Task 12

Machine Learning

Upload .py or Ipynb extension file on GitHub public repo “100DaysofBytewise" and share the link in the submission form by 23 July 2024.

1. Predicting Employee Attrition Using Logistic Regression
   Dataset: HR Analytics Employee Attrition Dataset
   Preprocessing Steps:
   - Handle missing values if any.
   - Encode categorical variables (e.g., one-hot encoding for department, gender, etc.).
   - Standardize numerical features.
   Task: Implement logistic regression to predict employee attrition and evaluate the model using precision, recall, and F1-score.

2. Classifying Credit Card Fraud Using Decision Trees
   Dataset: Credit Card Fraud Detection Dataset
   Preprocessing Steps:
   - Handle missing values if any.
   - Standardize features.
   Task: Implement a decision tree classifier to classify credit card transactions as fraud or not and evaluate the model using ROC-AUC and confusion matrix.

3. Predicting Heart Disease Using Logistic Regression
   Dataset: Heart Disease Dataset
   Preprocessing Steps:
   - Handle missing values (e.g., fill missing values with mean).
   - Encode categorical variables (e.g., one-hot encoding for gender, chest pain type, etc.).
   - Standardize numerical features.
   Task: Implement logistic regression to predict heart disease and evaluate the model using accuracy and ROC-AUC.

4. Classifying Emails as Spam Using Decision Trees
   Dataset: Spam Email Dataset
   Preprocessing Steps:
   - Handle missing values if any.
   - Standardize features.
   - Encode categorical variables if present.
   Task: Implement a decision tree classifier to classify emails as spam or not and evaluate the model using precision, recall, and F1-score.

5. Predicting Customer Satisfaction Using Logistic Regression
   Dataset: Customer Satisfaction Dataset
   Preprocessing Steps:
   - Handle missing values (e.g., fill missing values with median).
   - Encode categorical variables (e.g., one-hot encoding for region).

- Standardize numerical features.

Task: Implement logistic regression to predict customer satisfaction and evaluate the model using accuracy and confusion matrix.