

4MCA-A&B Advanced Data Analytics(MCA472A)

20-06-2024(Thursday)

Regular lab Questions

P1: Implementation of Correlation and Regression

Objective: Understand the basics of various regression and correlation analysis techniques using the given “shopping_trends” dataset. You can use Python libraries such as pandas, scikit-learn, matplotlib, and seaborn to perform these tasks. **Interpret your result using meaningful, relevant terms.**

1. Data Loading and Exploration

Load the dataset and plot any 5 different graphs using a variety of variables

2. Data Preprocessing:

Encode all categorical variables using Label encoding.

Handle any missing values or outliers.

3. Correlation Analysis:

- Calculate the correlation matrix for all variables.
- Visualize the correlation matrix using a heatmap.

4. Linear Regression

- Fit a simple linear regression model to predict Purchase Amount (USD) based on Age. Interpret the regression coefficients.
- Determine the goodness-of-fit (R-squared value).
- Visualize the regression line on the scatter plot. Make predictions using the regression model for new ages 25, 35, and 45.

5. Multiple Linear Regression

- Fit a multiple linear regression model to predict Purchase Amount (USD) based on Age, Gender, Location, and Review Rating.
- Interpret the regression coefficients and their significance.
- Determine the goodness-of-fit (R-squared value).
- Make predictions using the regression model for specific values of the predictors.

6. Logistic Regression

- Fit a logistic regression model to predict Subscription Status based on Age, Gender, and Review Rating.
- Interpret the regression coefficients.
- Evaluate the model using accuracy, precision, recall, and F1-score.
- Create a confusion matrix to visualize the model's performance.

7. Decision Tree Regression

- Fit a decision tree regression model to predict Purchase Amount (USD) based on Age, Gender, and Previous Purchases.
- Visualize the decision tree.
- Evaluate the model using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
- Compare the decision tree model with a multiple linear regression model.

Evaluation Rubrics		
Rubrics	Evaluation Criteria	Marks
R1	Correctness and Clarity	5 Marks
R2	Complexity and Validation	3 Marks
R3	Result Interpretation	2 Marks

Submission Guidelines

- Generate the single .pdf file for the given question. The file name should be your register number followed by program No: Example: 2347101_P2
- Upload the .pdf files in Google Classroom on or before the deadline mentioned.