Report of Project 2 and Project 3

Project :2
Simple Linear Regression Housing Prices

Introduction

The Boston Housing dataset is a classic dataset used for regression tasks. It contains information about housing in Boston, including features such as average number of rooms, property tax rate, and house prices by selection and normalization.
In this project, we build a simple linear regression model to study the relationship between the average number of rooms per dwelling (RM) and the median house price (MEDV).

Methodology

->Load the dataset (Boston Housing in csv file) and explore features.
->Select two key variables:
1. RM → Average number of rooms.
2. MEDV → Median house price.
->Split the data into training and testing sets (80% / 20%).
->Implement the Least Squares Method to compute slope (m) and intercept (c).
->Train a regression line to predict house price from the number of rooms.
->Evaluate performance using R^2 Score and Mean Squared Error (MSE).
->Visualize the regression line against actual values.

Results

The regression line shows a <u>positive correlation</u> between number of rooms and house price.
As the number of rooms <u>increases</u>, the predicted house price also <u>increases</u>.
The evaluation metrics (R^2 Score, MSE) indicate how well the model fits the data.

Conclusion

The simple linear regression model demonstrates that the average number of rooms (RM) is a significant predictor of house price in Boston. However, housing prices are influenced by many other factors. A more advanced regression model with multiple features would yield higher accuracy.

Project :3
Loan Eligibility Prediction

Introduction

Banks and financial institutions often face challenges in determining whether a loan application should be approved or not approved. Automating this process with machine learning helps in making faster and more consistent decisions.
In this project, we build a Loan Eligibility Prediction Model using machine learning algorithms.

Methodology

->Load the dataset (Loan Dataset in csv file).
->Preprocess the data:
->Remove unnecessary column (Loan_ID).
->Handle missing values with mode (categorical) and median (numerical).
->Encode categorical variables using LabelEncoder.
->Split dataset into training and testing sets.
->Train two models:
1. Logistic Regression (with feature scaling).
2. Random Forest Classifier.
->Evaluate models using accuracy, classification report, and confusion matrix.
->Visualize:
1. Confusion matrix heatmap.
2. Feature importance (Random Forest).
3. Save best-performing model (loan_eligibility_model.pkl).
->Show sample predictions with Approved / Not Approved status.

Results

Logistic Regression Accuracy: ~78–79%
Random Forest Accuracy: ~75–76%
Confusion matrix shows that most approved loans were correctly predicted.
Feature importance highlights which features like applicant income, loan amount, credit history influence loan approval the most.

Conclusion

The project demonstrates how machine learning can be applied to predict loan eligibility.
Logistic Regression performed slightly better in this dataset compared to Random Forest.
The model can assist banks in making faster and more reliable loan approval decisions.
With more data and advanced techniques e.g., hyperparameter tuning, ensemble methods, accuracy can be further improved.