and variance inflation factor (VIF).

Split the dataset into training and test sets.

Tools: pandas, scikit-learn for splitting data.

Week 4: Model Building

Objective: Guide students in building a regression model.

Tasks:

Explain the concept of linear regression and how it works.

Train a simple linear regression model on the training dataset.

Evaluate the model using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

Tools: scikit-learn for regression modeling and evaluation.

Week 5: Model Evaluation and Reporting

Objective: Finalize the project by evaluating the model and preparing a final report.

Tasks:

Evaluate the model on the test dataset.

Discuss model performance and possible improvements (e.g., adding more features, trying other algorithms like Decision Trees).

Problem definition.

Data exploration and cleaning process.

Model building and results.

Conclusion and next steps.

Tools: scikit-learn, Jupyter Notebook for reporting.

Deploy the Model:

Flask or Streamlit for deploying the model as a web application.