

Assignment: 3

Problem Statement:

Apply appropriate ML algorithm on a dataset collected in a cosmetics shop showing details of customers to predict customer response for special offers.

Software Library Package:

Python with pandas , matplotlib , seaborn and scikit-learn

1. Theory

1.1 Methodology

The implemented program applies machine learning (ML) algorithms to predict customer responses for special offers using a dataset collected from a cosmetics shop. The methodology involves several steps including data preprocessing, model training, and evaluation.

1.2 Advantages and Applications

- Advantages:

- Data Preprocessing: Imputing missing values and standardizing features enhance model performance.

- Model Selection: Logistic Regression offers simplicity and interpretability, suitable for this classification task.

- Evaluation Metrics: Accuracy, confusion matrix, and classification report provide comprehensive insights into model performance.

- Applications:

- Personalized marketing strategies tailored to individual customer responses.

- Efficient allocation of resources and special offers based on predicted customer behavior.

- Insights into customer preferences and market trends for strategic decision-making.

1.3 Limitations

- Data Quality: The accuracy of predictions heavily relies on the quality and representativeness of the dataset.
- Model Generalization: Logistic Regression may not generalize well to unseen data if the dataset is not diverse enough.
- Interpretability: Complex ML algorithms may lack interpretability, making it difficult to understand the reasoning behind predictions.

2. Working/Algorithm

The implemented program follows these steps:

1. Data Loading: The dataset containing customer details is loaded using pandas.
2. Feature Selection: Relevant features ('Price' and 'Rank') are selected for prediction.
3. Data Preprocessing:
 - Imputation: Missing values are imputed using the mean strategy with `SimpleImputer`.
 - Standardization: Features are standardized using `StandardScaler` to ensure uniformity in scale.
4. Model Training: Logistic Regression algorithm is chosen for its simplicity and interpretability, implemented with `LogisticRegression`.
5. Model Evaluation:
 - Train-Test Split: Data is split into training and testing sets using `train_test_split`.
 - Model Fitting: The logistic regression model is trained on the training data.
 - Prediction: Predictions are made on the test set.
 - Evaluation Metrics: Accuracy score, confusion matrix, and classification report are computed using `accuracy_score`, `confusion_matrix`, and `classification_report`.
6. Prediction: The model is used to predict customer response for a hypothetical case with given input values.

7. Visualization: Confusion matrix is visualized using seaborn heatmap.

3. Conclusion

The implemented program demonstrates the application of machine learning algorithms, specifically logistic regression, in predicting customer responses for special offers based on cosmetics shop data. Despite its simplicity, logistic regression achieves an accuracy of [insert accuracy here], providing valuable insights into customer behavior. Further improvements could involve exploring more complex algorithms and enhancing the dataset quality to improve model performance and generalization. This program serves as a foundational tool for developing advanced decision-making systems in cosmetics retail businesses.