A Contemporary Ensemble Aspect-based Opinion Mining Approach for Twitter Data

Satvika[1]
PhD Scholar, CSE Department
Amity University
Gurugram, India

Dr. Vikas Thada[2]
Associate Professor, CSE
Department, Amity University
Gurugram, India

Dr. Jaswinder Singh[3]
Associate Professor, CSE
Department, GJUS and T
Hisar, India

Abstract—**Aspect-based opinion mining is one among the thought-provoking research field which focuses on the extraction of vivacious aspects from opinionated texts and polarity value associated with these. The principal aim here is to identify user sentiments about specific features of a product or service rather than overall polarity. This fine-grained polarity identification about myriad aspects of an entity is highly beneficial for individuals or business organizations. Extricating these implicit or explicit aspects can be very challenging and this paper elaborates copious aspect extraction techniques, which is decisive for aspect-based sentiment analysis. This paper presents a novel idea of combining several approaches like Part of Speech tagging, dependency parsing, word embedding, and deep learning to enrich the aspect-based sentiment analysis specially designed for Twitter data. The results show that combining deep learning with traditional techniques can produce excellent results than lexicon-based methods.**

Keywords—*Aspect-based sentiment analysis; dependency parsing; long short-term memory (LSTM); part of speech (POS) tagging; term frequency-inverse document frequency (TF-IDF)*

I. Introduction

The past decade is undoubtedly dominated by data and its analytics. Now, anyone with an immense amount of data related to a domain and with the right tools to mine this colossal mountain of data is considered powerful. Surveys (either offline or online) are obsolete now as now more and more user is giving their opinions about myriad products and services on various social networking websites like Facebook, Instagram, LinkedIn, QQ, Telegram, Twitter, WhatsApp, WeChat, etc.[1]. All these social media platforms have been quite popular for exchanging opinions and sentiments and hence provide much-needed feedback about the specific products, services, vital events, organizations, or persons with the active participation of users. A lot of time, effort, and money has been put into analyzing user sentiments from a plethora of social sites especially Twitter due to the reason that tweets are limited to 140 characters[2]. This character limitation makes Twitter an easy and efficient tool for sentiment analysis, which can be pretty useful for organizations, who want to analyze related documents and make optimum changes to better suit their targeted customers. However, the sentiment analysis generally determines the overall opinion and hence may not be able to extract the precise essence needed to review that particular product or service.

This could be easily understood by the following example from a review about Cars: "Ford Mustang is awesome to drive, but the price is too high". This particular sentence basically talks about two different features of the subject (Ford Mustang) here i.e., "Driving Comfort" and "Price". Another notable point to comprehend is that a sentence may contain more than one aspect and the polarity of each aspect can be diverse. In the example shown above, "Driving comfort" is an undoubtedly positive polarity aspect while the "Price" aspect is declined more towards negative polarization. Hence to properly analyze the available data, a more detailed-level approach is needed, which is stated as the aspect-centered sentiment analysis. This methodology provides better insight for mining user opinions about the data under review. Overall, the sentiment analysis can be done at manifold stages that are document level, sentence phase, and aspect level[3]. Thus, in a nutshell, it can be stated that aspect-based sentiment analysis produces more fine-grained data analysis of user data.

In this research, an aspect-based sentiment mining problem is discussed on Twitter data and an efficient hybrid approach for the same has been proposed. The rest of this paper is organized as follows: Section II describes the numerous techniques used for aspect terms extraction. Section III gives an understanding of the projected framework. Section IV presents the outcomes of the proposed framework and evaluates them. Finally, Section V entails the conclusion and future work in the same direction.

II. Aspect Extraction Techniques

Sentiment analysis is usually defined as the computational study of a person's opinions or emotions or views about a particular entity. Nevertheless, an entity has certain features associated with it, which must be considered with utmost precaution as these characteristics define opinions at the atomic level. Aspects can be briefly stated as the attributes or traits of a product or service [4]. An Aspect extraction is a primary process of identifying these significant attributes and a considerate argument here is that aspect extraction aiming at no particular aspect or target is of circumscribed usage [5]. Hence aspect extraction is the most imperative step for pulverized opinion mining. There are two sorts of aspects, namely: implicit and explicit aspects. The explicit aspects are mentioned overtly in the opinionated sentence whereas the implicit aspects are expressed indirectly and that is why more challenging to discover. For example: "Mustang is a classic car from Ford, renowned as being the muscle car". In the

above sentence, "Ford" as the "brand" of the car is an explicit aspect, while "muscle car" refers to the implicit aspect "performance". The subsequent step to aspect extraction is finding the sentiment polarity of recognized features; like in the above example "brand" aspect is neutral and the "performance" aspect is positive.

Some of the key techniques employed for finding aspects are shown in Fig. 1.

1) Frequency-based approaches: Considered as the most traditional approach, it chiefly employees tracing out the most commonly occurring words. One such technique is Part-of-Speech (POS) tagging which basically finds out the classification of words based on the grammar within a text like a noun, pronoun, verb, adjective or adverb, etc. POS tagging is not only the simplified syntactically tagging taught in school education, but it also tries to find the relationship between the word under consideration and the adjacent words. Thus, POS tagging is not generic but depends upon the sentence to sentence and finally, it can be said that it tries to extract speech tags based on the sentence context. Also, an imperative point is that POS tagging does not actually do the Natural Language Processing (NLP), but it is the precondition to handle a lot of NLP tasks. The main characteristic of POS tagging is that it finds the recurrent noun and noun phrases from the reviews as they are usually the aspects [7]. POS tagging on a review is demonstrated in Fig. 2.

However, it must be understood that some of the aspects identified by this approach may not be the key aspects and must be rendered. Another frequency-based tactic is Pointwise Mutual Information (PMI), which is figured as the variance of the common information between the feature and the related word [8]. Last but not the least is Term Frequency-Inverse Document Frequency (TF-IDF) methodology which evaluates how significant a word is to a document in a group of the corpus by calculating two terms. The first term (TF) is the frequency of a word appearing in a document, divided by the overall words' tally in that document. The second term (IDF) is computed as the logarithm of the whole documents in the corpus divided by the number of documents where the particular word appears [9]. In simple words, TF-IDF is the statistical tool that keenly discovers the most vital words in the articles. An important feature of TF-IDF is that some words seem noteworthy due to their huge frequency in the text, but actually, they are insignificant like the, this, that, etc. [8]. Hence, TF-IDF can be described as the amalgamation of two algorithms that diminishes the weight of stopwords and recognizes the high opinionated words, which actually influence the sentiments contained in the documents.

2) Relation-based approaches: These techniques are based on discovering relations between features and the opinion words to identify aspects clearly. One such approach is Dependency Parsing – which is based on extracting the grammatical (or syntactic) structure of a sentence and finding the relation between vital words and the words which modify these vital words [10]. The second practice is called Double Propagation, which disseminates information among opinion words and targets back and forth. This method comes under the semi-supervised category as it takes opinion word as the seed to start the process [11].

3) Supervised machine learning: The two most prominent aspect mining techniques Hidden Markov Model (HMM) and Conditional Random Fields (CRF) come under this category. HMM are a set of probabilistic visualization model that allows forecasting a sequence of hidden variables from a set of perceived variables. It is a great approach for tracing the aspects like it may predict the type of weather by observing the type of clothes worn by a person [12], as depicted in Fig. 3.

On the other hand, CRF is a discriminative model employed for predicting sequences where multiple variables are dependent on each other by imputing contextual information from previous labels [14]. Its main applications are in POS tagging and Named Entity Recognition (NER).
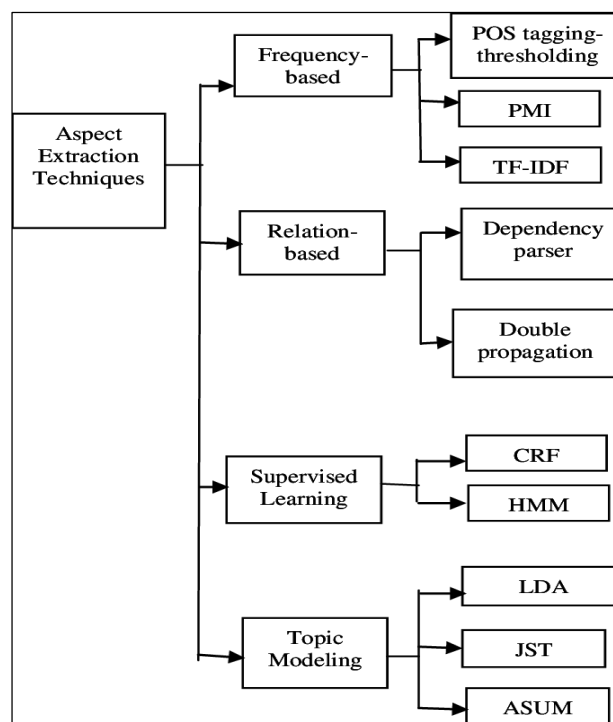

Fig. 1.   Various Aspect-Extraction Approaches [6].
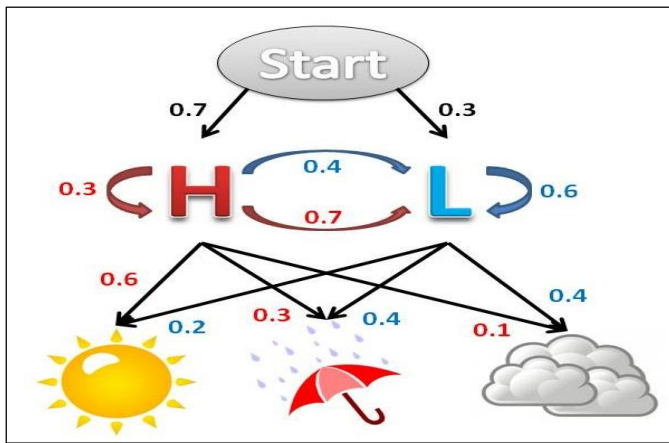

Fig. 2.   POS Tagging Output.

Fig. 3.   Working of Hidden Markov Model [13].

4) Topic modeling: As the name suggests, it is a category of techniques used to automatically detect topics present in the text corpus. It comes under unsupervised learning for observing topics (a group of words) in a colossal amount of text clusters. Latent Dirichlet Allocation (LDA) is used to categorize text in a document to a certain topic by using Bag-of-Word (BOW) approach[15]. The prime objective of LDA is to match all the documents to the topics such that words in every document are customarily classified by those imagined topics. The next approach is Joint Sentiment Topic (JST) is the exquisite unsupervised technique that detects sentiments as well as the topics simultaneously from the text corpora by considering their mutual relations[16]. The next methodology categorized under topic modeling is the Aspect and Sentiment Unification (ASUM) model which automatically determines aspects and different sentiments towards these recognized aspects in one go. This model ascertains aspect-sentiment pair called 'senti-aspects' in an unsupervised way, explaining how much a particular word is related to aspect and sentiment [17].

III. PROPOSED FRAMEWORK

The focal emphasis is to conduct an exploration of aspect-based sentiment analysis conducted on Twitter data. The planned work consists of the following steps:

1) Data collection from Twitter.
2) Data pre-processing and preparation.
3) Aspect extraction from the data.
4) Aspect selection & polarity detection.
5) Tweet level Sentiment classification & Performance evaluation.

The first and foremost step is collecting data from Twitter, which is a social media networking site. There are numerous ways to extract data from Twitter; among which the most popular one is Twitter API Search, which allows us to retrieve the latest tweets about any topic. For doing so, a Twitter development account must be created first. Earlier, this method had a limitation on the number of tweets and also on the number of requests received per hour from a particular IP address [18].

After the data collection process from Twitter, the next work is to prepare the data for analysis and this process involves cleaning and pre-processing of data. The data from Twitter is not in a particular format and it must be cleansed beforehand like removing the links, emoticons without clear sentiment, new lines, hashtags and symbols like @, punctuation marks, etc. For effective aspect discovery, tweets, users who tweeted them, followers of users, count of retweets, date of the tweet, etc. are chosen from the whole data. After data collection and pruning, the next step is to pre-process it. The data pre-processing starts with the removal of stopwords i.e. removing the commonly occurring words like a, an, the, and, or, is, etc. as these words do not convey meaningful opinions. Then, all the words in the corpus are changed to lowercase for enhanced understanding and investigation. Afterward, the corpus is subjected to stemming and lemmatization i.e., words are reduced to their basic or root form. It is a highly recommended part because it reduces derivationally related forms of a word. For example: "programmable" is changed to "program" and "running" is converted to "run". One example of text before and after pre-processing is shown in Fig. 4.

Original Tweet:

Tecney C900 Black Car DVD Sale, Price &
Reviews | Gearbest https://t.co/oM6s9jp8nn

Clean Tweet:

Tecney C900 Black Car DVD Sale, Price amp
Reviews | Gearbest

Fig. 4.   The Output obtained after Applying Pre-processing on Data.

Now, the data is ready for opinion mining and the foremost thing to do is finding aspects in these tweets. It can also be done in myriad ways as discussed in the above section. The aspect-extraction followed here is a hybrid one as multiple methodologies are combined for doing so including POS tagging, Dependency, etc. POS tagging is used to find the frequent nouns and noun phrases which tend to be the explicit aspects, while depending on parsing is brilliant for finding the implicit tags. For bigrams and trigrams, tokens are identified by checking words on the left side for 'proper Nouns' identified earlier. If the dependency is a compound or adjective modifier, then it is regarded as the aspect term. If the token is an 'adjective' or 'verb', the words on both sides (left and right) are examined for adverbial modifiers, open-clausal components, and 'auxiliary' dependencies and if it happens, they are regarded as aspect terms. For negation modification, words to the left of verbs and words on both left and right are diagnosed for adjectives and auxiliaries and in this case, the polarity of aspect is changed. Thus an ensemble of techniques is employed here to give the best output and providing every key aspect from the text corpora obtained after data preprocessing. The aspects can be ranked in decreasing order of frequency and selected on this basis. The next phase is detecting the polarity of each aspect and there are three types of polarities: positive, negative, and neutral. Positive polarity refers to the constructive and affirmative comments about the

aspect under consideration while negative polarity denotes the adverse and undesirable user views regarding that precise aspect and if the opinions do not fall under any of the above categories, it is considered as neutral polarity. Also, an imperative deliberation is that if an aspect has an equal number of positive and negative opinions, then it can be ignored. The algorithm for aspect-terms extraction can be described as the following:

1) Apply Tokenization first.

2) Prepare a list of stopwords, so that aspects, if identified on the stopword list, can be ignored.

3) Apply POS tagging for finding aspects.

4) The prime candidates for aspects are 'Nouns' or 'Proper Nouns' individually or in pair with Verbs/ Adverbs or Adjectives (explicit aspects).

5) For finding implicit aspects, dependency parsing is employed, along with bigrams and trigrams:

a) For a noun, check its dependency on its L.H.S. and if it is found to be 'compound', with an adjective or adjective modifier, then it should be an aspect-term.

b) For an adjective, check its dependency parsing on both L.H.S & R.H.S., if its dependency is auxiliary on L.H.S and 'negation modifier' on R.H.S, then is also identified as an aspect-term.

c) For a verb, if an adjective appears to its left or right and their dependency is 'adverbial modifier', then also it should be an aspect. Also, if an adjective appears on its L.H.S and dependency relation is negation modifier, the aspect is recognized.

Finally, the tweet level sentiments are analyzed using Vader Sentiment and Recurrent Neural Networks, and the results determined are verified in regard to state-of-art technologies. The overall process can be described by the Fig. 5 underneath.
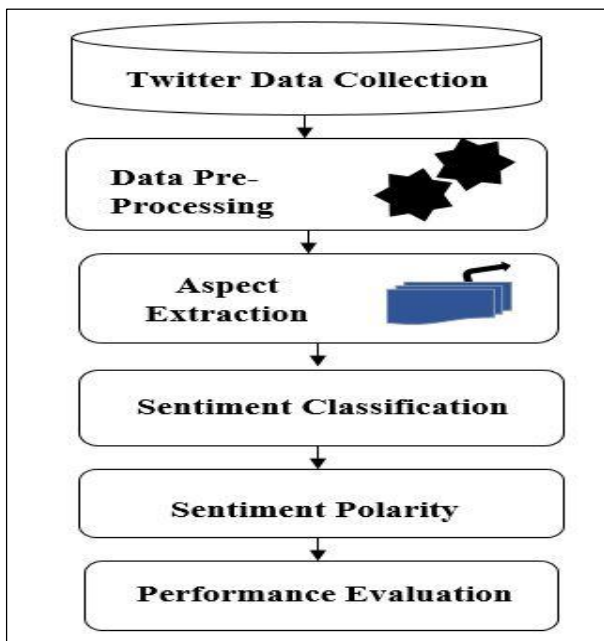
IV. RESULTS AND EVALUATION

The ensemble methodology for aspect-based sentiment analysis is a novel approach that does excellent work for up to 49493 tweets data corpus. However, more data means more training and better results. The new algorithm is also employed on 'Scraped_Car_Review_ford.csv', which can be downloaded from Kaggle for verification. This dataset has 49493 reviews about ford cars. Fig. 6 shows the first five tweets downloaded from the Twitter site.

The above image Fig. 7 shows the comparison between tweet frequency and the number of followers for verified Twitter users. The data downloaded from Twitter is pre-processed as discussed in the above section and sample output is shown in Fig. 4 above. After pre-processing of data including tokenization and stemming, POS tagging is applied. After obtaining POS tags for each token, some special words with peculiar tags are selected and these are Nouns, Noun Phrases (like NNP, NNS & NNPS), Adjectives, and Adverbs because these tags are extremely helpful while identifying aspects.

The wordcloud is primarily an innovative tool that generates an aesthetic amalgamation of frequently occurring words. It puts together recurrent textual data into a beguiling visual representation and shows them in bigger and bolder fonts according to their frequency [19]. Fig. 8 shows the potential aspects identified by POS tagging and frequency distribution:

Fig. 9 demonstrates the top 20 frequent words occurring in the tweets according to their frequencies arranged in descending order.

While Noun and Noun phrases are the potential aspects, adjectives and adverbs support their claim when the dependency is checked between them. This is identified by applying dependency parsing between them, as illustrated below in Fig. 10.

After all the aspects are extracted, their polarity needs to be checked and for that Sentiwordnet can be used to check the opinion sentiment for all the mined aspects according to the tweets they are appearing in. Finally, the overall sentiment classification can be calculated using Vader Sentiment, which is a proficient tool for mining the sentiments. Fig. 11 shows the overall sentiment distribution of the downloaded tweets.

Fig. 5. Aspect-Based Sentiment Analysis Process.

Fig. 6. Sample of Tweets Downloaded from Twitter.
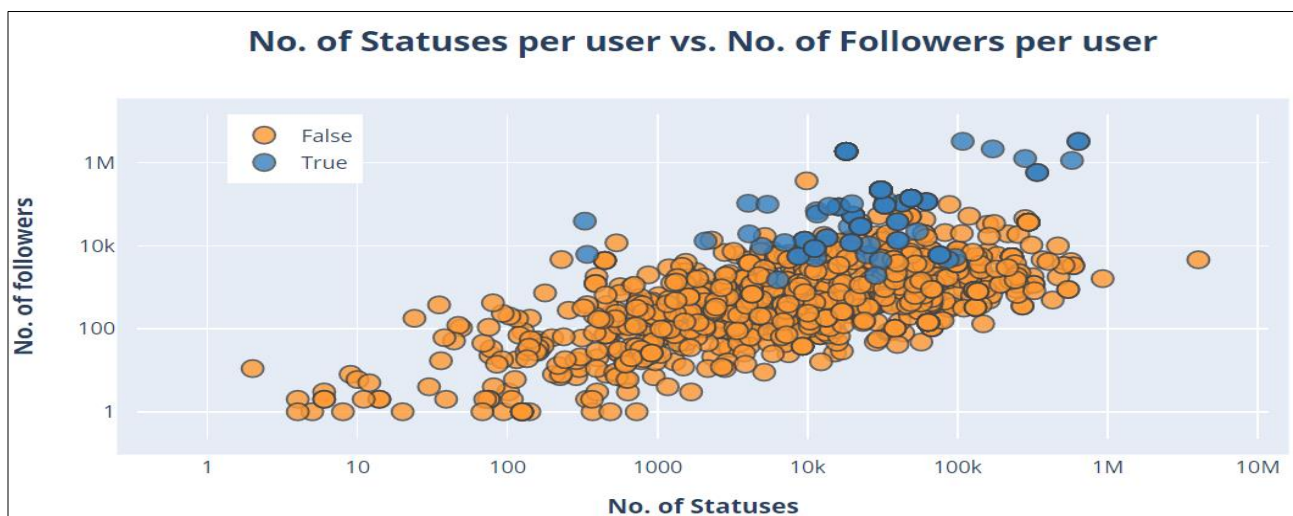
Fig. 7.   No. of Statuses vs. No. of Followers.

Fig. 8.   Wordcloud Showing Potential Aspects.
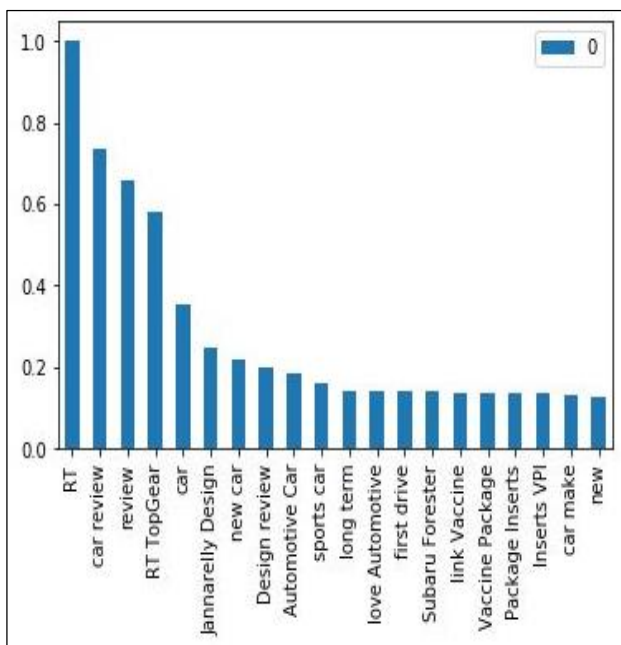

Fig. 10. Dependency Graph Example.

Fig. 9.   Frequency Distribution of Potential Aspects.
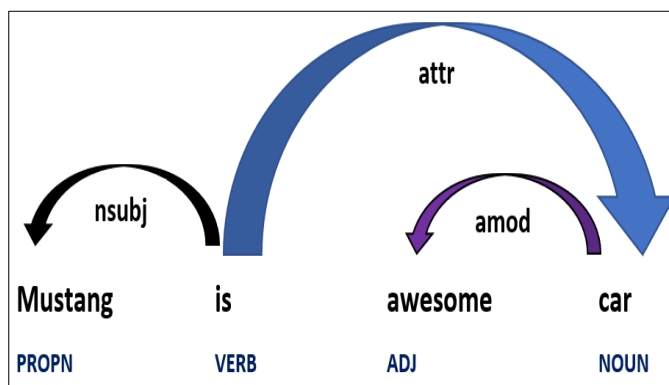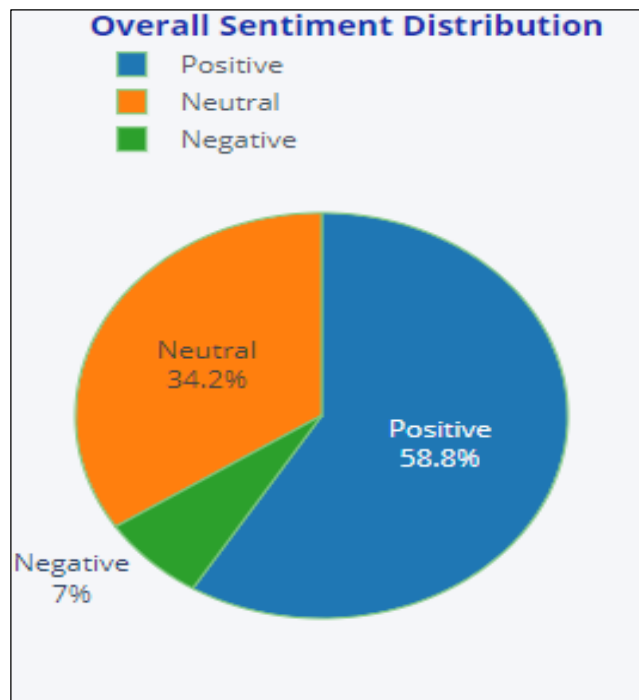

Fig. 11. Overall Sentiment Distribution.

An alternate sentiment categorization can be done using deep neural networks with the aid of TensorFlow and Keras. The model proposed in this paper is quite simple and contains three layers: embedding layer, a layer with LSTM functionality, and last but not least an output layer. The embedding layer will learn word embeddings for each, and every word contained in the corpus through pre-trained word embeddings provided by Google's word2vec. The second layer is of Long Short-Term Memory Networks (LSTM), which is an exceptional branch of Recurrent Neural Networks that are skilled in finding and learning relationships between features of an input sequence. An LSTM layer with 100 memory units is engaged in the model. Lastly, the dense layer provides uses a sigmoid activation function as it needs to classify the polarity of a tweet as either positive or negative. The summary of LSTM model used in this study is displayed in the Fig. 12.

The model once built, must be tested for accuracy to check if it is doing the intended work with efficacy or not. Table I lists the results of the proposed model, primarily the model score and its accuracy on the train and test data.

The validation loss and validation accuracy can also be identified for checking the model's performance. As evident from the Fig. 13, the model gave the best output till 5 epochs and then its performance decreases abruptly till 15 epochs. Hence, the epochs can be decreased to give a superior result.

As apparent from Fig. 14 beneath, the validation loss is quite high and hence the model overfits.

To conclude, the proposed model is evaluated for train and test data. The model shows 82% accuracy for training data but 99% for test data, which shows overfitting of the model.

Fig. 12. Proposed Model's Summary.

TABLE I. RESULTS OF MODEL EVALUATION

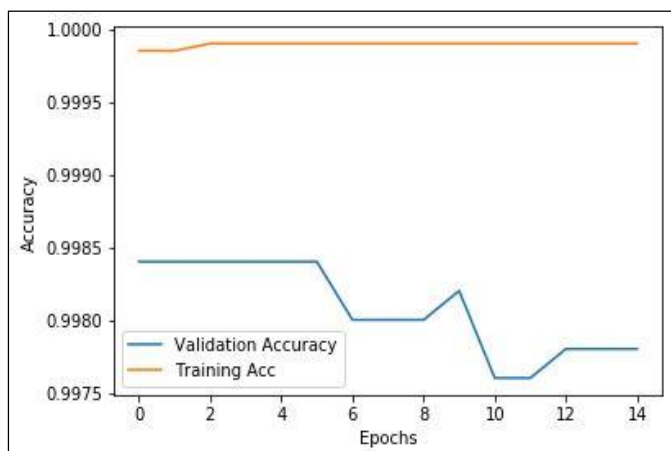| Model Score	1.014088016230464
Model Accuracy (Train Data)	82.11%
Model Accuracy (Test Data)	99.95%

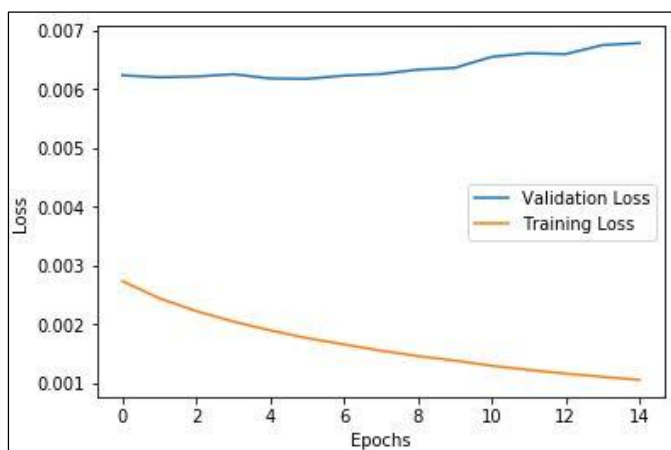Fig. 13. Validation Accuracy of Proposed Model.

Fig. 14. Validation Loss of Proposed Model.

V. CONCLUSION AND FUTURE WORK

The opinion mining of user reviews about a specific product or service has been under the researcher's hammer for quite a long time now. Most of the work revolves around finding sentence-level sentiment analysis, which gives a bigger but not so clear picture of user's opinions. Hence a fine-grained approach that focuses more on the aspect-level sentiment classification is needed. This paper proposes an ensemble aspect extraction methodology based on POS tagging, dependency parsing, and dense neural networks which yields better outputs for sentiment analysis instead of using solo traditional methods. This research will be valuable for the individuals or commercial entities that may employ it for fine graining the existing users' opinion about specific features of their product or service and also targeting the potential customers. On the flip side, the model overfits the Twitter dataset under consideration. The future work of this research will be manifold.

Firstly, some parameters of the model must be modified to avoid overfitting. Secondly, several machine learning and deep learning classifiers can be employed for enhanced sentiment analysis. Finally, the most important prospect will be deploying it for more human languages like French, Spanish, Hindi or Tamil, etc.

REFERENCES

[1] A. Gural, B. B. Cambazoglu, and P. Karagoz, Sentiment Focused Web Crawling, *CIKM'12*, *ACM Transactions on the web*, Maui, HI, USA, 2012.

[2] O. Alqaryouti, N. Siyam, A. A. Monem, and K. Shaalan, Aspect-Based Sentiment Analysis Using Smart Government Review Data, *Applied Computing and Informatics,* 2019.

[3] K. Bafna, and D. Toshniwal, Feature based summarization of customers' reviews of online products, In: *Proc. Comput Sci 22*, 2013, pp. 142–151.

[4] Y. Zhang, and W. Zhu, Extracting implicit features in online customer reviews for opinion mining, In*: Proc. 22nd international conference on World Wide Web companion. International World Wide Web Conferences steering committee,* 2013, pp. 103–104.

[5] R. Kumar, and R. Vadlamani, A Survey on Opinion Mining and Sentiment Analysis:Tasks, Approaches and Applications, *Knowledge Based Systems*, Vol. 89, No. 1, pp. 14-46, 2015.

[6] P. More, and A. Ghotkar, A Study of Different Approaches to Aspect-based Opinion Mining, 2016.

[7] B. Liu, and L. Zhang, A survey of opinion mining and sentiment analysis, *In: Proc. Mining Text Data. Springer US*, 2012, pp. 415–463.

[8] A. M. Popescu, and O. Etzioni, Extracting product features and opinions from reviews, *Natural language processing and text mining. Springer,* pp. 9–28, 2007.

[9] C. Quan, and F. Ren, Unsupervised product feature extraction for feature-oriented opinion determination, *Inf. Sci. 272*, pp. 16–28, 2014.

[10] K. Liu, L. Xu, and J. Zhao, Co-extracting opinion targets and opinion words from online reviews based on the word alignment model, *IEEE Trans Knowl Data Eng.*, Vol. 27, No. 3, pp. 636–650, 2015.

[11] G. Qiu, B. Liu, J. Bu, and C. Chen, Opinion word expansion and target extraction through double propagation, *Comput. Linguist*, Vol. 37, No. 1, pp. 9–27, 2011.

[12] W. Jin, and H. H. Ho, A novel lexicalized hmm-based learning framework for web opinion mining, In: *Proc. 26th annual International Conference on Machine Learning,* 2009.

[13] L. Rabiner, Predicting The Weather with Hidden Markov Models, [Blog], Retrieved from URL *http://guizzetti.ca/blogs/lenny/2012/04/ predicting-the-weather-with-hidden-markov-models*, 2012.

[14] L. Chen, L. Qi, and F. Wang, Comparison of feature-level learning methods for mining online consumer reviews. *Expert Syst Appl.*, Vol. 39, No. 10, pp. 9588–9601, 2012.

[15] B. Ma, D. Zhang, Z. Yan, and T. Kim, An LDA and synonym lexicon based approach to product feature extraction from online consumer product reviews, *J Electron Commer Res*,. Vol. 14, No. 4, pp. 304–314, 2013.

[16] F. Li, M. Huang, and X. Zhu, Sentiment analysis with global topics and local dependency, In: *Proc. Twenty-Fourth AAAI Conference on Artificial Intelligence,* 2010.

[17] B. Liu, and L. Zhang, "A survey of opinion mining and sentiment analysis", *Mining text data. Springer, Boston, MA*, pp. 415-463, 2012.

[18] N. Zainuddin, A. Selamat, and R. Ibrahim, Hybrid sentiment classification on twitter aspect-based sentiment analysis, *Applied Intelligence*, Vol. 48, No. 5, pp. 1218-1232. 2018.

[19] Y. Jin, Development of word cloud generator software based on python, *Procedia engineering*, Vol. 174, No. 1, pp. 788-792, 2017.