

Reviews & Rating Sentimental Analysis

Submitted By :

ASHUTOSH CHAUDHARY

ACKNOWLEDGMENT

I would like to express my gratitude towards my internship mentor Ms. Swati Mahaseth for helping me in completion of the project.

BUSINESS PROBLEM

From my understanding the problem is about scraping the reviews & rating data & understand the sentimental analysis

OBJECTIVE FOR PROBLEM UNDERTAKEN

We have to study the reviews of the products & then we need to make prediction of ratings on the basis of reviews that were written to understand the sentiments of public towards the product.

ANALYTICAL PROBLEM FRAMING

- **Origin of dataset & datatypes**

We created dataset of two columns by doing webscraping and one data types of column was categorical & other was continuous.

- **Mathematical/Analytical Modelling**

For visualization we only use one plot which was histogram & for modelling purpose we use two models which are random forest classification & Logistic Regression models

- **Assumptions related to problem statement**

No assumptions were made while working on dataset

- **Libraries & Tools used**

We used numpy, pandas, matplotlib.pyplot, seaborn, sklearn, pickle, keras, warnings libraries for this task.

STEPS TAKEN FOR THE TASK

1. Importing Libraries for the task

Numpy, pandas, matplotlib.pyplot, seaborn, sklearn, keras, pickle & warnings were imported for the task to get completed.

2. Importing Dataset

Imported the dataset using pandas library in jupyter notebook.

3. Checking dimension of dataset

Dataset contains 24118 rows & 2 columns

4. Checking for null values

There were 2 null values present which got filled using random technique

5. Preprocessing

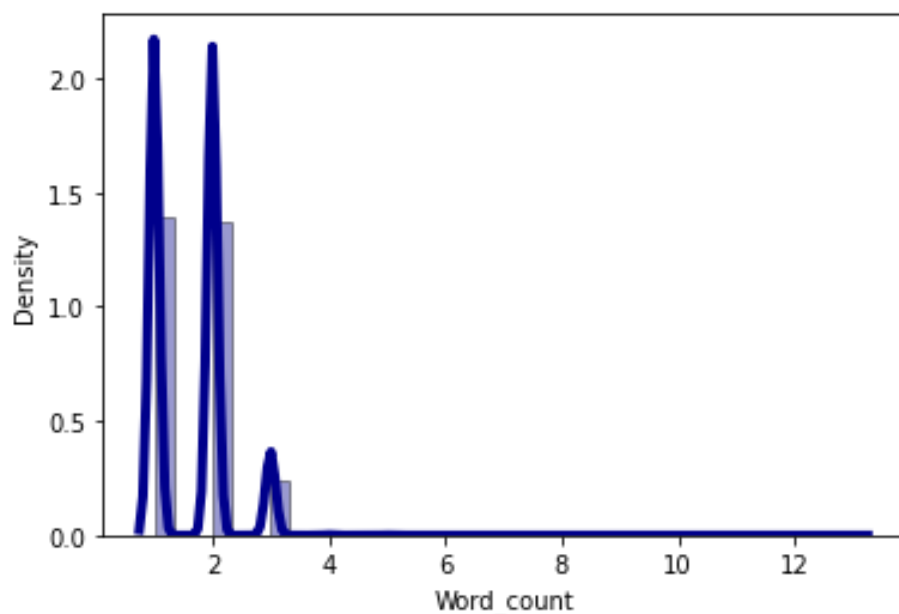
After this we performed preprocessing on the Review column like applying Stopwords, Tokenization, Lemmatization, etc to make data more refined for further use

6. Creating a new column

We created a new column named `word_count` in which the number of words used in that rows paragraphs were counted & written.

7. Performing EDA

Word_Count



From visualization we concluded that most paragraphs or sentences have only one word or two words in them

8. Performed Vwectorization

Then we performed vectorization of Review column to convert text data into meaningful data so that we can use it for sentimental analysis.

9. Splitting data

Then we splitted data into train & test using
train_test_split

10. Applying models

We applied two models (Random Forest Classification model & Logistic Regression model) & performance of both models were coming out the same while checking the accuracy score of both the model which was same and the amount was 0.5569

11. Making prediction & saving model

We randomly select the model which comes out to be Random Forest Classification Model & made prediction using that model & after that we saved the model using pickle library.

CONCLUSION

In the end we can conclude that sentiment analysis of product depend on the satisfaction that product provided to customer & that satisfaction gets reflected in the form of words that are written by the public for product in review section of various sites.