# Aspect Based User Reviews Classification

Roch DANISTAN, Thulasika ARUNAKIRINATHAN, Archchana SIVARAJAH, Yanusha MEHENDRAN, Jayalath EKANAYAKE

(*Dept. of Computer Science and Informatics*, *Faculty of Applied Sciences*, *Uva Wellassa University*, *Passara Rd. Badulla*, 90000, *Sri Lanka*)

**Abstract:** At present online shopping is very popular as it is very convenient for the customers. However, selecting smartphones from online shops is bit difficult only from the pictures and a short description about the item, and hence, the customers refer user reviews and star rating. Since user reviews are represented in human languages, sometimes the real semantic of the reviews and satisfaction of the customers are different than what the star rating shows. Also, reading all the reviews are not possible as typically, a smartphone gets thousands of reviews in popular online shopping platform like Amazon. Hence, this work aims to develop a recommended system for smartphones based on aspects of the phones such as screen size, resolution, camera quality, battery life etc. reviewed by users. To that end we apply hybrid approach, which includes three lexicon-based methods and three machine learning modals to analyze specific aspects of user reviews and classify the reviews into six categories--*best, better, good* or *somewhat* for positive comments and for negative comments *bad* or *not recommended*--. The lexicon-based tool called *AFINN* together with Random Forest prediction model provides the best classification F1-score 0.95. This system can be customized according to the required aspects of smartphones and the classification of reviews can be done accordingly.

**Key words:** User-reviews Classification, Aspects of Smartphones, Reviews of Smart Phones, Classification Algorithms, Lexicon-based Methods, Sentimental Analysis

## 1 Introduction

When a person is looking for an item to buy, the experiences of those who have already made that decision matters. What did the other users like? What did they dislike? And would they choose the same if they could do it again? The user reviews are very important for both the consumer and the business [1-4]. The customers get insights from the previous users about their feelings towards a product whereas for the businessmen or manufacturers of the product, positive reviews effect as a form of encouragement and negative reviews provide feedback on required improvements.

A large number of different brands of smartphone are available in markets. Selecting a phone from a large collection is tricky as different brands have different attributes and qualities. Specially, choosing a smartphone from online shops is a challenging task as little information such as images and small description about the product is available for the customers. Besides this information, user reviews about the smartphone are available. However, reading all the user reviews is not a trivial task as one product gets thousands of reviews specially, in popular e-commerce platforms like Amazon. Also, it is very difficult to manually gather and analyze the large amount of information due to human mental and physical limitations. Normally the star rating of a product is a popular measure for smartphones evaluation. The average number of stars received for a smartphone is used to evaluate the overall smartphone quality. However, the start rating is calculated using a mathematical expression rather than using actual sentimental analysis [5-6].

In recent years, sentiment analysis is a major field in Natural Language Processing (NLP) that studies methods for identifying and extracting opinions from written text, such as product reviews, discussion forums etc. [7]. This excellent source of information is useful to gather opinions about a particular product. Sentiment analysis has found interest in commercial applications as it can use in almost every business purpose. Opinion on different perspectives of a product is very helpful for the customers to evaluate the product. Hence, the aspect-based sentiment analysis is essential as it gives summary of online reviews in a descriptive way such as positive, negative or neutral sentiment scores for each aspect level [8-9]. This information is useful for consumers and manufacturers.

At present smartphones are produced as different mobile brands in the world. People are willing to buy trending smartphones by analyzing each and every aspect of different brands and hence, it becomes very easy for customer to choose a smartphone if they get reviews or opinions on particular feature of the product. For example, a smartphone contains many attributes such as screen size, resolution, processor speed, storage, camera quality, battery life and so on. Some users prefer to use their phone to watch videos. Accordingly, they are more concerning on screen size, resolution, processor speed and battery life whereas the users enjoy selfie focus more on camera quality. Thus, when they write their opinions, they tend to describe their feeling towards the features relevant to them. These opinions are very useful for the future customers as well as the manufacturers to evaluate the products from the perspective of different aspects and make a decision to buy or to further improve the quality of the products.

Several researchers carried out projects similar to this project. Among them Nguyen [10] and Shaheen [11] are closely related to our project. However, they classified user reviews into two categories--positive and negative--whereas this project classifies into six categories. Predicting into six categories is somewhat a challenging task than predicting into two categories. Furthermore, our models outperform the above models providing the prediction accuracy of 0.95 F1-score.

This manuscript proposes a method to analyze the user reviews on different aspects of smartphones and classify them into six categories--*best, better,* good, or *somewhat* for positive comments and for negative comments *bad* or *not recommended*--. This information is very useful for the customers to evaluate the performances of the phones on different features and hence, they can decide whether to buy that phone.

## 2 Related Works

Chauhan et al. [2] explored the effect of adverbs in user reviews on the classification of sentiments. They evaluated over 50,000 reviews of two products, office products and musical DVDs collected from Amazon and found that two general superlative adverbs and degree-wh adverb have more impact than the other forms of adverbs.

Wang et al. [12] proposed a method to analyze user reviews of online food delivery platform called Meituan language correlation function. They used featured words and the emotions in the reviews and concluded that the food delivery service is cheap, economical, convenient and fast.

Mathew et al. [13] analyzed sentiments of user reviews using the K-Nearest Neighbor (KNN) algorithm. First, the reviews were tokenized and the clusters were formed using KNN algorithm, and then the testing data points were assigned to the nearest clusters based on the majority votes. The prediction quality of the models is 83.65%. Further, the Naïve Bayes model was trained from the same dataset and the accuracy is 75.77%, which is less than the KNN model.

Lee et al. [14] developed a method named as MPM (mining perceptual map) to generate perceptual maps and radar charts from user reviews of smartphones. The outcome of the study shows that using MPM, valuable information, which is useful for smartphone companies, can be mined from the posted reviews.

Kim et al. [15] conducted an empirical analysis on user reviews of smartphones over a 10-year period for more than 300 brands of 32 manufacturers. They used a hierarchical Bayes model and found that the appearance aspects of mobile phones are declined whereas the technology aspects increase.

Arora et al. [16] illustrated important of sentiment analysis on user reviews about popular smartphone brands. They analyzed the reviews posted on Tweeter about some features of smartphones such as the battery life, screen quality, and performance of mobile OS. The results of the study show that the Twitter data provides users' opinions on some brands but does not provide sufficient information on some other brands and hence, conducting a comprehensive analysis is somewhat challenging task.

Zhong et al. [17] proposed a method to analyze consumer purchasing patterns in divers' culture such as British, American and Indian from the consumer reviews. They used lexicon-based approach to analyze the reviews on product features and classified them as positive or negative. Kaushik et al. [3] discussed the approaches and tools in semantic analysis and further discussed the important of sentiment analysis in several domains.

Nguyen et al. [10] developed a method for sentimental analysis from user reviews. They applied supervised machine learning algorithms--Logistic Regression (LR), Support Vector Machine (SVM), and Gradient Boosting--and three lexicon-based algorithms--VADER, Pattern, and SentiWordNet--. The user reviews are classified into two classes positive and negative. The SVM together with VADER outperformed the other two algorithms providing 0.94 F1-score. This work is pretty similar to our work. However, we used Naïve Bayes, SVM and Random Forests as classification algorithms and TextBlob, VADER and AFINN as sentimental analysis tools. Further, we classified the user reviews into six categories and achieve a comparatively higher F1-score value 0.95.

Another similar project was conducted by Shaheen et al. [11]. They trained seven classifiers; Gradient Boosting, SGD, Multinomial NB, LSTM, Random Forest, NB-SVM and CNN for opinion mining on user reviews of mobile phones. According to the results, the random forest classifier outperforms all the other classifiers with 84% accuracy whereas our approach preforms comparatively better.

Researchers conducted research on user reviews analysis in different domains such as restaurants. Mubarok et al. [18] developed an approach for aspects-based sentiment analysis on user reviews of products and services related to restaurants. The approach contained three phases; data preprocessing, features selection and classification. The Naïve Bayes classifier was used for classification. The results show that the classification accuracy is 78% of F1-measure.

We obtained information from above references for this work. However, our hybrid approach for user-reviews classification enhances the accuracy. Further, the reviews are extracted based on different aspects of smartphones and the classification is conducted accordingly.

## 3 Methodology

### 3.1 Data Collection

The user reviews of smartphones available at many sites. However, we use the reviews in amazon. We use the Scrapper Tools; Selenium and Scrap Hero to collect reviews for modern smartphones from amazon website. These tools collect review texts based on required parameters such as Product Name, Brand Name, Rating, Reviews, and Reviews Votes. Finally, we collected around 50,000 reviews text on different brands on trending smartphones from Amazon website. However, the collected reviews texts are unclassified, hence, we label them using a lexicon-based approach, which is described in flowing sections.

### 3.2 Data Pre-processing

#### 3.2.1 Data Cleaning

Data cleaning is essential as many of the real-world data are mixed with some form of noisy. Without proper data quality, our final analysis would suffer inaccuracy, or we could potentially arrive at a wrong conclusion.

Any person can enter a review. Some of the reviews are not relevant for this study as the reviews are not described any of the considered aspects of smartphones. Such reviews are filtered out as noisy data. To that end, we develop a function in Python codes to extract the review texts containing the considered properties of smartphones.

#### 3.2.2 Data Tokenization

Tokenization splits longer strings of text into smaller pieces, or tokens at the places of delimits. A delimit may be a space or a punctuation mark. Larger chunks of text are often tokenized into sentences; sentences are often tokenized into words, etc. We make sequence of review texts, which contain different aspects into pieces like words, keywords, phrases, symbols and other elements called tokens.

#### 3.2.3 Stop-word Removal

A textual description may contain words such as the, is, at, which, etc. These words have minimal lexical meaning and hence can be removed from a description without a significant change in context or semantic of the description. There are many stop-word removal techniques [19-22]. The approach presented by Fox [19], generates stop word lists mentioning differences between stated conventions and actual instances. This project uses a list of stop words generated by Fox. The stop words are filtered out as they are considered to be uninformative or meaningless when tokenizing a textual description.

#### 3.2.4 Stemming

Data Stemming is Text Normalization (or sometimes called Word Normalization) technique in the field of Natural Language Processing (NLP) that is used to prepare text, words, and documents for further processing. Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes. By using data stemming we reduced different forms of a word and removed suffix, prefix in sentences. It helped us to accomplish our next step too.

Following functions in Python are used for tokenization, stop words removal and stemming.

*tokenize = sent_tokenize*

*stopwords = set(stopwords.words("english")*

*stemmer = PorterStemmer()*

*stemmer = SnowballStemmer('english')*

### 3.3 Aspect Extraction

Our main objective is to classify the smartphones based on the aspects reviewed by the users. So, we consider the aspects mentioned in review texts as the features of smartphones. The main aspects, which the users discussed about, are battery, screen, storage, price and camera of smartphones, and hence, we select those aspects as the main features of the smartphones. So, we select customer reviews, which contain the above set of words for further processing. Due to this fact the final dataset contained about 30,000 review texts.

### 3.4 Data Labeling

As we train supervised classification models, the dataset should be labeled. To that end we use three lexicon-based approaches--VADER, TextBlob and AFINN--to calculate sentiment scores for each user review. Subsequently, the user reviews are labeled as *best, better,* good, or *somewhat* for positive comments and for negative comments *bad* or *not recommended*, based on the sentiment scores.

**Table 1 Class Distribution of Datasets**

| Class | VADER | AFINN | TextBlob |
|---|---|---|---|
| better | 2211 | 5887 | 4107 |
| best | 4706 | 3103 | 2954 |
| good | 4505 | 3867 | 3975 |
| somewhat | 12233 | 11312 | 12941 |
| bad | 3391 | 3759 | 1632 |
| Not recommended | 882 | 3588 | 2319 |

Table 1 shows the class distribution of the datasets extracted from each lexicon-based tool.

#### 3.4.1 VADER (Valence Aware Dictionary for Sentiment Reasoning)

VADER is a lexicon-based approach to categorize user review texts. VADER provides score values for Vader sentiment, Vader sentiment positive and Vader sentiment negative. The sentiment values are ranging from −1 to +1. In here, we have divided into multiple classes. Then we categorized the score to assign a dictionary word and labelled them as in Table 2.

**Tabel 2 Categorization of Sentimental Scores in VADER**

| Score | Class | Label |
|---|---|---|
| >= 0.75 | best | 4 |
| >= 0.5 | better | 3 |
| >= 0.25 | good | 2 |
| >= 0 | somewhat | 1 |
| >= (–0.25) | bad | –1 |
| Otherwise | not recommended | –2 |

#### 3.4.2 TextBlob

TextBlob is another lexicon-based approach to calculate sentiment scores. TextBlob calculates the text sentiment and subjectivity scores for each review. Similar to VADER, the sentimental scores of TextBlob are ranging from −1 to +1. However, TextBlob uses different algorithms to calculate the sentimental scores. Hence, the sentiment scores are categorized same as in VADER (as in the Table 2).

#### 3.4.3 AFINN

The third lexicon-based approach is AFINN. The calculated sentimental value is known as AFINN score and the value is ranged from -15 to 15. The sentimental score is categorized and labeled as in Table 3.

**Table 3 Categorization of Sentimental Scores in AFINN**

| Range | Class | Label |
|---|---|---|
| >=5 | best | 4 |
| >=3 | better | 3 |
| >=1 | good | 2 |
| >=0 | somewhat | 1 |
| >=–1 | bad | –1 |
| otherwise | not recommended | –2 |

### 3.5 Count Vectorizer

The Count Vectorizer provides a simple way for both tokenize a collection of texts and build a vocabulary of known words. Based on this vocabulary a new text is encoded. The encoded vector is returned with a length of the entire vocabulary and an integer count for the number of times each word appeared in the text. Finally, the text is represented as set of keywords and the numerical values represent the frequency of keywords appeared in the text.

### 3.6 Sentiment Classification

Sentiment classification is a technique to classify the texts into defined classes. The classification is conducted using classifiers. First, the classifiers are trained and then tested to evaluate their performances. To that end the labeled dataset is divided into two parts as training and testing. Two third of the dataset is used for training and the rest is used for testing the models. A classifier trains itself using training data and checks its accuracy on testing data. There are different types of classifiers, which can be utilized in classification. In this project we use Naïve bays, Random Forests and Support Vector Machine (SVM). The accuracy of the models is measured using accuracy, precision, recall and F1-score.

Our approach for user review classification is hybrid as first; we label the reviews using lexicon-based approach and then apply classification models for categorizing the reviews.

VADER, TextBlob and AFFIN are implemented using the relevant libraries linked with python code. The SVM, Random Forest and Naïve Bayes algorithms are also implemented in python using the library Tensor Flow. Each classification algorithm is trained and tested using the datasets created by VADER, TextBlob and AFFIN. Consequently, we developed nine models and measured the accuracy of each model using accuracy, precision, recall and F1-score.

## 4 Results and Discussion

All the combinations of the lexicon-based labeling tools and the classification models generate nine models. Table 4 shows the summary of the models' performances. However, classification accuracy alone cannot be trusted when presenting imbalanced datasets like in our case (see Table 1). Therefore, a detail evaluation of each model is presented in Tables 5-13.

**Table 4 Prediction accuracy of models**

| Lexicon-based Tool | Classification Model | Accuracy |
|---|---|---|
| VADER | Naïve Bayes | 73% |
| VADER | Support Vector Machine | 83% |
| VADER | Random Forests | 90% |
| AFINN | Support Vector Machine | 89% |
| AFINN | Random Forests | 92% |
| AFINN | Naïve Bayes | 79% |
| TextBlob | Support Vector Machine | 85% |
| TextBlob | Random Forests | 90% |
| TextBlob | Naïve Bayes | 75% |

**Table 5 VADER and Naïve Bayes**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –2 | 0.82 | 0.22 | 0.34 | 244 |
| –1 | 0.79 | 0.67 | 0.72 | 1016 |
| 1 | 0.84 | 0.88 | 0.86 | 3713 |
| 2 | 0.61 | 0.61 | 0.61 | 1315 |
| 3 | 0.55 | 0.74 | 0.63 | 1378 |
| 4 | 0.80 | 0.39 | 0.52 | 692 |
| Accuracy | | | 0.73 | 8358 |
| Macro avg | 0.73 | 0.58 | 0.61 | 8358 |
| Weighted avg | 0.74 | 0.73 | 0.72 | 8358 |

**Table 6 VADER and SVM**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –2 | 0.78 | 0.55 | 0.65 | 244 |
| –1 | 0.84 | 0.81 | 0.82 | 1016 |
| 1 | 0.88 | 0.96 | 0.92 | 3713 |
| 2 | 0.72 | 0.71 | 0.72 | 1315 |
| 3 | 0.77 | 0.72 | 0.74 | 1378 |
| 4 | 0.84 | 0.70 | 0.76 | 692 |
| Accuracy | | | 0.83 | 8358 |
| Macro avg | 0.81 | 0.74 | 0.87 | 8358 |
| Weighted avg | 0.83 | 0.83 | 0.83 | 8358 |

**Table 7 VADER and Random Forest**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –2 | 0.92 | 0.72 | 0.80 | 244 |
| –1 | 0.92 | 0.86 | 0.89 | 1016 |
| 1 | 0.92 | 0.98 | 0.95 | 3713 |
| 2 | 0.85 | 0.87 | 0.86 | 1315 |
| 3 | 0.88 | 0.86 | 0.87 | 1378 |
| 4 | 0.97 | 0.79 | 0.87 | 692 |
| Accuracy | | | 0.90 | 8358 |
| Macro avg | 0.91 | 0.85 | 0.87 | 8358 |
| Weighted avg | 0.91 | 0.90 | 0.90 | 8358 |

**Table 8 AFFIN and SVM**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –1 | 0.91 | 0.87 | 0.89 | 1106 |
| 1 | 0.92 | 0.97 | 0.94 | 3380 |
| 2 | 0.81 | 0.82 | 0.82 | 1151 |
| 3 | 0.87 | 0.82 | 0.84 | 1818 |
| 4 | 0.86 | 0.81 | 0.84 | 924 |
| Accuracy | | | 0.89 | 8379 |
| Macro avg | 0.88 | 0.86 | 0.87 | 8379 |
| Weighted avg | 0.89 | 0.89 | 0.89 | 8379 |

**Table 9 AFINN and Random Forest**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –1 | 0.98 | 0.84 | 0.91 | 1106 |
| 1 | 0.91 | 0.99 | 0.95 | 3380 |
| 2 | 0.88 | 0.85 | 0.87 | 1151 |
| 3 | 0.90 | 0.92 | 0.91 | 1818 |
| 4 | 0.95 | 0.82 | 0.88 | 924 |
| Accuracy | | | 0.92 | 8379 |
| Macro avg | 0.92 | 0.89 | 0.90 | 8379 |
| Weighted avg | 0.92 | 0.92 | 0.92 | 8379 |

**Table 10 AFINN and Naïve Bayes**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –1 | 0.88 | 0.78 | 0.83 | 1106 |
| 1 | 0.89 | 0.88 | 0.88 | 3380 |
| 2 | 0.72 | 0.66 | 0.69 | 1151 |
| 3 | 0.65 | 0.81 | 0.72 | 1818 |
| 4 | 0.71 | 0.58 | 0.64 | 924 |
| Accuracy | | | 0.79 | 8379 |
| Macro avg | 0.77 | 0.74 | 0.75 | 8379 |
| Weighted avg | 0.79 | 0.79 | 0.79 | 8379 |

**Table 11 TextBlob and SVM**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –2 | 0.80 | 0.71 | 0.76 | 686 |
| –1 | 0.85 | 0.80 | 0.83 | 471 |
| 1 | 0.88 | 0.93 | 0.90 | 3918 |
| 2 | 0.76 | 0.66 | 0.70 | 1188 |
| 3 | 0.81 | 0.79 | 0.80 | 1201 |
| 4 | 0.86 | 0.93 | 0.90 | 915 |
| Accuracy | | | 0.85 | 8379 |
| Macro avg | 0.83 | 0.80 | 0.81 | 8379 |
| Weighted avg | 0.84 | 0.85 | 0.84 | 8379 |

**Table 12 TextBlob and Random Forest**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| –2 | 0.90 | 0.81 | 0.85 | 686 |
| –1 | 0.92 | 0.79 | 0.85 | 471 |
| 1 | 0.92 | 0.96 | 0.94 | 3918 |
| 2 | 0.89 | 0.80 | 0.85 | 1188 |
| 3 | 0.86 | 0.95 | 0.91 | 915 |
| 4 | 0.86 | 0.87 | 0.86 | 1201 |
| Accuracy | | | 0.90 | 8379 |
| Macro avg | 0.89 | 0.86 | 0.88 | 8379 |
| Weighted avg | 0.90 | 0.90 | 0.90 | 8379 |

**Table 13 TextBlob and Naïve Bayes**

| Class Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| −2 | 0.80 | 0.48 | 0.60 | 686 |
| −1 | 0.87 | 0.50 | 0.63 | 471 |
| 1 | 0.78 | 0.88 | 0.83 | 3918 |
| 2 | 0.64 | 0.56 | 0.60 | 1188 |
| 3 | 0.67 | 0.74 | 0.70 | 1201 |
| 4 | 0.77 | 0.78 | 0.78 | 915 |
| Accuracy | | | 0.75 | 8379 |
| Macro avg | 0.76 | 0.66 | 0.69 | 8379 |
| Weighted avg | 0.75 | 0.75 | 0.74 | 8379 |

According to Table 5, AFINN together with Random Forests Algorithm provides the best classification accuracy 92%. Also, as in Table 9 the F1-score of the Random Forest classifier is ranging from 0.87 to 0.95 for each class. This indicates that the Random Forest equally well performs in predicting the classes even the dataset is highly skewed. More importantly the Random Forest provides 0.95 F1-score for classifying the class label 1 (category *somewhat*), which contains the highest number of instances (3380 test cases).

Further, AFINN in generally produces comparatively decent classification accuracy with all the three classification models. The lexicon score of AFINN is ranging from −15 to +15 whereas TextBlob and VADER are ranging from −1 to +1.

The Random Forest classifier together with AFINN, VADER and TextBlob provides overall

classification accuracies 92%, 90% and 89% respectively. Hence, the prediction quality of the Random Forest classifier is comparatively better than the other two classifiers; Naïve Bayes and SVM. According to Tables 7,9 and 12, the prediction

quality of the Random Forest in each class (−2 to +4) is above 0.80 F1-score, which indicates the Random Forest equally well performs in predicting the classes even the dataset is highly skewed.

According to the results, the Naïve Bayes classifier produces the lowers accuracy. But the Naïve Bayes is well known for text classification for a long time. However, this project shows that the Random Forest is also a good candidate for text classification as Naïve Bayes.

According to the litreature the apporoach proposed by Nguyen et al. [10] provided the highest accuracy 0.94 F1-score. However, our method obtained 0.95 F1-score for the class representing the *somewhat* category, which contains the highest number of instances. This shows that our approach further enhances the user-reviews classification accuracy.

## 5 Conclusion

Sentiment analysis is a technique to identify sentiments in terms of polarities. Aspect level sentiment analysis is a technique, which focuses on particular important features in the user-reviews. The aim of this project is to classify user-reviews based on different aspects of smartphones. The user-reviews on smartphones were collected from amazon website using the scrapper tools, Selenium and Scrap Hero and labeled them into eight categories based on sentimental scores calculated by three lexicon-based tools. Three classifiers together with three lexicon-based tools generated nine different combinations of classifiers. AFINN together with Random forest classifier outperformed all other combinations providing 92% accuracy. Further, the Random Forest classifier equally preformed on each class though the class distribution is highly imbalance. Hence, the users can use this method to evaluate the overall quality of smartphones.

This project extracts the user reviews only about smartphones from the Amazon site. Also, the reviews are on perticuler aspects of the phones. However, in future studies, the reviews will be extrcated from other online shopping platfomes. Also, the project will be extended to analyse the reviews of other items and services.

Summarizing, this project shows that the user-reviews can be extracted based on different aspects and then classified into pre-defined categories with a decent accuracy. Further, the Random Forest classifier is also a good candidate for text classification as Naïve Bayes.

## References


[1] Abhishek K., Anchal K., Sudhanshu, N. (2015). A Study on Sentiment Analysis:Methods and Tools, *International Journal of Science and Research (IJSR). Volume 4 Issue 12*.

[2] Arora, D., Li, K.F. and Neville, S.W. (2015). March. Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study. *In 2015 IEEE 29th International Conference on Advanced Information Networking and Applications,* pp. 680-686, IEEE.

[3] Cao, Q., Duan, W. and Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, *50*(2), pp.511-521.

[4] Chauhan, U.A., Afzal, M.T., Shahid, A., Abdar, M., Basiri, M.E. and Zhou, X. (2020). A comprehensive analysis of adverb types for mining user sentiments on Amazon product reviews. *World Wide Web*, pp.1-19.

[5] Chong, T.Y., Banchs, R.E. and Chng, E.S. (2012). An empirical evaluation of stop word removal in statistical machine translation. *In Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation,* pp. 30-37.

[6] Fox, C. (1989). A stop list for general text. *In Acm sigir forum (Vol. 24, No. 1-2, pp. 19-21). New York, NY, USA: ACM*.

[7] Heidi, N. (2018). Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches. *ISSN: 0975 – 6760| Nov 12 to Oct 13, volume – 02, issue – 02*

[8] Hu, H., Bezemer, C.P. and Hassan, A.E. (2018). Studying the consistency of star ratings and the complaints in 1 & 2-star user reviews for top free cross-platform Android and iOS apps. *Empirical Software Engineering, 23(6),* pp. 3442-3475.

[9] Kaushik, A. and Naithani, S. (2015). A study on sentiment analysis: methods and tools. *International Journal of Science and Research (IJSR) ISSN (Online)*, pp. 2319-7064.

[10] Kim, S., Katsumata, S. and Akiike, A. (2020). Dynamic Impacts of Aspects of Appearance and Technology on Consumer Satisfaction: Empirical Evidence from the Smartphone Market. The Review of Socionetwork Strategies, pp.1-21.

[11] Lackermair, G., Kailer, D. and Kanmaz, K. (2013). Importance of online product reviews from a consumer’s perspective. *Advances in economics and business, 1(1),* pp.1-5.

[12] Lee, A.J., Yang, F.C., Chen, C.H., Wang, C.S. and Sun, C.Y. (2016). Mining perceptual maps from consumer reviews. *Decision Support Systems, 82*, pp.12-25.

[13] Mathew, L., James, A.D., Shaji, A. and Sebastian, M.S.M. (2020). Opinion Mining/Sentiment Analysis for User Reviews. *(No. 2682), EasyChair*.

[14] Mubarok, M.S., Adiwijaya and Aldhi, M.D. (2017). August. Aspect-based sentiment analysis to review products using Naïve Bayes. *In AIP Conference Proceedings (Vol. 1867, No. 1, p. 020060). AIP Publishing LLC*.

[15] Nguyen, H., Veluchamy, A., Diop, M. and Iqbal, R., 2018. Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches: *SMU Data Science Review, 1(4), p.7*.

[16] Pimpalshende, A. and Mahajan, A.R. (2017). September. Test model for stop word removal of devnagari text documents based on finite automata. *In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI*), pp. 672-674 IEEE.

[17] Plante, T.B., O’Kelly, A.C., Macfarlane, Z.T., Urrea, B., Appel, L.J., Miller III, E.R., Blumenthal, R.S. and Martin, S.S. (2018). Trends in user ratings and reviews of a popular yet inaccurate blood pressure-measuring smartphone app. *Journal of the American Medical Informatics Association*, 25(8), pp.1074-1079.

[18] Shaheen, M., Awan, S.M., Hussain, N. and Gondal, Z.A.

(2019). Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques. *International Journal of Modern Education & Computer Science, 11(7)*.

[19] Silva, C. and Ribeiro, B. (2003). The importance of stop word removal on recall values in text categorization. *In Proceedings of the International Joint Conference on Neural Networks, Vol. 3,* pp. 1661-1666, IEEE.

[20] Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.F. and Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing,* 65, pp.3-14.

[21] Wang, Y.J., Yu, G.H., Sun, L.N. and Liu, P.G. (2020). June. Text Mining and Analysis of Meituan User Review Text. *In International Conference on Green Communications and Networking (pp. 151-155). Springer, Cham.*

[22] Xu, H., Liu, B., Shu, L. and Yu, P.S. (2019). Bert post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232*.

[23] Xue, W. and Li, T. (2018). Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*.

[24] Zhong, Q., Liang, S., Cui, L., Chan, H.K. and Qiu, Y. (2019). Using online reviews to explore consumer purchasing behaviour in different cultural settings. *Kybernetes*.

## Author Biographies

**Roch DANISTAN,** is currently pursuing his B.Sc. (Sp) Degree in Computer Science & Technology from Uva Wellassa University of SriLanka. His main research interests include Sentiment Analysis and Machine Learning.

Email: rochdanistan123@gmail.com

**Thulasika ARUNAKIRINATHAN,** is currently pursuing her B.Sc.(Sp) Degree in Computer Science & Technology from Uva Wellassa University of SriLanka. Her main research interests include Sentiment Analysis and Machine Learning.

Email: thulasika95@gmail.com

**Archchana SIVARAJAH,** is currently pursuing her B.Sc. (Sp) Degree in Computer Science & Technology from Uva Wellassa University of SriLanka. Her main research interests include Sentiment Analysis and Machine Learning.

Email: archchu95@gmail.com

**Yanusha MEHENDRAN,** received Master's in Computer Science degree from University of Colombo School of Computing in 2020. She is currently working as Lecturer (Probationary) in Uva Wellassa University. Her main research interests include computer vision and machine learning.

Email: Yanusha@uwu.ac.lk

**Jayalath EKANAYAKE,** is currently working as a lecturer in Computer Science at the Uva Wellassa University, Sri Lanka. His main research interest is pattern recognition.

Email: jayalath@uwu.ac.lk