INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

TWO LEVEL LSTM FOR SENTIMENT ANALYSIS WITH LEXICON EMBEDDING AND POLAR FLIPPING

[1]B. RAJANI, [2]MUDUPU YASHASWINI REDDY, [3]BELLY SAI, [4]CHILIVERI SHRAVANI

[1]Assistant Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

rajani.bashaboina@gmail.com

[2,3,4]BTech Student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad

mudupuyashaswini@gmail.com, bellisanju05@gmail.com, shravanichiliveri3@gmail.com

Abstract : *Sentiment analysis is a key component in various text mining applications. Numerous sentiment classification techniques, including conventional and deep-learning based methods, have been proposed in the literature. In most existing methods, a high-quality training set is assumed to be given. Nevertheless, constructing a high-quality training set that consists of highly accurate labels is challenging in real applications. This difficulty stems from the fact that text samples usually contain complex sentiment representations, and their annotations subjective. We address this challenge in this project by leveraging a new labeling strategy and utilizing a two-level long short-term memory network to construct a sentiment classifier.*

KEYWORDS: *Sentiment analysis, Labeling, Training, Encoding, Dictionaries, Logic gates, neural networks*

I. INTRODUCTION

Text is important in many artificial intelligence applications. Among various text mining techniques, sentiment analysis is a key component in applications, such as public opinion monitoring and comparative analysis. Sentiment analysis can be divided into three problems according to input texts, namely, sentence, paragraph, and document levels. This study focuses on sentence and paragraph levels Text sentiment analysis is usually considered a text classification problem. Almost all existing text classification techniques are applied to text sentiment analysis. Typical techniques include bag-of-words (BOW)-based, topic model-based, deep learningbased, and lexicon based (or rule-based) methods.

ISSN: 2057-5688

Although many achievements have been made and sentiment analysis has been successfully used in various commercial applications, its accuracy can be further improved. The construction of a high-accuracy sentiment classification model usually entails the challenging compilation of training sets with numerous samples and sufficiently accurate labels. The reason behind this difficulty is two-fold. First, the sentiment is somewhat subjective, and a sample may receive different labels from different users. Second, some texts contain complex sentiment representations, and a single label is difficult to provide. We conduct a statistical analysis of public Chinese sentiment text sets in GitHub. The results show that the average label error is larger than 10%. This error value reflects the degree of difficulty of sentiment labeling.

Negation and interrogative sentences are difficult to classify when deep-learningbased methods are applied. Although lexicon-based methods can deal with particular types of negation sentences, their generalization capability is poor. We address the above issues with a new methodology. First, we introduce a two-stage labeling strategy for sentiment texts. In the first stage, annotators are invited to label a large number of short texts with relatively pure sentiment orientations. Each sample is labeled by only one annotator. In the second stage, a relatively small number of text samples with mixed sentiment orientations are annotated, and each sample is labeled by multiple annotators. Second, we propose a two-level long short-term memory (LSTM) [6] network to achieve two-level feature representation and classify the sentiment orientations of a text sample to utilize two labeled datasets. Third, in the proposed twolevel LSTM network, lexicon embedding is leveraged to incorporate linguistic features used in lexicon based methods. Finally, the labels in a word-polar dictionary usually contain noise and the polarity of a word can also change in different contexts. A flipping model is proposed to model the sentiment polarity flipping of a word in a sentence. Three Chinese sentiment datasets are compiled to investigate the performance of the proposed methodology. The experimental results demonstrate the effectiveness of the proposed methods.

The motivation behind using a two-level LSTM for sentiment analysis lies in the

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

ability of this architecture to capture both local and global dependencies within the text, leading to improved sentiment understanding and prediction. Sentiment analysis often requires understanding sentiments at different levels of granularity, such as 3 sentiment expressed by individual words or phrases, as well as the overall sentiment of the entire text. A two-level LSTM provides a hierarchical representation of the input text, where the lower-level LSTM focuses on capturing local sentiments, while the higher-level LSTM integrates these local sentiments to infer the overall sentiment. This hierarchical structure allows the model to learn and reason about sentiment at multiple levels simultaneously. The motivation behind using a two-levelLSTM for sentiment analysis is to leverage its hierarchical structure, long-term dependency modeling capabilities, and feature extraction power to capture both local and global sentiment dependencies within the text. By doing so, the model can achieve a more nuanced understanding of sentiment and make more accurate predictions in sentiment analysis tasks.

II. LITERATURE SURVEY

Microblogging today has become a very popular communication tool among Internet users. Millions of users share opinions on different aspects of life everyday. Therefore microblogging web-sites are rich sources of data for opinion mining and sentiment analysis. Because microblogging has appeared relatively recently, there are a few research works that were devoted to this topic. In our paper, we focus on using Twitter, the most popular microblogging platform, for the task of sentiment analysis. We show how to automatically collect a corpus for sentiment analysis and opinion mining purposes. We perform linguistic analysis of the collected corpus and explain discovered phenomena. Using the corpus, we build a sentiment classifier, that is able to determine positive, negative and neutral sentiments for a document. Experimental evaluations show that our proposed techniques are efficient and performs better than previously proposed methods. In our research, we worked with English, however, the proposed technique can be used with any other language. Probabilistic topic models have been widely used for sentiment analysis. However, most of existing topic methods only model the sentiment text, but

ISSN: 2057-5688

6 do not consider the user, who expresses the sentiment, and the item, which the sentiment is expressed on. Since different users may use different sentiment expressions for different items, we argue that it is better to incorporate the user and item information into the topic model for sentiment analysis. In this paper, we propose a new Supervised User-Item based Topic model, called SUIT model, for sentiment analysis. It can simultaneously utilize the textual topic and latent user-item factors. Our proposed method uses the tensor outer product of text topic proportion vector, user latent factor and item latent factor to model the sentiment label generalization. Extensive experiments are conducted on two datasets: review dataset and microblog dataset. The results demonstrate the advantages of our model. It shows significant improvement compared with supervised topic models and collaborative filtering methods. We report on a series of experiments with convolutional neural networks (CNN) trained on top of pre-trained word vectors for sentence-level classification tasks. We show that a simple CNN with little hyperparameter tuning and static vectors achieves excellent results on multiple benchmarks. Learning taskspecific vectors through fine-tuning offers further gains in performance. We additionally propose a simple modification to the architecture to allow for the use of both task-specific and static vectors. The CNN models discussed herein improve upon the state of the art on 4 out of 7 tasks, which include sentiment analysis and question classification. We present a lexicon-based approach to extracting sentiment from text. The Semantic Orientation calculator (SO-CAL) uses dictionaries of words annotated with their semantic orientation (polarity and strength), and incorporates intensification and negation. SO-CAL is applied to the polarity classification task, the process of assigning a positive or negative label to a text that captures the text's opinion towards its main subject matter. We show that SO-CAL's performance is consistent across domains and in completely unseen data. Additionally, we describe the process of dictionary creation, and our use of Mechanical Turk to check dictionaries for consistency and reliability. Sentiment lexicons have been leveraged as a useful source of features for sentiment analysis models, leading to the state-of-the-art

ISSN: 2057-5688

accuracies. On the other hand, most existing methods use sentiment lexicons without considering context, typically taking the count, sum of strength, or maximum sentiment scores over the whole input. We propose a context-sensitive lexicon-based method based on a simple weighted-sum model, using a recurrent neural network to learn the sentiments strength, intensification and negation of lexicon sentiments in composing the sentiment value of sentences. Results show that our model can not only learn such operation details, but also give significant improvements over state-of-the-art recurrent neural network baselines without lexical features, achieving the best results on a Twitter benchmark.

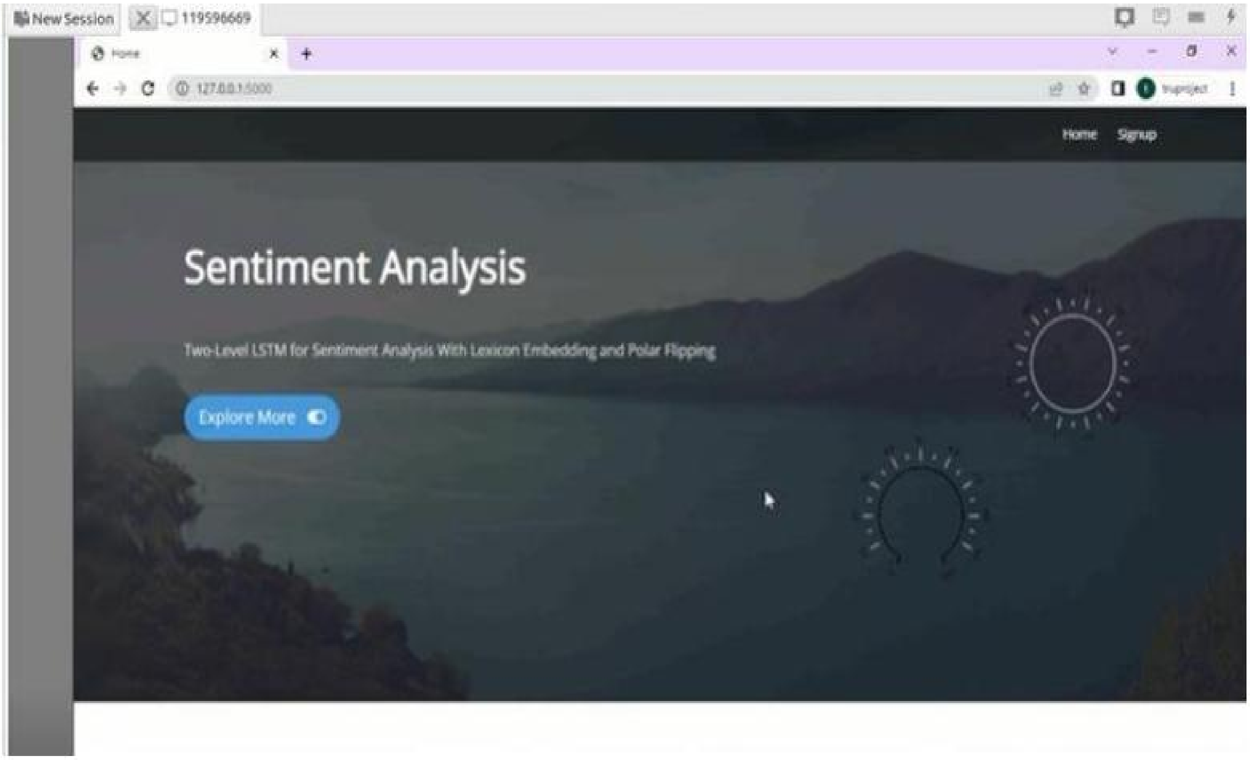
III. PROPOSED SYSTEM

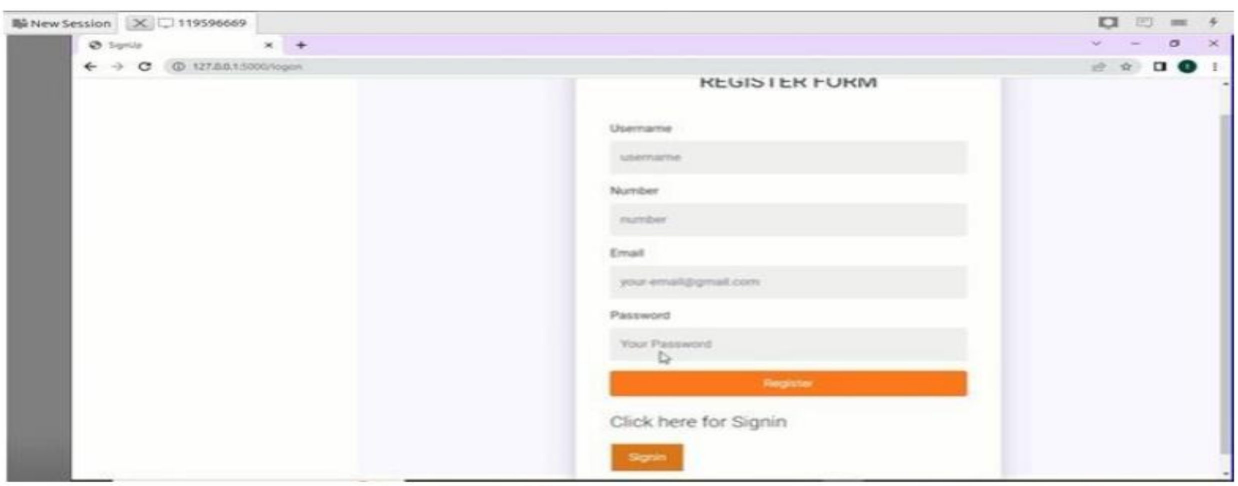
We address the above issues with a new methodology. First, we introduce a two-stage labeling strategy for sentiment texts. In the first stage, annotators are invited to label a large number of short texts with relatively pure sentiment orientations. Each sample is labeled by only one annotator. In the second stage, a relatively small number of text samples with mixed sentiment orientations are annotated, and each sample is labeled by multiple annotators. Second, we propose a two-level long short- term memory (LSTM) network to achieve two-level feature representation and classify the sentiment orientations of a text sample to utilize two labeled datasets. Third, in the proposed twolevel LSTM network, lexicon embedding is leveraged to incorporate linguistic features used in lexiconbased methods. Finally, the labels in a word-polar dictionary usually contain noise and the polarity of a word can also change in different contexts. A flipping model is proposed to model the sentiment polarity flipping of a word in a sentence

IV. RESULTS

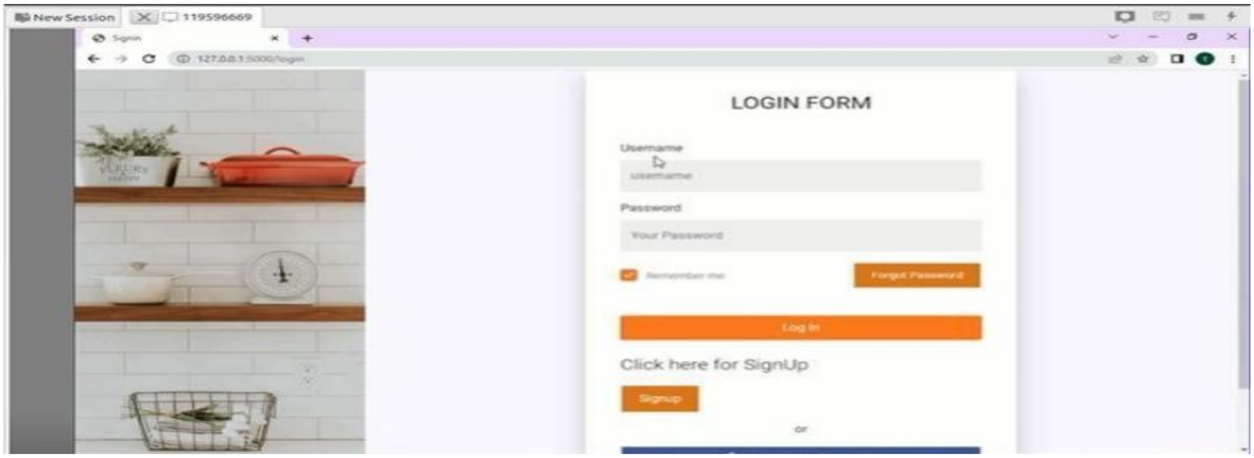
INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

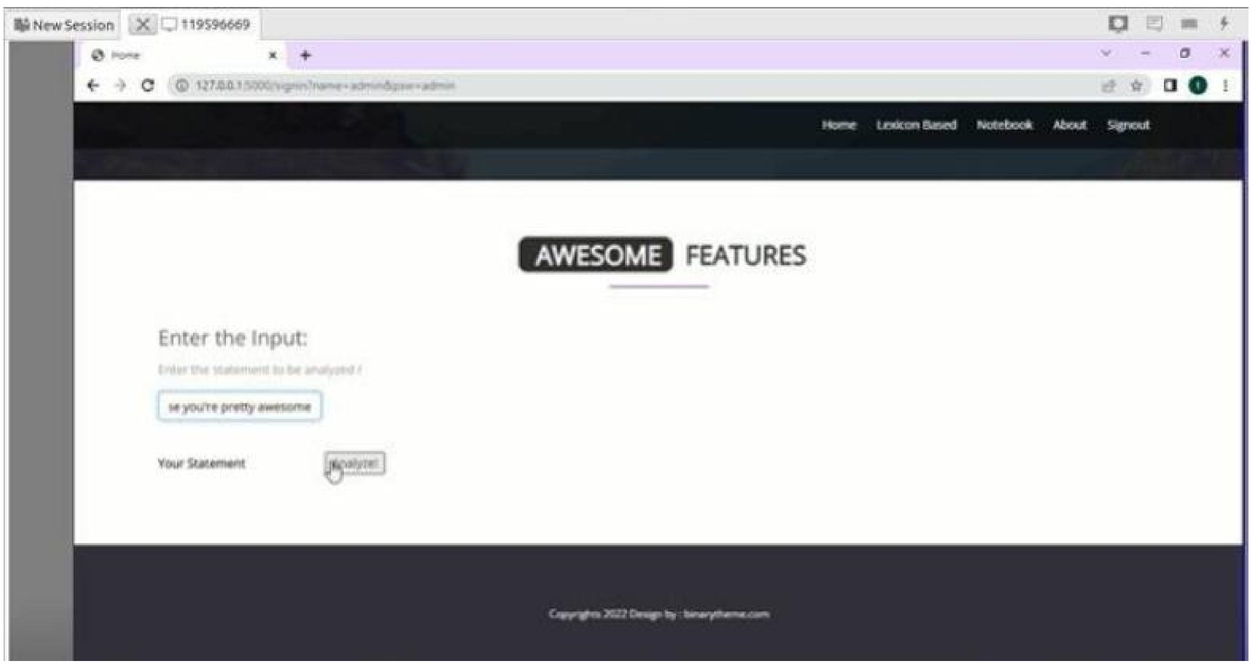
ISSN: 2057-5688



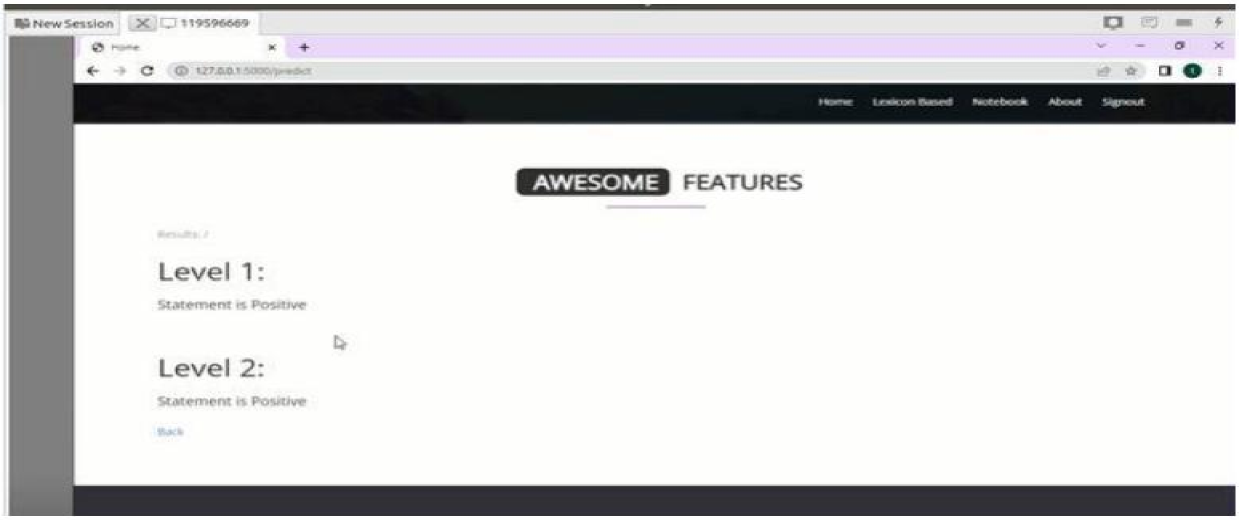


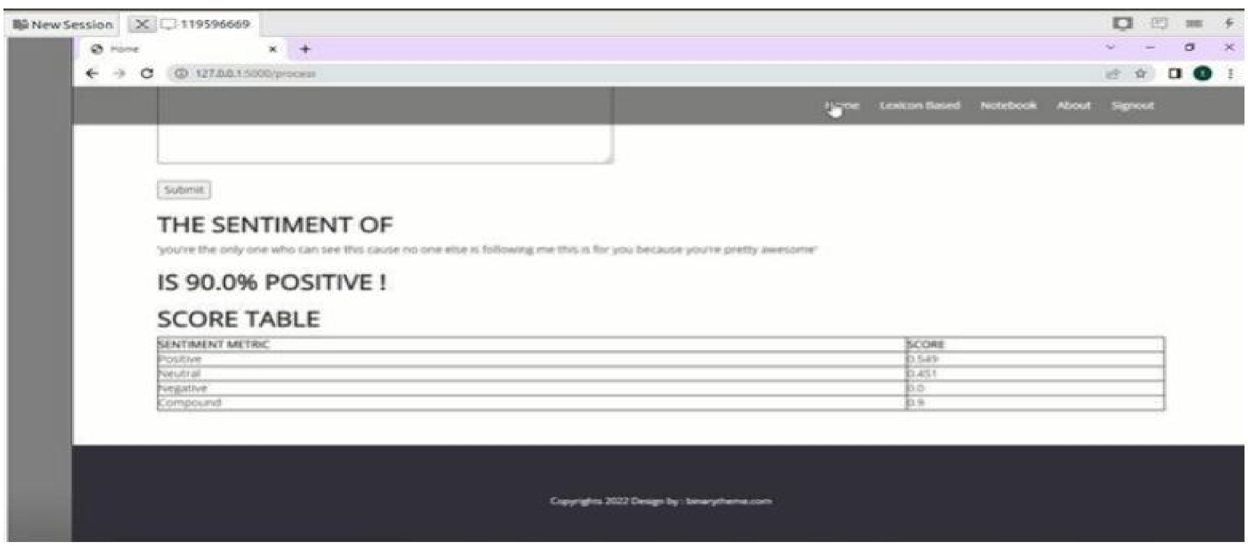
ISSN: 2057-5688





INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688





V. CONCLUSION

High-quality labels are crucial for learning systems. Nevertheless, texts with mixed sentiments are difficult for humans to label in text sentiment classification. In this study, a new labeling strategy was introduced to partition texts into those with pure and mixed sentiment orientations. These two categories of texts were labeled using different processes. A two-level network was accordingly proposed to utilize the two labeled data in our two-stage labeling strategy. Lexical cues (e.g., polar words, POS, and conjunction words) are particularly useful in sentiment analysis. These lexical cues were used in our two-

ISSN: 2057-5688

level network, and a new encoding strategy, that is, ρ-hot encoding, was introduced. P-hot encoding was motivated by one-hot encoding. However, the former alleviates the drawbacks of the latter. Due to labeling noise or context, the polarity of a word varied in different texts. A flipping model was proposed to model the polarity flipping process. Three Chinese sentiment text data corpora were compiled to verify the effectiveness of the proposed methodology

REFERENCES

[1] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[2] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proc. Int. Conf. Lang. Resources Eval. (LREC), 2010, pp. 17–23.

[3] F. Li, S. Wang, S. Liu, and M. Zhang, "SUIT: A supervised user-item based topic model for sentiment analysis," in Proc. AAAI Conf. (AAAI), 2014, pp. 1636–1642.

[4] Y. Kim, "Convolutional neural networks for sentence classification," in Proc. Int. Conf. Empirical Methods Nat. Lang. (EMNLP), 2014, pp. 1746–1751.

[5] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexiconbased methods for sentiment analysis," Comput. Linguist., vol. 37, no. 2, pp. 267–307, 2011.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[7] Z. Teng, D.-T. Vo, and Y. Zhang, "Context-sensitive lexicon features for neural sentiment analysis," in Proc. EMNLP, 2016, pp. 1629–1638.

[8] Q. Qian, M. Huang, J. Lei, and X. Zhu, "Linguistically regularized LSTMs for sentiment classification," in Proc. Annu. Meeting Assoc. Comput. Linguist. (ACL), 2017, pp. 1679–1689.

[9] W. Zhao et al., "Weakly-supervised deep embedding for product review sentiment analysis," IEEE Trans. Knowl. Data Eng., vol. 30, no. 1, pp. 185–197, Sep.2018. 40

[10] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD), 2004, pp. 168–177.

ISSN: 2057-5688

[11] T. Mullen and N. Collier, “Sentiment analysis using support vector machines with diverse information sources,” in Proc. Int. Conf. Empirical Methods Nat. Lang. (EMNLP), 2004, pp. 412–418.

[12] S. Liu, X. Cheng, F. Li, and F. Li, “TASC: Topic-adaptive sentiment classification on dynamic tweets,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 6, pp. 1696–1709, Dec. 2015.

[13] N. Srivani, Dr Prasadu Peddi, “Efficient Fr a Geometrical-Model-Based Face Segmentation and Identification in Terms of Identification the Face ”, *JFCR*, pp. 1283-1295, Jun. 2022.