Feature-Based Opinion Mining on Cellphone Reviews

Suraj Suresh Sajjan
Applied Modelling and Quantitative Methods: Big Data Analytics
Trent University
Peterborough, Canada
surajsureshsajjan@trentu.ca
0649311

***Abstract* - In the world we live in today, more people are getting introduced to the digital way of life than ever before. In such an ever-changing world, shopping on e-commerce websites online is one of the domains with a rising scope and importance where people can purchase almost everything they need at the convenience and comfort of their house. As more people rely on online shopping, the importance of customer reviews on products increases significantly. Opinion mining is one of the disciplines which tackles the analysis of such customer reviews. Feature-based opinion mining is one of the various tasks related to opinion mining. In this report, feature-based opinion mining is being focused upon.**

Keywords - SentiWordNet; Machine Learning; Linguistic Patterns; Opinion Mining; Feature/Aspect.

1. Introduction

Opinion Mining is a study in which people's sentiments, opinions, perceptions and emotions regarding entities, events and the properties associated with them are extracted [1]. In recent times, online shopping has become quite popular with the increasing number of people who prefer to purchase products online. Prior to the purchase, people tend to read the reviews and comments left by other customers to gather information about the product. These reviews are generally long and unstructured. It also makes it difficult for a potential buyer to read through thousands of reviews, hence making it difficult to make a decision on product purchasing. In a similar manner, it is also inconvenient for a company to know the customer's opinion about the products they produce which could help them develop market strategies and placement of products in the market. Knowing customer’s opinions could also help the companies realize the likes and dislikes of people which could help them better their products.

Opinion mining classifies the reviews on three separate levels: Document-level, sentence-level and feature level [1]. At the document level, the entire document is classified as either positive or negative. This level of classifications is referred to as text classification problems. In objective/subjective analysis, the review document is classified into classes that are previously defined. Further, classifying the subjective reviews into a positive and negative class is referred to as sentiment analysis. In sentence level, each of the sentences is classified as either positive or negative. Aspect level opinion mining refers to a fine-grain analysis of the product. This could help both future customers as well as the manufacturing company to know which features of the product are mostly disliked or liked by the reviewers.

Sentences are of two types, subjective and objective. The subjective sentences are of relevance and present the view, belief, and feelings of the user. Objective sentences provide factual information and are generally irrelevant to the study. Based on this idea, we filter out the objective sentences and only keep the subjective sentences. Machine learning algorithms such as SVM, Random Forest Classifier and Naïve Base algorithms are used for the purpose of classification of the sentences into objective and subjective. Further, linguistic patterns are employed to detect the aspect words and opinion words. Certain rules have been utilized to extract a combination of feature and aspect words to determine the polarity of the opinions towards the aspects. SentiWordNet is a lexicon that is used to determine the polarity of each opinion word. Finally, a summary of the percentage positive and percentage negative opinion towards each of the identified aspect is generated.

2. Literature Review

The information available around us can be mainly classified into two categories, opinions and facts.
Opinions are the statements we consider as subjective reviews that reflect the opinion of the people’s sentiment or their view with regard to the events or objects. Facts are the statements about entities and events, we consider as objective reviews. [2]
Before the presence of the World Wide Web, most of the research was done on factual information from text. If an individual needed an opinion to make a decision on a particular product, he had to ask his friends/family. When a company needed to find the opinion of the public about its product or service, it had to conduct surveys and focus groups. With the growth of the internet, there has been a steady rise in user-generated content. Now, users can post reviews of product the merchant’s website or express opinion on any platform such as blogs and social media. [3]
Customer reviews are classified into three different levels: Document-level, Sentence level, and Feature level. In the

document level, the entire document is classified into positive or negative. Bo Pang and Lee [4] made use of standard machine learning techniques for movie review classification. Sentence level classification refers to the task where the sentences are classified into two categories [5] [6]. Feature level opinion mining is proposed in [7]. There are different methods for feature-based opinion mining, namely Statistical, rule-based, NLP, etc. Hu and Liu in [7] performed the task for mining and summarizing customer reviews in three steps: Extraction of project features from the reviews, polarizing the opinion sentences for each aspect denoting positive or negative opinion and summary generation by utilizing the information which is discovered. In [8] the lexicon-based method has been utilized for opinion analysis.

3. Methodology:

The architecture of the system:

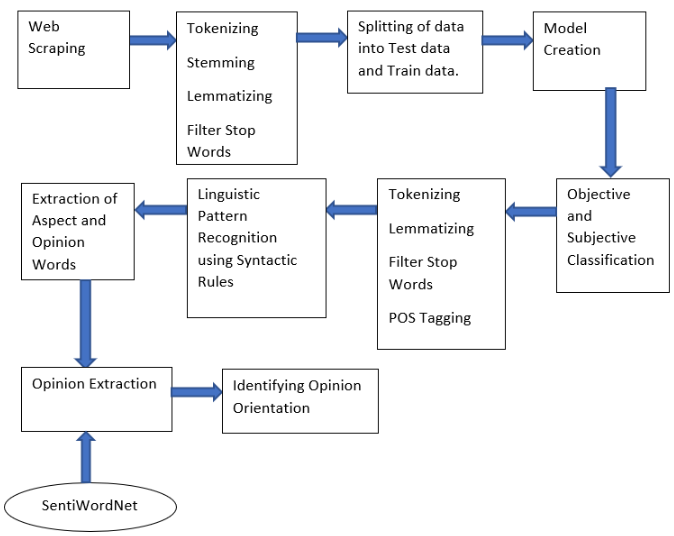

Figure 1. The architecture of the system.

3.1 Web Scraping of Data

The data is scrapped for the two devices namely Samsung Galaxy S10 and Apple iPhone XR from their respective review pages on Amazon using an extension for Google Chrome called "Web Scraper". The data is saved in the form of a JSON file which is further converted into a CSV file.

3.2 Exploratory Data Analysis

This part of the analysis deals with exploratory data analysis where the distribution of star ratings for both the devices is plotted, followed by the months in which the reviews were posted by the customers. Further, some text pre-processing is performed to plot trigrams of the data to look at interesting 3 words pairs.

3.3 Text Pre-processing

Some of the steps involved in text pre-processing include stemming, lemmatization, tokenizing, pos tagging of the data.

Stemming: Stemming is a process in which the algorithm works by cutting off the beginning or the end of a word taking into account common suffixes and prefixes that are usually found in an inflected word. This form of cutting can be successful in some scenarios but not all. Hence, we affirm that there are some limitations to this approach. [9]

For example,

Consider the word "studies". The suffix, in this case, is "es". Removing the letter "es" results in the word "studi" which does not make much sense. Similarly, when we consider "studying", the suffix here is "ing". Removing the suffix results in "study" which is a proper word.

Lemmatization: Lemmatization is a process in which the algorithm considers the morphological analysis of the words. In this case, it is necessary to have a dictionary of details with regard to all the words. The algorithm looks through the dictionary to find the base or the root word of the word considered.

For example,

Consider the words "studies" and "studying" again, here the lemma or the base word for both "studying" and "studies" is study.

Tokenization: Tokenization is a method in which the words in the document are represented by numbers. There are numerous methods to perform this task. The method considered here is the TF-IDF method. TF-IDF stands for Term Frequency – Inverse Document Frequency. [10]

Term Frequency refers to the count or the frequency of occurrence of the word in the document. This helps in filtering unwanted words by setting a cap of a certain number on the count. Document Frequency refers to the frequency of occurrence of the word in the entire text corpus taken into consideration. Hence taking the ratio of TF/DF or TF * IDF assigns a score to every word which ensures the selection of most relevant words.

3.4 Classification of sentences into Objective and Subjective sentences.

Subjective sentences: The sentence can be referred to as a subjective sentence if it describes an opinion or a feeling by a user towards an item/entity.

For example, consider the sentence “The phone has a great camera”. Here, the word great is being used to describe the camera on the phone.

Objective sentences: The sentences can be referred to as objective sentences if it provides factual information which can be easily verified.
For example, consider the sentence “The iPhone XR has a 12mp camera sensor”. Here, the fact that the camera on the iPhone XR is 12mp cannot be disputed. It is a fact and can be verified.

Since the data is scraped from amazon, it is unlabelled data. We use a library called TextBlob [11] to get a subjectivity score for each of the sentences. Further, we categorize sentences with scores greater than 0.5 as 1 and those with lesser than 0.5 as 0. Where “1” stands for Subjective and “0” stands for an objective. Using this labeled data, we create test and train data sets and use them to model our classifier.

Three different classification algorithms were used for the purpose of classification of the data, namely Random Forest Classifier, Support Vector Machine, and Naïve Base model. The scikit-learn package in the python programming language was utilized for this task. We chose Random Forest and SVM as the algorithms of chose over Naïve Base model due to their better performance.

3.5 Opinion Mining

3.5.1 POS Tagging

POS tagging which is short for Part Of Speech Tagging is a process in which every word in the text corpus is identified based on the parts of speech the word represents. [12] Some of the parts of speech used for tagging are as follows:

NN noun, singular ‘car’
NNS noun plural ‘cars’
NNP proper noun, singular ‘James’
NNPS proper noun, plural ‘Canadians’
JJ adjective ‘small’
JJR adjective, comparative ‘smaller’
JJS adjective, superlative ‘smallest’
RB adverb ‘quietly’
RBR adverb, comparative ‘better’
RBS adverb, superlative ‘worst’

For example, consider the phrase – “great camera”. Here, “camera” is the noun tagged as ‘NN’ and “great” is the adjective tagged as ‘JJ’. Python package NLTK was utilized for this task.

3.5.2 Pattern Extraction:

Syntactic rules [13] are utilized for the appropriate combination of product features and opinion words describing the feature. There are generally four types of sentiment words and whose combinations can express opinions on product features. These are nouns (NN), adjectives (JJ), verbs (VB) and, adverbs (RB). Mostly opinion words are a combination of adjectives or adjectives along with adverbs. Sometimes, they are combinations of verbs and nouns.

3.5.3 Product aspect and Opinion words Extraction:

First, we extract the features of the products. These features/aspects may occur as explicit mention or implicit mention. For example, consider the sentence “The camera is amazing”. Here, the camera is explicitly mentioned in the sentence. Now, consider the sentence “it takes a second to unlock”. Here the reviewer is speaking about the fingerprint sensor of the camera but there is no explicit mention of the fingerprint reader. Hence it is an implicit mention. Implicit mentions are relatively rare. In our case, we only take into account the features which are explicitly mentioned. We make use of linguistic patterns to identify and extract the feature words which are generally nouns. We filter the list of known based on the number of occurrences. We filter out nouns that occur less than twice. Further, we filter only the nouns which are relevant to the product in case such as the camera, fingerprint reader, etc. by creating a list of useful features and performing string matching to remove irrelevant words.

Next, we extract the opinion words. We use the extracted features to determine the nearest opinion words to the feature which are the sentiment words such as adjectives or adjectives paired with adverbs etc. These are the words used by the reviewer to express positive or negative feelings.

For example,
The fingerprint reader is very good.
The battery is amazing.

In the first sentence, “fingerprint” and “reader” are the two nouns that feature. “very” is an adverb and “good” is an adjective, both being opinion words describing the features. Similarly, in the second sentence, “battery” is the noun and “amazing” is the opinion word.

3.5.4 Opinion Extraction (Summarizing)

Once all the opinion words are extracted, we extract the opinion with the help of SentiWorldNet. [14] SentiWordNet is a lexicon that contains the positive and negative score associated with each opinion word. For all the aspect words, we determine the count of positive and negative opinions and determine a percentage positive and percentage negative opinion. This is stored in a data frame and printed in the form of a summary.

4. Data and Experiment

The data being used is the amazon product review data consisting of attributes:

Date: The date on which the review was posted.
Rating: The rating provided by the customer on a scale of 1 to 5.
Content: The text content of the review.

4.1 Distribution of the rating given for the two devices:

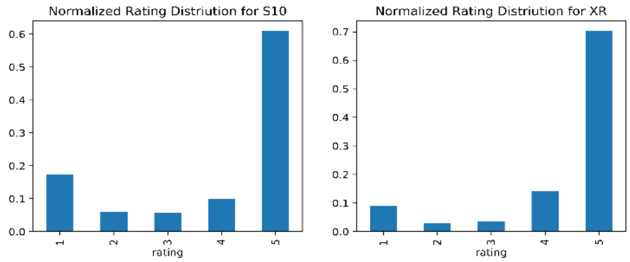

Figure 2. Normalized rating distribution for S10 and XR.

Since the count of the number of reviews is different for the two devices, the count has been normalized to find values lying between 0 and 1. From figure 2, it is observed that the iPhone XR receives a slightly higher number of 5-star rating and a slightly lower number of 1-star rating compared to the Samsung Galaxy S10.

4.2 Distribution of the reviews across various months of the year:

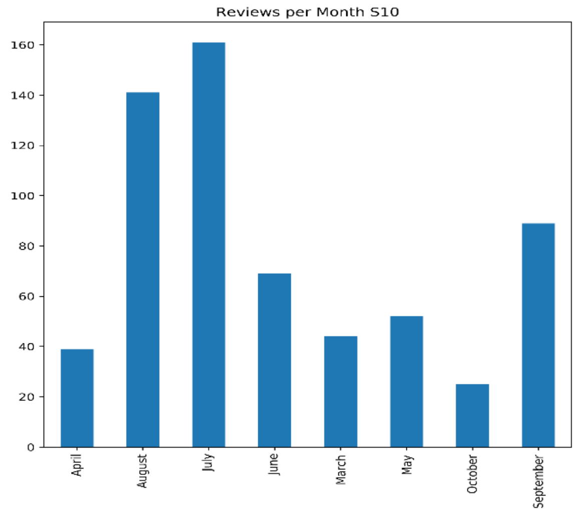

Figure 3. Month-wise distribution of reviews for S10.

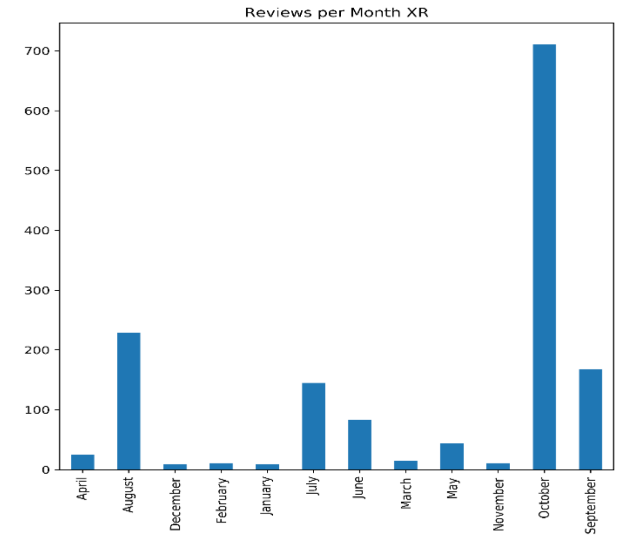

Figure 4. Month-wise distribution of reviews for XR.

It is observed that the Samsung Galaxy S10 received the highest number of reviews in the month of July compared to that being the month of October for the iPhone XR.

4.3 Most frequent trigrams for both the devices:

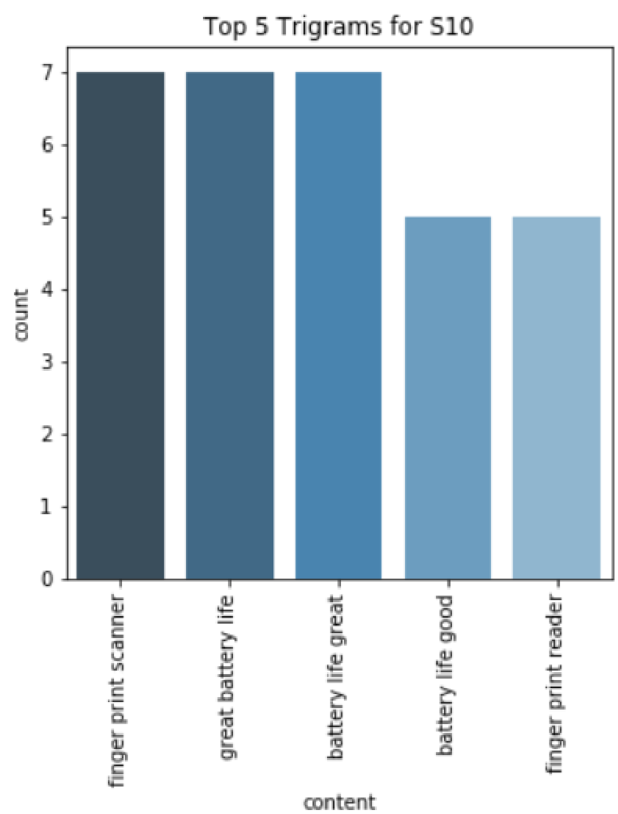

Figure 5. Top 5 Trigrams for S10.

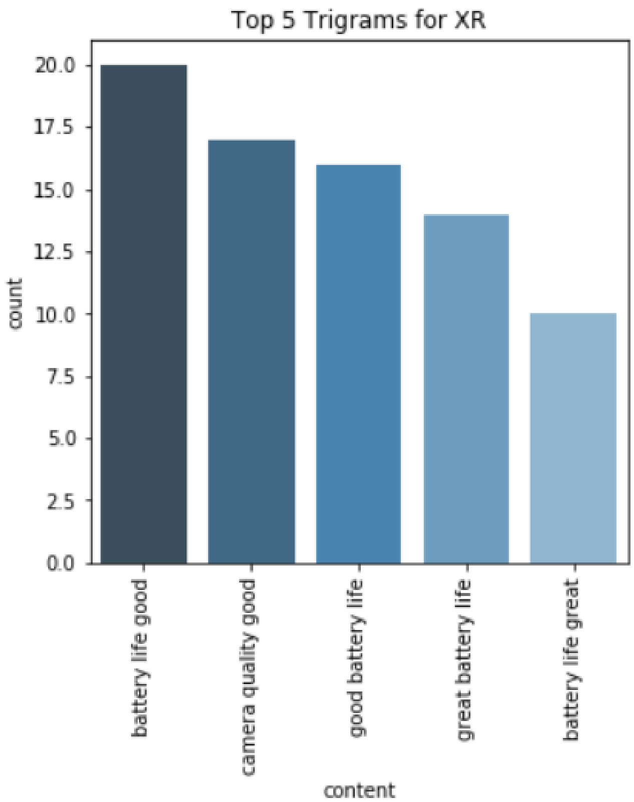

Figure 6. Top 5 Trigrams for XR.

It can be observed that the most frequent trigram for the iPhone XR is "battery life good". Similarly, for the Galaxy S10, it is, "finger print scanner", "great battery life" and "battery life great".

4.4 Evaluation measures:

We consider Accuracy, Precision and Recall as our evaluation measures for classification of sentences into Objective and Subjective sentences.

First, let us understand certain terms used: [15]

True Positive (TP): The cases whose actual class was true and was predicted as true.

False Positive (FP): The cases whose actual class was false but predicted as true.

False Negative (FN): The cases whose actual class was true but predicted as false.

True Negative (TN): The cases whose actual class was false and predicted as false.

Accuracy: The ratio of the number of true positives and true negatives to the total number of sentences.
Accuracy = (TP+TN)/Total
Precision: The ratio of the number of true positives to the number of predicted positives.
Precision = TP/(TP+FP)
Recall: The ratio of the number of true positives to the number of actual positives.
Recall = TP/(TP + FN)

4.5 Discussion:

We made use of the discussed models for the classification of the reviews into objective and subjective. The results are shown in Table 1.

	Samsung Galaxy S10			iPhone XR		
Model	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Support Vector Machine (SVM)	0.84	0.87	0.80	0.82	0.84	0.71
Random Forest Classifier	0.82	0.81	0.85	0.83	0.82	0.75
Naïve Base Method	0.76	0.72	0.83	0.81	0.82	0.71

Table 1. Tabulation of performance metrics for each model.

It can be seen that Random Forest and Support Vector Machine with accuracies of 0.83 and 0.84 respectively for the galaxy S10 and accuracies of 0.83 and 0.82 for the iPhone XR are better models suitable for text classification. These two models also have better precision and recall scores when compared to the Naïve Base model.

	Percentage Positive	Percentage Negative
PHONE	31.88	10.48
CAMERA	33.76	11.39
BATTERY LIFE	37.58	5.37
SCREEN	27.68	11.86
BATTERY	37.43	6.70
FINGERPRINT SCANNER	41.18	20.59
FINGERPRINT	26.42	17.61
FACE RECOGNITION	23.53	11.76
FINGERPRINT SENSOR	16.95	15.25
PROCESSOR	16.67	12.50
STORAGE	22.22	16.67
FINGERPRINT READER	34.21	15.79
DESIGN	38.46	15.38
FINGER PRINT READER	31.82	15.91
DISPLAY	33.33	15.38
SECURE	28.57	14.29
SCREEN PROTECTOR	40.00	5.00
PHONE CAMERA	31.82	11.82

Table 2. Aspect based opinions for Samsung Galaxy S10.

Table 2 above signifies a summary of the aspect-based opinions on various features. It can be observed that the percentage positive opinion is higher than the percentage negative opinion for the Samsung Galaxy S10 for almost all the features.

	Percentage Positive	Percentage Negative
PHONE	31.33	11.66
BATTERY LIFE	37.00	7.95
CAMERA	37.40	9.54
PERFORMANCE	36.26	11.70
BATTERY	34.58	8.47
BATTERY BACKUP	40.00	10.00
DISPLAY	33.08	10.77
CAMERA QUALITY	46.27	6.72
FACE RECOGNITION	36.11	11.11
PROCESSOR	30.77	2.56
INTERFACE	22.00	4.00
FACE DETECTION	30.43	8.70
FACE UNLOCK	31.67	3.33
PRICE	15.09	24.53
SCREEN	36.92	12.31
FRONT CAMERA	25.00	8.33
ADAPTER	25.00	25.00
BATTERY PERFORMANCE	30.00	12.50
CHIP	13.64	22.73
SPEAKER	29.63	7.41
CAMERA PICTURE QUALITY	58.33	8.33
ANDROID	13.39	20.54

Table 3. Aspect based opinions for iPhone XR.

From the above table, it can be noted that the positive percentage of opinion is high for features such as the battery, camera, etc. The percentage of negative opinions is seen to be higher for one aspect, the price of the phone. As we know, iPhones are generally more expensive.

5. Conclusion and future work

In this report, we presented a comparison of different machine learning models used to classify the sentences based on subjectivity. Using the rule-based method of opinion mining, opinion words and aspect words were successfully extracted and with the help of SentiWordNet, the polarity of the opinion words was identified. Percentage positive and percentage negative opinion for each of the aspects were successfully identified. A summary table of these opinions was produced. This summary table could provide potential buyers and the product manufacturers insights on the public opinion of their products, saving time and resources. In our work, we used the rule-based approach focusing on the explicit mention of features. Rule-Based Approach towards Opinion Extraction help in fundamentally understanding what exactly is happening inside the system. They are a little hard to maintain in case a new rule is written to support new vocabulary. Future work could include using more recent methods such as unsupervised machine learning to find opinions as well as work on sentences with implicit mention of features.

References

[1] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.

[2] J. Khan and B. S. Jeong, "Summarizing customer review based on product feature and opinion," *International Conference on Machine Learning and Cybernetics (ICMLC)*, 2016.

[3] K. Dave, S. Lawrence and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," 2003.

[4] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 2002.

[5] V. Hatzivassiloglou and J. M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," in *COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics*, 2000.

[6] J. M. Wiebe, R. F. Bruce and T. P. O'Hara, "Development and Use of a Gold-Standard Data Set for Subjectivity Classifications," in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland, USA, 1999.

[7] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *KDD '04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Seattle, WA, USA, 2004.

[8] X. Ding, B. Liu and a. P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *WSDM '08 Proceedings of the 2008 International Conference on Web Search and Data Mining*, Palo Alto, California, USA, 2008.

[9] C. García, J. P. Cabanilles and B. Ramirez., "What is the difference between stemming and lemmatization?," February 2018. [Online]. Available: https://blog.bitext.com/what-is-the-difference-between-stemming-and-lemmatization/. [Accessed December 2019].

[10] A. Panchal, "Text Summarization using NLTK: TF-IDF Algorithm," June 2019. [Online]. Available: https://towardsdatascience.com/text-summarization-using-tf-idf-e64a0644ace3. [Accessed December 2019].

[11] S. Li, "Introducing TextBlob," January 2019. [Online]. Available: https://towardsdatascience.com/having-fun-with-textblob-7e9eed783d3f. [Accessed December 2019].

[12] SubhadeepRoy, "Part of Speech Tagging with Stop words using NLTK in python," [Online]. Available: https://www.geeksforgeeks.org/part-speech-tagging-stop-words-using-nltk-python/ . [Accessed December 2019].

[13] S. Kat, "Implementing Opinion Mining With Python," June 2016. [Online]. Available: https://dzone.com/articles/opinion-mining-python-implementation. [Accessed December 2019].

[14] A. Esuli and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, 2006.

[15] M. Sunasra, "Performance Metrics for Classification problems in Machine Learning," November 2017. [Online]. Available: https://medium.com/thalus-ai/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b. [Accessed December 2019].