RATING PREDICTION

Submitted

by:

SANDES P

ACKNOWLEDGMENT

I used informative tutorials to follow some steps in the task from websites like geeksforgeeks, stackoverflow, etc.

INTRODUCTION

- Business Problem Framing

We have a client who has 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. So, we have to build an application which can predict the rating by seeing the review.

Data Collection Phase

You have to scrape at least 20000 rows of data. You can scrape more data as well, it's up to you. more the data better the model

In this section you need to scrape the reviews of different laptops, Phones, Headphones, smart watches, Professional Cameras, Printers, Monitors, Home theater, Router from different ecommerce websites. Basically, we need these columns

1) reviews of the product.

2) rating of the product.

You can fetch other data as well, if you think data can be useful or can help in the project. It completely depends on your imagination or assumption.

Hint: • Try to fetch data from different websites. If data is from different websites, it will help our model to remove the effect of over fitting.

• Try to fetch an equal number of reviews for each rating, for example if you are fetching 10000 reviews then all ratings 1,2,3,4,5 should be 2000. It will balance our data set.

• Convert all the ratings to their round number, as there are only 5 options for rating i.e., 1,2,3,4,5. If a rating is 4.5 convert it 5.

Model Building Phase

After collecting the data, you need to build a machine learning model. Before model building do all data preprocessing steps involving NLP. Try different models with different hyper parameters and select the best model.

Follow the complete life cycle of data science. Include all the steps like1. Data Cleaning

- Exploratory Data Analysis
- Data Preprocessing
- Model Building
- Model Evaluation
- Selecting the best model

- Business Goal:

  You are required to create a model that predicts the rating of a product from its review.

- Conceptual Background of the Domain Problem

  This domain is related to the online shopping experience. Some users, while purchasing a product, leave their feedback on e-commerce applications. So other users can know more about the actual condition of the product other than advertising information. Without reading all reviews of a product, a customer can understand the quality of a product by looking to the average rating of the product. Here we are trying to create a model that predicts the rating of a product which varies from 1 to 5.

- Review of Literature

  We were not given any data. So the first task is to get some rating information from an e-commerce website. We are setting a minimum of 20000 data for training the model.

- Motivation for the Problem Undertaken

  We took the rating information of different electronic products like laptops, earphones, etc from the e-commerce website Flipkart. Our objective is to create a model that predicts rating from the scraped data

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  Here we tried to use different classification modeling techniques such as Naive Bayes, DecisionTree, and RandomForest to find the best accuracy.

- Data Sources and their formats

  We scraped around 30000 rating information from www.flipkart.com for the product stated in the problem. The scraped data is saved as a CSV file which is used to train the model

  The description of the columns

  1) reviews of the product.

  2) rating of the product.

- Data Preprocessing Done
  - Removed all things that are not letters
  - converted all letters to lower case
  - Lemmatized to get the root word of all words
  - Stopwords are not used since this is a type of sentiment analysis
  - The corpus of the word is given to a count vectorizer to transform the data

- Hardware and Software Requirements and Tools Used

The model building is done on a computer with specifications as follows

- 8GB RAM
- i5 7th gen

processor

Software requirements

- Python
- Jupyter notebook

Libraries

- Numpy
- Pandas
- Matplotlib
- Seaborn
- SKLearn
- Selenium
- re
- NLTK

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)
  - Scraping the dataset from an e-commerce website
  - Word Encoding
  - Training using ML models
  - Finding out the best model

- Testing of Identified Approaches (Algorithms)
  - MuliNomial NB
  - Decision Tree
  - Random Forest

- Key Metrics for success in solving problem under consideration

  Here we used accuracy score since this is a classification problem

CONCLUSION

Created a ML model that predicts a rating about a product from a review.

- Learning Outcomes of the Study in respect of Data Science

  Got the opportunity to work with different Ml algorithms Naive bayes, Random Forest, Decision tree etc

- Limitations of this work and Scope for Future Work

  The fewer resources of the computer led to stop the scripts in-between.