Regression Project

Salary Prediction

Project Introduction:

Predict employee salary based on their years of experience using Simple Linear Regression.

Dataset:

- **Name:** Salary Dataset
- **Features:** Years Experience
- **Target:** Salary
- **Use Case:** Best suited for learning linear regression concepts.

Preprocessing Steps:

- Data reading using pandas
- Missing value & duplicate removal
- Outlier detection and cleaning

- Standardization
- Dimensionality reduction
- Handling
- Train-test splitting

Modeling:

- Model: LinearRegression from sklearn
- Fitted on training data to learn relationship between experience and salary.

Evaluation:

- **R^2 Score**: Measures how well data fits the model
- **MSE (Mean Squared Error)**: Measures prediction error
- Visualization of regression line on data

Key Findings:

- Strong positive linear relationship found
- Higher experience → higher predicted salary
- Minimal preprocessing was sufficient

Conclusion:

- Linear regression is effective for linearly correlated numeric data.
- Ideal for beginners to learn model evaluation & interpretation.

Classification Project

Loan Prediction

Project Introduction:

Classify whether a loan should be approved based on applicant information.

Dataset:

- Likely includes multiple features (e.g., gender, income, credit score)
- Target: Loan Status (Approved/Not Approved)

Preprocessing Steps:

- Data reading using pandas
- Missing value & duplicate removal
- Outlier detection and cleaning
- Standardization

- Dimensionality reduction
- Handling
- Train-test splitting

Modeling:

- Algorithms: Logistic Regression / Decision Tree
- Trained to classify loan approval outcome

Evaluation:

- **Accuracy**
- **Precision, Recall**
- **Confusion Matrix**

Key Findings:

- Decision Tree offered better interpretability

- Feature scaling improved KNN performance
- Imbalanced classes handled using undersampling

Conclusion:

- Classification models are effective for binary decision tasks
- Performance depends heavily on preprocessing and model choice