Rating Predictions

Submitted by:-

Akashdeep Singh Manral

Batch- 1834

ACKNOWLEDGEMENT

The project consists of Reviews and Ratings of some products from flipkart website.

I want to thank my intern mentor miss- Swati Mahaseth for providing assistance in solving my queries, with her help and guidance I was able to complete my project successfully

INTRODUCTION

1-Problem Statement-

1-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 e-commerce websites.

2-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.

Analytical Problem Framing

1-Introduction-

1-Our dataset(data) consists of around 20000 rows and 2 columns.

2-The data collected in from flip kart website.

3-Our dataset consists of Reviews and Ratings of different items in the flip kart web site.

4-We collected reviews and ratings for different products like- sunglasses, T.V, mobiles and merged them to form a single data set(data) containing 20000 rows and 2 columns.

5- We then created a Machine Learning model for predicting the ratings for different reviews.

2-Data Collection

1-We collected the data from flip kart website.

2-For the collection of data we made use of Selenium using Python library.

3-We collected data of reviews and rating of different items separately and later merged them to form a single dataset(data).

Model Evaluation and Selection

1-Model Pre-Processing-

1-We made use of Natural Language Processing tool kit for model pre-processing.

2-steps of pre-processing-

a-Lowering letters

b-Removing unwanted things other than letters and numbers.

c-Removing spaces

d-Removing stopwords

2-Model Selection-

1-We made use of TfidfVectorizer for converting the text into digits so that sysem could recognise it.

2-Data was then splitted into x(feature column-Reviews) and y(target column-Rating)

3-Use of train_test_split for splitting it into training and testing sets.

4-LogisticRegression,MultinomialNB, DecisionTreeClassifier, RandomForestClassifier, GradientBoostingClassifier were taken and the accuracy score was recorded.

5-The accuracy score of above was compared to their coss_val_score and after that LogisticRegression was taken out to be the most efficient classifier for the project.

6-LogisticRegression was used for prediction which showed 99% accuracy score for prediction.

3-Prediction-

1-The prediction of ratings was done through LogisticRegression with 99% accuracy score.

THANK YOU