International Journal of Innovative Computing XX(nn) pp-pp

INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Sentiment Analysis and Rating Prediction for an E-commerce Platforms in Malaysia Using CNN and LSTM

Abubakar Sadiq Muhammad
Faculty of Computing
Universiti Teknologi Malaysia,
81310 Skudai Johor, Malaysia,
Sadiq-20@graduate.utm.my

Assoc. Prof. Dr. Mohd Shahizan bin Othman
Faculty of Computing
Universiti Teknologi Malaysia,
81310 Skudai Johor, Malaysia,
Shahizan@utm.my

***Abstract*—This research explores sentiment analysis and rating prediction for Malaysian e-commerce platforms Shopee, Lazada, and Zalora. Data was collected from the Google Play Store and Apple Store using scraping APIs. The project, executed in three phases, involved data collection, preprocessing, and model implementation. Preprocessing included text cleaning, translation using the Malaya Python library, and handling of Manglish and abbreviations. Sentiment labels were generated using VADER, Sentiment Analyzer, and TextBlob, and data was balanced using SMOTE. In the second phase, word embedding techniques such as GloVe and Word2Vec were applied, and the data was used to train CNN and LSTM models. The CNN models, particularly with Word2Vec, achieved higher accuracy (94.47%) compared to CNN models with GloVe (93.39%). The LSTM models exhibited lower performance metrics with 74% accuracy. The final phase developed a comprehensive sentiment analysis dashboard to monitor public opinions and satisfaction levels. This study highlights the comparative effectiveness of CNN and LSTM models for sentiment analysis and rating prediction, providing valuable insights for understanding customer feedback and enhancing service quality in the Malaysian e-commerce landscape.**

Keywords — Sentiment Analysis, SMOTE, Malaya library, Convolutional Neural Network, Long Short-term Memory.

I. INTRODUCTION

In Natural Language Processing (NLP), sentiment analysis is used to determine the emotional tone from a given text, applicable across various domains such as movie and product reviews [3]. This technique helps understand customer opinions and preferences, though it faces challenges due to the variability and complexity of natural language, especially in long texts with multiple sentiments [16]

. To address this, deep learning techniques like recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) have been proposed, which can capture sequential and contextual information from text and learn semantic features, showing promising results in sentiment analysis applications (Alzahrani et al., 2022; Kurniasari & Setyanto, 2020).

However, applying these techniques to e-commerce platforms, where text reviews are often accompanied by numerical ratings, remains a gap. Ratings can provide additional sentimental information but are frequently inconsistent or noisy. This project aims to develop a sentiment analysis and rating prediction model for e-commerce platforms using RNNs and LSTMs. The model, trained and evaluated on a large dataset of reviews and ratings from the Google Play Store and Apple Store, will classify reviews into positive, neutral, or negative sentiments and generate rating scores from 1 to 5 based on the text. Performance will be measured using metrics such as accuracy, precision, recall, and F1-score, while also exploring the impact of various hyperparameters on the model's performance. This approach aims to enhance customer satisfaction, product quality, and recommendation systems for e-commerce platforms.

II. RELATED WORK

In the domain of sentiment analysis, numerous studies have focused on enhancing the accuracy of sentiment models, but questions remain about their effectiveness across various domains and contexts [9]. One significant challenge is developing models that can accurately predict ratings based on review text [7] [8]. While deeper, more complex models are often pursued, they risk overfitting and may struggle to generalize effectively [11] Several studies have explored hybrid models and machine learning algorithms, like Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, and Decision Tree, to address neutral sentiments and improve sentiment analysis accuracy [3] [13]. [13] found that SVM and Logistic Regression outperformed other algorithms with 80% accuracy. However, these methods often require manual feature selection, which can be a limitation [7].

Deep learning approaches, such as those integrating Convolutional Neural Networks (CNN) and attention-based Bidirectional Gated Recurrent Units (BiGRU), have shown promise by automatically extracting features [15]. Another research by [18] proposed the SLCABG model, combining sentiment lexicon and deep learning techniques, to overcome drawbacks in sentiment analysis of product reviews. Additionally, [17] demonstrated that a hybrid SVM-CNN model outperformed other machine learning algorithms in sentiment analysis of e-commerce data. [13] employed the VADER sentiment analysis algorithm, achieving a 15% improvement in accuracy. However, VADER's performance can decline with increasing data volume due to its reliance on rule-based analysis. These studies underscore the potential of hybrid and deep learning models in enhancing sentiment analysis, especially when leveraging both text and ratings for more comprehensive insights.

III. METHODOLOGY

In Fig. 1, the research process flow is summarized. The framework gives an approach to overcoming the difficulties in doing sentiment analysis and rating prediction. Beginning with problem creation, data collection, data preparation, modeling, performance evaluation, comparison and analysis, model generalization: evaluation, and data visualization, there are six steps that make up the operational framework.

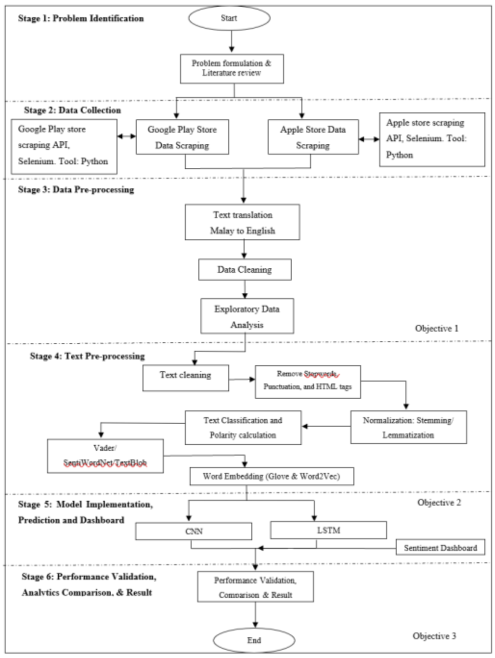

Fig 1 Project Framework

A. Problem Formulation

Here, the problem formulation and the literature review achieved as described in previous chapter aimed to explain the initial study before carrying out the research. The comprehension of the application of the sentiment analysis techniques on customer satisfaction, tools and the deep learning techniques selected to be executed in this project were discussed.

B. Data Collection

Google Play Store and Apple Store scraping involve extracting app data, including details, reviews, and ratings, for market research, app analysis, competitor monitoring, and sentiment mining. Various methods such as web scraping tools, custom scripts using libraries (like Python, Node.js, or Ruby), and APIs can be used for this purpose. Data from Shopee, Lazada, and Zolara app reviews were scraped from both stores. The geographical focus was on Malaysia, and the analysis covered the period from 2013 to 2023 to understand the market trends of these platforms.

C. Text preprocessing

The datasets collected in stage went through cleaning steps to be ready and used for deep learning. Data preparation is the most essential process in understanding data. The quality of the result depends on how the data is processed. To analyse any data, language must be translated to English. Then, for data cleaning, the steps involved filtering the data by removing and missing values, special characters, duplicates, and punctuation in review data. standardizing the reviews by removing stop words using lemmatization and converting it to lower case

D. Model Implementation

This phase used the pre-processed and clean data from stage four to implement Convolutional Neural Network and Long Short-Term Memory algorithms to perform sentiment analysis and predict user rating.

1. Tokenization

This procedure divides paragraph, sentence, phrase, or complete file into smaller parts, such as separated words [5]. Each of these units is referred to as tokens. For sentiment analysis tokenization is important because the importance of the sentence can be easily determined by examining the text in the words [1]. The nltk.tokenize library word tokenizer was employed to split a string at each space using a specified separator [8]. Stop words library from NLTK will be use in the English language in conjunction with tokenization to eliminate unnecessary words.

2. Lemmatization/Stemming

In this research, we utilized NLTK library called WordNet. Wordnet is a vocabulary database encompassing numerous of languages, facilitating the provision of semantic relationships among words [10]. This procedure considered the entire vocabulary of a language and conduct a morphological analysis on texts [10]. This analysis aims to return the word list form of text by eliminating the endings.

3. Initial Finding

The initial phase of data analysis is data exploration. During this phase, users scrutinize large datasets in an unstructured way to uncover initial trends, characteristics, and interesting points [6]. This process does not expose all the intricacies in the dataset, but it simplifies the understanding of significant trends and crucial points for subsequent research. Data exploration can utilize a combination of manual techniques and automated tools like data visualization, charts, and preliminary reports [4].

In any dataset, it's crucial to examine the nature of the data, such as whether it's qualitative or quantitative, and the scale of data measurement, such as nominal, ordinal, interval, and ratio. Depending on these factors, various statistical methods like descriptive statistics can be employed to scrutinize the dataset based on specific requirements and conditions. Brief descriptive coefficients are used to characterize a given dataset, which could represent a population or a sample of the population. Descriptive statistics use measures of central tendency and measures of variability. Measures of central tendency include the mean, median, and mode, while measures of variability include the standard deviation, variance, minimum and maximum values, as well as kurtosis and skewness (Agarap, 2020).

In this project, both categorical and numerical data are analyzed using descriptive statistics. The analysis generates pie charts, bar charts, histograms, and box plots to examine and compare the trends and relationships among each variable in the categorical and numerical data. To comprehend the meaning of the data, it's essential that it is presented in an easily understandable manner. Descriptive statistics can be used to present data in a more insightful way, facilitating a straightforward understanding of the results.

4. Polarity Classification

Three modules from the NLTK library, namely Vader (Valence Aware Dictionary), TextBlob, and SentiWordNet, will be employed to determine the polarity of unlabeled text data [12]. Vader sentiment analysis, which is based on lexicon and rules, accepts a complete sentence string and returns word list scores that range from -1 (negative) to 1 (positive), including a compound score [12].

TextBlob, another lexicon-based method, returns a polarity like Vader, ranging from -1 to 1, and a subjectivity score for a sentence, ranges from 0 to 1. The subjectivity in TextBlob measures the proportion of information and opinion in a text, with a score indicating a greater reliance on opinion over information.

SentiWordNet, derived from the WordNet database, is another lexicon-based approach where each opinion term is associated with numerical score (positivity, negativity, and neutral). SentiWordNet employs a random walk algorithm, and each score value ranges from 0 to 1, with the sum for each synset being 1.

To ensure that all three polarities have standard normal distribution properties before vectorization, all three polarity classifiers will be normalized using Min-Max parameter scaling from the sklearn library.

5. Vectorization

According to [10] vectorization refers to a procedure that transforms sentences into a numerical format suitable for machine learning or deep learning applications. Vectorization will be implemented on pre-processed, translated comments before they are input into the machine learning model. This is because vectorization enhances the speed of the machine learning algorithm and significantly reduces training time.

In this project, the sklearn library called Glove and Word2Vec will be utilized to turn the collection of processed text into a matrix of token.

6. Feature Engineering

Feature engineering involves creating and transforming features to improve model performance. In this study, word embeddings namely Word2Vec and GloVe are generated as features for the models. These embeddings capture the semantic relationships between words, allowing the models to understand the context and sentiment better. By experimenting with these advanced embedding techniques and evaluating their impact on the model performance, the most effective embeddings were selected for the final model training.

In the exploration of sentiment analysis and rating prediction, two distinct deep learning architectures were employed. Instead of traditional machine learning approaches like Naïve Bayes or SVM, the focus was on utilizing CNN, which processes sequential data, and LSTM, a specialized RNN variant capable of retaining long-term dependencies in data sequences. These models were trained and optimized to understand the sentiment expressed in text data and predict ratings more accurately, harnessing the power of deep learning and CNN ability to capture sequential patterns within the E-commerce context.

E. Performance Validation and Conclusion

In this part, following the data classification process, an assessment utilizing a confusion matrix is conducted to compare the outcomes achieved using Deep Learning models such as the CNN and LSTM. Table 1 illustrates the configuration of the confusion matrix, which is instrumental in evaluating the performance of these models. Accuracy, Precision, Recall, and F1-score are computed metrics based on the confusion matrix to determine the efficacy of the applied Deep Learning algorithms. This matrix serves as a concise representation of the model performance, offering insights into its accuracy concerning known true values within the test dataset. Additionally, the metrics - accuracy, precision, recall, and F1-Score - are computed from this confusion matrix to evaluate the performance of the CNN and LSTM models utilized in this experiment.

TABLE I. CONFUSION MATRIX

	Predicted	
Original	TrueNegatice	FalsePositive
	FalseNegative	TruePositive

To evaluate the performance of the algorithms, accuracy, precision, recall, and F1- score are measured. Below is the formula to measure the performance:

$$Accuracy = \frac{(TN + TP)}{(TN + TP + FN + FP)} \tag{1}$$

Precision measures the accuracy of a classification algorithm by calculating the ratio of correctly predicted positive reviews to the total number of reviews predicted as positive [16]. It quantifies the proportion of true positive reviews among all positive predictions. A higher precision value indicates fewer false positives, while a lower value suggests more false positives.

$$\text{Precision} = \frac{TruePositive}{TruePositive + FalsePositvie} \tag{2}$$

Recall, or sensitivity, is the ratio of correctly predicted positive reviews to the total number of positive reviews in the dataset. Higher recall indicates fewer false negatives, while lower recall indicates more false negatives [2].

$$\text{Recall} = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

The F1 score, or F-measure, is a composite metric that balances precision and recall. It ranges from 0 to 1, with 1 indicating perfect precision and recall. It provides a single metric to evaluate the effectiveness of a classifier, especially when there is a trade-off between precision and recall (Başarslan & Kayaalp, 2023).

$$\text{F1-score} = \frac{2\, x\, (precision)\, x\, (recall)}{precision + recall} \tag{4}$$

IV. Result and Discussions

A. Convolutional Neural Network Model

The figure 5.6 showcases two-line graphs that are evaluating the performance of our model over a series of epochs. The first graph, titled "Accuracy Over Epochs," plots the training and validation accuracy, both of which exhibit an upward trend, indicating an improvement in the model ability to correctly predict outcomes as training progresses. The training accuracy slightly surpasses the validation accuracy, suggesting that the model may be performing marginally better on the training data compared to the unseen validation data.

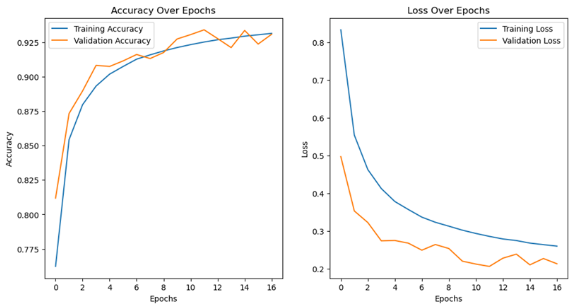

Fig 2 Accuracy and Epoch of CNN

B. Long Short-term Memory Model

The figure 3 above shows a line graphs that depict the performance of a Long Short-Term Memory (LSTM) neural network model over a series of training epochs. The first graph, "Accuracy Over Epochs (LSTM)," indicates an upward trend in both training and validation accuracy, indicating that the model is learning effectively. The training accuracy starts at approximately 0.66 and ends near 0.74, while the validation accuracy begins around 0.68 and ends just below 0.72. This suggests that the model is becoming better at making correct predictions as it processes more data.

The second graph, "Loss Over Epochs (LSTM)," presents a downward trend in both training and validation loss, which is a positive sign of the model's improving prediction capability. The training loss starts just below 1.1 and ends around 0.2, whereas the validation loss begins slightly above 1.0 but decreases more gradually, ending around 0.6.

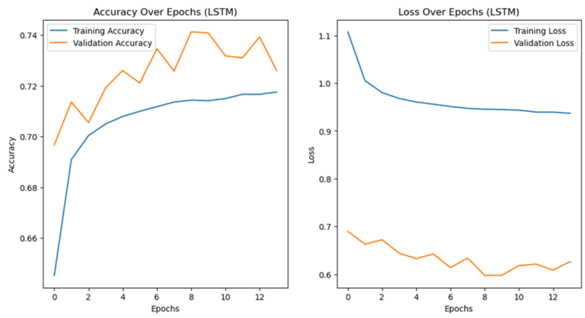

Fig 3 Accuracy and Epoch of CNN

The table compares the performance metrics of two models, CNN and LSTM, based on precision, recall, and F1-score across three classes (0, 1, and 2). The CNN model demonstrates superior performance across all metrics compared to the LSTM model. Specifically, CNN achieves higher precision, recall, and F1-scores for classes 0 and 1, indicating its ability to accurately classify instances across these classes. For class 2, LSTM shows higher recall (0.94) compared to CNN (0.99), with lower precision and F1-score. Overall, CNN achieves an accuracy of 0.93, outperforming LSTM which achieves an accuracy of 0.74. This comparison highlights CNN's stronger performance in this classification task, especially in handling class imbalances and achieving overall better accuracy and precision-recall balance across all classes [2].

TABLE II. MODEL COMPARISON

Model	Precision (0)	Recall (0)	F1-score (0)	Precision (1)	Recall (1)	F1-score (1)	Precision (2)	Recall (2)	F1-score (2)	Accuracy
CNN	0.96	0.88	0.92	0.93	0.93	0.93	0.92	0.99	0.95	0.93
LSTM	0.93	0.82	0.87	0.94	0.46	0.62	0.58	0.94	0.72	0.74

V. CONCLUSION

This research project is focused on employing Convolutioanl Neural Networks (CNN) and Long Short-Term Memory (LSTM) models to analyse sentiment in datasets related to three prominent online shopping platforms: Shopee, Lazada, and Zolara.

The initial datasets were gathered from these platforms and involved reviews and feedback from users. Preprocessing steps, including cleaning, tokenization, and lemmatization, were carried out using Python to enhance the quality of the data. The goal was to create a refined dataset that eliminates noise and ensures consistent input for the sentiment analysis models.

Subsequently, the sentiment analysis models, particularly CNN and LSTM, were implemented using the tokenized and pre-processed data. The models were trained on a subset of the data and tested on another subset, employing a train-test split of 80:20. The aim is to evaluate the models' effectiveness in discerning sentiments expressed in reviews on the respective e-commerce platforms.

Throughout the analysis, performance metrics such as accuracy, precision, recall, and F1 score were calculated to quantify the model success in predicting sentiment. With loss value of 0.035, Accuracy 93%, Val_loss of 0.0739, and val_accuracy of 93%. Graphical representations, including precision-recall curves, ROC curves, and confusion matrices, were generated to provide a visual understanding of the models' performance.

The goal of this research is to gain insights into the sentiment trends across different e-commerce platforms and to develop effective sentiment analysis models that can be applied to similar datasets in the future and predict the review rating.

ACKNOWLEDGMENT

This research paper is not funded. I would like to extend my gratitude to my supervisor for his guidance and mentorship that made this project possible.

REFERENCES

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example: [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in "[3]"—do not use "Ref. [3]" or "reference [3]". Do not use reference citations as nouns of a sentence (e.g., not: "as the writer explains in [1]").

Unless there are six authors or more give all authors' names and do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] Agarap, A. F. (2020). Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network (RNN) (arXiv:1805.03687). arXiv. http://arxiv.org/abs/1805.03687

[2] Balita, L. B. S., Degrano, K. M. A., Pamoso, A. D. A., & De Goma, J. C. (2023). Sentiment Analysis on Book Reviews Using Convolutional Neural Network (CNN) Long Short-Term Memory (LSTM) Hybrid. Proceedings of the 2022 6th International Conference on E-Business and Internet, 125–132. https://doi.org/10.1145/3572647.3572666

[3] Bhowmik, N. R., Arifuzzaman, M., Mondal, M. R. H., & Islam, M. S. (2021). Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary. Natural Language Processing Research, 1(3–4), 34–45. https://doi.org/10.2991/nlpr.d.210316.001

[4] Bu, J., Ren, L., Zheng, S., Yang, Y., Wang, J., Zhang, F., & Wu, W. (2021). ASAP: A Chinese Review Dataset Towards Aspect Category Sentiment Analysis and Rating Prediction (arXiv:2103.06605). arXiv. http://arxiv.org/abs/2103.06605

[5] Chamekh, A., Mahfoudh, M., & Forestier, G. (2022). Sentiment Analysis Based on Deep Learning in E-Commerce. In G. Memmi, B. Yang, L. Kong, T. Zhang, & M. Qiu (Eds.), Knowledge Science, Engineering and Management (Vol. 13369, pp. 498–507). Springer International Publishing. https://doi.org/10.1007/978-3-031-10986-7_40

[6] Cuizon, J. C., & Agravante, C. G. (2020). Sentiment Analysis for Review Rating Prediction in a Travel Journal. Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval, 70–74. https://doi.org/10.1145/3443279.3443282

[7] D'Souza, P., Naik, A. V., & Pai, B. (2023). Analysis Of Customer Sentiment Through Computational Study Of Product Reviews Using Machine Learning Approach. https://www.semanticscholar.org/paper/Analysis-Of-Customer-Sentiment-Through-Study-Of-D%E2%80%99Souza-Naik/8dd4a6f012e4f76b8bc1d8fdd5b6b80d6c172ce6

[8] Garg, S. B., & Subrahmanyam, V. V. (2022). Sentiment Analysis: Choosing the Right Word Embedding for Deep Learning Model. In M. Bianchini, V. Piuri, S. Das, & R. N. Shaw (Eds.), Advanced Computing and Intelligent Technologies (pp. 417–428). Springer. https://doi.org/10.1007/978-981-16-2164-2_33

[9] Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. IEEE Access, 11, 90367–90382. https://doi.org/10.1109/ACCESS.2023.3307308

[10] Muhammad, P. F., Kusumaningrum, R., & Wibowo, A. (2021). Sentiment Analysis Using Word2vec And Long Short-Term

Memory (LSTM) For Indonesian Hotel Reviews. Procedia Computer Science, 179, 728–735. https://doi.org/10.1016/j.procs.2021.01.061

[11] Ph.D, T. N., & Amalanathan, A. (2020). APPLICATIONS, APPROACHES, AND CHALLENGES IN SENTIMENT ANALYSIS (AACSA).

[12] Rabbimov, I., Kobilov, S., & Mporas, I. (2021). Opinion Classification via Word and Emoji Embedding Models with LSTM. Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings, 589–601. https://doi.org/10.1007/978-3-030-87802-3_53

[13] Raj P M, K., & Sai D, J. (2022). Sentiment analysis, opinion mining and topic modelling of epics and novels using machine learning techniques. Materials Today: Proceedings, 51, 576–584. https://doi.org/10.1016/j.matpr.2021.06.001

[14] Rajeswari, A. M., Mahalakshmi, M., Nithyashree, R., & Nalini, G. (2020). Sentiment Analysis for Predicting Customer Reviews using a Hybrid Approach. 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 200–205. https://doi.org/10.1109/ACCTHPA49271.2020.9213236

[15] Shrivash, B. K., Verma, D. K., & Pandey, P. (2023). An Effective Framework for Sentiment Analysis Using RNN and LSTM-Based Deep Learning Approaches. In M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, & T. Ören (Eds.), Advances in Computing and Data Sciences (pp. 340–350). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37940-6_28

[16] Sree, V. H., Hymavathi, V., Sathwika, A. V., & Rajarajeswari, P. (2023). Implementation of Text-Based Sentiment Analysis Using LSTM Model. 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), 1–5. https://doi.org/10.1109/ICCPCT58313.2023.10244988

[17] Suganya, E., & Vijayarani, S. (2020). Sentiment Analysis for Scraping of Product Reviews from Multiple Web Pages Using Machine Learning Algorithms (pp. 677–685). https://doi.org/10.1007/978-3-030-16660-1_66

[18] Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. IEEE Access, 8, 23522–23530. https://doi.org/10.1109/ACCESS.2020.2969854