Task 13

Machine Learning

Upload .py or Ipynb extension file on GitHub public repo "100DaysofBytewise" and share the link in the submission form by 24 July 2024.

1. Classifying Loan Status Using Decision Trees

Dataset:
- Lending Club Loan Data

Preprocessing Steps:
- Handle missing values if any.
- Encode categorical variables (e.g., one-hot encoding for loan grade, sub-grade, etc.).
- Standardize numerical features.

Task:
- Implement a decision tree classifier to classify loan status and evaluate the model using accuracy and ROC-AUC.

2. Predicting Hospital Readmission Using Logistic Regression

Dataset:
- Hospital Readmission Dataset

Preprocessing Steps:
- Handle missing values (e.g., fill missing values with mode for categorical variables).
- Encode categorical variables (e.g., one-hot encoding for hospital type, region, etc.).
- Standardize numerical features.

Task:
- Implement logistic regression to predict hospital readmission and evaluate the model using precision, recall, and F1-score.

3. Classifying Digit Images Using Decision Trees

Dataset:
- MNIST Dataset

Preprocessing Steps:
- Normalize pixel values.
- Reshape data as needed for model input.

Task:
- Implement a decision tree classifier to classify handwritten digits and evaluate the model using accuracy and confusion matrix.

4. Predicting Loan Approval Using Logistic Regression

Dataset:
- Loan Prediction Dataset

Preprocessing Steps:
- Handle missing values (e.g., fill missing values with mode for categorical variables).
- Encode categorical variables (e.g., one-hot encoding for gender, married status, etc.).

- Standardize numerical features.

Task:
- Implement logistic regression to predict loan approval and evaluate the model using accuracy and confusion matrix.

5. Classifying Wine Quality Using Decision Trees

Dataset:
- Wine Quality Dataset

Preprocessing Steps:
- Handle missing values if any.
- Standardize features.
- Encode categorical variables if present.

Task:
- Implement a decision tree classifier to classify wine quality (good/bad) and evaluate the model using accuracy and ROC-AUC.