ECMM447:MINI PROJECT

Amazon Product Review: Sentiment Analysis

Student candidate number: 710025535

Abstract—**The application problem of sentiment analysis of product reviews has lately gained a lot of traction in text mining and machine learning research. Sentimental analysis offers reputable ways for enhancing target data. Sentimental analysis is the subject of a considerable quantity of study, which emphasizes its importance. In this project, I want to study the correlation between the Amazon ratings and the product reviews given by the customers. Through the reviews, I analyze the sentiment and use various Machine learning algorithms to classify the positive, negative, or neutral reviews. Using machine learning models, I aim to predict the sentiment of the reviews and the accuracy of the forecast. I also explore the topic modelling techniques to identify the major topics of discussions.**

Keywords—Sentiment Analysis, NLP, Topic Modelling, Classification algorithms, Product reviews.

I. INTRODUCTION

In recent years, there has been an increase in research activities focused on analysing sentiment in textual data. In our daily lives, we come across a variety of products; but, on the digital media, we may swipe through hundreds of product options under one category. The customer will find it difficult to make a decision. Then there reviews, where buyers who have previously purchased the product submit a rating and a brief description of their experience in the form of reviews. Ratings, as we all know, may be readily sorted and used to determine if a product is excellent or terrible. When it comes to sentence reviews, however, we must examine each line carefully to ensure that the review delivers a favourable or negative message. Things like that have become more easier in the age of artificial intelligence thanks to Natural Language Processing (NLP) technology. This approach has a wide range of applications. Businesses, for example, are constantly interested in public or consumer thoughts and feelings about their products and services. Before using a service or purchasing a product, potential consumers want to know what other people think about it.

The goal of this project is to identify customers' positive and negative reviews across various products and to develop a supervised learning model to polarise enormous volumes of data. Our data collection is made up of consumer reviews and ratings obtained from Amazon's Consumer Reviews. I analysed the accuracy of various models and gained a deeper picture of the' polarising opinions on the products

II. RESEARCH CONTEXT

Sentimental analysis is a branch of NLP and machine learning that focuses on the polarity of a text's context as well as the emotions it contains. This assists in the human-level precise extraction of critical data from computers. In this project, I have implemented the Graded Sentiment Analysis. The following are the types of Sentiment Analysis which are also popular apart from Graded: Emotion detection, Aspect-based, and Multilingual Sentiment Analysis.

III. OBJECTIVE

1. Text Preprocessing and cleaning
2. Text Visualizations
3. Sentiment Analysis
4. Extracting Features using TF-IDF
5. Sentiment Modelling
6. Topic Modelling

IV. DATA AND RESOURCES

Our data is derived from Amazon Product Reviews by Customers. It consists of 4915 data points and 10 columns(features).
Description of columns in the file:

- reviewerID - ID of the reviewer, e.g. A2SUAM1J3GNN3B
- asin - ID of the product, e.g. 0000013714
- reviewerName - name of the reviewer
- helpful - helpfulness rating of the review, e.g. 2/3
- reviewText - text of the review
- overall - rating of the product
- summary - summary of the review
- unixReviewTime - time of the review (unix time)
- reviewTime - time of the review (raw)
- helpful_yes- was the product helpful

V. METHODS

A. Preprocessing and Tokenization

Reviewer names and review content both have null values. The names of the reviewers add no value to the project's goal (we have ids instead). So let's concentrate on the review text. Because there are only 1 null values, I don't anticipate dropping would be a problem, but instead I'm considered about imputing that as ‘missing’. The text in reviewText was cleaned and refined by removing HTMl tags and extra white spaces, converting it to lower case, replacing punctuations and numbers with space. I also removed urls from the text.I handled the time column by dividing the year, date, month into separate columns. The useful feature with values in list [a,b] format can be seen in the main dataframe. That review was useful to a out of b persons, according to the data. However, in that format, it is unlikely to offer value to the machine learning model and deciphering the meaning for the computer will be challenging. As a result, I intend to add a helpful rate feature that returns an a/b number from [a,b].I also performed Tokenization using Textblob library. Tokenization, also known as word segmentation, is a basic procedure that divides a corpus into little parts called tokens.

B. Text Normalization

Lemmatization appropriately lowers inflected words, guaranteeing that the underlying word belongs to the language. Lemma is the root word of Lemmatization. A lemma (plural lemmas or lemmata) is a collection of words in their canonical, dictionary, or citation form. Performed Lemmatization to normalize the text. The morphological analysis of the word is considered during the process.

Figure 1 shows the word cloud for the reviews, the words in bold are most frequent and important. Figure 2 depicts the count of reviews and ratings ranging from 1 to 5.

Figure 1: Word Cloud

From the word cloud, figure 1, we can analyse that the frequent and important words which come up in reviews are: Card,phone, gb,sd card,sandisk,memory card. We can deduce that a lot of reviews are about the phone products on amazon.

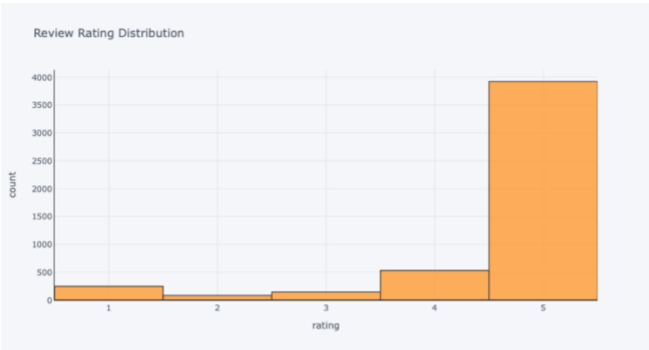

Figure 2: Count of Reviews and rating

From the above graph we can deduce that there are lot of 5 ratings from the customers and quite less 2,3 ratings.

VI. Sentiment Analysis

In the process of Sentiment analysis, the first step is using the SentimentIntensityAnalyzer from the NLTK library and calculated the polarity scores. I created a separate column named, Sentiment Score. The sentiment scores more than 0 is labelled 'positive', scores less than 0 are labelled 'negative' and those which are equal to 0 are 'neutral' were classified in a new 'Sentiment' column. Label encoding was also performed on the target variable which is the 'sentiment' column which classified the sentiments as follows:0 being negative, 1 as neutral and 2 as positive.

From Figure 3 graph we can conclude that there are a greater number of positive reviews with high helpful rate.

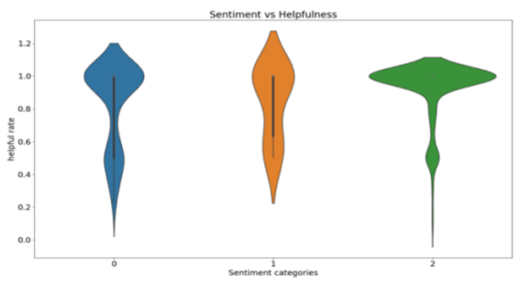

Figure 3: Sentiment vs Helpfulness

VII. Extracting Features

Because a computer cannot understand words or their sentiment, we must transform the review texts into vector form before we can develop the model for the sentiment analysis. To convert the texts in this project, we will utilise the TF-TDF technique. "Term Frequency — Inverse Document Frequency" is abbreviated as TF-IDF. This is a strategy for quantifying a word in a text; we assign a weight to each word that represents its value in the document and corpus. In the fields of information retrieval and text mining, this method is commonly employed. To formalize in mathematical terms, the TF-IDF score for the word t in the document d from the document set D is calculated as follows:

$$tf\,idf\,(t,\,d,\,D) = tf\,(t,\,d)\;.\;idf\,(t,\,D)$$

Where:

$$tf\,(t,\,d) = \log\,(1 + freq\,(t,\,d))$$

$$idf\,(t,\,D) = \log\,\left(\frac{N}{count\,(d \in D : t \in d)}\right)$$

Handling Target Feature Imbalance using SMOTE:

- I discovered that we received more positive sentiments than negative and neutral sentiments in our target feature. In such a situation, it is critical to maintain a class balance. To balance out the skewed dataset problem, I used SMOTE (Synthetic Minority Oversampling Technique). Its goal is to achieve class parity by adding minority class examples at random.
- SMOTE creates new minority instances by combining existing minorities. For the minority class, it uses linear interpolation to create virtual training records.

VIII. Model Selection

Using the train-test split method, the dataset was divided in the ratio 80% train and 20% test data.

After processing the text data, we can start building the model for predicting classes in target feature from sparse matrix.

Using cross validation, first I chose the best performing model. Let's have a look at all the classification algorithms and go through the model selection procedure. Below table summarizes the result from the algorithms:

S.no	Model Name	Accuracy
1	Logistic Regression	0.7989
2	Decision Tree	0.7031
3	KNN	0.6773
4	SVC	0.7849
5	Naïve Bayes	0.7519
6	Random Forest	0.7904

The findings show that logistic regression outperformed the other techniques. Therefore, the final algorithm chosen was Logistic Regression to tune the hyperparameters.

For hyperparameter tuning, I chose regularization and penalty (L1, L2). The following were the results obtained:

Logistic Regression (C=10000.0, random_state=0)

The mean accuracy of the model is: 0.94207181

There was a significant increase in the model accuracy which was due to hyperparameter tuning.

A. *Classification Metrics:*

After applying logistics regression on the test set, I plotted the confusion matrix and calculated the F1 score for accuracy:

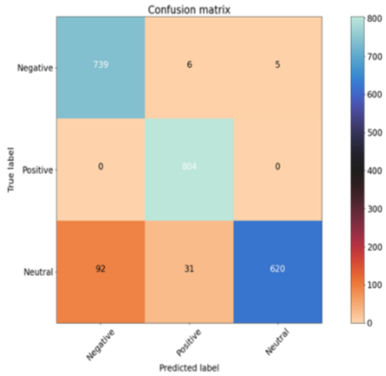

Figure 4: Confusion Matrix

```
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.99      0.93       750
           1       0.96      1.00      0.98       804
           2       0.99      0.83      0.91       743

    accuracy                           0.94      2297
   macro avg       0.95      0.94      0.94      2297
weighted avg       0.95      0.94      0.94      2297
```

As can be seen, it obtained the high score in all the categories.

B. *ROC-AUC Curve:*

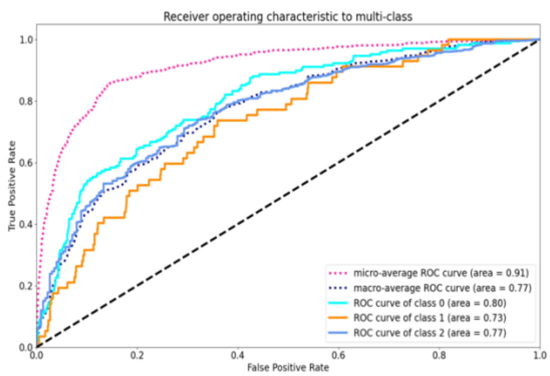

Figure 5: ROC-AUC Curve

Figure 5 depicts a critical curve since it determines the threshold to establish based on objective criteria. I plotted ROC for various classes to assist us in determining which class was better categorized. On the roc curve, I also plot micro and macro averages. Class 2 and 0 have been categorized rather well, since their area under the curve is high, according to the ROC curve for classes. To obtain the ideal number of TPR and FPR, we might choose any threshold between 0.6 and 0.8. When it comes to the micro and macro averages, the micro average performs exceptionally well, while the macro average scores poorly.

IX. Topic Modelling Using LDA

In NLP, topic modelling is a technique that assigns a subject to a corpus based on the words present. If a business's receives hundreds of reviews, for example, it is critical for the company to understand which kind of evaluations are more significant and vice versa.

LDA, short for Latent Dirichlet Allocation. LDA proposes that the documents are formed via a statistical generative process, with each document consisting of a mixture of subjects and words. Figure 6 shows the example of topics for few reviews in the form of word cloud:

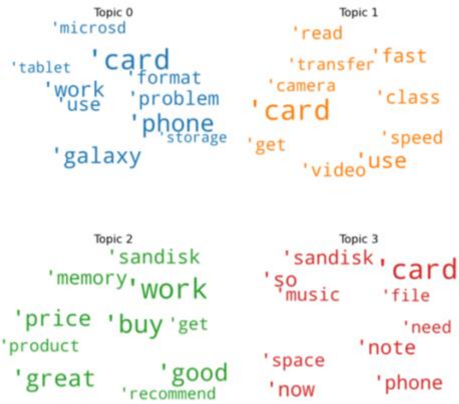

Figure 6: Word Cloud for topic modelling

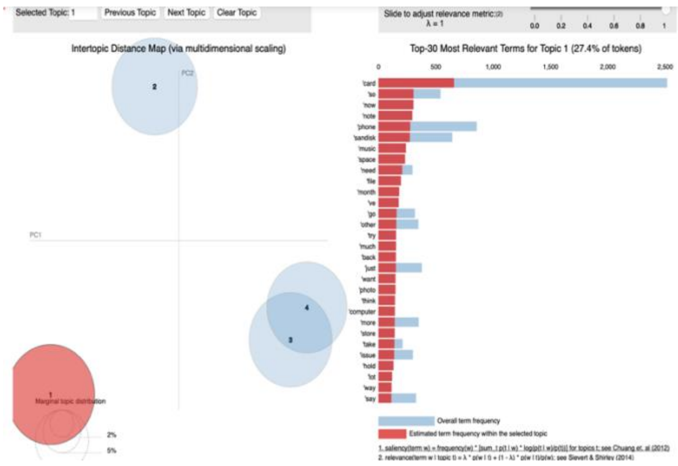

Figure 7: Intertopic Distance map

Figure 7 shows the estimated term frequency within the selected topic. In the figure, the topic selected is 1 and the red colour highlights the estimation of frequency.

X. Future Scope

As part of future scope of this project, one can analyse the emotion sentiment analysis, which further classifies the emotion of each review whether it's anger, disgust, joy, anticipation, surprise, trust, sadness, or fear, which will be an interesting take on the reviews.

XI. Conclusion

We should focus on our f1 score, which averaged 94 percent in sentiment analysis. Balancing the target feature is important. Most of the neutral reviews were actual critic of the product and the company can consider these as feedback. We did not remove stopwords as they might affect the results.

REFERENCES

[1] S. Hota and S. Pathak. Knn classifier based approach for multi-class sentiment analysis of twitter data. In International Journal of Engineering Technology, pages 1372–1375. SPC, 2018.

[2] C. Rain. Sentiment analysis in amazon reviews using proba- bilistic machine learning. *Swarthmore College*, 2013.

[3] Y. Xu, X. Wu, and Q. Wang. Sentiment analysis of yelps ratings based on text reviews, 2015

[4] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web, pages 519–528. ACM, 2003.

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews,“ Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining, 2004.

[6] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs.,“ICWSM, 2007.