TITLE : PRODUCT DEMAND PREDICTION

LIBRARIES/MODULES USED:

PANDAS for dataset manipulation,

NUMPY for encoded variables manipulation,

SEABORN for visualization,

SCI-KIT LEARN for the core,

MODEL SELECTION

- train_test_split for the division within the dataset

PRE-PROCESSING

- Label_encoder for encoding strings and categorical variables

LINEAR MODEL

- Logistic_Regression for the core

METRICS

- confusion_matrix,
- accuracy_score and
- classification_report for the evaluation.

CONCEPT USED:

The model uses a Logistic regression library as its core to fit in the given dataset.

This library is referred by the model to observe the underlying pattern within the given dataset.

In a brief it tries to apply the formula,

$$Y = b0 + b1x1 + b2x2 + b2x2 + ... + bnxn$$

Where,

b0...bn denotes coeffecients of the input features and

x1...xn denotes the input features.

But this approach uses a transformed version of this formula to specifically find a probabilistic value for a positive outcome such as,

$P(Y = 1/x) = 1 / (1 + \exp(-(b0 + b1x1 + b2x2 + b2x2 + ... + bnxn)))$

The above formula can be described as the probability of the outcome being 1 given the x features.

Where,

b0...bn denotes coeffecients of the input features and

x1...xn denotes the input features

DESIGN PROCESS:

1) MODEL DEFINITION

Step 1: The '.csv' files are loaded into the model.

Step 2: The dataset is pre-processed by handling any missing values and removing the least relevant features

Step 3: The dataset is classified as Y for the prediction and X for the features.

Step 4: The dataset is then split into training and testing sections.

Step 5: The parameters for randomness are defined.

Step 6: The dataset is observed for underlying patterns.

Step 7: The dataset is then fit into the model.

2) TUNING

Step 8: A test metric is used to evaluate the model's accuracy.

Step 9: Based on the accuracy the model's hyper-parameters are tuned.

Step 10: Step 9 is repeated until the model reaches the desired accuracy.

3) TESTING

Step 11: A new dataset is loaded.

Step 12: The dataset is then pre-processed and cleaned.

Step 13: The dataset is then loaded into the model for prediction.