



REVIEW RATING PREDICATION

Submitted by:
Pratiksha Potghan

ACKNOWLEDGMENT

I would like to convey my heartfelt gratitude to Flip Robo Technologies for providing me with this wonderful opportunity to work on a Machine Learning project “Review Rating Prediction” and I also want to thank my SME “Mohd Kashif” for providing the dataset to complete this project.

This project would not have been accomplished without their help and insights.

Working on this project was a great experience.

INTRODUCTION

- **Business Problem Framing**

Our customer has a website where users can post various product reviews for technical items. The reviewer will now be required to include stars (ratings) along with the review on their website, which is a new feature they are currently adding. There are only 5 alternatives available, and the ranking is out of 5. 1, 2, 3, 4, and 5 stars, respectively. They are attempting to forecast ratings for past reviews that have not yet received one. Therefore, we must create a programme that can gauge the rating from the review.

- **Conceptual Background of the Domain Problem**

In general, shoppers utilise two straightforward heuristics to determine whether to make a final purchase of a product: ratings and pricing. The total star ratings of the product reviews, however, frequently do not accurately reflect the polarity of the opinions. Due of the possibility of varying customer ratings for a given review, rating prediction becomes a challenging topic. For instance, one person might give a product a 5-star rating and rate it as nice, but another user might write the same comment and only give it a 3-star rating.

Additionally, reviews could include anecdotal data, which is not informative and makes forecasting more difficult. Users may select different ways to express their sentiments. For instance, some users might use the word "good" to describe a merely passable product, while others might use it to describe a top-notch one. In addition to user bias, there is product bias. For example, the opinion word "long" can express a "positive" feeling for a cell phone's battery life but a "negative" feeling for a camera's focus time. We may use different opinion words to review different products, or even the same opinion word to express different sentiment polarities for different products. For the purpose of forecasting review ratings, it is crucial to take into account both the relationships between the review authors and the target products.

- **Review of Literature**

According to the Lackermair, Kailer and Kanmaz (2013), product reviews and ratings represent an important source of information for consumers and are helpful tools in order to support their buying decisions. They also found out that consumers are willing to compare both positive and negative reviews when searching for a specific product. The authors argue that customers need compact and concise information about the products. Therefore, consumers first need to pre-select the potential products matching their requirements.

With this aim in mind, consumers use the star ratings as an indicator for selecting products. Later, when a limited number of potentials products have been chosen, reading the associated text review will reveal more details about the products and therefore help consumers making a final decision. It becomes daunting and time-consuming to compare different products in order to eventually make a choice between them. Therefore, models able to predict the user rating from the text review are critically important (Baccianella, Esuli & Sebastiani, 2009).

Pang, Lee and Vaithyanathan (2002) approach this predictive task as an opinion mining problem enabling to automatically distinguish between positive and negative reviews. In order to determine the reviews polarity, the authors use text classification techniques by training and testing binary classifiers on movie reviews containing 36.6% of negative reviews and 63.4% of positive reviews. On the top of that, they also try to identify appropriate features to enhance the performance of the classifiers. Dave, Lawrence, and Pennock (2003) also deal with the issue of class imbalance with a majority of positive reviews and show similar results. SVM outperforms Naïve Bayes with an accuracy greater than 85% and the implementation of part-of-speech as well as stemming is also ineffective. However, this work demonstrates that bigrams turn out to be more successful at capturing context than unigrams in the specific situation of their datasets, despite earlier research having produced better results with unigrams to capture the weights of such characteristics by minimising the mean square error.

- **Motivation for the Problem Undertaken**

The main drivers behind this were the chance to apply my skill set to a real-world problem and the exposure to data from the actual world. The data needed for this project must be scraped from an e-commerce site and cleaned up. Its associated star ratings are predicted using features collected from textual evaluations. To do this, the prediction issue is turned into a task requiring multi-class classification, where reviews are assigned to one of five categories based on their star rating.

Gaining a general understanding of a text review may enhance the user experience. However, the reason I decided to do this project was because it is relatively a new field of research.

Analytical Problem Framing

• Mathematical/ Analytical Modeling of the Problem

In order to apply text classification, the unstructured format of text has to be converted into a structured format for the simple reason that it is much easier for computer to deal with numbers than text. This is mainly achieved by projecting the textual contents into Vector Space Model, where text data is converted into vectors of numbers. Documents are frequently handled like a Bag-of-Words (BoW) in the field of text categorization, which means that each word is distinct from the other words that are present in the text.

They are scrutinised without consideration for grammar or word order. In this model, the classifier is trained using the term-frequency (the frequency with which each word occurs) as a feature. However, the use of the word frequency suggests that all concepts are given equal weight. The word frequency, as its name implies, does nothing more than weight each term according to how frequently it occurs; it does not take the discriminatory potential of terms into consideration. Each word is given a term frequency inverse document frequency in order to handle this issue and penalise words that are used excessively (tf-idf) score which is defined above.

• Data Sources and their formats

Data Sources and their formats Data is collected from Amazon using selenium and saved in CSV file. Around 40000 Reviews are collected for this project.

This is multi-classification problem and Rating is our target feature class to be predicated in this project. There are five different categories in feature target i.e., The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars.

There are some missing values in product review. The datatype of Product review is object while datatypes of Ratings is int.

• Data Preprocessing Done

The dataset is large and it may contain some data error. In order to reach clean, error free data some data cleaning & data pre-processing performed data.

Missing Value Imputation: Missing value in product reviews are replace with 'Review Not Available'.

Data is pre-processed using the following techniques: Convert the text to lowercase
Remove the punctuations, digits and special characters
Tokenize the text, filter out the adjectives used in the review and create a new column in data frame
Remove the stop words
Stemming and Lemmatising
Applying Text Vectorization to convert text into numeric.

• **Data Inputs- Logic- Output Relationships**

The dataset consists of 2 features with a label. The features are independent and label is dependent as our label varies the values (text) of our independent variable's changes. Using word cloud, we can see most occurring word for different categories.

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

- Processor — Intel i3 processor with 2.4GHZ
- RAM— 4GB
- GPU — 2GB AMD Radeon Graphics card Software utilised
- Anaconda – Jupyter Notebook
- Selenium – Web scraping
- Google Colab – for Hyper parameter tuning
- Libraries Used – General library for data wrangling & visualization

Model/s Development and Evaluation

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

First part of problem solving is to scrap data from amazon which we already done. Second is performing text mining operation to convert textual review in ML algorithm useable form. Third part of problem building machine learning model to predict rating on review. This problem can be solve using classification-based machine learning algorithm like logistics regression. Further Hyperparameter tuning performed to build more accurate model out of best model.

- **Testing of Identified Approaches (Algorithms)**

The different classification algorithm used in this project to build ML model are as below:

1. Random Forest classifier
2. Decision Tree classifier
3. Logistics Regression
4. AdaBoost Classifier
5. Gradient Boosting Classifier.

- **Run and Evaluate selected models**

- Logistics Regression Train-test split is used to split data into training data & testing data. Further best random state is investigated through loop.

- Decision Tree Classifier
- Random Forest Classifier
- Ada Boost Classifier
- Gradient Boosting Classifier

5-fold Cross validation performed over all model. We can see that Random Forest Classifier gives us good Accuracy and maximum f1 score along with best Cross-validation score.

Hyperparameter tuning is applied over Random Forest model and used it as final model. Final model is built using best parameter in hyper parameters tuning.

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

Precision can be seen as a measure of quality; higher precision means that an algorithm returns more relevant results than irrelevant ones. Recall is used as a measure of quantity and

high recall means that an algorithm returns most of the relevant results. Accuracy score is used when the True Positives and True negatives are more important.

Accuracy can be used when the class distribution is similar. F1-score is used when the False Negatives and False Positives are crucial. While F1-score is a better metric when there are imbalanced classes. Cross validation Score: To run cross-validation on multiple metrics and also to return train scores, fit times and score times. Get predictions from each split of cross-validation for diagnostic purposes. Make a scorer from a performance metric or loss function.

- **Visualizations**

Comment:

1. Around 67% customer given 4- star rating followed by 20% customer given lowest 5-star rating.
2. Average Rating is 4.

Word Cloud:

Word Cloud is a visualization technique for text data wherein each word is picturized with its importance in the context or its frequency. The more commonly the term appears within the text being analysed, the larger the word appears in the image generated.

The enlarged texts are the greatest number of words used here and small texts are the smaller number of words used

CONCLUSION

- **Key Findings and Conclusions of the Study**

Final Model is giving us Accuracy score of 96% which is slightly improved compare to earlier Accuracy score of 96%.

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

Hands on chance to enhance my web scraping skillset. In this project we were able to learn various Natural language processing techniques like lemmatization, stemming, removal of Stop words. This project has demonstrated the importance of sampling effectively, modelling and predicting data.

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

More input features can be scrap to build predication model. There is scope for application of advanced deep learning NLP tool to enhanced text mining operation which eventually help in building more accurate model with good cross validation score.

References use in this project:

- SCIKIT Learn Library Documentation
- Blogs from towardsdatascience, Analytics Vidya, Medium
- Andrew Ng Notes on Machine Learning (GitHub)
- Data Science Projects with Python Second Edition by Packt
- Hands on Machine learning with scikit learn and tensor flow by Aurelien Geron
- Lackermair, G., Kailer, D. & Kanmaz, K. (2013). Importance of online product reviews from a consumer's perspective. Horizon Research Publishing, 1-5. doi: 10.13189/aeb.2013.010101
- Baccianella, S., Esuli, A. & Sebastiani, F. (2009). Multi-facet rating of product reviews. Proceedings of the 31st European Conference on Information Retrieval (ECIR), 461- 472