

Sentiment Analysis Framework using Deep Active Learning for Smartphone Aspect Based Rating Prediction

Rathan Muralidhar, Vishwanath R Hulipalled *

Abstract. Social media are a rich source of user generated content where people express their views towards the products and services they encounter. However, sentiment analysis using machine learning models are not easy to implement in a time and cost effective manner due to the requirement of expert human annotators to label the training data. The proposed approach uses a novel method to remove the neutral statements using a combination of lexicon based approach and human effort. This is followed by using a deep active learning model to perform sentiment analysis to reduce annotation efforts. It is compared with the baseline approach representing the neutral tweets also as a part of the data. Considering brands require aspect based ratings towards their products or services, the proposed approach also categorizes predicting ratings of each aspect of mobile device.

Keywords: Sentiment Analysis, Opinion Mining, Twitter, Active Learning, Deep Learning

1. Introduction

Unstructured data in social media is growing at a rapid pace. This is mainly due to the ease at which everyone can create a post by clicking a picture, recording a video, or posting text in social media sites like Facebook, Twitter, and others [22]. Twitter alone, has registered 353 million monthly active users on an average for Quarter 3 of 2020 [50].

Businesses often listen to social media to understand thoughts of the masses about their products or services offered. By listening to the positive as well as negative comments posted online by the masses, business aim to improve and provide a better service or product, in turn attracting more customers and growing the company [18].

*Faculty of School of Computing & Information Technology, REVA University, Bangalore, Karnataka, India, {Rathan Muralidhar, rathan.muralidhar}@gmail.com

Companies actively track their brand name or their product name over social media sites to observe, respond, and improve customer interaction positively to provide what the customers feel are the most required to stay ahead in the game [15].

With this, it's pretty easy to imagine that people post everything they may like or dislike about, on an everyday basis. Whether it's the time of elections and a person wants to express their view or whether it's a new pair of shoes that's just been bought and a person want to express how comfortable they are, people express often over social media [49]. With a lot of personal views about a product posted online, consumer reviews often play an effect on potential buyer behaviour when it comes to online purchasing [52]. There has been a lot of research performed in various fields such as identifying rumors in Twitter [44], identifying bots in Twitter [16] and bias detection in politics [6] given it's public access to data. Twitter is one such social media site that is of major focus by companies to perform research on their target audience and understand the opinion of the masses. A lot of organizations invest within brand monitoring and listen to the conversations on Twitter whether in Direct Messages or as a tweet posted by tagging the organization [27]. As Bill Gates has quoted, "Your most unhappy customers are your greatest source of learning". Taking this quote into practice, the major companies across the world have millions of followers and keep track of the negative comments on their product, and also strive to respond to these comments [26]. For instance, if a customer tweets that a product is not of the standard expected or has a defect, the company responds by asking the customer either to send a direct message or provide a solution as a response.

Observing at such instances, the paper looks at obtaining the tweets by the masses on the topic of mobile devices. The paper aims to perform sentiment analysis of these tweets based on aspects of the mobile device such as audio, battery, display, network, and more and show the results in form of a ratings scale between 1 to 10 for each aspect with 1 indicating least liked and 10 indicating most liked. Sentiment analysis is performed by using deep learning model. To provide an approach for annotation of training data selectively which is time and cost efficient, active learning method is utilized. This helps strategically select important sentences to perform annotation in an efficient manner.

The research paper also explores the idea of filtering and utilizing only positive and negative statements to build the machine learning model instead of utilizing neutral statements as well. Neutral statements are removed using a combination of lexicon based approach and human effort. We compare the results of the sentiment analysis models which incorporate as well as remove neutral statements. The research paper further explores the usage of pre-trained models in cases of limited training data which is a concern in many real time systems. The idea of limited training data coupled with active learning and a pipeline of techniques in pre-processing helps improve the performance in terms of accuracy.

The structure of the rest of the paper is as follows. Section 2 looks into related work in terms of research work that has been performed to create sentiment analysis framework using various methods leading to the proposed approach. A detailed specification of the proposed approach is provided in Section 3. Section 4 looks into the results obtained by the proposed approach with comparison to alternate approaches

utilized currently. Section 5 provides a brief conclusion about the proposed approach, its advantages and limitations which could lead to development of more efficient sentiment analysis frameworks.

2. Related Work

Sentiment analysis can be performed by utilizing a rule based approach i.e. using a lexicon to perform word by word comparison, machine learning approaches by building models using algorithms or a hybrid approach which involves a combination of lexicon and machine learning approaches.

2.1. Rule Based Approach

A rule based approach consists various techniques such as stemming [25], tokenization [55], lemmatization [37], parts of speech tagging [53], and usage of generalized or domain specific lexicons. A lexicon is a list of words which consist of a sentiment along with the words in form of scores. For instance, a lexicon will consist of the word “good” with a score of +0.5, the word “better” with a score of +0.75 and the word best with a score of +1.0. The lexicon will also consist of words with negative emotions and scores such as “bad” with -0.5, “worse” with -0.75 and “worst” with -1.0. When a sentence is provided to the lexicon such as “The book was really good”, the word “good” has a positive score. The rest of the words do not have any score and it is considered to be a positive sentence. Negation handling is often done by lexicon based approaches by reversing the polarity of the word nearest to the negation word such as “not”, “aint” and so on. A lot of such rule based methods utilize the concept of stemming or lemmatizing to reduce certain words to their root words. Balakrishnal et al. [7] have performed a comprehensive comparison between stemming and lemmatization and have concluded that lemmatization provides the better result of the two. Stemming tries to reduce the word to the root word by trimming off the end of each word. For instance, “Universe” and “Universities” become “univers” by trimming off the ‘e’ and “ities” at the end. This may be ambiguous if we have the two words in our corpus. To avoid this, lemmatization can be used which utilizes parts of speech to ensure that each word is reduced to its root word accurately. For instance, the word “better” and “best” can be reduced to the word “good” and there would be a requirement of only one word in the lexicon instead of three words i.e. “good”. Lexicon based sentiment analysis is considered as a naïve method and it is not feasible to utilize this in various domains as each domain would require various changes in the rules which maybe too complex to implement. For instance, when considering the field of stock market, the rules for many words would need to be updated. “Bull” and “bear” are just a few instances which have domain specific meanings as “Positive” and “Negative” but a generalized lexicon would not score these words as either positive or negative [34, 30].

2.2. Machine Learning Approach

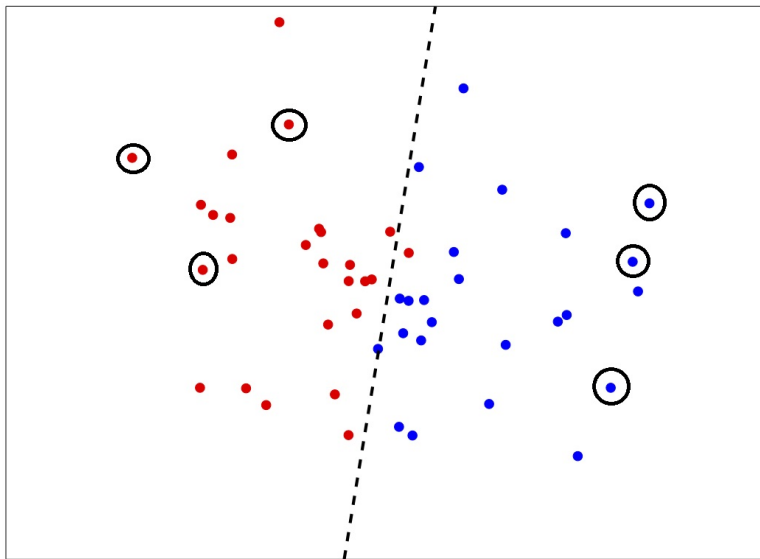


Figure 1. Two Class Linear Classifier with least informative data samples selected

Analysing thousands of tweets manually to identify certain scores in order to create the best possible lexicon and classify whether the tweets are positive, negative or neutral in sentiment using the lexicon is a time consuming and a cumbersome task. Besides, the lexicon based method is also prone to provide false neutral sentiments with complex sentences. To avoid this, computing technologies and advancements have created various methods. Machine learning is a form of designing computer algorithms that learns by experience over time. Machine learning algorithms are divided into supervised and unsupervised algorithms. For instance, a supervised learning model requires previously labelled data and is often termed as classification [1]. In our case, it learns how to identify sentiment of a tweet by experience. Experience here, is defined by the amount of tweets provided to the machine learning model which is already labelled with each tweet or review as positive, negative or neutral. Once the machine learning model has enough reviews labelled and learned from, it can classify a new tweet into one of the three classes i.e. positive, negative or neutral. In case of unsupervised learning models often termed as clustering, it is utilized for the purpose of grouping tweets into different categories. For instance, the tweets particularly expressing something about shoes and about mobile phones can be divided into two categories as such without the requirement of labelled data particularly. This works by identifying

words which are similar in tweets and grouping such tweets together.

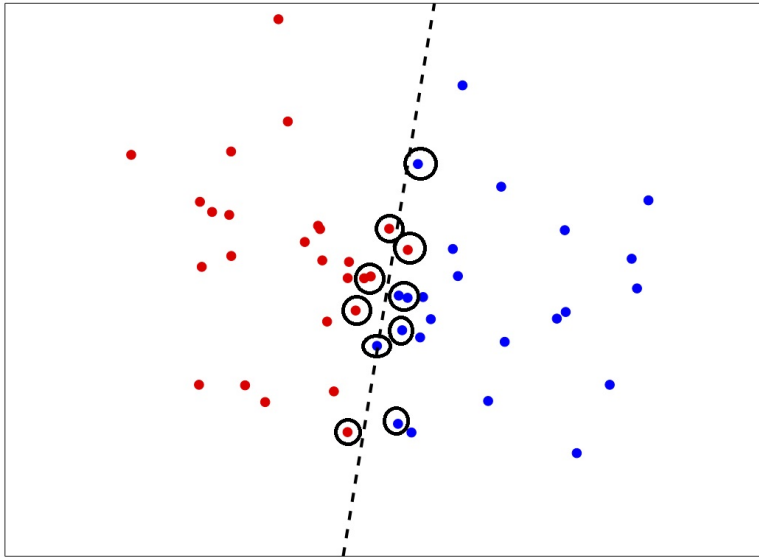


Figure 2. Two Class Linear Classifier with most informative data samples selected

Schumaker et al.[40] have performed Twitter based sentiment analysis to predict winner of premier league matches. The authors have also utilized Twitter to perform NFL match outcome predictions and wager decisions to obtain a higher payout [41].

Agarwal et al. [2] have created a machine learning based approach to perform sentiment analysis on twitter data utilizing the short sentence format used within Twitter. For instance, the usage of the word “gr8” instead of “great” and emoticon usages have been analyzed to feature into identification of the sentiments. Over the years, there have been a huge amount of research performed on sentiment analysis over Twitter data to develop insights. Pimprikar et al. [36] have created a Twitter based stock market price predictor and concluded that they obtained the highest accuracy using LSTM based method compared to Support vector machines and linear regression. Bose et al. [11] have utilized a deep learning model to perform analysis of tweets to predict the winner of Gujarat legislative assembly elections of 2017. Ansari et al. [4] utilized LSTM model to perform sentiment analysis over Twitter data to predict the results of general elections of India in the year 2019. The authors have compared the LSTM results with other models and have concluded that although, LSTM model takes the highest amount of time for training, it’s results are the most accurate. Fan et al. [19] have performed analysis of Twitter data during world cup matches to identify sentiments and emotions of both sets of fans during a match.

They conclude that Twitter data closely reflects the real world emotions during the match. There have been various applications of Twitter based sentiment analysis ranging from election winner prediction to stock market prediction.

In most cases, labelling data for classification in machine learning is a time consuming and expensive process. For instance, if a machine learning model is to be developed to classify texts into positive, negative, neutral about a particular product or domain, an oracle or a person who is proficient in the language as well as expert of the domain is required to classify the texts and label them manually especially when they are noisy [57, 58]. Machine learning algorithms often require a sufficient amount of data classified in order to classify a new tweet. There have also been various ensemble models developed to perform sentiment analysis. Lango et. al. [28] have looked into imbalanced data which often is present in real world where the expectation of near equal amount of data for each class is rare. The authors identify that imbalanced data is a common feature of streaming data and utilize various pre-processing steps and ensemble models to achieve higher accuracy. Random forest models with k-fold cross validation as bagging models are utilized by the authors here. Gradient boosting has also been utilized in performing sentiment analysis in various cases. Hew et. al. [24] have performed learner satisfaction in MOOC's with a six aspect based sentiment analysis. If the number of tweets labelled initially are less, the machine learning model may not have enough learning knowledge to classify new tweets which leads to an under-fit model. Hence, it's necessary to have an oracle and spend time in order to label tweets so that the machine learning performs with a good level of accuracy. There have been advancements in machine learning to reduce the amount of time required for an oracle to perform labelling by using active learning methodologies instead of passive learning.

2.3. Passive Machine Learning vs Active Machine Learning

Passive Machine Learning is where an oracle takes a chunk of input data at random and starts labelling it. Once the data is labelled by the oracle, majority of it is provided to the machine learning model for training. The remaining minority of the data is provided to machine learning model for validating the results of the machine learning model against the results labelled by the oracle. To understand why passive machine learning annotation is costly as well as time consuming, we need to look at the Figure 1. Let us assume the oracle labelled a certain random input samples which are marked in circles.

The figure 2 indicates a scenario of active learning where the reviews classified are the closest to the decision boundary. The positive reviews are indicated with the red dots and the negative reviews are indicated with the blue dots. The decision boundary separating the two classes i.e. positive and negative is indicated with the dotted line. It can be inferred that the most informative reviews required by the classifier to be labelled by the annotator to make better decisions are ones that are closest to the decision boundary separating the two classes. If passive machine learning method is chosen where all the data available is labelled by the oracle, we can observe the

reviews which are far away from boundary line as marked in circles within Figure 1 also classified by the oracle. These reviews are not as useful as the reviews near the decision boundary which possess higher informative value for the machine learning model to classify the reviews. Since passive learning has the oracle select reviews randomly and label them, the reviews which possess the most as well as the least informative value are labelled [35]. This is where semi-supervised learning in the form of active learning comes into the picture to reduce the time and effort spent by the oracle. Active learning is a specialized form of machine learning where the machine learning model provides similar accuracy by choosing significantly lesser reviews to be labelled by the reviewer instead of the whole dataset [48].

Hiring an oracle can be expensive sometimes as well. Consider the scenario of labelling malign and benign tumour images. It would be quite expensive to have a doctor be paid to go through tens of thousands of images. This is where active machine learning excels in reducing the annotation time and making sure that there isn't a significant compromise in accuracy. There have been various studies performed to incorporate active learning and reduce the requirement of annotation effort [5]. Carse et al. [13] have utilized active learning to perform digital pathology using CNN. Their aim is to reduce the annotation costs required for building the model.

When it comes to data annotation, there are various challenges faced in the world of machine learning. Data annotation requires an expert in the field, often termed as the oracle. Often, it's an expensive process to hire an oracle to generate ground truth for a large amount of data [21]. Furthermore, if it is a subjective annotation such as review or recommendations, the view of two experts may disagree with each other [38]. As the training data scales, so does the requirement of annotators which is time and cost sensitive. This in turn, may lead to annotator fatigue, leading to incorrect annotations as well. Considering these factors, it's imperative in many cases to ensure that annotators are not fatigued, given the time and cost constraints of various projects, annotators might be available for a limited time. Active learning methodologies take leverage this limited availability of oracle to annotate most useful data, ensuring these setbacks do not occur.

This type of an interactive machine learning system queries the oracle with the input data to be labelled with the highest informative value. The rest are annotated by the model itself without the need of the oracle. However, it has also been touched upon that due to the requirement for only the immediate and local improvements in each iteration of active learning based model building, the model may have lower accuracy and fail to prove it's purpose as well [43]. The model may in such cases suffer from over-fitting.

2.4. Learning Scenarios for Active Learner

There are three major methods to provide highly informative input data to the oracle in order to be labeled and provided back to the interactive active learning model:

- Membership Query Synthesis

- Stream Based Selective Sampling
- Pool Based Sampling

2.4.1. Membership Query Synthesis

This method synthesizes i.e. creates new reviews and queries them to the oracle in order to obtain labels for the same. To create new unlabelled data within the input space, the automated sentence generator must generate sentences in a meaningful way by considering parts of speech tagging, synonyms, antonyms, and grammar of the language. Schumann et al. [42] have created an active learning model based on member query synthesis unlabelled data to be provided to the oracle. Variational auto-encoders are used to create the unlabelled data. They show that the results in terms of accuracy is comparable to the pool based method and the time spent in query generation with informative sentences is considerably lesser. However, in cases where the synthesizing query generation strategy is of lower quality, the inputs to model would also be compromised.

2.4.2. Stream Based Selective Sampling

Stream based Selective Sampling method however takes a different approach. A non-interrupting stream of non-annotated data is provided to the active learning model which needs to figure out which of these non-annotated input data sentences are to be provided to the oracle for manual annotation. Stream based selective sampling require a stable data stream with distribution of data to be uniform. Improvements upon stream based selective sampling is performed by Caciarelli et. al. [12].

2.4.3. Pool Based Selective Sampling

Pool based sampling method is different in contrast compared to stream based selective sampling where non-annotated data is already available for the active learning model [20]. The model just needs to select the most useful data required to order to improve the classifier's accuracy and provide the same to the oracle for annotation. However, a pool of available data is not guaranteed in cases of streaming data based platforms such as twitter.

2.5. Sampling Strategies for Active Learner

The method of selecting the right input data to provide to the oracle can be done using various techniques as stated below.

Methods of sampling data:

- Committee Based Method / Region of Disagreement Method

- Large Margin Based Method
- Posterior Probability Based Method

2.5.1. Committee Based Method / Region of Disagreement Method

This method requires multiple machine learning models to be created. Once the initial sample of random data within the input space is annotated, the model is trained on the same and tested against another sample chunk of the input data [59]. The first sample of data is analysed by each of these ‘n’ machine learning models which provide their own labelling for each of the testing input data. The input data that is considered to be most informative for labelling by the oracle is the input data which gets the most varied results across the n machine learning models. Several machine learning models are trained on a single set of data annotated. They are then provided with the same set of unlabelled data to see the disagreement between the models. For instance, there maybe multiple models using SVM, Naïve Bayes, Logistic Regression and Decision Trees. When these models provide different results in terms of classification for a particular input text, the input is provided to the annotator for classification manually. The same approach can also be performed using a single machine learning model with different features and parameters as well. The same method has been used by Mackowiak et al. [33] who were able to reduce the annotation required to 17% to create annotated data. The committee based method has also served as an inspiration for Wang et al. [54] to perform particle swarm optimization.

2.5.2. Large Margin Based Method

A model based on Support Vectors can be used to perform large margin based method. If the input data are far away from the hyperplane, they are considered to be confidently analyzed by the model. Any input data sample close to the hyper plane or within the support vectors are considered to be most informative and are provided to the annotator for manual labelling [8, 39]. This is applicable for a binary classifier. However, for a multi-class classifier, the distance from an input sample to each of the class hyperplane is calculated. The largest and second largest distance to hyperplanes are selected and are subtracted. If the resulting value is deemed to be a very low value, the input data is considered to be useful for annotation and provides the highest informative value to the model to learn from.

Posterior Probability Based Method Posterior probability distribution is used to figure out the probability of an input data belonging to a particular class using values such as confidence. This helps decide whether an input data is to be provided to the annotator for labelling or not. Posterior probability distribution method can be implemented using least confidence method for a binary classifier, smallest margin

method for multi-class classifier. However it would be computationally intensive since posterior probability needs to be calculated for each iteration.

Least Confidence Method In the least confidence method, let us consider a binary classifier. The dataset tested belongs to one of the two classes by some confidence level between 0.0 to 1.0. If a data sample is classified with a confidence of 0.9, it means that the model is 90% confident that the data sample belongs to a certain class. If the data sample however is classified with a confidence level of 0.5 in a binary classifier, it means that the data sample could equally belong to either of the classes. Such data samples tested with lower confidence levels are provided back to the annotator for labelling. The remaining data samples tested with relatively higher confidence levels such as greater than 0.8, are considered to be least informative to the machine learning model in terms of usefulness to improve the accuracy of the model [29]. Song et al. [46] have successfully created a crowd sourced confidence based active learning method to perform annotation with better accuracy than the baseline method compared with.

Smallest Margin Method Least confidence level doesn't do justice when it comes down to a multi-class classifier. To identify the most informative data samples to be annotated using smallest margin method, the confidence levels of two most probable predictions are subtracted. If this margin i.e. the difference between the confidence levels of the two most probable predictions is low, it is considered to be an informative data sample which needs to be annotated manually. There are other methods utilizing entropy and variance reduction that are utilized in posterior probability based method as well. Begum et al. [9] have performed breast cancer detection using smallest margin based method to reduce annotation costs and have obtained F-Measure using Least Confidence, smallest Margin and Entropy Based Active Learning of 95%.

Wu et. al[56] look into the issue of over reliance in the data labelled by the machine learning model during the iterations in active learning and provide possible solutions to solve the issue. They indicate that the oracle's knowledge in the language as well as understanding of the domain in order for the features to be selected for sentiment analysis are two important factors defining the active learning model's accuracy. They have also considered SentiWordNet for polarity analysis. With SentiWordNet being a generic lexicon, it doesn't offer the best result in a domain specific problem. For instance, Oscar is considered as a positive sentiment in a sentence "This movie is Oscar worthy." But at the same time, SentiWordNet doesn't state that it is a positive word as it is a generalized lexicon rather than a movie domain specific lexicon.

2.6. Aspect Based Sentiment Analysis

There is a huge increase of purchases in e-commerce sites as of late and a lot of people write reviews within e-commerce sites such as Amazon, Flipkart as well as social media sites such as Twitter and FaceBook. These reviews are hugely important to

the company to understand the sentiment of the masses. The reviews however are complex and generally consist of opinions on various aspects of the product [17]. Aspect based sentiment analysis has been performed in various domains such as hotel review analysis [3], e-commerce data [51] and mobile device reviews [23]. For instance, the reviews of mobile phone consist of opinions on various aspects such as battery, camera and so on. Analysing the reviews based on each aspect helps the company understand the most liked and disliked feature of the mobile device easily. To analyze the reviews effectively, each aspect based opinion must be extracted. For instance, consider the following sentence “The battery is terrible but the camera is great. So it averages out.” This review consists of two aspects. One of battery and the other of camera. The review itself is neutral in a compound sense as it has two words “terrible” and “great”. However, when the sentence is divided appropriately into phrases, we can obtain the following three phrases, “The battery is terrible”, “the camera is great” and “So it averages out”. Only the first two phrases consist of the terms aspect terms i.e. “battery” and “camera”. Their opinions based on the words around them are “negative” and “positive” respectively. Considering the importance of such social media data as well as e-commerce data for companies to identify aspect based opinions of the masses towards their products, it is vital perform effective aspect based sentiment analysis.

Gupta et al. [23] have utilized lexicon to perform aspect based sentiment analysis for the domain of mobile devices. There are however certain issues within a lexicon based method and it doesn’t work well when the input has statements such as “I wish the battery was better.” This is due to the fact that the word “better” is considered to be positive and there is no negation performed due to the presence of the word “wish”. Classification could prove effective when there is a machine learning model which understands the sequence of words around it. The model created searches for the aspect keyword and searches five words either side of the aspect keyword to find an opinionated word as adjective or adverb or verb. For straightforward statements, the model however does well as long as the lexicon consists of the words with their opinions assigned for each aspect. The lexicon used in the model is a generalized lexicon obtained from MQPA and SentiWordNet. Within the real world scenario, a single word often takes many forms providing polarized meanings. For instance, the sentiment of the word “long” within “The battery lasts too long” is very different in polarity compared to “The battery takes too long to charge”. Similarly, the review “battery charges quickly” and the review “The battery needs to be charged often” are completely opposite in sentiment and it is clear that a generalized lexicon would not be able to analyse all these kinds of reviews by just words associated with scores. Hence, it shows that there’s a need for sentiment analysis using different methodologies to analyse such texts more accurately. A sentence such as “The battery seems to be barely average at best days as well” is a negative sentence but it consists of the word “best” which is positive and no negative words along with it. Such a sentence is considered to be positive which influences the results in the wrong way. Considering such sentences, it’s important to understand the order in which words are written in a sentence to analyse it. RNN and LSTM have a significant advantage in these scenarios.

2.7. Deep Learning in Sentiment Analysis over Social Media Data

The state of the art methods in sentiment analysis recently are based on machine learning approaches rather than lexicon based approaches. Methods using algorithms such as support vector machines which create a clear boundary between two classes i.e. positive and negative are widely used. However, there have been various advancements in deep learning techniques involving neural networks to improve the accuracy of sentiment analysis models. Particularly, convolution neural networks and long short term memory models for sentiment analysis have been used hugely as of late. Convolution neural networks and recurrent neural networks fall short compared to long short term networks due to their gradient vanishing [10]. When the gradient carries information over a longer series of data, it loses the initial series of data remembering only the recent series. Alternatively, these networks can also face exploding gradient problem as well. This often happens when error gradients amass and create a large change in the weights during training of a model. With the limitation of learning long term dependencies, long short term memory networks (LSTMs) are utilized which overcome the problems faced by RNNs. Generally RNNs have a single tanh activation layer whereas LSTMs consist of four layers interacting with each other. The “forget gate” layer which decides the sequences of data to forget particularly, ensures to keep all the sequences of data that matter over a longer term of sequences [45]. The presence of the forget gate layer ensures that the network controls gradient values in a better way [32]. Convolution neural networks have also been used to perform aspect based sentiment analysis in various cases. Liang et. al. [31] utilized Sentic-Net based graph convolution network to capture sentiment dependencies. In GCN’s the graph nodes are represented by the words and edges represent similarity of these words. GCN’s often help capture complex relationship between words which is beneficial in the context of sentiment analysis. Combination of artificial neural networks is also utilized to perform aspect based sentiment analysis such as GCN and Bi-LSTM by Sun et. al. [47] to identify contextual as well as dependency information. Similarly, Cem et. al. [14] have used a combination of recursive and recurrent neural networks for dependency parsing as well as sentiment analysis. Our approach utilizes LSTM instead of RNN to perform aspect based sentiment analysis.

2.8. Key Contributions

The main contributions of the paper are as follows:

- Creation of a sentiment analysis framework based on Twitter data to help consumers decide the right smartphone to purchase by viewing ratings of each aspect of the smartphone.
- Creation of a deep active learning based sentiment analysis framework for Twitter data whose main aim is to reduce the time and cost towards human annotation.
- Demonstrate the significance of removal of neutral statements to improve the

accuracy of the sentiment analysis model as a binary classifier considering only positive and negative statements.

- Present an aspect based sentiment analysis framework with different model for each aspect which can be utilized for real time opinion mining about a product or service using social media data.
- Demonstrate the effectiveness of a lexicon based pre-processing approach to remove neutral statements with human intervention.
- Comparison with other approaches to show the efficiency of the proposed approach.

3. Proposed Approach

In this section, the proposed approach i.e. a sentiment analysis framework utilizing active learning to reduce human annotation is described in detail. This framework is aimed towards aspect-based opinion analysis of mobile phone reviews with preprocessing steps that include removal of neutral statements for each aspect. The model developed is a combination of rule based approach as well as machine learning approach. It considers certain rules as well as input data to be equally important to perform sentiment analysis. For training the model, dataset of mobile phone reviews are obtained from Amazon which were of various languages including English, French and Russian. Only English reviews were retained, and the other language reviews were discarded for the generation of training dataset. These reviews were identified to be mostly from the following countries: USA, Australia, India, England, and Canada.

Each review goes through certain steps of processing before terms are searched for within the reviews and placed into aspect based categories. The preprocessing steps help in correcting and reducing the sentences to phrases. This helps to place the reviews into aspect based categories. The steps are also specified within the Algorithm 1.

3.1. Pre-Processing

The preprocessing steps begin with converting each review into lower-case. Followed by this, url's are removed from each review as they are not deemed to be useful. Segmentation is performed using Ekphrasis tool which is built on social media corpus such as Facebook and twitter. Since Twitter users generally overlook spelling mistakes, spacings and jargons, it becomes important to create a meaningful sentence out of the input samples. A lengthy text of words without spaces is split into constituent words using this segmenter. A text such as “#batterysucks” is segmented into “battery sucks” where the words separated provide meaningful information for sentiment analysis. The corpora used for the Ekphrasis tool consists of 330 million tweets as well as thousands of articles from Wikipedia. The corpora chosen for our preprocessing is

Algorithm 1 Pre-processing of Tweets to Create Training Dataset for Active Learning based Sentiment Analysis Model

Require: The following are inputs for pre-processing stage.

- 1: Tweets = $\{t_1, t_2, t_3, \dots, t_n\}$, n tweets as input data
- 2: Aspects = $\{a_1, a_2, a_3, \dots, a_m\}$, m aspects of mobile devices
- 3: Keywords = $\{a_{i1}, a_{i2}, a_{i3}, \dots, a_{ij}\}$, an arbitrary number of j keywords for each aspect a_i of mobile devices

Ensure: The following are the outputs of the pre-processing stage.

Tweets pre-processed and divided into reviews. These reviews are further categorized into different aspect based reviews.

Reviews = $\{ra_{i1}, ra_{i2}, ra_{i3}, \dots, ik\}$ an arbitrary number of k reviews belong to each aspect a_i as a result of the algorithm.

- 1: **for** ($i \leftarrow 1$ to n) **do**
- 2: Discard tweet t_i if the tweet does not consist of English words and move to the next iteration i.e. t_{i+1}
- 3: Convert t_i to lowercase and remove URLs, usernames, hashtags. The remaining tweet t_i consists of f words.
- 4: **for** ($z \leftarrow 1$ to f) **do**
- 5: Perform segmentation for each word, w_z in tweet t_i to split non space separated words such as hashtags
- 6: Perform spell correction on word, w_z and replace it with the corrected spelling if applicable
- 7: Perform lemmatization on word, w_z and replace it with the lemmatized root word if applicable
- 8: **end for**
- 9: Replace each emoji within the tweet t_i with a lexicon equivalent positive or negative word
- 10: Replace each emoticon within the tweet t_i with a lexicon equivalent positive or negative word
- 11: Divide the tweet t_i into x reviews by separating parameter as period, exclamation mark, comma and conjunctions.
- 12: **end for**
- 13: **for** ($i \leftarrow 1$ to x) **do**
- 14: **for** (**each** a in aspect) **do**
- 15: **for** (**each** key in keywords of a) **do**
- 16: **if** key exists in review r_i **then**
- 17: Place the review r_i within the category of aspect a
- 18: Perform lexicon based sentiment analysis using custom lexicon to identify polarity of the review r_i as $polarity_{ri}$
- 19: **end if**
- 20: **end for**
- 21: **end for**
- 22: **end for**
- 23: **for** (**each** a in aspect) **do**
- 24: **for** (**each** r_i in reviews belonging to aspect a) **do**
- 25: **if** $polarity_{ri}$ is neutral **then**
- 26: Add the words of r_i and it's frequency to list of neutral words for aspect a called $neutralwords_a$
- 27: **else**

```

31:   for (each 100 most frequent words  $nw_i$  in  $neutralwords_a$ ) do
32:       Allot a score either as positive, negative or neutral for the word  $nw_i$ 
33:       Add the word  $nw_i$  containing either positive or negative score to the
           customized lexicon for aspect  $a$  and repeat the training algorithm to
           generate new annotated training dataset
34:   end for
35: end for

```

of twitter. The reviews are not searched for any stopwords to be removed. Instead, they're retained. Followed by this, spell correction is performed to make sure that the words in the tweets match the keywords for each aspect of mobile device in the next steps. Any word such as "batteri" is replaced with "battery" or "processar" is replaced with "processor". Certain keywords which are important to understand an aspect of the mobile device are not spell checked. Words such as '4g', 'lte', 'cdma' are not spell checked in order to retain the same words for aspect identification. This ensures that valuable user reviews to analyze sentiment of the masses is not discarded.

Each word of the review goes through the process of lemmatization. Albeit stemming is considered as a faster step to reduce a word to its root form, lemmatization is considered here. This is because stemming returns the root word which sometimes is meaningless. Considering we have created keywords for each aspect which are to be searched and carry meaning, lemmatization is preferred instead using wordnet with parts of speech tagging. When porter stemmer was used, the root word obtained for the word 'batteries' was 'batteri', similarly the root word obtained for the word 'charges' was 'charg'. To avoid this and provide the keywords for aspects as intended by keeping their meanings, lemmatization is used. With lemmatization, the root word obtained for 'batteries' was 'battery' and the root word for 'charges' was 'charge'. With the data obtained from amazon reviews and twitter, there are a lot of emojis and emoticons within the text. These need to be converted to appropriate words to be analyzed for sentiments. For instance, the following emoticon (>.<) is often used to indicate sadness or troubled reactions whereas certain emojis such as faces with heart eyes are used to indicate happiness or liking. These emojis when coupled with just the phone name in twitter don't mean much by itself unless we decode their meaning. All the emojis and emoticons are converted to either happy or sad words that are useful in annotation. Each twitter review is divided into phrases using conjunctions, comma, exclamation mark, and period.

These reviews are assigned to different categories based on the aspect of the mobile phone such as "battery", "display", "performance", "audio", "communication" and more. Each review is identified to belong to a particular aspect or discarded depending on the terms the review consists of. The terms based on which a review is placed into the category of an aspect is provided in the table 1.

The proposed approach consists of removal of neutral tweets as an essential part of the work before sentiment analysis. Many models consider neutral tweets along with positive and negative tweets as well and create a three class classifier. It is effective to analyze between three classes as long as there are no overlapping sequence

Table 1. Aspect Based Keywords

Aspect	Terms / Keywords
Network	gsm, cdma, hspa, evdo, lte, network, call quality, connectivity, calls, reception, signal, 2g, 3g, 4g.
Build	dimension, build, design, plastic, weight, gorilla glass, frame, durable, plastic, aluminum, steel.
Display	screen, resolution, multitouch, glass, touchscreen, display, led, lcd, tft, brightness, contrast, nits, refresh rate, megapixel, amoled, aspect ratio, bezel, ppi, hdr10, hdr10+.
Performance	processor, speed, cache, startup, boot, perform, os, slow, hang, struck, perform, response, respond, cpu, gpu, ram, ufs, animation, switch, app, memory, freeze, game.
Camera	primary camera, secondary camera, front camera, back camera, cam, video, image, pic, photo, flash, hdr, panorama, night mode, dual camera, dual cam, triple cam, triple camera, lens, zoom, recording, slow motion, landscape, portrait, wide angle, ultra wide, telephoto, periscope, monochrome, depth sensor, focal length, aperture, autofocus, focus, optical stabilization, selfie.
Audio	sound, audio, speaker, 3.5mm, stereo, dual speaker, mic, headphone, headset, earphone, microphone, music, bass, treble.
Connectivity	wifi, hotspot, airdrop, router, bluetooth, device pair, nfc, near field, gps, usb.
Features	fingerprint, finger print, sensors, face unlock, accelerometer, compass, features, ir blaster.
Battery	battery, backup, charge, power, standby, back up, adapter, mah, charging coil.
Overall	budget, price, cost, money, system, phone, appearance, handset, look, device.

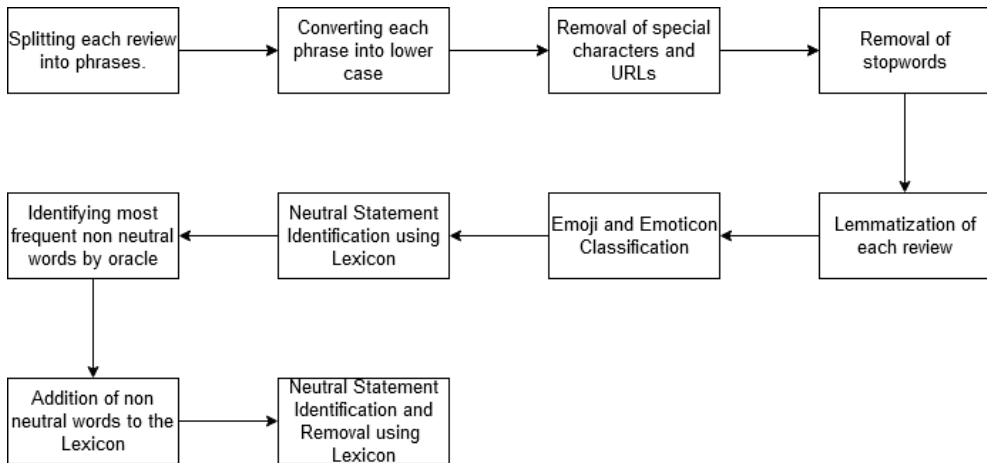


Figure 3. Pre-Processing Steps of the Proposed Approach

of words within any two classes which may cause misclassification. It is natural that neutral statements are an integral part of sentiment analysis as there are subjective as well as objective statements in reality. Having a two-class classifier with input data consisting of positive, negative and neutral tweets may have objective statements classified either as negative or positive. To resolve this, it's important to either create a three class classifier or remove neutral tweets. If undone, neutral sentences generally lie near the boundary, they either are considered as negative or positive which provides objective statements as either positive or negative. The terms that are objective are misclassified in these cases. Many models which have utilized machine learning approaches as well as lexicon based approaches have not considered these scenarios.

3.2. Model Selection

In our proposed model, we have considered to ignore neutral tweets in training as well as testing data. This is performed by using lexicon-based sentiment analysis with human effort to improve the accuracy in order to remove neutral tweets. A combination of Vader, AFINN and SentiWordNet lexicons hereafter termed as custom lexicon is used to perform sentiment analysis. For each aspect, the neutral words may differ. Considering this, it is vital to remove only neutral words and retain non-neutral tweets which carry sentiment for a particular aspect. For instance, words such as “lasts” and “long” are considered neutral with Vader sentiment analysis whereas they are positive terms when considered with the aspect “battery”. Such words are added to the lexicon. The process of finding these terms which are neutral according to Vader lexicon but not for the aspect considered, is as follows:

Tweets belonging to a single aspect are stored together. There are certain tweets which consists of more than a single aspect. Such tweets are stored separately for each

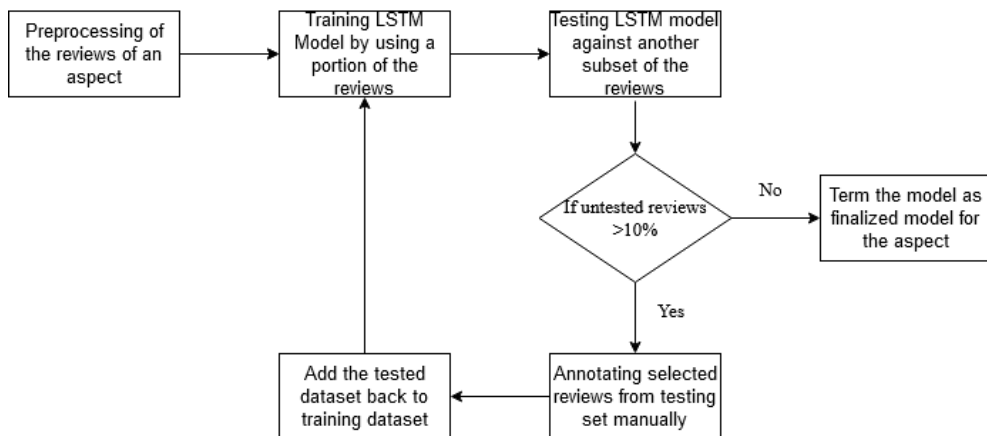


Figure 4. Active Learning in Proposed Approach

aspect. For instance, “Camera battery are great” is a tweet which speaks about the camera as well as battery with the same sentiment in a single statement. However, if the sentiments were different, the tweet would be, “Camera is great, battery is terrible.” This tweet would be divided into two phrases respectively.

To perform removal of neutral tweets, training data belonging to a single aspect are taken up. Vader sentiment analysis is utilized with additional words from AFINN and SentiWordNet lexicons. This set of words and their positive or negative scores are utilized to perform sentiment analysis on the set of tweets belonging to a single aspect. Followed by this, all the positive and negative tweets are stored separately from all the neutral tweets. The top 100 most frequently used words from the neutral tweets are retrieved and shown to the human annotator. The human annotator who is an expert in the domain decides if there are any false neutral words present. For instance, false neutral words such as “heavy”, “consume”, “small” and “beast” were found in the top 100 most frequently used words within neutral tweets for the aspect “battery”. These false neutral words are then added back to the lexicon by assigning scores from the annotator and sentiment analysis is performed again. The remaining neutral tweets in the second round of sentiment analysis are discarded. This ensures to keep the opinionated tweets for the binary classifier which classifies as positive or negative. These remaining tweets are considered as training data for the model to be built. These steps are repeated for each aspect by performing sentiment analysis using the lexicon created. The 100 most frequently used words within the neutral tweets are analyzed by the human annotator. The false neutral words are added to the lexicon separately for each aspect to be analyzed by the lexicon based sentiment analysis model again. The neutral tweets resulting here are discarded. Opinionated training dataset for each aspect is obtained using these steps which can be fed into the active learning based deep learning model. The steps involved within pre-processing of the dataset performed are depicted in Figure 3.

The deep active learning model is based on long short term memory model. The

Algorithm 2 Active Learning based Sentiment Analysis Model Creation

Require: The following are inputs for active learning based sentiment analysis model.

Tweets = $\{t_1, t_2, t_3, \dots, t_n\}$, n tweets as testing data

Ensure: The following are the outputs of the active learning based sentiment analysis model.

Checkpoints of Aspect Based Sentiment Analysis Models = $\{ac_1, ac_2, ac_3, \dots, ac_m\}$,
for m aspects of mobile devices

```

1: for ( $i \leftarrow 1$  to  $n$ ) do
2:   For each input tweet  $t_i$ , perform pre-processing steps as per Algorithm 1 from
     steps 1 to 12 resulting in  $x$  reviews for all the tweets
3: end for
4: for (each  $r$  in review) do
5:   for (each  $a$  in aspect) do
6:     for (each  $k$  in keywords of  $a$ ) do
7:       if  $k$  exists in review  $r$  then
8:         Place the review  $r$  within the category of aspect  $a$ 
9:       end if
10:    end for
11:  end for
12: end for
13: for (each review  $r_{ia}$  in aspect  $a$ ) do
14:   Manually annotate first 5000 reviews  $r_{1a}, r_{2a}, \dots, r_{5000a}$  within the aspect  $a$ 
     and train model  $ac_a$ 
15:   for (each next 5000 reviews  $r_i$  of aspect  $a$ ) do
16:     Classify the review  $r_i$  as either positive or negative using the model  $ac_a$ 
17:     for (each review  $r_i$  with a confidence level of 0.75 or lesser) do
18:       Classify the review  $r_i$  manually by the expert and store the result
         back with confidence level of 1.0
19:     end for
20:   end for
21: end for
22: for (each model generated  $ac_a$  for aspect  $a$ ) do
23:   Checkpoint the best weighted model for the aspect  $a$  as  $ac_a$ 
24: end for

```

steps within building the model are described within Algorithm 2. The aim of the active learning model is to ensure that the deep learning model will provide an accuracy of at least 70% in the initial run. The initial run consists of the human annotator manually annotating all the tweets provided to him / her.

3.3. Inner workings of the pipeline

This initial run consists of a small part of training data. The tweets annotated by the oracle are considered as the first set of training data for active learning and hence provides selective tweets in the upcoming runs ensuring that the annotated tweets are less. For instance, consider that the number of tweets to be analyzed manually is 50000, the oracle annotates 5000 tweets manually in the initial run, these tweets are considered as training data within the active learning method. The next 5000 tweets of the 50000 tweets considered to be training data are not provided to the oracle for annotation. Instead, a LSTM model is created considering the first 5000 tweets annotated by the oracle as training data is built. This model considers the next 5000 tweets as testing data and performs the classification. Within the classified tweets i.e. the test data of 5000 tweets, we consider all the tweets classified with a confidence of greater than or equal to 75% as accurately classified. Only the remaining tweets whose confidence levels lie to be lesser than 75% are taken into consideration and provided to the oracle for annotation. This effectively reduces the amount of tweets to be analyzed by the oracle considerably. Albeit there could be misclassifications to a certain degree, the aim of the research project is to reduce the human effort required during annotation. Let us assume out of the 5000 tweets tested by the model, around 3000 have confidence level greater than 75% (This is 60% of the tested tweets). The rest of the tweets (2000 in number) is analyzed manually by the oracle as it's confidence level lies below 75%. This essentially means that the tweets manually annotated lie closer to the decision boundary. These tweets are the most important tweets to be classified correctly as they provide the highest informative value to the model to learn and hence, improve the accuracy. These 5000 tweets are added to the initially analyzed 5000 tweets as the new set of training data. This new set of training data is used to train the LSTM model and classify the next 5000 tweets. Of the tested data, the tweets which have lesser than 75% confidence levels are annotated manually by the oracle and added back to the training dataset.

The same steps are repeated until all the tweets considered as training data are selectively annotated based on the informativeness level and the final training dataset is created to perform testing in real time. This is performed for the training data within each aspect. The best weighted models are saved as a snapshot using checkpointing. This eliminates the need for the model to be trained every time a real time classification is to be performed. The active learning methodology utilized is described within Figure 4.

Table 2. Number of Reviews per Aspect for Training the Model

Aspect	Number of Reviews (Neutral Statements Removed)	Number of Reviews (With Neutral State- ments)
Network	41592	64353
Build	77809	102133
Display	192369	253241
Performance	156871	194352
Camera	237651	296547
Audio	88265	101238
Connectivity	47813	63941
Features	81308	97324
Battery	212997	269768
Overall	146783	165934

4. Results

Mobile phone reviews were collected from e-commerce sites namely Amazon and Flipkart. Followed by this, aspect based search was performed in Twitter with the keywords as mobile phone name, the aspect to be searched for with “and” filter. For instance, “iPhone” and “battery”. This provided us with millions of reviews in total to consider as the training dataset. The amount of tweets and reviews obtained were then pre-processed to reduce them to phrases consisting of aspect based keywords. A total of 1415135 reviews were obtained from Amazon to be utilized as training dataset in total. Followed by this, using Twitter (Twint Package), a total of 203081 tweets were obtained. Non-English tweets were discarded during the process. All of these tweets were taken together to perform pre-processing in order to split them into phrases, remove neutral statements, perform pre-processing steps such as conversion to lower case, removal of special characters and stopwords, lemmatization, emoji and emoticon classification. Only the reviews consisting of aspects are retained. The number of reviews which are retained after the pre-processing steps stands at 1283458. The same process is repeated to obtain the tweets without performing neutral statement remove. The number of reviews obtained in this case stands at 1608831 as indicated in 2.

These are divided into aspects as shown in Table 2. A separate model is created for each of the aspect using deep active learning technique. The reviews were provided in the streams of 5000 at a time to the annotator. For each aspect, 5000 tweets were provided to the annotator for labelling. These were considered as the training dataset to test against the next 5000 tweets. Only the tweets having a confidence level lower than 75% were provided back to the user for analysis ensuring that the most informative reviews were provided to the annotator for labelling. Based on this, testing dataset was taken from twitter by searching for phone name and aspect as the keyword on twitter. The test keyword used i.e. the phone name as “iphone se” and

Table 3. Number of Reviews per Aspect for Testing Data

Aspect	Number of Reviews	Number of Positive Classifications	Number of Negative Classifications	Rating
Network	11083	6027	5056	5.44
Build	13532	8135	5397	6.01
Display	21043	16470	4573	7.83
Performance	12035	7302	4733	6.07
Camera	28793	22329	6464	7.76
Audio	12935	8402	4533	6.5
Connectivity	9892	5023	4869	5.08
Features	6950	4054	2896	5.83
Battery	37953	14043	23910	3.7
Overall	7521	5421	2100	7.21

Table 4. Number of Reviews per Aspect for Testing Data

Aspect	Number of Reviews	Number of Positive Classifications	Number of Negative Classifications	Number of Neutral Classifications
Network	11083	5967	4987	129
Build	13532	7094	5012	1426
Display	21043	14954	4328	1761
Performance	12035	6785	4134	1116
Camera	28793	19475	5938	3380
Audio	12935	7943	4183	809
Connectivity	9892	4012	4297	1583
Features	6950	3985	2285	680
Battery	37953	13723	19719	4511
Overall	7521	5286	1937	298

“battery”.

Table 5. Comparison of Results between different methodologies

		Confusion Matrices		Precision	Results	
		Pos	Neg		Recall	F-Measure
SVM based	Pos	68850	28356	87.52	70.83	78.29
Active Learning	Neg	9817	54714	65.86	84.79	74.14
LSTM based	Pos	75094	22112	93.27	77.25	84.50
Active Learning	Neg	5417	59114	72.78	91.61	81.11

A total of 300000 tweets were collected using twint package for the mobile device “iPhone SE” over a long period of time. This ensured that we had enough tweets in each aspect to test ensuring that the sampled testing data isn’t too low that may adversely affect the accuracy of the model. These tweets are clubbed together to perform preprocessing steps as a single tweet can have more than one aspect despite searching for a single aspect within the tweets. The number of tweets obtained belonging each aspect are listed along with the results of our approach is listed in Table 2. The total reviews which are aspect based that remain are 161737.

The model built using active learning with LSTM was recreated using different machine learning algorithms. Utilizing the same initial 5000 reviews annotated by the experts as training data, the next 5000 reviews were used as testing data to identify the most informative reviews to be annotated by the experts. This was performed using each machine learning algorithm separately as the decision boundary by each algorithm maybe different. To identify the accuracy, confusion matrix is created with the predicted output classes against the tested output classes for each model utilized. The comparison of Precision, Recall and F-Measure are provided in the Table 5

The same approach has also been applied to the three class classifier developed. The testing data consists only of positive and negative tweets to analyse the predictive capabilities of the three class classifier model. Deep active learning with LSTM was utilized with streams of 5000 tweets provided as input to the model. 9.7% of the polarized tweets were incorrectly classified as neutral which provides motive to the remove neutral tweets and perform sentiment analysis, leading to a higher accuracy.

The final step in calculation of results is to rate the aspect on a scale of 1 to 10, 1 being the lowest and 10 being the highest. This is calculated as the number of reviews that are positive divided by the total number of tweets.

$$Rating = (\sum PositiveReviews / TotalReviews) \times 10$$

The above formula provides a rating between the scale of 1 to 10. Assuming the number of reviews classified using for the aspect “battery” using the deep learning model are 15000. Of the 15000 reviews, let us assume that 10000 are positively classified. This provides the resulting rating using the formula as 6.6. We utilize this as the rating system to identify that 6.6 out of 10 people rate the aspect “battery” positively. The ratings for each aspect is shown in tables 3 and 4 . We are able to observe from our results that the people liked the display as the primary feature, camera as the secondary feature and overall features next. We also observe that the people voiced

their opinions towards disliking battery the most followed by connectivity the second most.

The accuracy for the SVM based model averaging F-Measures stands at 76.215 and active learning based model at 82.805. The accuracy difference between the two has a significant difference when considering these would be utilized to make business decisions. These results demonstrate a significant improvement in the performance of our proposed sentiment analysis framework compared to the baseline model. The proposed model achieves a higher F-measure and precision for both positive and negative sentiments, which indicates that it can accurately classify tweets as either positive or negative with a higher degree of precision.

5. Conclusion and Future Enhancements

The research performed demonstrates the utility of active learning within text classification which significantly reduces the effort in terms of time and money required to hire annotators. The research performed also shows that removal of neutral reviews should be considered to create a binary classifier instead of three-class classifier. The research performed shows that deep learning models show good accuracy in text classification when the initial training data is sufficient for the active learning model in the first run. We have observed that the model is under-fitted when a small amount of data samples i.e. around 100 are provided in the initial run for the model.

Further research can be performed utilizing different algorithms, parameter tuning and approaches to perform a comparative study. This would help identify the pitfalls and advantages of each approach. Furthermore, the sentiment analysis framework created considers neutral reviews to be removed by using a lexicon based method. The framework can be improved upon by utilizing the model to ensure that false neutral tweets are not removed and a three class classifier can be created to perform a comparative analysis. We have encountered certain scenarios where positive or negative tweets are identified as neutral and the words which are misrepresented as neutral do not fall into the top 100 most frequent words within the neutral tweets. Further research in these areas may lead to development of a more accurate model.

Research can also be performed by utilizing datasets of different domains to examine the effectiveness of the pipeline of techniques utilized. Instead of models for each aspect, a pre-trained model can be fine-tuned per aspect. Furthermore, the computational costs are higher in re-training the model in the proposed approach which can be replaced by fine-tuning the initial version of the model trained. This can be performed by freezing the existing layers in order to retain the information and add new layers on top of it.

The number of tweets selected by our active learning approach is 5000 random tweets within the training dataset as the initial training set for the deep active learning model. For further rounds of training, the data utilized in training set are the tweets that have a confidence level lesser than 75%. We have not performed research by changing this value as we have observed manually that most tweets that have confidence level greater than 75% were high in number and this was chosen arbitrarily

to reduce the number of annotations. However, further research can be performed by identifying different confidence levels or choosing the tweets that consist of words that are not available as embedded vectors in the model.

Instead of keeping sentiment analysis for two classes i.e. positive and negative, the same can be improved by making the categorization system such as extremely positive, positive, negative and extremely negative. This would ensure that two sentences such as "Greatest phone ever" and "Good phone" would have significantly different ratings such as possibly 10 and 7 respectively. This approach can be utilized to create a better rating system for customer understanding purposes.

References

- [1] Agarwal B, Mittal B, "Machine learning approach for sentiment analysis", in: Prominent feature extraction for sentiment analysis, Springer, 2016, pp. 21–45.
- [2] Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R.J., "Sentiment analysis of twitter data", in: Proceedings of the workshop on language in social media (LSM 2011), 2011, pp. 30–38.
- [3] Akhtar N , Zubair N, Kumar A, Ahmad T, "Aspect based sentiment oriented summarization of hotel reviews", *Procedia computer science* 115 (2017) 563–571.
- [4] Ansari M.Z., Aziz M, Siddiqui M, Mehra H, Singh K., "Analysis of political sentiment orientations on twitter", *Procedia Computer Science* 167 (2020) 1821–1828.
- [5] Arora S, Nyberg E, Rose C, "Estimating annotation cost for active learning in a multi-annotator environment", in: Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing, 2009, pp. 18–26.
- [6] Arruda G, Roman N, Monteiro A, "Analysing Bias in Political News.", *JUCS - Journal of Universal Computer Science*(2020), 173–199. <https://doi.org/10.3897/jucs.2020.011>
- [7] Balakrishnan V, Lloyd-Yemoh E, Stemming and lemmatization: "A comparison of retrieval performances", *Lecture Notes on Software Engineering* 2.
- [8] Balcan F.M. , Broder A, Zhang T, "Margin based active learning", in: International Conference on Computational Learning Theory, Springer, 2007, pp. 35–50.
- [9] Begum S, Bera S, Chakraborty D, Sarkar R, "Breast cancer detection using feature selection and active learning", in: Computer, communication and electrical technology: proceedings of the international conference on advancement of computer communication and electrical technology (ACCET 2016), West Bengal, India, 21-22 October 2016, The British Institute of Radiology., 2017, pp. 43–8.
- [10] Bianchi F.M., Maiorino E, Kampffmeyer M.C., Rizzi A , Jenssen R, "Recurrent neural network architectures", in: Recurrent Neural Networks for Short-Term Load Forecasting, Springer, 2017, pp. 23–29.

- [11] Bose R, Dey R.K., Roy S, Sarddar D, “Analyzing political sentiment using twitter data”, in: *Information and Communication Technology for Intelligent Systems*, Springer, 2019, pp. 427–436.
- [12] Cacciarelli D., Kulahci M., Tyssedal J, “Stream-based active learning with linear models.” *Knowledge-Based Systems*. 254 pp. 109664 (2022)
- [13] Carse J, McKenna S, “Active learning for patch-based digital pathology using convolutional neural networks to reduce annotation costs”, in: *European Congress on Digital Pathology*, Springer, 2019, pp. 20–27.
- [14] Aydin C., Güngör T, “Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations.” *IEEE Access*. 8 pp. 77820-77832 (2020)
- [15] Culnan M.J., McHugh P.J., Zubillaga J.I., “How large us companies can use twitter and other social media to gain business value.”, *MIS Quarterly Executive* 9 (4).
- [16] Daouadi K, Rebaï R, Amous I, “Real-Time Bot Detection from Twitter Using the Twitterbot+ Framework”, *JUCS - Journal of Universal Computer Science*(2020), 496–507. <https://doi.org/10.3897/jucs.2020.026>
- [17] Do H.H., Prasad P, Maag A, Alsadoon A, “Deep learning for aspect-based sentiment analysis”: a comparative review, *Expert Systems with Applications* 118 (2019) 272–299.
- [18] Enginkaya E, Yilmaz H, “What drives consumers to interact with brands through social media? a motivation scale development study”, *Procedia-Social and Behavioral Sciences* 148 (2014) 219–226.
- [19] Fan M, Billings A, Zhu X, Yu P, “Big data analysis of english national team fans during the 2018 fifa world cup”, *Communication & Sport* 8 (3) (2020) 317–345.
- [20] Freund Y, Seung S.H. , Shamir E , Tishby N, “Selective sampling using the query by committee algorithm”, *Machine learning* 28 (2) (1997) 133–168.
- [21] Friguet C, Dambreville R, Kijak E, Laroze M, Lefèvre S, “Data annotation with active learning: application to environmental surveys”, *JDS Journées De Statistique De La Société Française De Statistique (SFdS)*. pp. 1-6 (2020)
- [22] Gundecha P, Liu H, “Mining social media: a brief introduction”, *Institute for Operations Research and the Management Sciences (INFORMS), TutORials in Operations Research*. (2014) 1–17.
- [23] Gupta V, Singh V.K., Mukhija P, Ghose U, “Aspect-based sentiment analysis of mobile reviews”, *Journal of Intelligent & Fuzzy Systems* 36 (5) (2019) 4721–4730.
- [24] Hew K, Hu X, Qiao C, Tang Y, “What predicts student satisfaction with MOOCs; A gradient boosting trees supervised machine learning and sentiment analysis approach”, *Computers & Education*, 145 pp. 103724 (2020)

-
- [25] Jivani A.G., et al., “A comparative study of stemming algorithms”, *Int. J. Comp. Tech. Appl* 2 (6) (2011) 1930–1938.
- [26] Ibrahim N.F., Wang X, “A text analytics approach for online retailing service improvement: Evidence from twitter”, *Decision Support Systems* 121 (2019) 37–50.
- [27] Ibrahim N.F., Wang X, Bourne H, “Exploring the effect of user engagement in online brand communities: Evidence from twitter”, *Computers in Human Behavior* 72 (2017) 321–338.
- [28] Lango M, “Tackling the Problem of Class Imbalance in Multi-class Sentiment Classification: An Experimental Study”, *Foundations Of Computing And Decision Sciences*. 151-178 (2019), <https://doi.org/10.2478/fcds-2019-0009>
- [29] Li M , Sethi K.I., “Confidence-based active learning”, *IEEE transactions on pattern analysis and machine intelligence* 28 (8) (2006) 1251–1261.
- [30] Li Q, Shah S, “Learning stock market sentiment lexicon and sentiment-oriented word vector from stocktwits”, in: *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, 2017, pp. 301–310.
- [31] Liang B, Su H, Gui L, Cambria E, Xu R, “ Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks”, *Knowledge-Based Systems*. 235 pp. 107643 (2022), <https://www.sciencedirect.com/science/article/pii/S0950705121009059>
- [32] Luan Y, Lin S, “Research on text classification based on cnn and lstm”, in: *2019 IEEE international conference on artificial intelligence and computer applications (ICAICA)*, IEEE, 2019, pp. 352–355.
- [33] Mackoviak R, Lenz P, Ghori O, Diego F, Lange O, Rother C, “Cereals-cost-effective region-based active learning for semantic segmentation”, *arXiv preprint arXiv:1810.09726*.
- [34] Oliveira N, Cortez P, Areal N, “Stock market sentiment lexicon acquisition using microblogging data and statistical measures”, *Decision Support Systems* 85 (2016) 62–73.
- [35] Olsson F, “A literature survey of active machine learning in the context of natural language processing”, *SICS Technical Report T2009:06* ISSN: 1100-3154.
- [36] Pimprikar R, Ramachadran S, Senthilkumar K, “Use of machine learning algorithms and twitter sentiment analysis for stock market prediction”, *Int J Pure Appl Math* 115 (6) (2017) 521–526.
- [37] Plisson J, Lavrac N, Mladenic D, et al., “A rule based approach to word lemmatization”, in: *Proceedings of IS*, Vol. 3, 2004, pp. 83–86.

- [38] Reidsma, D, Akker, R, “Exploiting ‘subjective’ annotations”. *Coling 2008: Proceedings Of The Workshop On Human Judgements In Computational Linguistics*. pp. 8-16 (2008)
- [39] Roth D, Small K, “Margin-based active learning for structured output spaces”, in: *European Conference on Machine Learning*, Springer, 2006, pp. 413–424.
- [40] Schumaker R.P., Jarmoszko A.T., Labedz Jr C.S., “Predicting wins and spread in the premier league using a sentiment analysis of twitter”, *Decision Support Systems* 88 (2016) 76–84.
- [41] Schumaker R.P., Labedz Jr C.S., Jarmoszko A.T., Brown L.L, “Prediction from regional angst—a study of nfl sentiment in twitter using technical stock market charting”, *Decision Support Systems* 98 (2017) 80–88.
- [42] Schumann R, Rehbein I, “Active learning via membership query synthesis for semi-supervised sentence classification”, in: *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 472–481.
- [43] Simard P, Chickering D, Lakshmiratan A, Charles D, Bottou L, Suarez C.G.J, Grangier D, Amershi S, Verwey J, Suh J, “Ice: enabling non-experts to build models interactively for large-scale lopsided problems”, *arXiv preprint arXiv:1409.4814*.
- [44] Slimi H, Bounhas I, Slimani Y, “Adapting Pre-trained Language Models to Rumor Detection on Twitter”, *JUCS - Journal of Universal Computer Science*(2021), 1128–1148. <https://doi.org/10.3897/jucs.65918>
- [45] Smagulova K, James A.P., “A survey on lstm memristive neural network architectures and applications”, *The European Physical Journal Special Topics* 228 (10) (2019) 2313–2324.
- [46] Song J , Wang H, Gao Y, An B, “Active learning with confidence-based answers for crowdsourcing labeling tasks”, *Knowledge-Based Systems* 159 (2018) 244–258.
- [47] Sun K, Zhang R, Mensah S, Mao Y, Liu X, “Aspect-level sentiment analysis via convolution over dependency tree”, *Proceedings Of The 2019 Conference On Empirical Methods In Natural Language Processing And The 9th International Joint Conference On Natural Language Processing (EMNLP-IJCNLP)*. pp. 5679-5688 (2019)
- [48] Thrun S, “Exploration in active learning”, *Handbook of Brain Science and Neural Networks* (1995) 381–384.
- [49] Tsimonis G, Dimitriadis S, “Brand strategies in social media”, *Marketing Intelligence & Planning* 32 (2014) 328–344.
- [50] URL - Twitter User Details: <https://backlinko.com/twitter-users>

-
- [51] Vanaja S, Belwal M, “Aspect-level sentiment analysis on e-commerce data”, in: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2018, pp. 1275–1279.
- [52] Von Helversen B, Abramczuk K, Kopeć W, Nielek R, “Influence of consumer reviews on online purchasing decisions in older and younger adults”, *Decision Support Systems* 113 (2018) 1–10.
- [53] Voutilainen A, “Part-of-speech tagging”, *The Oxford handbook of computational linguistics* (2003) 219–232.
- [54] Wang H, Jin Y, Doherty J, “Committee-based active learning for surrogate-assisted particle swarm optimization of expensive problems”, *IEEE transactions on cybernetics* 47 (9) (2017) 2664–2677.
- [55] Webster J.J., Kit C., “Tokenization as the initial phase in nlp”, in: *COLING 1992 Volume 4: The 15th International Conference on Computational Linguistics*, 1992.
- [56] Wu T, Weld D.S., Heer J, “Local decision pitfalls in interactive machine learning: An investigation into feature selection in sentiment analysis”, *ACM Transactions on Computer-Human Interaction (TOCHI)* 26 (4) (2019) 1–27.
- [57] Yan Y, Rosales R, Fung G, Schmidt M, Hermosillo G, Bogoni L, Moy L, Dy J, “Modeling annotator expertise: Learning when everybody knows a bit of something”, in: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, 2010, pp. 932–939.
- [58] Yan Y, Rosales R, Fung G, Subramanian R, Dy J, “Learning from multiple annotators with varying expertise”, *Machine learning* 95 (3) (2014) 291–327.
- [59] Zhou Z.H., Li M, “Semi-supervised learning by disagreement”, *Knowledge and Information Systems* 24 (3) (2010) 415–439.

Received 07.11.2022, Accepted 18.05.2023