NAME OF THE PROJECT

RATINGS PREDICTION

Submitted by:

Yogesh.C.Mudliar

ACKNOWLEDGMENT

It is a genuine pleasure to express my deep sense of thanks and gratitude to my guide, Ms.Khushboo Garg, for allowing me to work on this project. It was a great way to expose myself to the actual research environment.

I thank FLIPROBO for permitting me to work with them.
I take this opportunity to say heartfelt thanks to Dr. Deepika Sharma, VP-learning And development DataTraind for her overall dedication, devotion, and support towards me. I convey my sincere regards to all the DataTraind team thanks for supporting me during academic years of my post-graduation course in data science.

I express my profound sense of gratitude to my mentor Ms.Khushboo Garg, FLIPROBO for her guidance at every step of my research work.

Apart from the project, I learned a lot from her, she gave me valuable thought- "To think"; that I will benefit from, for a long time to come. I am indebted to her more than she knows.

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.

- ## Conceptual Background of the Domain Problem

  One of the most important problems in e-commerce is the correct calculation of the points given to after-sales products. The solution to this problem is to provide greater customer satisfaction for the e-commerce site, product prominence for sellers, and a seamless shopping experience for buyers. Another problem is the correct ordering of the comments given to the products. The prominence of misleading comments will cause both financial losses and customer losses. In solving these 2 basic problems, e-commerce site and sellers will increase their sales, while customers will complete their purchasing journey without any problems.

- ## Review of Literature

  -The better you know your customers, then the better original and unique content you can provide for them. Collecting reviews and ratings enables you to gain insight into what user-generated content your audience prefers and provides search engine crawlers with the content needed to properly evaluate a website.

  -Take control of your reputation and utilize the power of social proof. Consumers now, more than ever, are looking for ratings and reviews regarding the services and products of companies prior to making purchasing decisions. Through online reviews, companies can help influence purchasing decisions by promoting and marketing customer feedback.

-Especially since 92% of consumers now read online reviews, E-Commerce businesses need to rethink their strategies for implementing the proper digital strategy in order to take advantage of the viral effects that customer reviews have. Generating brand reputation and product value by aligning your feedback with your marketing strategies is the perfect way to convert visitors into loyal customers.

- **Motivation for the Problem Undertaken**

-5 Reasons Why You're E-Commerce Company Needs Ratings & Reviews:

Generate Trust: Today's shoppers trust the reviews of strangers more than they trust the advertisement from a company. Valuable brands are taking notice of this trend and are capitalizing on the power of social proof by implementing review-centric marketing strategies.

Strengthen Marketing Efforts: By taking advantage of your customer reviews & ratings, your business can monitor and improve brand awareness, reputation, and loyalty.

To Drive More Revenue: The main goal of any business is to generate revenue by providing a product or service. By collecting and publishing authentic customer feedback, businesses can help drive higher SEO, generate new customers, and increase sales.

Increase Visibility on Google: Google is extremely important for the collection and advertisement of customer reviews and ratings. Through numerous platforms and services, Google has innovated the entire digital advertising industry with ease while enabling businesses to reach specific target audiences of millions of people.

Optimize Strategy: By companies collecting feedback from customers, they are also collecting a vast amount of customer intelligence as well. Through reviews, businesses can utilize this meaningful and useful information from their own customers to help create more effective strategies.

Analytical Problem Framing

- Data Sources and their formats

  This dataset containing Amazon Product Data includes product categories and various data. The product with the most comments in the electronics category has user ratings and comments.

- Data Preprocessing Done

  1. Read dataset and make it in proper format.
  2. Encode labels
  3. Convert all cases to lower
  4. Remove punctuations
  5. Remove Stop words
  6. Check stats of messages
  7. Convert all texts into vectors
  8. Import classifier
  9. Train and test
  10. Check the accuracy/confusion matrix

- Data Inputs- Logic- Output Relationships

The data contain review Text and overall ratings given by customers from amazon webs site, The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars.

- Hardware and Software Requirements and Tools Used

-from sklearn.feature_extraction.text import TfidfVectorizer

-from sklearn.model_selection import train_test_split

-from sklearn.svm import LinearSVC

-from sklearn.metrics import classification_report

Model/s Development and Evaluation

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

1. Read dataset and make it in proper format.
2. Encode labels

3. Convert all cases to lower: I will now convert the text to all lowercase. Our machine learning algorithms recognize words that start with a capital letter as different words, and we will convert them to lowercase. Thus, our machine learning algorithms will not perceive words that start with a capital letter as a different word.

4. Remove punctuations

5. Remove Stop words

6. Check stats of messages

7. Convert all texts into vectors

8. Import classifier

9. Train and test

10. Check the accuracy/confusion matrix

- # Testing of Identified Approaches (Algorithms)

-from sklearn.feature_extraction.text import TfidfVectorizer

Transforms text to feature vectors that can be used as input to estimator. Vocabulary Is a dictionary that converts each token (word) to feature index in the matrix, each unique token gets a feature index.

-from sklearn.model_selection import train_test_split

The train_test_split function of the sklearn. model_selection package in Python splits arrays or matrices into random subsets for train and test data, respectively.

-from sklearn.svm import LinearSVC

The objective of a Linear SVC (Support Vector Classifier) is to fit to the data you provide, returning a "best fit" hyperplane that divides, or categorizes, your data. From there, after getting the hyperplane, you can then feed some features to your classifier to see what the "predicted" class is.

-from sklearn.metrics import classification_report

A classification report is a performance evaluation metric in machine learning. It is used to show the precision, recall, F1 Score, and support of your trained classification model.

- Run and Evaluate selected models

-I remove all the unnecessary data from the dataset

In [5]:
```
# Replace email addresses with 'email*
df['reviewText'] = df['reviewText'].str.replace(r'^.+@[^\.].*\. [a-z]{2,}$',
                                'emailaddress')

#Replace URLs with 'webaddress"
df['reviewText'] = df['reviewText'].str.replace(r'^http\://[a-zA-Z0-9\-\.]+\.[a-zA-Z]{2,3}(/\S*)?$',
                                'webaddress')


#Replace money symbols with 'moneysymb* (£ can by typed with ALT key + 156)
df['reviewText'] = df['reviewText'].str.replace(r'£|\$','dollers')

#Replace 10 digit phone numbers (formats include paranthesis, spaces,no spaces, dashes) with 'phonenumber
df['reviewText'] = df['reviewText'].str.replace(r'^\(?[\d]{3}\)?[\s-]?[\d]{3}[\s-]?[\d]{4}$',
                                'phonenumber')


#Replace numbers with numbr
df['reviewText'] = df['reviewText'].str.replace(r'\d+(\.\d+)?','numbr')
```

- Removed punctuation and removed leading and trailing whitespace

In [6]:
```
# Remove punctuation (^ = this symbol is mean not equals to)
df['reviewText'] = df['reviewText'].str.replace(r'[^\w\d\s]',' ')

# Replace whitespace between terms with a single space
df['reviewText'] = df['reviewText'].str.replace(r'\s+',' ')

# Remove Leading and trailing whitespace
df['reviewText'] = df['reviewText'].str.replace(r'^\s+|\s+?$',' ')
```

- **Key Metrics for success in solving problem under consideration**

-Read dataset and make it in proper format.

- Encode labels

- Convert all cases to lower: I will now convert the text to all lowercase. Our machine learning algorithms recognize words that start with a capital letter as different words, and we will convert them to lowercase. Thus, our machine learning algorithms will not perceive words that start with a capital letter as a different word.

-Remove punctuations

- Remove Stop words

-Check stats of messages

- Convert all texts into vectors

- Import classifier

- Train and test

-Check the accuracy/confusion matrix

- Interpretation of the Results

  -We want to predict ratings for the reviews which were written in the past and we don't have a rating. So, we have to build an application which can predict the rating by seeing the review. We build an application that can predict ratings according to the given reviews.

  -The better you know your customers, then the better original and unique content you can provide for them. Collecting reviews and ratings enables you to gain insight into what user-generated content your audience prefers and provides search engine crawlers with the content needed to properly evaluate a website.

CONCLUSION

- Key Findings and Conclusions of the Study

-We build an application that can predict ratings according to the given reviews.

```
In [18]:
1 x = 'this product is really bad. i do not like it'
2
3 vec = tfidf.transform([x])
4 clf.predict(vec)
```

```
Out[18]: array([1.])
```

```
In [19]:
1 x = 'this product is really good. i like it'
2
3 vec = tfidf.transform([x])
4 clf.predict(vec)
```

```
Out[19]: array([5.])
```

```
In [20]:
1 x = 'this product is good,but i dont want to buy it'
2
3 vec = tfidf.transform([x])
4 clf.predict(vec)
```

```
Out[20]: array([3.])
```

-Especially since 92% of consumers now read online reviews, E-Commerce businesses need to rethink their strategies for implementing the proper digital strategy in order to take advantage of the viral effects that customer reviews have. Generating brand reputation and product value by aligning your feedback with your marketing strategies is the perfect way to convert visitors into loyal customers.

- ## Learning Outcomes of the Study in respect of Data Science

-Defining the question

First I define the problem statement

-Collecting the data

-Cleaning the data

-Removing major errors, duplicates

-Removing unwanted data points

-Bringing structure to your data

-Filling in major gaps

-Analyzing the data

-Sharing the results

- ## Limitations of this work and Scope for Future Work

  -The collection of ratings and reviews can be used as an immensely powerful tool for any E-commerce business.

  -This digital feedback acts as an incredible way to increase traffic, improve PPC, and build any online reputation while increasing the confidence of consumers.

  -Modern shoppers today are well-informed, influential, and expect to have their needs and want to be met before making purchases.

  -Through social proof, companies can capture the powerful voices of their customers while surrounding their brand with positivity through efficient experience management.

  -As consumers become more influential in the future, E-Commerce companies need to take advantage of ratings & reviews to start seeing the benefits.