EXPERIMENT:8

AIM: Illustrate data science lifecycle for selected case study. (Prepare case study document for the selected case study) sample case studies :
1. Customer Segmentation
2. Fraud Detection
3. House Price prediction
4. Product Recommendation
5. Stock price prediction
6. Weather prediction

Theory:

Case Study: House Price Prediction

1. Introduction
In this case study, we aim to predict house prices using a dataset containing various features such as the size of the house, location, number of bedrooms, etc. House price prediction is a common problem in real estate and finance industries, and accurate predictions can aid buyers, sellers, and real estate agents in making informed decisions.

2. Data Collection
The dataset used for this analysis is sourced from a real estate database or public repository.
It includes features like square footage, number of bedrooms, number of bathrooms, location (e.g., city, neighborhood), year built, and sale price.
Ensure data quality by handling missing values, outliers, and inconsistencies.

3. Data Preprocessing
- Handling Missing Values: Impute missing values using methods such as mean, median, or predictive modeling.
- Outlier Detection: Identify and handle outliers that may skew the model's predictions.
- Feature Engineering: Create new features or transform existing ones to better represent the underlying patterns in the data.
- Encoding Categorical Variables: Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.
- Feature Scaling: Scale numerical features to ensure all features contribute equally to the model.

4. Exploratory Data Analysis (EDA)
Perform descriptive statistics to understand the distribution and relationships between variables.
Visualize distributions, correlations, and trends using techniques such as histograms, scatter plots, and heatmaps.
Identify key insights and patterns that may inform feature selection and model building.

5. Feature Selection
Use techniques like correlation analysis, feature importance scores, or domain knowledge to select the most relevant features for the prediction task.
Remove redundant or irrelevant features to improve model efficiency and interpretability.

6. Model Selection
Choose appropriate regression algorithms for house price prediction, such as linear regression, decision trees, random forests, or gradient boosting models.
Split the data into training and testing sets to evaluate the performance of different models.
Experiment with various hyperparameters and model architectures to optimize performance.

7. Model Training
Train selected models on the training dataset using techniques like cross-validation to assess generalization performance.
Tune hyperparameters using techniques like grid search or randomized search to improve model performance.

8. Model Evaluation
Evaluate model performance on the testing dataset using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).
Compare the performance of different models and select the best-performing one for deployment.

9. Deployment
Deploy the trained model into a production environment, either as a standalone application or integrated into an existing system.
Monitor model performance in real-time and update as necessary to maintain accuracy and reliability.

10. Conclusion
House price prediction is a valuable application of data science in the real estate industry. By leveraging machine learning techniques and a structured data science lifecycle, accurate predictions can be made to assist buyers, sellers, and real estate professionals in making informed decisions.

This document outlines the data science lifecycle for the house price prediction case study, covering data collection, preprocessing, exploratory data analysis, model selection, training, evaluation, deployment, and conclusion. Following this lifecycle ensures a systematic approach to solving the problem and delivering actionable insights.

Hence, We have successfully illustrated data science lifecycle for house price prediction case study