

Module 5: Text Classification 2

Case Study

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Case Study

You are given 3 CSV Files for the following 3 questions respectively.

- HouseData.csv
- Marketing.csv
- Results.csv

1. Write a function which accepts a CSV file location and split the data randomly in the file into 60% and 40% and save it in two different files namely “First.csv” and “Second.csv”.

Run this function for HouseData.csv

2. Write a function which accepts a CSV and performs Label Encoding for the “Class” column of that file, creates a new Column called “Label” and saves the result in the same file.

Run this function for Marketing.csv

3. Write a function which accepts a CSV which has two columns: “ActualValues” and “PredictedValues”. You need to design a confusion matrix and give values for the following:

- Accuracy
- Misclassification Rate
- True Positive Rate
- False Positive Rate
- Specificity
- Precision
- Null Error Rate

Also plot the ROC Curve

Provide your understanding for each of them in a sentence or two.

Run this function for Results.csv