

Sentiment Analysis of Customer Reviews for Predictive Product Development in e-Commerce

Suresh Sankara Palli

Independent Researcher, USA.

ARTICLE INFO

Received: 10 Jan 2025

Revised: 20 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

The recent expansion of online transactions, such as e-commerce, has led to the offering of many products and services via online platforms. Because there are so many items accessible while buying online, users struggle to choose the one that best fits their needs. The complex interactions between user and item properties have been the subject of several research in deep learning-based recommender systems (RSs). Since e-commerce has grown so rapidly in recent years, it has become more important to use user evaluations as a basis for purchasing decisions. In addition to assisting prospective clients in making well-informed judgements, reviews create trust and provide companies useful information. Sentiment analysis is a method used to analyse product evaluations, marketing campaigns, and consumer sentiment. Decisions about future marketing initiatives, product and service development, and customer service improvements may be influenced by this useful data. Predicting ratings on social media is often used to forecast product ratings based on user feedback. Convolutional neural networks (CNN), recurrent neural networks (RNN), and bi-directional long short-term memory (Bi-LSTM) are among the deep learning models that are extensively benchmarked in our research. These models are assessed using a variety of word embedding methods, including Word2Vec, FastText, and bi-directional encoder representations from transformers (BERT) and its variations. In this study, we examine and compare the performance indicators of neural network-based models for consumer sentiment prediction using a dataset of product evaluations from consumers of an online women's clothing company.

Keywords: E-Commerce, Recurrent Neural Networks (RNN), Anticipate Product, Convolutional Neural Networks (CNN), Sentiment Analysis, Neural Network, FastText, And Word2Vec, Recommender Systems (RSs), Deep Learning.

INTRODUCTION

Nowadays, the best approach for any company to learn what the public thinks of its products and services is via the internet. After reading a few reviews, many consumers will decide on a product [1]. Every day, reviews are produced, and handling and evaluating such a vast amount of data may be difficult. Particularly during the present COVID-19 epidemic, when several nations have enforced stay-at-home orders, the increase in internet buying has been especially noticeable [2, 3]. Online shopping has become the main method that consumers make purchases as a result of the closure of several physical retail locations and concerns about the development of COVID-19. Online shops now often use textual evaluations and ratings to get feedback from their customers.

These evaluations, which are widely available on social media and internet platforms, are very important in influencing consumer choices to buy and providing merchants with insightful information. Sentiment analysis offers a rapid and simple method for classifying reviews, [3, 4], giving businesses and consumers useful information about what consumers are saying about goods and services.

Online marketplaces have been more and more popular over the last several decades, which has led to a trend of asking customers for feedback on the things they have purchased in an effort to improve the overall customer experience. Every day, millions of internet reviews of different goods, services, and locations are published [5, 6]. As a result, opinions and information on products and services are now mostly available online. But with so many

product evaluations and so many different points of view to consider, the decision-making process may become even more complicated, leading to uncertainty and confusion [5, 6].

After example, when a product has many reviews and ratings, it might be difficult for consumers to make an educated choice [5, 6]. To get feedback on their goods and to assist customers in making well-informed selections, e-commerce enterprises should assess this data. Technologies for sentiment analysis will play a significant role in the next years. Using opinion processing, we can differentiate between subpar and superior material [6, 7]. With today's technology, we can find out if a movie receives more favourable or bad reviews and why. Much of the early research in this area was based on user feedback, such as reviews on Amazon.com that categorised opinions as neutral, positive, negative, or [8, 9]. The majority of sentiment analysis research conducted nowadays uses social networking sites like Facebook, Twitter, and IMDB, which need suitable methods to satisfy text demand.

Access to e-commerce portals and online purchasing has become the new marketplaces for society as a result of rapid urbanization around the world and increasing internet penetration with the use of smart computation devices. Consumers evaluate products or services based on different evaluations. Evaluation can be specifications, ads or reviews [8, 9]. Reviews are one of the most influential factors affecting the sales of products and services. Reviews help alleviate the fear of being cheated and raise the confidence between consumers and businesses in the e-Commerce industry [9, 10]. Using Natural Language Processing (NLP), users can predict the type of review and what is the experience of the product. Due to the prevalence of fraudulent or two-word reviews on e-commerce websites, it is crucial to conduct a thorough study and analysis.

Customers may use natural language processing (NLP) to assess a service or product's quality without reading every review [11,12]. When there are several comparable items with evaluations, it might take a long time for humans to consider them all, and choosing the one that will provide the answer is crucial. When it comes to analysing text written in a variety of languages, NLP has garnered a lot of interest [13]. NLP benefits greatly from the contributions of computer vision, machine learning, and deep learning. Machine learning and deep learning are components of artificial learning (AI), and machine learning is transforming human thought [14].

Additionally, natural language processing (NLP) is a crucial component of artificial intelligence, and some models or algorithms align with machine learning and deep learning. NLP is helpful not just for text analysis but also for audio and video analysis [15]. The capacity of NLP to analyse emotions in speech and text may be used to tackle a number of problems. Numerous new opportunities and capabilities are made possible by NLP.

These days, businesses want to know what their customers think of their products because opinions can range from negative to neutral to positive [13], with some customers liking a product's quality but not its intended design. These opinions can cover a variety of topics, including the product's price, quality, and colour. However, understanding the general perception of the product does not give the company enough information to identify its strengths and weaknesses, which can be valuable to the entity [16]. For this reason, it is important to know how buyers perceive the specific product. By analysing this type of data, the company can improve the quality of its products, which is crucial for its survival in a highly competitive market. In the worldwide e-commerce business, more and more online retailers are using advanced personalisation algorithms [18]. Others utilise recommendation algorithms at retail establishments similar to yours, and some e-commerce pioneers have access to Amazon.com.

However, sentiment analysis is a contextual text mining technique that seeks to find and extract subjective information from the source material. This helps a business monitor online opinions and understand the social sentiment of the product or service it offers [17]. Thanks to recent advancements in machine learning techniques, these techniques are an effective way to delve deeper into the topic [17]. In sentiment analysis, consumers express their opinions about the goods or services they have bought by leaving comments. Therefore, reading other customers' reviews is crucial to gaining a more thorough understanding of the goods or services, also referred to as "word of mouth" (WOM) [17].

Play a vital part in letting other prospective customers learn about the goods, services, and sellers; thus, the more convincing the evaluations are [17], the more confidence they will inspire in prospective customers, encouraging them to choose to purchase the good or service. However, prospective customers may find the increase in reviews

to be tiresome [19], since they have to go over each comment and thoroughly assess the goods or service before deciding.

Therefore, there is a lot of research on machine learning and the use of artificial intelligence in the field of product sentiment analysis [17]. Used a multiread attention network in conjunction with the LSTM algorithm to create a sentiment-based text prediction while accounting for a collection of Chinese social media data [19]. Similarly, a convolutional model and a deep neural network were used to identify phoney positive and negative ratings in an Amazon dataset.

Using sentiment analysis learning techniques and data from online purchases, they created a supervised machine linear regression model in their research to identify customer sentiment. Furthermore, a number of research have concentrated on creating models to handle the growing complexity of big data while expanding the analysis across a range of applications, including illness diagnosis, financial forecasts, [18], and other fields like entertainment.

Nevertheless, there aren't many research that compare machine learning models to provide actual proof of how well they work [19]. According to a Nature article, some authors have built machine learning models based on neural networks with more than three layers using the ideas of Natural Language Processing (NLP). The majority of these studies have found that these models are very accurate at identifying emotions in a variety of contexts [18]. These models are exemplified by CNN and deep neural networks as well as attention-based bidirectional models that integrate CNN and RNN.

This procedure employs a variety of techniques, such as machine etymology and IR. The foundation of sentiment analysis is determining and characterising the polarity of text or short conversations. Characteristics include "negative," "impartial," [17], and neutral opinion polarity. It's crucial to remember that emotion mining might include the following three processes [18].

- 1) Document-level sentence classification: a sentence may now be fully classified as either "positive" or "neutral."
- 2) Sentences should be classified as "yes," "no," or "unbiased."
- 3) Sensitivity classification of dimensions or kinds of features: statements or documents may now be classified as "positive," "negative," or "non-party" based on the aspects of words or archives that are generally acknowledged as "view grouping of the viewpoint stage" [19].

The consumer behaviour during sentiment analysis is shown in Figure 1 below [20],



Fig. 1 Sentiment analysis of customer behaviour. [22]

Determining the sentiment bias (positive, neutral, or negative) of textual data is a common task for sentiment analysis [14]. This facilitates better decision-making across a range of industries, including retail, digital payment services, goods, and the stock and financial markets [11]. With higher scores indicating more favourable comments, researchers that investigate sentiment analysis based on text communication sometimes try to calculate sentiment evaluations using a 1–5 or 10-point scale [18].

Although there are signs that sentiment analysis often employs a variety of machine learning methods, deep learning has gained popularity recently and shown encouraging outcomes. Furthermore, researchers have investigated a number of word embedding techniques, including well-known techniques like Word2Vec and sophisticated transformer-based pre-trained models like bi-directional encoder representations from transformers (BERT) [18].

Lexicon-based methods and machine learning are used in hybrid approaches, which often heavily include sentiment lexicons [18]. A taxonomy of deep learning-based sentiment analysis techniques is shown in Figure 1 [17].

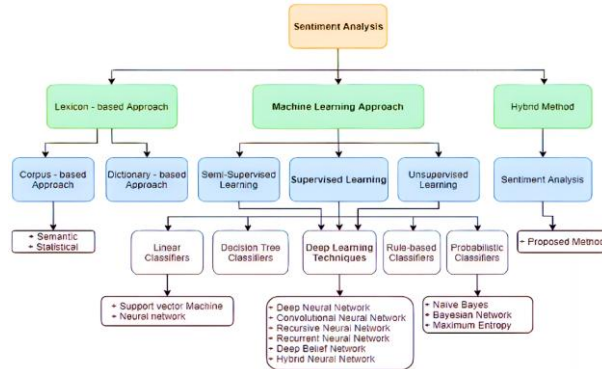


Fig. 2 Methods of sentiment analysis taxonomy [9].

METHODOLOGY

With the rise in popularity of e-commerce, actual order taking from individual customers is no longer necessary. To sell items directly to consumers, businesses use online platforms that make it simple for clients to place orders via the website [15]. Among the most well-known e-commerce giants are Amazon, Flipkart, Myntra, [22], Paytm, and Snapdeal.

By analysing user feedback to provide tailored recommendations, this initiative seeks to enhance the Amazon customer experience [23]. By examining evaluations, the system discovers user preferences and assists consumers in selecting items that suit their interests [22]. The objective is to use machine learning and natural language processing to rethink how consumers browse and interact with Amazon's extensive product selection.

Determining whether or not a product is recommended is the aim of the project's sentiment analysis. To increase prediction accuracy, a number of machine learning methods have been used. Examples of traditional machine learning techniques used in classification include random forest, XGBoost, Ada Boosting, CatBoost, naive Bayes, logistic regression, support vector machine (SVM), [23], and random forest. Furthermore, the BERT algorithm and other deep learning methods have been used [25]. The Woman Clothing Reviews dataset, which is accessible on Kaggle, was utilised. Figure 2 [28] shows the processes used to analyse the data.

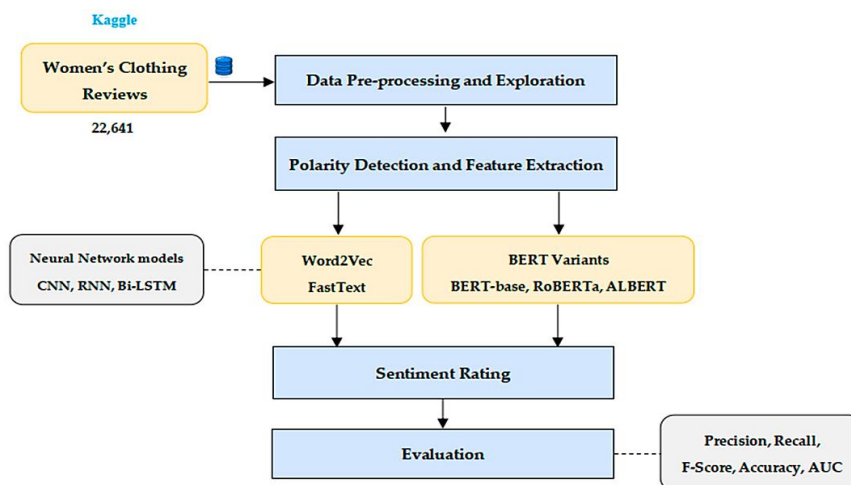


Fig. 3 Overall, the suggested technique. [14]

There are 22,641 rows and 8 column variables in the dataset, and each row contains additional customer information and a written remark [18]. Each row contains the following feature information variables and is a customer review:

- **Clothing ID:** An integer category variable that identifies the specific object being studied.
- **Age:** A positive integer variable is used to denote the reviewer's age.
- **Title:** The title of the review is included in a string variable [18].
- **Review Text:** A text variable that holds the substance of the review [20].
- **Rating:** A positive ordinal integer variable that represents the client's product score, with a range of 1 (worst) to 5 (best).
- **Recommended IND:** The customer's recommendation of the product is indicated via a binary variable (1 = recommended, 0 = not recommended) [25].
- **Number of Positive Comments:** A positive figure that indicates the proportion of other customers who found this review to be favourable [29].
- **Division Name:** A category name that denotes the highest-level division of the product.
- **Department Name:** A categorisation name that identifies the product department.
- **Class Name:** Determines the class of the product [22].

The textual reviews and polarity as determined by TextBlob are shown in Table 1.

Table 1 A dataset of sentiment and polarity for textual reviews. [23]

	Rating	Class_Name	Recommendation_IND	Text	Text_Lenght	Text_Polarity	Sentiment
0	4	Intimates	1	Perfectly lovely, smooth, seductive, and cosy	51	0.659	Positive
1	3	Dresses	1	I adore this outfit; it's very lovely.	302	0.629	Positive
2	4	Dresses	0	A few significant design errors I was so excited about this clothing.	542	0.621	Positive
3	5	Pants	1	My favourite purchases. This outfit is something I adore.	141	0.696	Positive
4	2	Blouses	1	The front tie is adjustable, and the shirt is quite flattering.	209	0.329	Positive
5	4	Dresses	1	Not for the gorgeous outfits by Tracy Reece, which I adore.	511	0.219	Positive
6	5	Knits	0	Fun with Cagrc coal Shimmer This was the final item I put in my cart.	532	0.632	Positive
7	2	Knits	1	Flattening I adore this outfit. It runs a bit, but I generally get an xs.	518	0.149	Positive
8	3	Dresses	0	What a wonderful dress! I purchased 124 lbs and I'm 55.	212	0.219	Positive

BERT is a transformer-based model that captures the contextual links between words in a phrase by using self-attention processes [23]. It can catch long-range relationships since it analyses the full text at once rather than sequentially. The BERT architecture is shown in Figure 8.

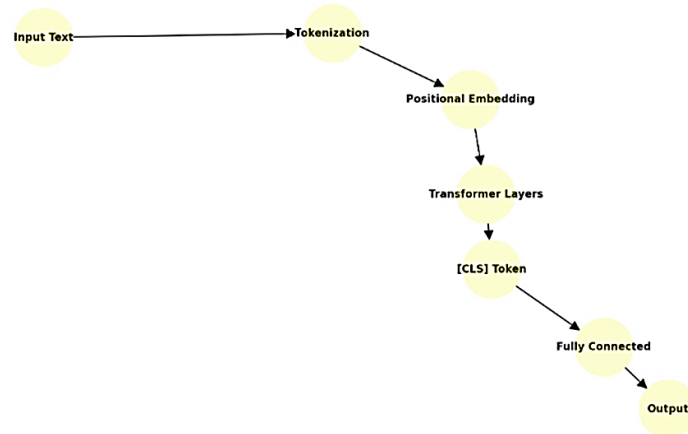


Fig. 4 BERT architecture. [23]

Three well-known neural network (NN) methods are described in the scholarly literature: bidirectional long short-term memory (Bi-LSTM) networks, recurrent neural networks (RNNs), and convolutional neural networks (CNNs) [22, 20]. Artificial neurones arranged in layers, comprising input (predictors), output (predictions), and hidden layers, make up these NN architectures [23]. Figure 5 [21] shows a feed-forward multilayer NN model.

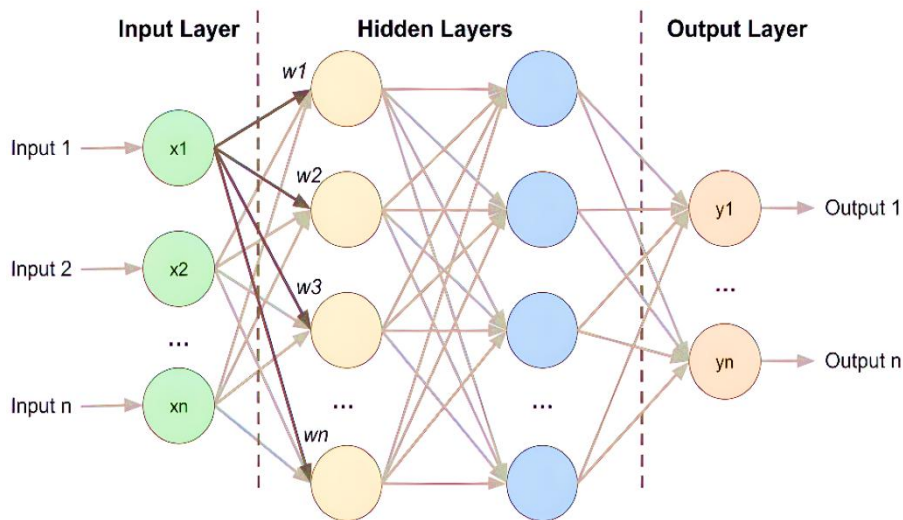


Fig. 5 All filler concentrations' Weibull distribution. [18]

It is important to remember that we also conducted further tests to compare the performance of deep learning models to standard machine learning techniques [20]. We specifically selected five popular machine learning methods that are often used in sentiment analysis research [12]: logistic regression, random forest, naive Bayes, support vector machine (SVM), and decision tree.

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F - score = 2 \times \frac{(precision \times Recall)}{(precision + Recall)}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

EXPERIMENTAL RESULT AND DISCUSSION

This section describes the tests that were carried out to determine if the proposed sentiment analysis approach is suitable for predicting product evaluations in the context of e-commerce recommendations [10]. Two popular datasets were employed in this procedure.

Table 1 displays the outcomes of every neural network model's experiment with the dataset in 5- and 3-class configurations using the Word2Vec and FastText approaches [3].

Table 1 Neural network model performance on the dataset, expressed as percentages (%): Word2Vec against FastText. [18]

Feature Extraction	Model	Precision	Recall	F-score	Accuracy	AUC
3-class						
Word2Vec	CNN	75.96	74.85	21.98	36.96	41.29
	RNN	89.64	94.58	48.98	42.94	26.69
	Ri-LSTM	54.98	63.99	25.89	63.96	68.69
FastText	CNN	87.98	14.89	47.89	49.62	45.69
	RNN	36.98	59.89	59.89	47.99	79.68
	Ri-LSTM	98.98	47.89	64.85	21.65	59.65
5-class						
Word2Vec	CNN	79.54	98.96	25.96	49.68	21.98
	RNN	48.96	49.84	36.94	65.95	54.96
	Ri-LSTM	71.59	41.95	25.96	48.96	48.96
FastText	CNN	87.98	29.85	48.99	45.21	58.96
	RNN	79.85	47.96	69.89	74.89	48.94
	Ri-LSTM	42.89	63.54	47.89	21.98	83.89

Table 2 displays the BERT variants' performance results, demonstrating a trend that consistently favours RoBERTa in our dataset and setups [19].

Table 2 BERT model performance on the dataset as percentages (%): 5-class vs 3-class configurations. [18]

Class	Model	Precision	Recall	F-Score	Accuracy	AUC
3-Class	BERT	58.96	56.98	79.88	89.84	79.89
	ALBERT	56.98	54.89	69.85	48.98	98.88
	RoBERT	54.89	52.48	74.89	58.98	87.98
5-Class	BERT	48.96	58.96	58.98	53.28	54.89
	ALBERT	56.89	54.89	57.89	51.59	56.96
	RoBERT	58.98	59.68	58.96	54.89	52.18

Table 3 [19] displays the results for ensemble models using Word2Vec and the ideal dataset setup, namely the 3-class scenario.

Table 3 Ensemble model performance on the dataset expressed as percentages (%). [18]

Model	Precision	Recall	F-Score	Accuracy	AUC
CNN-RNN	98.8	87.9	84.9	48.96	48.96
CNN_Bi-LSTM	87.9	85.8	89.5	47.95	47.86
RNN-Bi-LSTM	99.8	86.9	89.6	96.89	65.89
CNN-RNN-Bi-LSTM	96.8	84.4	48.9	48.98	48.95

Finally, to compare our results with the machine learning technique, we used the identical setup with other machine learning models, as shown in Table 4 [17]. All of the models performed poorly, as shown by at least a 22% discrepancy in accuracy results when compared to the deep learning models [9].

Table 4 Machine learning model performance expressed as a percentage (%). [14]

Model	Precision	Recall	F-Score	Accuracy	AUC
Naïve Bayes	42.96	39.89	39.60	32.99	63.25
Support Vector Machine	39.59	35.69	31.58	25.96	54.59
Logistic Regression	41.52	35.66	30.69	14.58	51.99
Decision Tree	43.58	34.89	39.65	36.96	32.69
Random Forest	46.96	34.59	32.59	54.59	58.96

CONCLUSION

By using a variety of embedding techniques and deep learning algorithms, our study advances the area of online consumer review analysis. Our 3-class vs 5-class comparative results show that all prediction models perform better when using fewer, more well calibrated classes. Furthermore, we found that Word2Vec performed better than FastText in the setting of context-free embeddings, although the differences were not statistically significant. The most promising results were obtained by RoBERTa, which outperformed BERT and ALBERT. Additionally, our work emphasises the advantages of ensemble models over solo models. Our study's findings were produced exclusively in American English, following its certain terminology and phrases as well as its spelling.

Popular neural network models and alternative embedding techniques, such the more complex BERT and its variants, were used in our experiments. However, additional approaches might be explored, such as lexicon integration together with neural network and BERT-variant models, such as lexicon-enhanced BERT and lexicon-RNN. It is also significant to note that, while this element has been included in several previous research projects, the proportion of polysemous words in BERT was not considered in this study. These experiments suggest that representations produced by BERT might reflect the degree of polysemy and sense partition capacity of a word. Future research on this topic would be intriguing if it took into consideration the proportion of polysemous terms for BERT variants.

As such, future research initiatives may explore other ensemble boosting approaches or optimisation procedures to enhance the predictive capacities of models. Furthermore, given the significant influence of the COVID-19

pandemic on the global purchasing environment and the rise in popularity of online shopping, using real-time data and applications to predict review ratings may be a compelling and significant path.

REFERENCES

- [1] Sharma P, Kumar R, Gupta M (2021) Impacts of Customer Feedback for Online-Offline Shopping using Machine Learning. Proceedings - 2nd International Conference on Smart Electronics and Communication.
- [2] Al Qahtani ASM (2021) Product Sentiment Analysis for Amazon Reviews. Int J Compute Sci Inf Technol 13: 30.
- [3] George K, Joseph S (2014) Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature. IOSR J Comput Eng 16: 34-38.
- [4] Deepu S, Raj P, Rajaraajeswari S (2016) A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction. Int J Adv Netw Appl.
- [5] Bouazizi M, Ohtsuki T (2017) A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter. IEEE Access.
- [6] Sun C, Qiu X, Xu Y, Huang X (2019) How to Fine-Tune BERT for Text Classification? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 194-206.
- [7] Alaparathi S, Mishra M (2021) BERT: a sentiment analysis odyssey. J Mark Anal 9: 118-126.
- [8] Kamath, U., Liu, J. & Whitaker, J. Deep Learning for NLP and Speech Recognition (Springer, 2019). 17. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).
- [9] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.
- [10] Peng, Yifan, Shankai Yan, and Zhiyong Lu. Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. arXiv preprint arXiv:1906.05474 (2019).
- [11] E Cambria, Q liu, S decherchi, F xing, K kwok. Senticnet 7: a commonsense-based neurosymbolic ai framework for explainable sentiment analysis. Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022): 3829-3839.
- [12] Trueman, T. E. & Cambria, E. A convolutional stacked bidirectional LSTM with a multiplicative attention mechanism for aspect category and sentiment detection. Cogn. Comput. 13(6), 1423–1432 (2021).
- [13] Cambria, Erik, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. Affective computing and sentiment analysis. In A practical guide to sentiment analysis, pp. 1-10. Springer, Cham, 2017.
- [14] Wu, J.J.; Chang, S.T. Exploring customer sentiment regarding online retail services: A topicbased approach. J. Retail. Consum. Serv. 2020, 55, 102145.
- [15] Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a Naïve Bayes continuous learning framework. Inf. Process. Manag. 2020, 57, 102221.
- [16] Kabir, A.I.; Ahmed, K.; Karim, R. Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. Inform. Econ. 2020, 24, 55–71.
- [17] Balakrishnan, V.; Lok, P.Y.; Rahim, H.A. A semi-supervised approach in detecting sentiment and emotion based on digital payment reviews. J. Supercomput. 2021, 77, 3795–3810.
- [18] Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment analysis for E commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE Access 2020, 8, 23522–23530.
- [19] Carosia, A.E.; Coelho, G.P.; Silva, A.E. Investment strategies applied to the Brazilian stock market: A methodology based on sentiment analysis with deep learning. Expert Syst. Appl. 2
- [20] Zad, S.; Heidari, M.; Jones, J.H.; Uzuner, O. A survey on concept level sentiment analysis techniques of textual data. In Proceedings of the 2021 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 10–13 May 2021; IEEE: New York, NY, USA; pp. 0285–0291.
- [21] Jing, N.; Wu, Z.; Wang, H. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. Expert Syst. Appl. 2021, 178, 115019.

- [22] He, Kai, Rui Mao, Tieliang Gong, Chen Li, and Erik Cambria. Meta-based Self-training and Re-weighting for Aspect-based Sentiment Analysis. *IEEE Transactions on Affective Computing* (2022).
- [23] Mao, Rui, Qian Liu, Kai He, Wei Li, and Erik Cambria. The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection. *IEEE Transactions on Affective Computing* (2022).
- [24] Alharbi, Najla M., et al. Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews. *Mathematical Problems in Engineering* 2021 (2021).
- [25] Labhsetwar, S. R. Predictive analysis of customer churn in telecom industry using supervised learning. *ICTACT Journal on Sof Computing* 10(2), 2054–2060 (2020).