Paper No: SC-31 **Smart Computing**

Overall and Feature Level Sentiment Analysis of Amazon Product Reviews Using Machine Learning Techniques and Web-Based Chrome Plugin

V. R. Welgamage
Department of Industrial Management
University of Kelaniya, Sri Lanka
vimukthi.welgamage@gmail.com

U. A. C. Senarathne
Department of Industrial Management
University of Kelaniya, Sri Lanka
harithasenarathne@gmail.com

N. H. A. C. Madhubhashani
Department of Industrial Management
University of Kelaniya, Sri Lanka
madhuhangawatta@gmail.com

T. C. Liyanage
Department of Industrial Management
University of Kelaniya, Sri Lanka
thathsara92@gmail.com

P. P. G. Dinesh Asanka*
Department of Industrial Management
University of Kelaniya, Sri Lanka
dineshasanka@gmail.com

2022 International Research Conference on Smart Computing and Systems Engineering (SCSE) | 978-1-6654-7375-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/SCSE56529.2022.9905125

Abstract - One of the critical tasks of Natural Language Processing (NLP) is sentiment analysis or opinion mining. Sentiment analysis has gained much attention in recent years. It collects data on each user's views, feelings, and opinions regarding a particular product to determine whether they have a positive, neutral, or negative attitude toward it. This study aims to address the categorising sentiment polarity, which is one of the essential issues in sentiment analysis and with extensive process descriptions. The key contribution of this study is to introduce feature-wise sentiment analysis for online products considering the customer reviews and star ratings using the modified web-based chrome plugin. Finally, we share some insight into our future sentiment analysis efforts. The research was based on the categorisation of sentiment polarity in online product reviews from Amazon.com

Keywords - classification, feature review, Natural Language Processing (NLP), polarity, product review, sentiment analysis, web-based chrome plugin

I. INTRODUCTION

Research projects analysing sentiment in textual materials have become more prevalent in recent years. In recent years, there has been an increase in the number of articles published on sentiment analysis, as shown by the statistics from the Web of Knowledge in Fig. 1. Sentiment analysis or opinion mining is the core of this research, in which we may computationally investigate people's opinions, assessments, attitudes, and feelings regarding entities, individuals, situations, events, topics, and their attributes given a quantity of text. This approach is applicable in a wide range of scenarios.

Businesses, for instance, always seek consumer or public perceptions of their products and services. Potential customers want to know what other people think of a service or product before using it or buying it. In addition, researchers [5] use this data to conduct in-depth analyses of market patterns and consumer sentiment, which could lead to improved stock market forecasting. Due to the abundance of various sites, discovering and monitoring online opinion sites and collecting the information they provide remain challenging tasks.

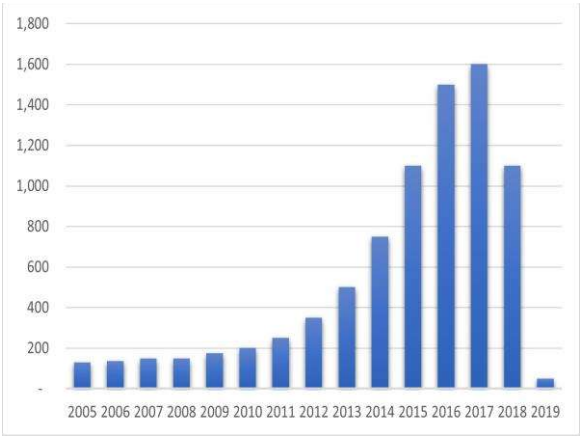

Fig. 1. Published Papers on Sentiment Analysis

There are frequently a lot of opinionated languages that are difficult to read on lengthy forum discussions and blogs. The average human reader will have trouble locating relevant sites and accurately summarising the information and viewpoints [1]. Furthermore, since computers cannot yet think like humans, teaching one to grasp sarcasm will be a challenging and time-consuming endeavour. This research aims to classify customer evaluations into three categories: positive, neutral, and negative, and to develop a supervised learning model to polarise massive amounts of data. We also create an unsupervised learning model to polarise feature-wise reviews. Customers' reviews and ratings, which we obtained from Amazon's Consumer Reviews, make up our dataset. We retrieved the features from our dataset and used them to build multiple supervised models. These models comprise not only conventional algorithms such as Random Forest, Multinomial Na¨ıve Bayes, Complement Na¨ıve Bayes, Bernoulli Na¨ıve Bayes, but also VADER Sentiment Analysis. The sentiment analysis tool VADER (Valence Aware Dictionary and Sentiment Reasoner), which is lexical and rule-based and specifically tailored to the sentiments expressed in social media, also performs well on texts from other domains. We better investigated the models' accuracy to understand people's divisive sentiments regarding the products.

II. LITERATURE REVIEW

Several notifiable related works can be mentioned under the area of study, Amazon Review Analytic. A qualitative analysis of the review content by polarity [7] was conducted to classify the review comment as either good, bad, or medium using TF-IDF word frequency count. The authors have measured the subjectivity of a review comment under the same study. The study also aims to determine a buyer's actual thinking about a product. The reviews are mainly done under four aspects: quality, service, comfort, and appearance. Considering these aspects, all the review comments are divided into five categories, pure praise, widespread praise (having some dissatisfaction), middle evaluation (more objective), negative evaluation due to some unexpected service attribute, and negative evaluation due to project attribute.

The study conducted for sentiment analysis of Amazon reviews and ratings [8] has explained a way to measure the compatibility between reviews with their corresponding ratings using a recurrent neural network (RNN) with a gated recurrent unit (GRU) that learned low- dimensional vector representation of reviews using paragraph vectors and product embedding. Fixed length feature vectors of the product reviews have been constructed using paragraph vectors that have been grouped, and each group has been trained with RNN with GRU. The final layer of the RNN produces what are known as "product embeddings," which are fixed-length vectors paired with paragraph vectors to train the support vector machine. These vectors capture crucial information like product attributes and temporal relationships across reviews. The provided user interface has facilitated the user to input the review, and as a result, the real rating of the review is generated, and any mismatch between the original rating is calculated and indicated to the user with a warning message.

An investigation [9] was conducted to create a deep neural network model (DNN) that includes an embedding layer, a long short-term memory (LSTM) part, and an output layer to predict the probability of a review becoming negative, neutral, and positive, and the test was conducted for 12 categories of Amazon product reviews. Reviews were constructed by concatenating the review body and summary and removing duplicates considering the user id and product id (when multiple comments were found for a product by the same user, the latest comment was used). To account for class imbalances, the initial output from the DNN model was recalibrated using the Conformal Prediction (CP) layer. Mondrian Conformal Prediction (MCP) does this recalibration independently for each class (negative, neutral, and positive, respectively).

III. METHODOLOGY

We have used customer comments on electronic products on Amazon, including 142.8 million reviews as per our training data set. This data set includes,

- reviewerID - ID of the reviewer
- overall - rating of the product
- ASIN - ID of the product
- reviewerName – The name of the reviewer
- reviewText - Text of the review
- summary - Summary of the review
- main cat – Related items of the review

In the study, several pre-processing and classification techniques were performed on top of the data set.

A. Data Pre-processing

To acquire the cleanest possible text data, we have done the basic data cleaning and pre-processing methods such as lemmatization, removal of unclean data (tags, stopwords, punctuation), and tokenization.

As the first step, the data cleaning process removed duplicates, missing values, and inconsistent data. Inconsistent data are primarily the records with incorrect ASIN id which does not follow the ASIN format. We decided the relevance of columns and dropped/kept them in accordance, primarily keeping only text data since the number of rows containing NANs was negligible compared to the overall number of rows we discarded.

The review text column storing textual data goes through multiple pre-processing stages before we run it through the model.

- Removing Punctuation – The removal of punctuation is a typical practice because it tends to mislead the models rather than aiding them in predicting the correct class. The punctuation removal and removal of unnecessary characters like emojis will help treat each text equally, reducing its randomness and bringing it closer to a predefined standard referred to as text normalization. This helps us reduce the with and improve efficiency.
- Removing stopwords - Stopwords are generally words appearing in any text corpus after looking at their frequencies in singular tokens. Articles (a the, an), Pronouns (you, them, they, me), and Prepositions are generally considered stopwords with little or no significance. To retain words with the most meaning and context, they are typically eliminated from the text during the pre-processing stage.
- Lemmatization - Lemmatization converts all word inflectional forms into a single common base or root.
- Tokenization - Words, sentences, phrases, and other portions of a written document are examples of tokens. Tokenization is the process of dividing a string of text into a token inventory.

B. Feature Extraction

The primary columns which are considered in the analysis are the overall column (rating of the product) and the review text column (textual description and product review). The reviews are not structured. In other words, the text lacks organization. However, sentiment analysis enables us to make sense of all this unstructured material by automatically categorizing it and processing massive amounts of data efficiently and cost-effectively.

The rating values in the training data set were converted into three classes considering the respective values. Ratings

5 and 4 in the training data set were labelled positive, and a new label was assigned as 2. Ratings 1 and 2 in the training data set were labelled negative, and a new labrandas assigned 0. Rating 3 in the training data set was labelled neutral, and a new label was assigned as 1. These classes are used as target classes for reviews in building a supervised model.

Numerical representation of textual data – We used TF-IDF Vectorizer and Count Vectorizer to transform textual information into a numerical form that our models can use. In Bag of words (Count Vectorizer), only count the number of times a word appears in the document, which results in bias with the most frequent words. The TF-IDF (Term Frequency-Inverse Document Frequency) method reveals the rarity of a word in the corpus by combining term frequency and inverse document frequency. If a term is uncommon, it is typically a distinctive word for a specific idea or piece of information.[4].

C. Model Selection

The data set (80% of the data set) was trained using four models with different parameters, and the best classification model that showed higher accuracy was selected. With the help of unigrams, TF-IDF, and Count Vectorizer, we used Random Forest and three Naive Bayes algorithms: Multinomial, Complement, and Bernoulli.

Table 1 shows a comparative study of different models, and we have used Random Forest Classifier with TF-IDF Vectorizer model for continuing with the sentiment analysis.

TABLE I. COMPARISON OF THE MODELS ACCURACY

Classification Model	**Accuracy**	
	CountVectorizer	*TF-IDF*
Random Forest	83.28%	82.75%
Multinomial Na¨ıve Bayes	82.69%	80.59%
Complement Na¨ıve Bayes	82.28%	78.31%
Bernoulli Na¨ıve Bayes	75.75%	74.92%

The authenticity of the reviews in the training data set was confirmed after manually evaluating the selected sample prior to training the models. This confirmed that the reviews describe the product features accordingly.

D. Sentiment Prediction

1) Overview Sentiment: In the Amazon product search web page, a popup window will be displayed in the right corner upon the user's click on the plugin, and the user can find the Overview sentiment of the product option is pre-selected. It gives an overview of sentiment with Positive, Negative, and Neutral sentiment percentage values in a chart.

2) Feature Sentiment: Pre-described popup gives an option to analyze the Feature sentiment of the product. Upon selecting the option, the system is capable of extracting sentiments regarding predefined features (Price, RAM, Processor, Battery, Display) of a product or service from user reviews. The rationale is that customers might cite various aspects in their reviews and have competing viewpoints on those features.

Our study separately used a predefined feature set related to product types (Camera and Computer). For example, Camera – {Price, Quality, ISO, Sensor} and Computer – {Price, RAM, Processor, Display}. We assumed that sentiment-bearing words such as adjectives are likely to be located close to a feature word within the distance of three tokens to either side.

To extract sentiment-bearing words for a feature word, we propose extracting a window of words around the latter, as shown in Fig. 2, and we used the VADER model to obtain and return the sentiment scores for full reviews. By adding the intensity of each word in a text, one can determine the sentiment score of that text. The model is included in the NLTK package and can be used to analyse unlabelled text data straightaway. [8]

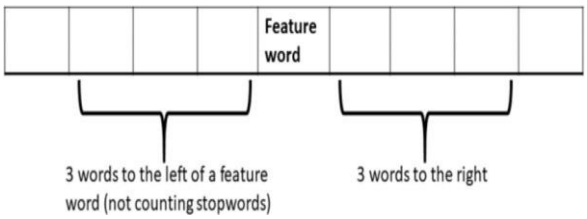

Fig. 2. Two windows each of three words on either side of the feature word

E. Component Architecture and Process Flow

Fig. 3 describes the component architecture of the solution. The solution is comprised of client-side implementation and a server-side implementation. The plugin development has been done as a part of the client-side implementation where Amazon specific plugin pop-up window has been integrated with the Chrome browser, which facilitates to trigger sentiment calculation request to the server end providing current product ASIN loaded in the browser (Client-side process is further described related to Fig. 4).

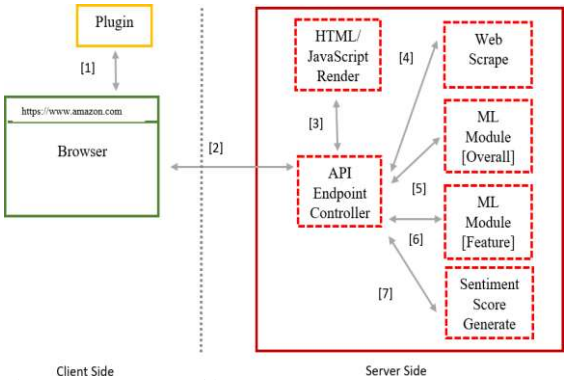

Fig. 3. Component Architecture

An endpoint has been established at the server-end accepting the client request and calculating the sentiment score, which is sent back as the response. The server-end solution comprises several sub-modules and calls sequences for each submodule from the endpoint controller, which has been annotated in Fig. 3. Upon receiving the client request, renders the Amazon product page based on the received ASIN with the ender module. Flask service deployed on a Docker is used for this purpose. The main requirement is rendering HTML components by executing required Java scripts.

The web scraping module accepts rendered HTML pages generated under step (3) and utilizes the Beautiful Soup library to perform scraping review comments. This is performed by targeting specific HTML class tags for reviews, and both review contents and the review titles are extracted. When reviews are expanded over multiple pages, each page is rendered under step (3), and scraping is performed. The output of this module is the extracted review list.

We chose a random forest classifier with TF IDF vectorizer as our machine learning algorithm after evaluating all the possible alternatives and the accuracy results (Table 1). This pre-trained model accepts each review collected under the step [4] It outputs the sentiment class of the review, whether positive, Neutral, or Negative feedback, as the result of the overall sentiment analysis. For feature sentiment analysis, an unsupervised approach has been used where an unsupervised Vader module has been incorporated into the solution to calculate a feature compound score.

Sentiment score generation is performed considering the total sentiment classes predicted in step (5) and using each class frequency. For example, when ten reviews are input to the ML module, which returns six positives, three negatives, and one neutral, the final scores are 60% positive, 30% negative, and 10% neutral. An average compound calculation is performed for feature analysis in step (6), and as the output, it provides the intensity of positive or negative sentiment for each feature.

Generated scores are sent back to the client-end, which displays the scores in the form of a Bar Chart. In this process flow, all the client-server back and forth communication is performed using JSON data format.

F. Client-side Implementation

The browser plugin comprises several scripts to maintain the back-and-forth communication between the plugin and the browser. Fig. 4 describes the front-end component linkage.

The plugin comprises a popup.html file and three JavaScript files named contentscript.js, popup.js, and background.js. Popup.js is responsible for the visual functionalities of the extension (popup.html) and interacts with background.js. Background.js is the extension's event handler where it listens to the extension's events, passes through the events to the contentscript.js, and passes the responses back to the extension. Contentscript.js is attached to the mainpage.html, and it accepts events from the background.js and is responsible for invoking the back end service endpoint and return the response received.

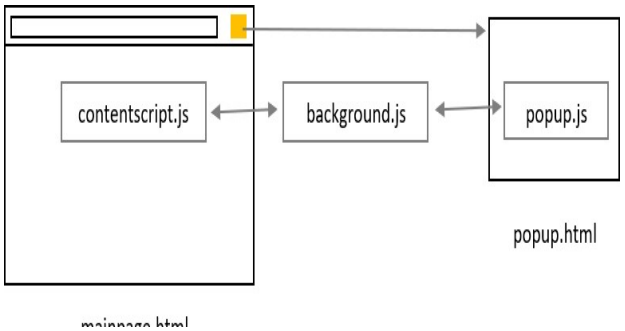

Fig. 4. Frontend component setup

G. Results

Fig.5 shows how the plugin icon has been aligned on the address bar and front-end popup window, which appear upon clicking on the plugin. The Popup window provides two options, selecting overview sentiment as the default option. Users can also go for feature sentiment which product they want to analyse.

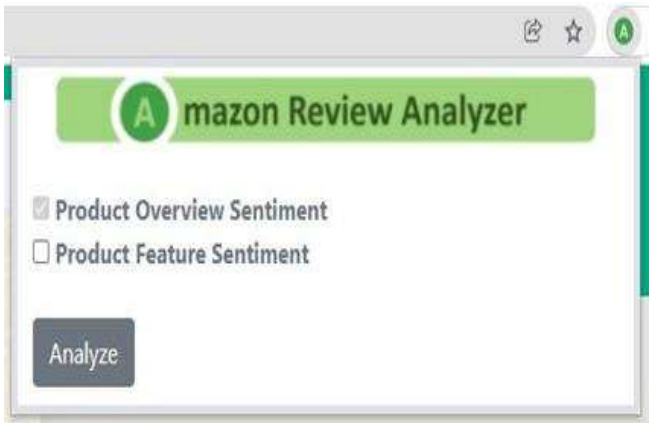

Fig. 5. Popup window

Fig. 6 shows the product overview sentiment output of an Amazon product. Negative sentiment shows in the red bar, while positive sentiment and neutral sentiment are shown in the green and yellow bars, respectively.

Fig. 6. Product overview sentiment output

Fig. 7 shows the product feature sentiment output of an Amazon product which conveys a more detailed idea to the customer. As it provides drilled-down feature-wise categorization, customers get the opportunity to be aware of a product's most prominent features and lagging features. The output stacked bar chart consisted of different colour schema with percentages representing the categories for each feature. , conveying a visual and informative message to the customer, which helps him decide whether to buy the product.

Fig. 7. Product feature sentiment output

IV. Conclusion

The proposed solution is comprised of a supervised model and an unsupervised model. After building several supervised models, the Random Forest Model with TF-IDF vectorization was selected. For supervised model training, customer reviews and ratings were used as the target classes. The unsupervised model was based on the VADER algorithm. The supervised model was used for overall sentiment analysis, while the unsupervised model was used for feature sentiment analysis. The customer is provided with plugin implementation attached to the Chrome browser providing a popup where the customer can interact, and results are shown that are supported to the customer's decision-making process whether to buy or not the product. This, in our opinion, will be extremely helpful in a variety of product and user relationship study contexts. A possible usage for this would be in recommendation systems, where consumers may be grouped based on the similar reviews they post on websites like Amazon.

V. Future Works

Results can be improved in terms of accuracy by using BERT (Bi-directional Encoder Representation of Transformer) , a machine learning technique developed by Google based on the Transformers mechanism and pre-trained model commonly used in NLP. Due to its bi-directional approach, the model can quickly identify the context of a word in a sentence depending on the words that came before it in the sentence.[6]. Since BERT demands significant computational resources and takes a long time to train on the data frame, we used the DistilBERT model and removed some rows from our data set to shorten the training period. It gave almost 89% accuracy, which is a considerably higher value.We used pre-defined features for implementing the feature sentiment solution. Instead, we can use topic modeling to recognize the features from the reviews. LDA (Latent Dirichlet Allocation) algorithm can be used as a topic modeling approac Moreover, we plan to develop the system as a recommendation system on top of the current solution. This solution is capable of proposing related products with features that have the highest ratings and are preferred by the customer.

References

[1] B.Liu, . & L. Zhang. "A survey of opinion mining and sentiment analysis", 2012.
[2] S. Brownfield, J. Zhou. "Sentiment Analysis of Amazon Product Reviews". vol 1295. Springer, Cham, 2020. Page ??
[3] D. Gamal, et al. "Analysis of Machine Learning Algorithms for opinion mining in different domains". Machine Learning and Knowledge Extraction 1, no. 1: 224-234. 2019. https://doi.org/10.3390/make1010014
[4] T.U Haque, N.N Saber, N.N., & F.M Shah. "Sentiment analysis on large scale Amazon product reviews". pp. 1-6, 2018. doi: 10.1109/ICIRD.2018.8376299.
[5] K. Dave, S. Lawrence, & Pennock, D.M. "Mining the peanut gallery: opinion extraction and semantic classification of product reviews", 2003.
[6] Geetha, M.M., & Karthika R. D. "Improving the performance of aspect based sentiment analysis using fine-tuned BERT base uncased model", 2021.
[7] Y. Xiao, C. Q. & Leng H. "Sentiment analysis of Amazon product reviews based on NLP," 2021, pp.1218-1221, doi: 10.1109/AEMCSE51986.2021.00249.
[8] Nishit, S., & Fatma, N. "Deep Learning sentiment analysis of Amazon.com reviews and ratings", 2019. https://doi.org/10.48550/arXiv.1904.04096

[9] Norinder, U. & Norinder, P. "Predicting Amazon customer reviews with deep confidence using Deep Learning and conformal prediction", 2022. https://doi.org/10.1080/23270012.2022.2031324
[10] M. Munikar, S. Shakya & A. Shrestha. "Fine-grained sentiment Classification using BERT," 2019, pp. 1-5, doi: 10.1109/AITB48515.2019.8947435.
[11] Ashima Y. & Dinesh K. V. "Sentiment analysis using deep learning architectures", 2020. https://doi.org/10.1007/s10462-019-09794-5
[12] Z. Gao, A., Feng, X. S. & X. Wu, "Target-dependent sentiment classification with BERT". vol. 7, pp. 154290-154299, 2019. doi: 10.1109/ACCESS.2019.2946594.
[13] https://gilfink.medium.com/using-messaging-in-chrome-extension-4ae65c0622f6 [Accessed 10 02 2022]
[14]https://medium.com/tech-tajawal/build-a-simple- google-chrome-extension-in-few-minutes-1f13b600e83e [Accessed 12 02 2022]