Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766160

PREDICTION BASED ANALYSIS OF ONLINE PRODUCT REVIEWS USING DEEP LEARNING MODELS

K.Jayabharathi[1], T.Velmurugan[2]
[1]Assistant Professor, [2]Associate Professor
Department of MCA[1], PG and Research Department of Computer Science[2]
[1]Gurunanak College, Chennai, India , [2]Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, India
[1]jayabharathikannan4@gmail.com,[2]velmurugan_dgvc@yahoo.co.in

Abstract: The goal of product reviews is to identify any positive or negative undertones in a given text document. It is widely utilized in a variety of commercial fields to enhance goods or services by learning what customers think about them. Deep learning produces cutting-edge findings in a variety of difficult disciplines. Due to deep learning's success, several researches have presented deep-learning-based sentiment classification models, which outperformed traditional machine learning models. The challenge of modeling and predicting the helpfulness of online reviews look at the variables that affect review helpfulness and attempts to forecast it with accuracy. In this research, It is compared several deep learning-based sentiment classification model topologies to draw useful predictions for the analysis of sentiment based customer reviews. Also, it is provided that a benchmark comparison of many deep learning models such as the Recurrent Neural Network, Attention Mechanism, and the Bi-Directional Long Short Term Memory and evaluated using several word embedding techniques such as FastText and Word2Vec. Each model was evaluated using one of two different setups. The performance of algorithms are compared and suggested for the better use real time applications.

Keywords: Deep Learning Methods, Sentiment Analysis, Online Review Data, Bidirectional RNN Model, Bi-Directional Long Short Term Memory.

1. Introduction

This research work is carried with the advancements that have been made in the fields of computers and the internet, online marketing has also evolved as an area that is at the forefront of technological innovation [1]. When purchasing anything online, the ratings and reviews left by previous customers have a significant amount of weight. It has an impact on the reputation of businesses as well, particularly those who utilize the internet market to promote their products [2]. Because user evaluations are, for the most part, unstructured, the process of extracting user feedback is a time-consuming and laborious one [3]. As a result, techniques from sentiment analysis are utilized. Analysis of people's attitudes and opinions regarding things such as goods, events, themes, and so on is referred to as sentiment analysis.

The extraction of aspects and feelings is an important endeavor [4]. Display and camera quality are two elements of Samsung that are evaluated in the statement "The Samsung J7's display is good, but its camera quality is awful." This sentence serves as an example of an evaluation of Samsung (entity). The feedback about the display is favorable; however, the feedback regarding the camera quality is unfavorable [5]. The relative polarity of the text can be determined by sentiment analysis. It analyses the provided text to determine if it is good, negative, or neutral. Opinion mining is another name for this process because it uncovers the viewpoint or perspective of the person being interviewed [6]. Users of social networking and shopping websites can function as a medium by posting reviews of items on these websites, which can then be utilized for categorization. Because it influences the degree of competition in marketing and the fact that people's requirements might vary greatly, a significant amount of study is being conducted in this area. Since views based on review texts may be used to determine the most effective management methods for maintaining a competitive edge over other market participants, online product review analysis has become a key study issue in retail firms [7]. Since it is impossible to manually inspect the ever-increasing number of customer reviews, deep learning methods are utilized to analyze this data.

The following is how this research was organized. A literature review is covered in Section 2. The precise problem and dataset from Section 3 are provided in the problem specification. Section 4 discusses the methodology for this research project. The result and discussion are shown in section 5. Section 6 of this work, contains its conclusion.

2. Literature review

This research work [7] proposes a method for forecasting consumer recommendations based on reviews and ratings of airline services found on the internet. The XGB model, enhanced by Cuckoo Search (CS-XGB), is the model that is suggested by this study. The research [8] provides a brand-new, integrated method for ranking items based on online customer reviews. To retrieve the extremely high review representation, which, during the educational process, contains the most semantic data, it takes more use of convolutional neural networks, which is different from the approaches that are currently being used. They make modifications to the hierarchy of participating in a society that was only recently defined so that it can operate in the rankings region for this method. This network learns better feature descriptions of the products and their ratings hierarchically by utilizing attention-based encoders on two different levels. The goal of [9] was to present a unique way of processing the reviewed text to improve the precision of rating prediction using machine learning techniques.

This research work is carried to used a variety of techniques to train our model, thus we offer a combination model to forecast the product ratings that correlate to a specific review's content. To make it simple to forecast ratings when similar goods and reviews are present together, we first cluster the subjects and products using k-means and LDA. The objective of [10] was to provide deep learning techniques for examining patrons' perceptions of restaurant attributes and identifying reviews with inconsistent scores. Four commonly cited restaurant elements (such as service, cuisine, setting, and experience) were found by looking through 112,412 restaurant reviews on Yelp.com that were published between January and June 2020. These features were also given sentiment scores. The goal of the research is to investigate and make projections regarding the user ratings of various insurance products by utilizing a wide range of machine methods of learning. They gathered comments from customers on the Yelp webpage and restricted the initial data set to solely evaluations of insurance companies [11].

Have used either cross-category assumptions, where the model and the test set are from looking at the different categories, or in-category suppositions, in which the framework and the testing dataset belong to the same review category, [12] aimed to show that the combination of deep learning, which includes NLP (Natural Language Processing), and conformal predicting produces accurate estimate and efficient sequenced validation set point of view projections for 12 categories of Amazon reviews, where the test set is predicted using a model from a different review category. Study [13], several classification techniques, including "Support Vector Machine Network-based Deep Learning (SVM-DL)" and "Convolutional Neural Network-Based Deep Learning (CNN-DL)" is used to calculate characteristics.

To determine how effective the system is, it consults the data found on the TripAdvisor website, which is a popular one in the United States. The results of the experiments show that when compared to other classification algorithms, CNN-DL has a lower rate of errors and achieves more accurate classification. To predict the attitude of SaaS reviews, this study [14] suggests five models based on 5 methodologies: "the Support Vector Machine algorithm, the Naive Bayes algorithm, the Naive Bayes (Kernel) algorithm, the k-nearest neighbor's strategy, and the decision tree algorithm".

3. Problem Definition

Customers can obtain evaluations and voice their thoughts about a product through an online forum that has been permitted by online retailers. The consumer couldn't read all of the reviews for a popular product since there were hundreds or thousands of reviews accessible for that product. As a result, the client was unable to make an informed choice about whether or not to purchase the product. A skewed perspective on the product may also be formed if the customer just reads a subset of the reviews. The problem with the suggested method is that it is fully opaque, which makes it difficult for consumers to believe the recommended things. As a result of this, the customers find it difficult to trust the recommendations. This is because there is no information accessible to guide a user in the right direction on how they should go about purchasing an item. Users must have complete faith in the algorithm that underpins the suggested system. This results in consumer unease, which in turn harms online e-commerce income. The issue with meta-data is that it makes it harder for users to make decisions about products based only on the features of those products.

4. Materials and Methods

The methodology employed in this work is described in this part, along with the datasets utilized, data pre-processing techniques, feature extractions, classifications for product reviews and ratings, experimental settings, and assessments. Figure 1 depicts the overall methodology.

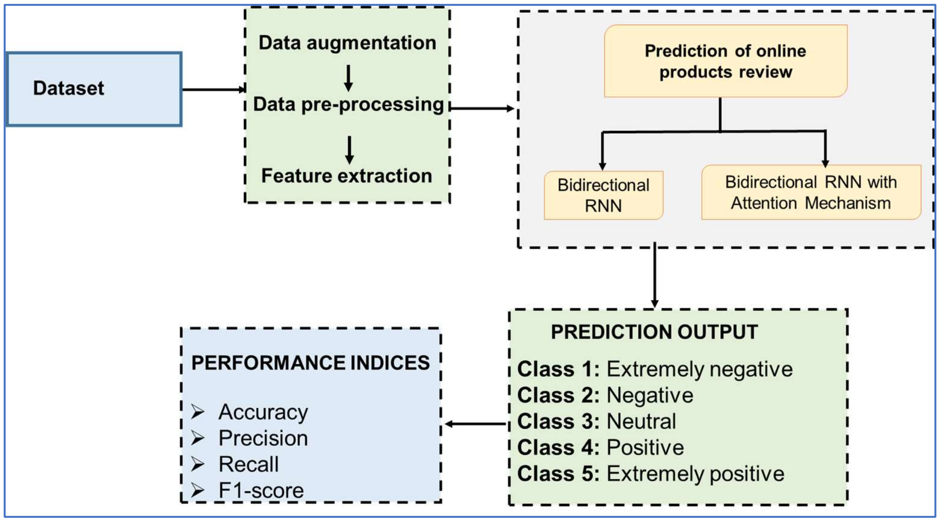

Figure 1: Architecture of Work Flow

4.1 Description of Dataset

This research work analyzed 13 review datasets gathered for diverse products and services to arrive at a broadly agreed empirical result. The enlarged and reduced dataset names are shown in the first and second columns, respectively. In proportions of 50%, 20%, and 30%, respectively, training, validation, and test datasets were created from all datasets. The usual length, vocab level, and positive/negative ratio are displayed in the following columns. Product reviews from Amazon are included in the first 10 datasets for a variety of categories. The eleventh dataset concerns Amazon's menu options. The “12th dataset is a hotel review dataset from Carnegie Mellon University, while the final dataset is a movie review dataset from Stanford University” [15].

4.2 Data augmentation

The practice of data augmentation has seen extensive application in the fields of computer science and voice processing, with a recent surge of interest in the augmentation of textual data. For deep learning models, it is widely employed to supplement the training dataset, making the resulting models more solid and resulting in better overall performance. Many data augmentation approaches, such as translations, question answering, synonym substitution, etc., have been suggested by academics since textual communications are intrinsically more complicated. The “Easy Data Augmentation (EDA)” method, which combines four Natural Language Processing (NLP) operations was one of the more modern techniques used in the current investigation. Because it doesn't require any predetermined datasets and frequently produces positive findings, EDA is prominent for its effortlessness and convenience of utilization. Since the augmentations were carried out by the previously reviewed and advised indices, the resultant substantially enhanced dataset resembles the original sentences, preserving the real labels and keeping the original data's meaning. The EDA technique, in addition to producing the initial dataset, produced a singular, enhanced dataset that was utilized for both training and testing the sentiment assessment prediction models.

4.3 Data pre-processing

Common “Natural Language Processing (NLP)” tasks were then included, such as canonicalization (changing text to lowercase, eliminating leading and trailing spaces, numerals, punctuation, and stop words). Stop words, which include "a," "an," "the," etc., are frequent English words that contribute nothing to understanding the context of the texts they appear in. After them, lemmatization occurred, which reduces words to their base forms (for example, "silky" to "silk," and "happen," to "happen"), and tokenization, which separates sentences into individual words. Using the Keras package, index encoding and zero padding were also carried out to guarantee that all the matrices had the same size.

4.4 Feature extraction

Features are distinct quantitative properties or dimensions that may be used by algorithms, and feature extraction is the process through which these texts are converted into a useful format. Typically, feature extraction strategies for sentiment analysis are determined by the prediction models used in the study. Through a variety of strategies in this work, word embeddings were recovered as features, including

Word2Vec: an already-trained model that discovers the connections between words in a corpus and produces an integrated matrix for every word found within the text.

Fast Text: a Word2Vec addon that converts words into n-grams (smaller components), such as "apple" to "app," to teach users the morphology of the words. For each word in the text, the model also provides a bag of embedded vectors.

4.5 Prediction of online product reviews using Bidirectional RNN

Parallel consideration of forward and backward features was achieved in this research by using two separate bidirectional RNNs (LSTM and GRU). Using two RNNs has the potential to improve performance on both big and small datasets. When used in conjunction with GRU, BiGRU may increase context information while also making the generalization process quicker or requiring fewer data. BiLSTM, on the other hand, builds the map's final features sequentially, which improves performance on big data. With n padding length or maximum feature length, the vectors acquired from the two channels in the preceding phases are used as inputs to these two separate Bi-RNN networks, which then generate the feature context matrix representation.

$$\vec{z}_{l_{lstm}} = \overrightarrow{LSTM}(B_m), m\epsilon[1, n] \tag{1}$$

$$\overleftarrow{z}_{p_{lstm}} = \overleftarrow{LSTM}(B_m), m\epsilon[n, 1] \tag{2}$$

Forward GRUs are represented by (1), while reverse GRUs are shown by (4).

$$\vec{z}_{l_{gru}} = \overrightarrow{GRU}(B_m), m\epsilon[1, n] \tag{3}$$

$$\overleftarrow{z}_{p_{gru}} = \overleftarrow{GRU}(B_m), m\epsilon[n, 1] \tag{4}$$

By joining the forward and the backward context in Eq. (5) and (6), we now have an annotated for every input vector.

$$z_{d_{lstm}} = LSTM\left[\vec{z}_{l_{lstm}}, \overleftarrow{z}_{p_{lstm}}\right] \tag{5}$$

$$z_{d_{gru}=GRU}\left[\vec{z}_{l_{gru}},\vec{z}_{p_{gru}}\right], \quad (6)$$

If $z_{d_{lstm}}$ represents the combined result of a forwarding operation extracting lengthy dependency from BiLSTM may be done both forward $\left(\vec{z}_{l_{lstm}}\right)$ and backward$\left(\overleftarrow{z}_{p_{lstm}}\right)$.To that aim, $z_{d_{gru}}$ is the combined result of forwards $\left(\vec{z}_{l_{gru}}\right)$and backward $\left(\overleftarrow{z}_{p_{gru}}\right)$extracts of a long dependence characteristic from BiGRU.The features $\left[\vec{z}_{l_{lstm}},\overleftarrow{z}_{p_{lstm}}\right], and \left[\overrightarrow{z}_{l_{gru}},\overleftarrow{z}_{p_{gru}}\right]$ are those retrieved from the whole sentence. As a result, both the present and past may be taken into account without having to choose between them.

4.6 Prediction of online product reviews using Bidirectional RNN with an Attention mechanism

We concentrate our attention on each Bi-RNN-produced feature to highlight the original intent of each phrase, given that various words have varied connotations. Both the word annotations, $z_{d_{lstm}}, and\ z_{d_{gru}}$, are fed into a single-layer perceptron to produce $w_{d_{gru}}$ as the hidden representation of $z_{d_{lstm}}$ for tth input and $w_{d_{gru}}$as the hidden representation of $z_{d_{gru}}$ for tth input, respectively.

$$w_{d_{lstm}} = tanh\left(u * z_{d_{lstm}} + p\right) \quad (7)$$

$$w_{d_{gru}} = tanh\left(u * z_{d_{lstm}} + p\right), \quad (8)$$

Where u is the weight and p is the bias, and afterward, we standardize w_d hence for with u_g, which really is initialized and jointly learned during testing, to measure word importance.

$$B_{d_{lstm}} = \frac{exp\left(w_{d_{lstm}} * u_g\right)}{\sum_{j=1}^{n} exp\left(w_{d_{lstm}} * u_g\right)} \quad (9)$$

$$B_{d_{gru}} = \frac{exp\left(w_{d_{gru}} * u_g\right)}{\sum_{j=1}^{n} exp\left(w_{d_{gru}} * u_g\right)}, \quad (10)$$

LSTM and GRU are denoted by $B_{d_{lstm}}$and$B_{d_{gru}}$, respectively, in Eqs. (9) and (10), n text length (.) is the increasing function.

$$B_d = B_{d_{lstm}} \oplus B_{d_{gru}} \quad (11)$$

$$C = \sum(B_d * z_d) \quad (12)$$

Then, we join the LSTM and GRU normalized weight $\ _{ed}$, and it may be written as Eq (11). By adding all of these B_d significance weights together in C, we have a single vector, which we may represent as Eq (12).

5. Experimental Results

In this research work has the following sets and situations were used during the experiments: The term "5-class" relates to the first rating system of 1 to 5. (Class 1: Extremely negative; Class 2 – Negative; Class 3 – Neutral; Class 4 – Positive; Class 5 – Extremely positive) The following are the situations for the experiments: Using enhanced datasets with

Word2Vec and FastText, RNN, Bi-LSTM, and Attention mechanisms were evaluated in 5-class configurations.

Performance analysis

All the models were evaluated using the performance measures for classification issues namely: **Accuracy** is defined as the ratio of the number of forecasts that are accurate across the board, taking into account all of the data collected (Eq. 13). Figure 2 depicts the comparison of accuracy for Bidirectional RNN and Bidirectional RNN with the Attention mechanism.

$$Accuracy = \frac{T_P - T_N}{T_P + F_P + F_N + T_N} \tag{13}$$

Table 1: Accuracy of Bidirectional RNN, Bidirectional RNN with Attention mechanism

Algorithms	Accuracy (%)
Bidirectional RNN	88.09
Bidirectional RNN with Attention mechanism	98.56

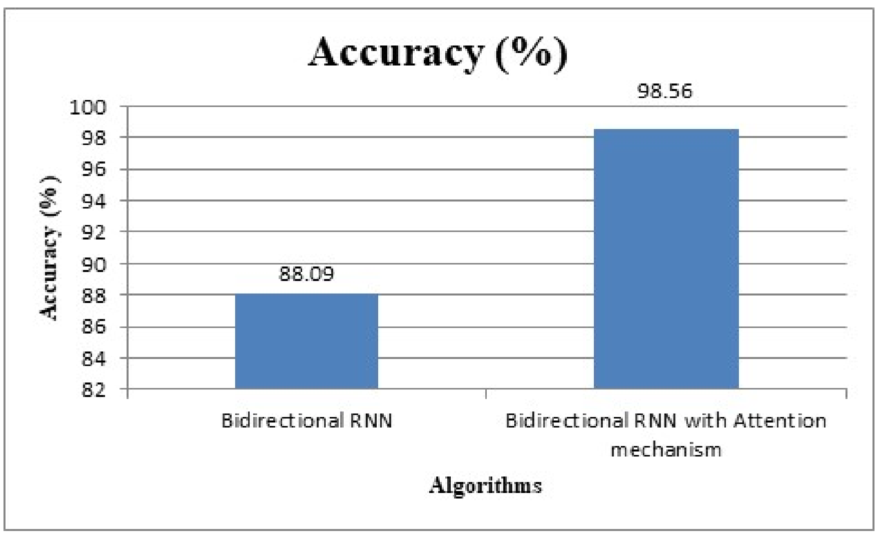

Figure 2: Accuracy of algorithms

Precision is defined as the ratio of true positive findings to the total number of positive predictions made by the model as shown in Eq. 14. Figure 3 depicts the comparison of precision for Bidirectional RNN and Bidirectional RNN with the Attention mechanism.

$$Precision = \frac{T_P}{T_P + F_P} \tag{14}$$

Table 2: Precision of Bidirectional RNN, Bidirectional RNN with Attention mechanism

Algorithms	Precision (%)
Bidirectional RNN	85.43
Bidirectional RNN with Attention mechanism	97.86

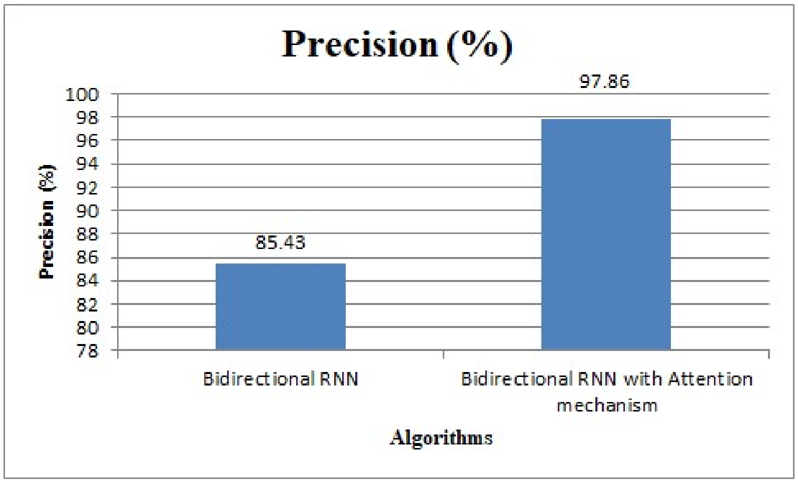

Figure 3: Precision of algorithms

Recall the percentage of genuine positive instances that are appropriately recognized depicted in Eq. 15. Figure 4 depicts the comparison of recall for Bidirectional RNN and Bidirectional RNN with the Attention mechanism.

$$Recall = \frac{T_P}{T_P+F_N} \quad (15)$$

Table 3: Recall of Bidirectional RNN, Bidirectional RNN with Attention mechanism

Algorithms	Recall (%)
Bidirectional RNN	85.01
Bidirectional RNN with Attention mechanism	98.20

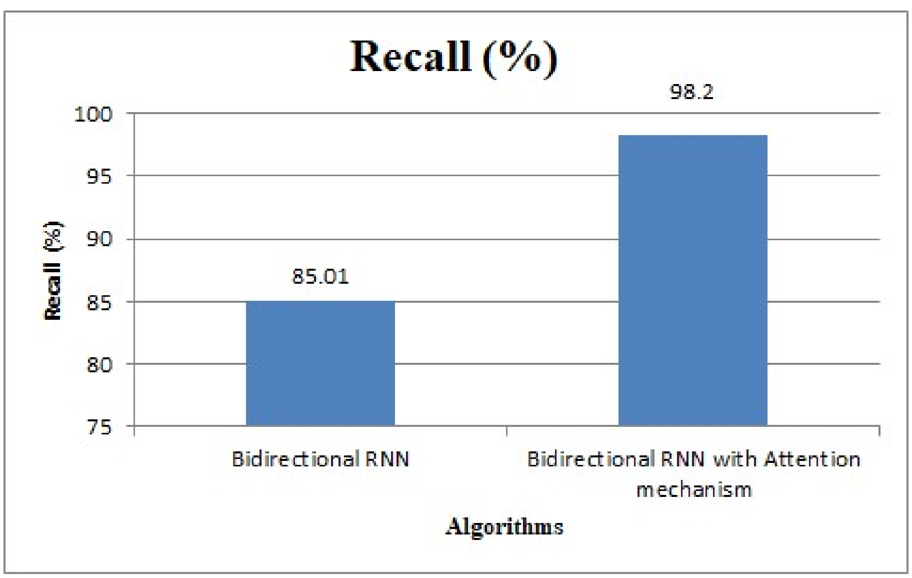

Figure 4: Recall of Algorithms

The F1-score is the harmonic mean of accuracy and recall, and the F-measure range is 0 to 1. A higher F-measure value suggests that the model is doing better. The formula for calculating F-measure is as follows: Figure 5 depicts the comparison of the F1 score for Bidirectional RNN and Bidirectional RNN with the Attention mechanism. Table 1 depicts the outcomes of Bidirectional RNN and Bidirectional RNN with Attention mechanism.

$$F - measure = 2 * \frac{(precision * Recall)}{(Precision + Recall)} \quad (16)$$

Table 4: F1 Score of Bidirectional RNN, Bidirectional RNN with Attention mechanism

Algorithms	**F1 Score (%)**
Bidirectional RNN	81.31
Bidirectional RNN with Attention mechanism	97.84

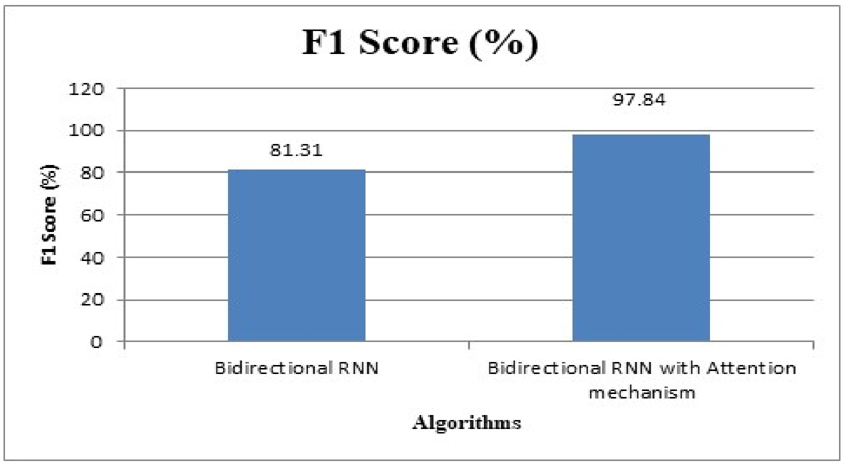

Figure 5: F1 score of Algorithms.

The findings for the ensemble approaches that rely on the configuration that performed the best (5-class) utilizing the expanded dataset and Word2Vec are shown in Table 1. This was done so that a comparison could be made between the accuracy of the integrated NN models and that of the individual models to predict the sentiment of evaluations. Our results show that all ensemble models outperform the individual NN models (see Table 1), with the Bidirectional RNN with Attention mechanism having the highest accuracy (98.56%), precision (97.83%), recall (98.20%), and F1-score (97.84%).

Table 5: Results of Bidirectional RNN, Bidirectional RNN with Attention mechanism

	Bidirectional RNN	**Bidirectional RNN with Attention mechanism**
Accuracy (%)	88.09	98.56
Precision (%)	85.43	97.83
Recall (%)	85.01	98.20

F1 score (%)	81.31	97.84

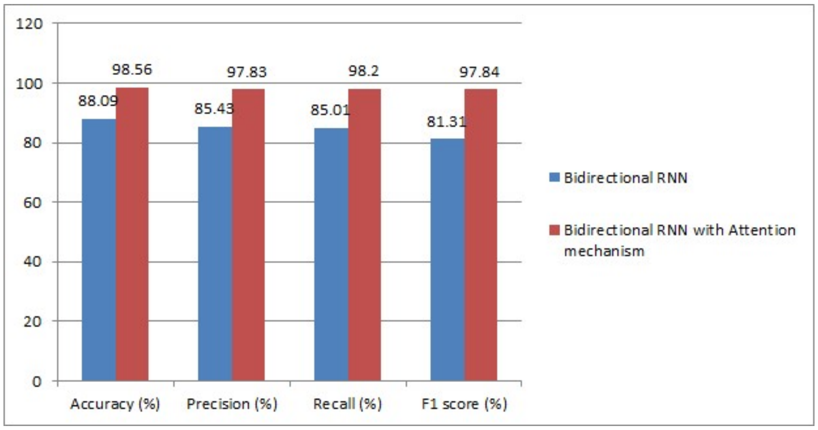

Figure 6: Results of Bidirectional RNN, Bidirectional RNN with Attention mechanism

This pattern of observation has also been observed in other investigations, where multi-models were shown to outperform individual models irrespective of the datasets that were utilized. Even though metric variations between the prediction methods do not meet the criteria for statistical significance, our research demonstrates that utilizing ensemble models (which make an effort to improve predictions) results in better help to review of forecasting consequences when compared to the conventional strategy of relying on deep learning models.

6. Conclusion

Using many deep learning algorithms based on a variety of embedding approaches, this work made a significant contribution to the field of research including online consumer evaluations. According to our results, all prediction models perform better in a setting with fewer and more granular classes (5-class), and the use of an enriched dataset enhances prediction performance over the original dataset. Word2Vec outperformed FastText in situationally anchoring tests, although by a little margin. The findings demonstrate that the Bidirectional RNN with Attention mechanism provides superior outcomes compared to the Bidirectional RNN. We identify a few drawbacks. Due to the fact that the dataset utilized in this research was not vetted for spam or fraudulent reviews, it is possible that this impacted the results. As a result, the pre-processing phase might be expanded to incorporate an extra step for automatically identifying spam and fraudulent reviews. Due to the study's exclusive focus on English-language reviews, it's possible that the suggested models and conclusions cannot be generalized to a more broadly linguistic context. Customers on the internet are said to come from all over the globe, and they often choose to converse in languages outside English. Extending the existing suggested framework to include languages than English might be a fruitful direction for future research. It's common knowledge that deep learning models outperform machine learning models, but they come at a high computational cost. Consequently, future research might investigate optimization strategies or use alternative ensemble boosting methods to enhance the prediction performance of the machine learning models.

Reference

[1] Deyalage, P.A. and Kulathunga, D., "Exploring Key Factors for Customer Satisfaction in Online Shopping", *A Systematic Literature Review*, 2020.

[2] König, T.M., Hein, N. and Nimsgern, V., "A value perspective on online review platforms: Profiling preference structures of online shops and traditional companies", *Journal of Business Research*, Vol. 145, pp.387-401, 2022.

[3] Dong, J., Chen, Y., Gu, A., Chen, J., Li, L., Chen, Q., Li, S. and Xun, Q., "Potential Trend for Online Shopping Data Based on the Linear Regression and Sentiment Analysis", *Mathematical Problems in Engineering*, 2020.

[4] Nofrialdi, R., "Online Shopping Behavior Model: Determining the Factors Affecting Repurchase Intention", *Journal of Law, Politic and Humanities*, 1(2), pp.88-97, 2021.

[5] Vijayakumar, S., Vidyashankar, G., Venkatesakumar, R., Madhavan, S. and Riasudeen, S., "Online Review Characteristics and Information Asymmetry Is it easy to switch between Online Shopping sites? A Case Study of Reviews from Amazon and Flipkart", *SDMIMD Journal of Management*, pp.27-39, 2021.

[6] Anh, K.Q., Nagai, Y. and Nguyen, L.M., "Extracting customer reviews from online shopping and its perspective on product design", *Vietnam Journal of Computer Science*, 6(01), pp.43-56, 2019.

[7] Jain, P.K., Yekun, E.A., Pamula, R. and Srivastava, G., "Consumer recommendation prediction in online reviews using Cuckoo-optimized machine learning models", *Computers and Electrical Engineering*, Vol. 95, p.107397, 2021.

[8] Lee, H.C., Rim, H.C. and Lee, D.G., "Learning to rank products based on online product reviews using a hierarchical deep neural network", *Electronic Commerce Research and Applications*, Vol. 36, p.100874, 2019.

[9] Kumar, P., Dayal, M., Khari, M., Fenza, G. and Gallo, M., "Nsl-bp: A meta classifier model based prediction of amazon product reviews", 2021.

[10] Luo, Y. and Xu, X., "Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic", *International Journal of Hospitality Management*, Vol. 94, p.102849, 2021.

[11] Hossain, M.S. and Rahman, M.F., "Customer sentiment analysis and prediction of insurance products' reviews using machine learning approaches", *FIIB Business Review*, p.23197145221115793, 2022.

[12] Norinder, U. and Norinder, P., "Predicting Amazon customer reviews with deep confidence using deep learning and conformal prediction", *Journal of Management Analytics*, 9(1), pp.1-16, 2022.

[13] Shoukry, A. and Aldeek, F., "Attributes prediction from IoT consumer reviews in the hotel sectors using conventional neural network: deep learning techniques", *Electronic Commerce Research,* 20(2), pp.223-240, 2020.

[14] Alkalbani, A.M., Ghamry, A.M., Hussain, F.K. and Hussain, O.K., "Predicting the sentiment of SaaS online reviews using supervised machine learning techniques", *International Joint Conference on Neural Networks (IJCNN)*, pp. 1547-1553, 2016.

[15] Seo, S., Kim, C., Kim, H., Mo, K. and Kang, P., "Comparative study of deep learning-based sentiment classification", *IEEE Access*, Vol. 8, pp.6861-6875, 2020.