**RATING PREDICTION PROJECT**

Submitted By:

Harneet kaur

**ACKNOWLEDGMENT**

I would like to thank Flip Robo Techniques and my SME Khusboo Garg who helped me in every possible way that she could and guide me through new things. without whom I won’t have been able to complete this project.

**INTRODUCTION**

**• Business Problem Framing:**

a website where people write different reviews for technical products. Now they are adding a new feature to their website i.e . The reviewer will have to add stars(rating) as well with the review. The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars. Now they want to predict ratings for the reviews which were written in the past and they don’t have a rating.

**• Review of Literature:**

In this project, I have used Flipkart as a source for scrapping the data and making the database. I have used webscraping(selenium) to collect my data. I have scraped the ratings and reviews of different laptops, Phones, Headphones, smart watches, Professional Cameras, Printers, Monitors, Home theater, Router for building dataset.

**• Motivation for the Problem Undertaken**

.Here I have builded a model with the help of NLP which predicts ratings on the basis of customers reviews . for the betterment of business and customers.

**Analytical Problem Framing**

**• Mathematical / Analytical Modelling Problems:**

The dataframe contains several rows and columns containing all the necessary information. To deal with NaN values present in the dataset and with replace method I have replaced the (-) with the least rating number . I have used the drop method and I have also used several statistical and exploratory data visualizations for better understanding .

**• Data Sources and their formats:**

The source of data is flipkart websites for scraping reviews and ratings of different laptops, Phones, Headphones, smart watches, Professional Cameras, Printers, Monitors, Home theater, Router . from Data is stored in a dataframe contains 20209 rows and 4 columns . containing all the details .

**• Data Preprocessing Done:**

There are NAN values present in the dataset and to deal with the NAN values I have used drop method. I have used replace method and replaced some blank data with some meaningful data. So that I can make the model and further take prediction.

**• Hardware and Software Requirements and Tools Used:**

For making the dataframe used webscrapping(selenium) to scrape the reviews and ratings of all the products from Flipkart. then I imported several libraries for further model building , EDA and data cleaning.

* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* %matplotlib inline
* import warnings
* warnings.filterwarnings('ignore')
* import seaborn as sns
* import re
* from sklearn.model\_selection import train\_test\_split
* from tensorflow.keras.preprocessing.text import Tokenizer
* from tensorflow.keras.preprocessing.sequence import pad\_sequences
* import tensorflow as tf

**Model/s Development and Evaluation**

**• Identification of possible problem-solving approaches (methods):**

* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* %matplotlib inline
* Import seaborn as sns

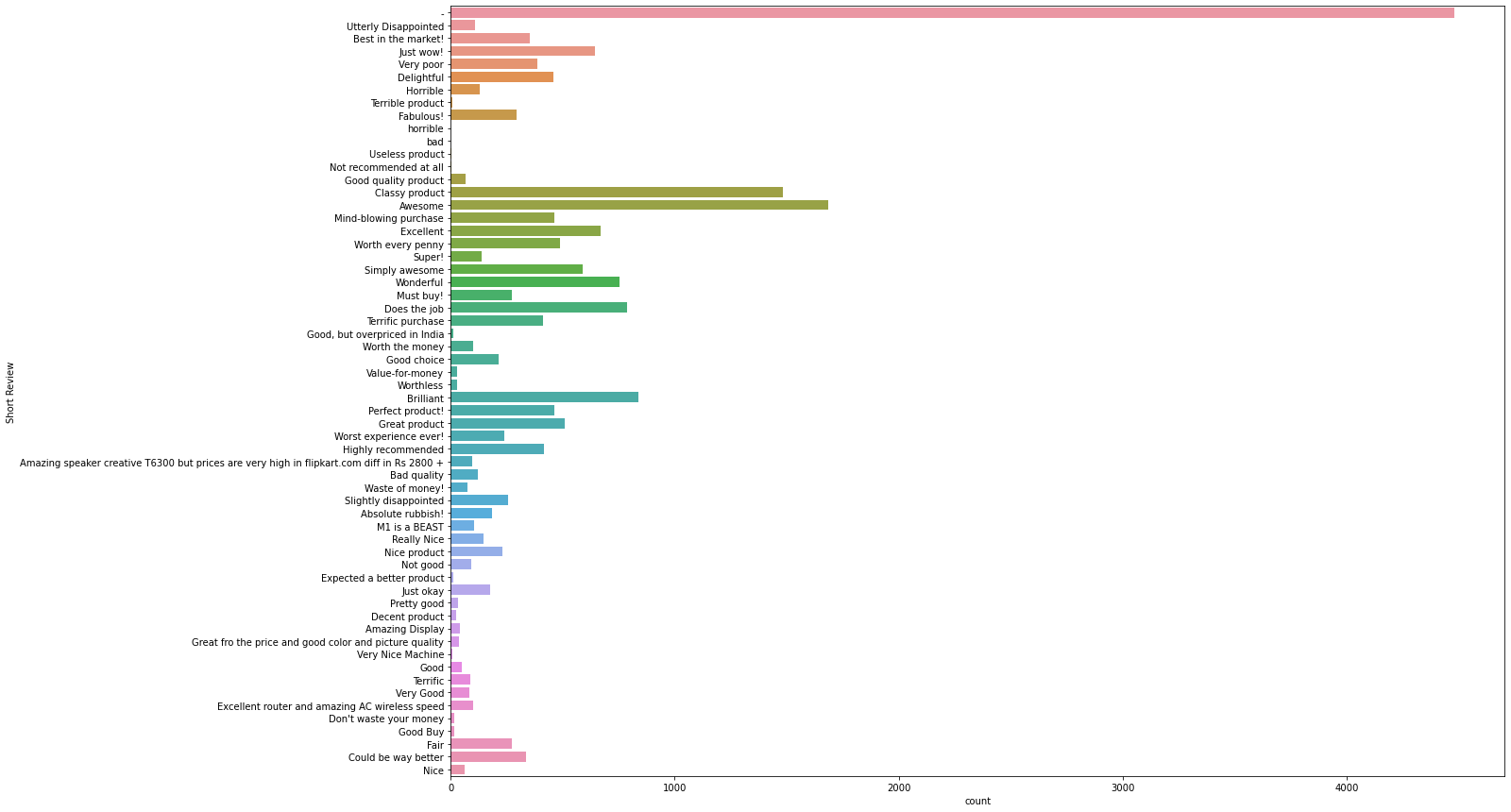
**•Testing of Identified Approaches (Algorithms):**

* from sklearn.model\_selection import train\_test\_split
* from tensorflow.keras.preprocessing.text import Tokenizer
* from tensorflow.keras.preprocessing.sequence import pad\_sequence
* import tensorflow as tf
* import re

**Visualizations**

****

Used counplot for visualizing star rating and 5 is the most given rating and 3 is the least given rating

****

Used Countplot showing all the short reviews

• **Interpretation of the Results:**

After Analysing and Visualizing I have concluded that 5star is the highest given rating by most of the customers and 3star is the least used rating. To remove NaN value I have used drop method and used Replace method to replace blank(-) value with some meaningful value. And used NLP techniques for model building.

**Conclusion**

**• Key Findings and Conclusions of the Study:**

After Analysing and Visualizing I have concluded that 5star is the highest given rating by most of the customers and 3star is the least used rating. And also used countplot to see different reviews used by the customers. To remove NaN value I have used drop method and used Replace method to replace blank(-) value with some meaningful value. And used NLP techniques for model building.

**• Learning Outcomes of the Study in respect of Data Science :**

The Dataframe contains NAN values. To deal with these NAN values I used the drop method. I have used replace method to replace some blank data(-) with some meaningful data. I have applied different statistical operations. And also applied Countplot for visualization and to understand it . I have also used NLP techniques for further model building.