Task 10

Machine Learning

Upload .py or Ipynb extension file on GitHub public repo “100DaysofBytewise" and share the link in the submission form by 18 July 2024.

1. Handling Missing Data in Titanic Dataset
   - Task:Identify and handle missing values in the Titanic dataset. Experiment with different strategies such as mean/median imputation, mode imputation, and dropping rows/columns.
   - Dataset: Titanic Dataset

2. Encoding Categorical Variables in a Car Evaluation Dataset
   - Task: Encode categorical variables in the Car Evaluation dataset using one-hot encoding and label encoding. Compare the results.
   - Dataset: Car Evaluation Dataset

3. Scaling Features in the Wine Quality Dataset
   - Task: Apply normalization and standardization to the features in the Wine Quality dataset. Analyze how scaling affects the distribution of data.
   - Dataset: Wine Quality Dataset

4. Handling Outliers in the Boston Housing Dataset
   - Task: Identify and handle outliers in the Boston Housing dataset using techniques like Z-score, IQR, and visualization methods.
   - Dataset: Boston Housing Dataset

5. Data Imputation in the Retail Sales Dataset
   - Task: Handle missing values in the Retail Sales dataset using advanced imputation techniques like KNN imputation and MICE.
   - Dataset: Retail Sales Dataset

6. Feature Engineering in the Heart Disease Dataset
   - Task: Create new features from existing ones in the Heart Disease dataset, such as age groups, cholesterol levels, and more.
   - Dataset: Heart Disease Dataset

7. Transforming Variables in the Bike Sharing Dataset
   - Task: Apply transformations like log, square root, and Box-Cox transformations to skewed variables in the Bike Sharing dataset.
   - Dataset: Bike Sharing Dataset

8. Feature Selection in the Diabetes Dataset
   - Task: Use techniques like correlation analysis, mutual information, and recursive feature elimination (RFE) to select important features in the Diabetes dataset.
   - Dataset: Diabetes Dataset

9. Dealing with Imbalanced Data in the Credit Card Fraud Detection Dataset
   - Task: Handle imbalanced data in the Credit Card Fraud Detection dataset using techniques like SMOTE, ADASYN, and undersampling.
   - Dataset: Credit Card Fraud Detection Dataset

10. Combining Multiple Datasets in the Movie Lens Dataset
    - Task: Combine and preprocess multiple related datasets from the Movie Lens dataset, such as ratings, user information, and movie metadata.
    - Dataset: Movie Lens Dataset