Linear Regression

Machine Learning Project Pipeline/ Workflow

1. Business Problem understanding (Domain Expertise)

2. Data ingestion/Collection → SQL

   (Raw data)

3. Data understanding → Rows (Records), Column (features)

85-90% time → 4. EDA (Exploratory Data Analysis)

(Statistics)

Data Normalisation/ Standardization ←

→ outliers
↳ deal

↓

Missing Values ←
└ Imputation

Quality data

1. Remove
2. focus on what kind of Cost function
3. To get the best optimal Value

Visualization
↳ Correlation heatmap

feature selection → heatmap

feature engineering

1000 Records

Data Split

Plots → seaborn, matplotlib

Train set (80%) 800 Records

Test set (20%) 200 Records

5 Model Training → which specific algorithm suits the most??

6 Model Testing ← unseen data for the model

Model Predictions $(\hat{y})$

Actual values (y)

7 Model Evaluation

Regression
- MSE
- MAE
- RMSE
- R-squared
- Adjusted R-squared

classification task

Confusion matrix
- Accuracy
- Precision
- Recall
- f1-score

8 Model Deployment

Data Standardisation

Biased Model

2

2500

'MedInc', 'HouseAge', 'AveRooms', 'AveBedrms', 'Population', 'AveOccup', 'Latitude', 'Longitude'

data will be within one specific range

↳ model to be unbiased

① Standard Scaler → standard normal form

↳ $\left\{ \begin{array}{l} \mu = 0 \\ \sigma = 1 \end{array} \right.$

$z\text{-score} = \frac{x_i - \mu}{\sigma}$

② Min-max Scaler

fit_transform → train data

(μ, σ)

transform → test data

(μ, σ)

Age	SNF
23	
45	
67	
49	
62	
22	

fit()

↳ μ, σ

transform()

↳ $z\text{-score} = \frac{x_i - \mu}{\sigma}$

$$\hat{y} = m_1x_1 + m_2x_2 + ------ + m_8x_8 + c$$

Coefficients = 8

intercepts = 1

MAE vs MSE

amplify error

MSE → $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

MAE → $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$ — Residuals

→ more robust to outliers

$$RMSE = \sqrt{MSE}$$

→ ideal → house price

→ 5 lakh

(high error

temp prediction in celsius

RMSE = 2°c → may be acceptable

0.5 lakh

→ decent

10°C → poor prediction

Lower RMSE → model prediction will be better