

Doi:10.32604/iasc.2025.058976

ARTICLE





AI-Driven Sentiment Analysis: Understanding Customer Feedbacks on Women's Clothing through CNN and LSTM

Phan-Anh-Huy Nguyen^{*} and Luu-Luyen Than

Faculty of Economics, Ho Chi Minh City University of Technology and Education, Ho Chi Minh City, 700000, Vietnam *Corresponding Author: Phan-Anh-Huy Nguyen. Email: huynpa@hcmute.edu.vn Received: 25 September 2024; Accepted: 28 February 2025; Published: 14 April 2025

ABSTRACT: The burgeoning e-commerce industry has made online customer reviews a crucial source of feedback for businesses. Sentiment analysis, a technique used to extract subjective information from text, has become essential for understanding consumer sentiment and preferences. However, traditional sentiment analysis methods often struggle with the nuances and context of natural language. To address these issues, this study proposes a comparison of deep learning models that figure out the optimal method to accurately analyze consumer reviews on women's clothing. CNNs excel at capturing local features and semantic information, while LSTMs are adept at handling long-range dependencies and contextual understanding. By integrating these two deep learning techniques, our model aims to achieve better performance in sentiment classification. The models were trained and evaluated on a dataset of women's clothing reviews sourced from Kaggle. The dataset was pre-processed to clean and tokenize the text data, and word embeddings were used to represent words as numerical vectors. The CNN component of the model extracts local features from the text, while the LSTM component captures long-range dependencies and contextual information. The outputs of the CNN and LSTM layers are then concatenated and fed into a fully connected layer for final sentiment classification. Experimental results demonstrate that the hybrid model outperforms traditional machine learning techniques and other deep learning models in terms of accuracy, precision, recall, and F1-score. By accurately classifying sentiment, identifying key themes, and predicting future trends, our model can provide valuable insights to businesses in the apparel industry. These insights can be used to improve product design, marketing strategies, and customer service, ultimately leading to increased customer satisfaction and business success.

KEYWORDS: AI-driven; sentiment analysis; RNN; LSTM; CNN; deep learning; e-commerce

1 Introduction

The advent of technology has significantly altered consumer buying behavior, with electronic word-ofmouth (eWOM) playing a crucial role in purchasing decisions [1]. Online reviews, both positive and negative, have become a primary source of information for consumers, often carrying more weight than companygenerated content [2]. These reviews, typically found on social networking sites and e-commerce platforms, allow users to share their experiences with products and services [3]. Positive reviews can accelerate sales and help build a brand reputation, while negative reviews tend to discourage purchases [4]. Importantly, consumers pay more attention to negative remarks when making purchasing decisions [1]. This shift in consumer behavior has prompted businesses to analyze customer satisfaction levels through eWOM data, enabling them to improve their offerings based on user trends, needs, and preferences [3]. This study strives to enhance the understanding and analysis of consumer reviews on social networks by applying a convolutional



neural network (CNN) algorithm, Evaluate the accuracy and limitations of the algorithm to improve the method of analyzing customer reviews in the future. Provide information about customers' needs for the business's products/services so that they can have more optimal business strategies and product development.

While numerous studies have explored sentiment analysis of online reviews, there remains a significant gap in leveraging deep learning techniques, particularly convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to accurately analyze customer reviews on social media platforms. This study aims to bridge this gap by:

- Proposing a hybrid CNN-LSTM model: This novel approach combines the strength of CNN in extracting local features with the ability of LSTM model to capture long-range dependencies in text.
- Evaluating model performance: Focusing on the model's accuracy and limitations will provide valuable insights for future improvements in sentiment analysis techniques.
- Providing actionable insights: By analyzing customer reviews, this study will uncover valuable information about customer needs, preferences, and pain points. This knowledge can be leveraged by businesses to develop more effective marketing strategies and product offerings.

Apparel and clothes are always essential products and have changed according to fashion trends year by year. In particular, women's clothing is a product category that is always attractive to businessmen. This paper used a dataset of 32,000 customer reviews of women's clothing to conduct an emotional analysis of customer reviews. The main training data of women's clothing product reviews was obtained from the open-source Kaggle. The deep learning algorithms are deployed and their performance is. The results of this paper can be applied to business decision-makers for their strategies in this competitive market of the AI era.

2 Literature Review

2.1 Sentiment Representation Models

CNN is a well-known Deep Learning algorithm. Using convolutional neural networks helps us build intelligent and highly accurate models. This algorithm excels at taking input images of data and assigning important values (including weights and machine-learnable features) about different aspects and objects in the image. Image and the result is to distinguish one image from another. Multiple studies employed neural network techniques, CNN and LSTM models, to classify sentiments in e-commerce reviews [5]. LSTM outperformed CNN in sentiment classification, achieving 91.69% accuracy [5]. Bidirectional LSTM models demonstrated high performance, with F1-scores of 0.88 for recommendation classification and 0.93 for sentiment classification [6]. Research revealed that product recommendations strongly indicate positive sentiment scores, while ratings are less reliable indicators [6]. These AI-driven approaches offer valuable insights for businesses to understand customer perceptions and improve their offerings. Studies show that emotional content in microblog posts can significantly influence product recommendations, with negative emotions potentially enhancing recommendation effectiveness during positive events [7].

CNN is applied in many problems such as classification, video analysis, MRI images, or articles in the field of NLP and this is the algorithm that is said to best solve these problems. The advantage to mention is that the pre-processing requirement in CNN is much lower than that of other classification algorithms. In similar methods, the filters are designed manually and have to undergo training many times, while CNN, is an algorithm capable of learning, understanding, and applying the filters optimally. RNN uses feed-forward neural networks and convolutional neural networks (CNNs), the network Regression neurons use training data to learn.

Sentiment analysis by deep learning were studied by many researchers in recent years. The LSTM neural network yields higher accuracy, with an LSTM of vector length 300 delivering over 89% accuracy, the

highest for any model. In [8], an approach and classification of business assessments using word embedding on a large scale is presented in this context. The authors compare word-based CNN model that uses an end-to-end vector for text evaluation classification. In [9], they proposed combining Joint Sentiment Topic detection with Naive Bayes classification, utilizing bigrams and a sentiment thesaurus to improve accuracy across multiple domains. In [10], they use different models for the semantic analysis of film reviews. Neural Regression, Recursive Neural, and Neural Convolutional algorithms were implemented on the Stanford Sentiment Treebank data set and compared with the baseline Naive Bayes classifier. These hybrid models have demonstrated high accuracy in sentiment classification across various datasets, with reported F1-scores ranging from 92.81% to 95.06% [11]. Multiple studies have explored the deployments of deep learning based models, such as Deep Belief Networks optimized with the Dragonfly Algorithm, to analyze investor sentiments from Twitter data and forecast stock prices [12,13].

2.2 Word Embeddings

Refining word embeddings is a key approach in recent research aimed at improving sentiment analysis performance. In [14], a refinement model was proposed that leverages sentiment lexicons to adjust word vectors, improving semantic and sentiment similarity. This method, applicable to any pre-trained embeddings, outperforms conventional and sentiment-specific embeddings in sentiment classification tasks. Similarly, in [14], they developed a refinement model that adjusts word representations to enhance sentiment information capture. A sentiment concept-based method is introduced in [15] that utilizing Microsoft Concept Graph and a sentiment lexicon that accounts for the diverse semantic contexts of words and their corresponding emotional intensities. These approaches address the limitations of traditional word embeddings in capturing sentiment information and demonstrate improved performance in domain-specific sentiment analysis tasks.

2.3 Deep Learning Methods for Customer Feedback Analysis

Recent research on deep learning methods for customer feedback analysis in women's clothing ecommerce has shown promising results. Various machine learning methods, such as Support Vector Machines (SVM), Logistic Regression (LR), and Artificial Neural Networks (ANN), have been applied to classify customer reviews as positive or negative, with ANN achieving the highest accuracy of 88% in [16]. Some NLP techniques combined with deep learning methods have been applied to analyze emotional tendencies and identify key product features in women's fashion reviews, enabling more personalized recommendations [17].

Graph neural networks and syntactic information are being explored in recent research to improve aspect-based sentiment classification (ASC). In [18], a dual graph neural network model incorporating part-of-speech guided syntactic dependency graphs and syntactic distance attention is proposed. Similarly, In [19], a new dual graph convolutional network model is developed, which effectively combines syntactic and semantic information. However, in [20], they questioned the necessity of explicit syntactic information, demonstrating that fine-tuned RoBERTa can outperform previous state-of-the-art models on ASC tasks across multiple languages and datasets. This aligns with the findings by [21], researchers discovered biases in pre-trained language models used for sentiment and emotion tasks. These biases stem from label counts, emotional word choices, prompt templates, and lexicon word forms, illustrating the complexity of affective computing with these models.

Realizing that applying deep learning is very beneficial to business development through analytics, user reviews, and ratings. But in the field of fashion, specifically, the data set on clothing has not appeared in previous studies. Considering the limitations of previous studies, the team chose to analyze user reviews

using a deep learning algorithm with a dataset of women's clothing. This paper uses an untapped dataset, with several 32,000 evaluations, and compares the results of the LSTM and CNN models. This research helps businesses understand customer needs and know to what extent their products can meet customers' needs. Contributing to optimizing their products/services, helping businesses maintain market share and increase profits.

3 Methodology

3.1 Tensorflow Overview

The Tensorflow library is an open-source end-to-end platform used to build and create machine learning applications in other words; It is used for numerical computation and machine learning at scale, is a symbolic math library that uses data flows and discriminant programming to perform a variety of tasks, but still focuses on deep neural network learning and inference. Thanks to Tensorflow, implementing machine learning and deep learning is not as difficult as it used to be as it contributes to making it easier to collect data, train models, predict, and refine future results. Future. The platform facilitates the training and execution of deep neural networks, including convolutional networks for image recognition, recurrent networks for sequence modeling, and seq2seq models for machine translation, as well as supporting natural language processing and partial differential equation-based simulations.

3.2 Tokenizer Algorithm

Tokenization is the process of breaking text into smaller units called tokens. These tokens can be words, characters, or subword units (n-grams). Tokenization methods are generally categorized into word-based, character-based, and subword-based approaches:

- Word tokenization is the primary technique for splitting text into words. It relies on a delimiter to identify word boundaries, and the tokens produced are influenced by this delimiter. Pre-existing word embeddings, like Word2Vec and GloVe, are built upon this word-based tokenization—Character encoding (Character Tokenization): dividing each piece of text into a set of characters. It overcomes more limitations than word encoding.

- Subword Tokenization: divide the text into sub-keywords (or n-gram characters). Words like lower can be segmented as low-er, smartest as smart-est, and so on. Models based on SOTA transformation in NLP-based on subkey coding algorithms for lexical preparation. The word separation algorithm is the most important step while modeling textual data.

3.3 Word Embedding and Word-to-Vector Algorithm

Word embeddings translate words into numerical vectors, making them understandable to computers. This method represents words with similar meanings as vectors close to each other in a multi-dimensional space. These vector representations are crucial for solving many natural language processing problems. Essentially, each word is transformed into a vector, with the vector values learned through a process similar to training a neural network. The Continuous Bag-of-Words (CBOW) model estimates the probability of a word appearing based on the words around it. It can use a single context word or multiple words. Skip-gram, conversely, predicts the context words surrounding a target word. The Gamma skip-gram model shares a similar process with CBOW, but its neural network architecture and weight matrix creation are distinct.

3.4 Global Vector for Word Representation

GloVe is an alternative method for generating word embeddings. It applied the matrix factorization technique on word context matrices. A large matrix of co-occurrence information is built and you count each "word" (rows) and how often we see this word in several "contexts" (columns) in a large corpus. To generate word vectors, we first analyze the words surrounding each term within a defined window, prioritizing nearby words. This process yields a large matrix of contextual information. We then reduce this matrix to a lower-dimensional form, where each row represents a word vector. This reduction is achieved by minimizing the error when reconstructing the original data, ensuring the vectors capture the essential patterns.

3.5 CNN Algorithm

Text and natural language classification is also the strength of the CNN algorithm. Since CNN can only output vectors of fixed size, this feature is perfectly suitable for application in many tasks such as Sentiment Analysis, Spam Detection. CNN will compare images in small arrays, for text, it will match each word together during data processing. Those subsections are called Features. When comparing each word, or character with an object labeled positively or negatively, we will find out what they have in common. Combine with convolution to get the standard model. Convolutional: When receiving a new text, or a new word, CNN will not know where they match to determine whether this sentence has a positive or negative meaning. Therefore, CNN will in turn try the words at each position in the sentence, this process creates Filters for the machine to continue learning. This process is done through the convolutional neural algorithm and it keeps repeating as we run the dataset. The data will flow through each layer (layer). The convolutional layer is the most important, taking care of most of the computational work as soon as we start running the data. The concepts included in this layer are covered: filter map, stride, padding, and feature map.

+ Filter map: As mentioned above, CNN will use filters to apply to the characters of the text, this is a 3-dimensional matrix of numbers.

+ Stride: is the shift of the filter map by a certain value from left to right of the text.

+ Padding: are the values added to the input.

+ Feature map: every time the filter map scans through the input data, a calculation process occurs called a feature map.

+ Relu layer: This is the activation function in the model, which simulates neurons with pulse transmission rate through the axon. When training the model, the advantage of the real layer is that it makes the computation faster.

+ Fully connected layer: This layer tells us the result of the operation. After the data is scanned through the convolutional layers, the pooling layer, the model has linked all the characteristic features of the text, thereby having a basis for classifying positive and negative reviews. To output those features we use the full association class. In addition, this class is not able to suggest which text clearly shows the positive and the negative. However, this rating system is not highly authentic.

3.6 Emotion Classification Algorithm

Convolutional Neural Networks (CNNs) have shown promising results in various Natural Language Processing tasks. The ability of CNNs to capture local dependencies and patterns in word sequences makes them a suitable starting point for analyzing the textual descriptions that is described in Algorithm 1. For Emotion Classification Algorithm, each vector of different sizes produces a feature part. These values are linked together to create the layer near the bottom of the figure. Finally, this value is filtered through a function and output the feature vector. Use this result to predict the trend of the text (for example to label positive or negative). The steps are as follows: (1) Sentence: example using the sentence "I like you very much!", i.e., the sentence has 5 words and one punctuation mark. We can count that punctuation as a word if it affects the final prediction result. Each word will be one vector.

(2) Filters for images, each filter will be a 2-dimensional vector, but for text, it is only 1-dimensional. We need to determine the number of filter dimensions for these consecutive strings of words.

(3) Feature Maps (feature matrix): to create a Feature Map, CNN will first perform a step-by-step convolution, as shown in the figure above, and multiply each element by a vector of size 2. And keep doing such calculations until the end of the sentence we will find the final result of the matrix.

(4) Imax: the size of the map is not fixed but depends on the text length as well as the filter size. To bring back the same size, we use the pooling layer, taking the largest value in the feature maps to produce the most representative feature. Using max pooling to get the maximum value of each feature map results in a matrix of the same size.

Algorithm 1: CNN text classification [22]

Input: Labeled training text as *D*, *l* be the total words in *D*, *m* be the dimension, *n* be the filter size—the filter is an n-gram, *z* be the total filters all sizes.

Training:

1. An embedding layer that embedded input texts to high dimensional space

 $\mathbf{x}_{1:l} = \{\mathbf{x}_{1}, \mathbf{x}_{2}, ..., \mathbf{x}_{l}\}^{T} \in \mathbb{R}^{l \times m}$

2. Convolutional layers with different kernel sizes are employed, combining convolution functions and ReLU activation

n filter $f_i \in \mathbb{R}^{n \times m} \to \mathbb{R}, j \in \{1, 2, \dots, z\}$

A convolution kernel sliding over the embedding layer.

$$c_k = f_j(x_{k:k+n-1}) = w_j \times x_{k:k+n-1} + b_j,$$

where
$$k \in \{1, 2, ..., l - n + 1\}, w_j \in R^n$$

and
$$c_j = \{c_1, c_2, \ldots, c_{l-n+1}\}$$

ReLu activation is applied over the convolution output

 $\boldsymbol{\gamma}_j = \{\gamma_1, \gamma_2, \ldots, \gamma_{l-n+1}\},\$

where $\gamma_k = \max(0, c_k)$.

3. A max-pooling layer takes the maximum value of convolution outputs

 $\theta_j = \max\{\boldsymbol{\gamma}_j\} = \max\{\gamma_1, \gamma_2, \ldots, \gamma_{l-n+1}\}.$

4. The final outputs of the model are produced by a feed-forward neural network

 $y = g(\boldsymbol{\theta}),$

where $\theta = \{\theta_1, \theta_2, ..., \theta_h\}$ is the max-pool layer output from all *z* convolutional filters.

End procedure:

Output: Prediction result D'

3.7 Some Measurements in Text Classification Algorithms

Accuracy: Accuracy refers to the ratio of correct predictions to the total number of cases. Therefore, the greater the accuracy, the better the model.

$$Accuracy = \frac{TN + TP}{Total \, sample} \tag{1}$$

Precision: This metric shows how many of the predictions are positive. The larger this number, the better it reflects our model.

$$Precision = \frac{TP}{Total \ predicted \ positive} = \frac{TP}{TP + FP}$$
(2)

Recall: gives the same result as precision but with a difference in the denominator. Recall calculates the correct ratio of the number of positive predictions to the total number of positive results. Apply to the model training data part, use recall, because the text is already labeled.

$$\operatorname{Recall} = \frac{TP}{Total \ actual \ positive} = \frac{TP}{TP + FN}$$
(3)

F1-score: The average index between Precision and Recall, this index is more objective because it represents both values. In case 1 of 2 indexes is 0, then we consider F1-score = 0

$$F1-score = \frac{2}{precision^{-1} + recall^{-1}}$$
(4)

4 Results

4.1 Dataset Description

The data are women's clothing product reviews in English produced by the team which is the opensource dataset of the Kaggle website. The original data set consists of 11 columns corresponding to the figure: Serial number, ID, Age, Title, Review body, Rating, Recommendation, Positive reviews, Category name, Part of the garment, and Types of clothes. We took 32,000 reviews and took two columns as Review Content, and Referral. In the Review Text column, which is renamed Review, the Recommended column will be renamed Sentiment as shown below.

4.2 Data Preprocessing

The data set includes 32,000 reviews, of which 16,000 are divided into positive reviews (Positive) and 16,000 negative reviews (Negative) based on whether it is Recommended (1) or Not recommended (0). The implementation team found that the positive reviews (Positive) will often have words like "nice" (nice), "fit" (medium), "love" (Favorite), and "cute" (cute)., "like" (like),... Negative reviews (Negative) will often have words like "hate" (hate), "disappointed" (disappointed), "awkward" (strange), etc...

Experimental setup is as follows: Hardware: Intel Core i7-12700K, 32 GB RAM, NVIDIA GeForce RTX 3090, Software: Windows 10, Python 3.9, TensorFlow 2.10, NumPy, Pandas, Scikit-learn.

4.3 Experimental Results

To be able to conduct evaluative emotion analysis, the implementation team has processed the data as follows:

• Step 1: Prepare the CSV format so that Python can process the file.

• Step 2: Use some functions in Python to remove redundant characters, stop-word has no meaning to make the data more compact.

• Step 3: To be able to train the data, the implementation team converted the text to vector using the word splitting algorithm (Tokenizer) supported library. At the same time, the team used GloVe (Global Vectors), which is an open-source set of Stanford University that converts English documents into vectors.

This study employs a deep learning-based sentiment analysis process to predict consumer trends regarding clothing products. The overall workflow of this process is illustrated in Fig. 1, which presents a flowchart summarizing the key steps.



Figure 1: Flowchart summarizing the process of deep learning based sentiment analysis

Fig. 1 depicts the following implementation process:

(1) Data Collection: The initial stage involves gathering ratings and reviews of clothing products. (2) Sentiment Analysis Using Deep Learning: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) algorithms are applied to analyze the emotions present in the collected data. (3) Consumer Trend Prediction and Business Solutions: Based on the sentiment analysis results, consumer trends are predicted, indicating whether consumers are likely to continue buying, stop buying, or require further consideration. This information is then used to propose actionable solutions for businesses. (4) Algorithm Improvement: Finally, measures are suggested to enhance the accuracy of the deep learning algorithms.

The research results derived from this process, as outlined in Fig. 1, will provide valuable insights into consumer evaluations.

The experimental setting is as follows:

Embedding Layer: Converts words into dense vectors of size 100. Convolutional Layers: Layer 1 applies 64 filters of size 3 to the input sequence, capturing local features. Layer 2 applies 128 filters of size 4 to the output of the first layer, capturing slightly larger features. Fully Connected Layer: Combines the extracted features into a 128-dimensional representation. Output Layer: Predicts the sentiment class using a softmax layer.

LSTM Architecture for Sentiment Analysis

LSTM Layer: 128 number of units Fully Connected Layer: Output dimension: 64 The activation function is ReLU The activation function is Softmax

Choosing the most suitable algorithm model

After training the data, the team evaluated the three models according to the scale and produced the following results:

Table 1 presents the accuracy and F1-score evaluation results of the LSTM model. As shown in Table 1, the model exhibits strong performance, particularly for class 0, with a high precision of 0.82 and recall of 0.97. This indicates the accurate identification of positive instances. However, the model's performance for class 1, as detailed in Table 1, is slightly lower, with a recall of 0.78, suggesting potential missed positive cases. The F1-score, a balanced measure of precision and recall, is 0.89 for class 0 and 0.86 for class 1, demonstrating overall good performance, as can be observed in Table 1. The model's accuracy of 88%, also documented in Table 1, further reinforces its effectiveness in correctly classifying instances.

	0	1	Accuracy	Weighted Avg
Precision	0.82	0.96	-	0.89
Recall	0.97	0.78	-	0.88
F1-score	0.89	0.86	0.88	0.87
Support	4860	4740	9600	9600

Table 1: Accuracy and F1-score evaluation results of LSTM

Table 2 presents the CNN's performance metrics, showcasing strong results across various evaluation measures. The CNN model demonstrates high precision for both classes, with 0.89 for class 0 and 0.97 for class 1, indicating a low rate of false positives. The model achieves excellent recall for class 0 at 0.98, signifying its ability to correctly identify nearly all positive instances of that class, though recall for class 1 is slightly lower at 0.88. The F1-score, a balanced measure of precision and recall, is consistently high for both classes, with 0.93 for class 0 and 0.92 for class 1, and an overall accuracy of 0.93. This consistent weighted average of 0.93 across precision, recall, and F1-score highlights the model's balanced and robust performance across the entire dataset. With a support of 4860 for class 0 and 4740 for class 1, the model's performance is shown to be effective across a balanced distribution of the two classes.

Table 2:	CNN's performance metrics

	0	1	Accuracy	Weighted Avg
Precision	0.89	0.97	-	0.93
Recall	0.98	0.88	-	0.93
F1-score	0.93	0.92	0.93	0.93
Support	4860	4740	9600	9600

In Fig. 2, a comparative analysis of the evaluation results for both models reveals significant differences in their performance metrics:



Figure 2: Performance comparisons

To provide a clear visual comparison of the models' performance, a column chart was generated in Table 3 as follows:

	Accuracy	Precision	Recall	F1-score
CNN	0.93	0.93	0.93	0.93
RNN (LSTM)	0.88	0.89	0.87	0.87

 Table 3: The evaluation results of the two methods

The evaluation results indicate that the CNN model outperforms the LSTM model across all metrics. The CNN achieved an accuracy of 0.93, precision of 0.93, recall of 0.93, and F1-score of 0.93, demonstrating strong performance in correctly classifying instances and identifying true positive cases. In contrast, the LSTM model achieved lower scores across all metrics, suggesting potential limitations in capturing complex patterns and long-range dependencies in the text data.

The scale evaluates the results of the two models in Fig. 3 as follows:



Figure 3: Comparing two models: CNN and LSTM

This paper performed a convolutional neural network model and LSTM to classify text emotions with two data sets with 1000 and 3000 reviews of women's clothing ratings.

The classification results are listed in Table 4 as follows:

LSTM	CNN		
Number of total reviews: 1000	Number of total reviews: 1000		
Number of positive reviews (LSTM): 281	Number of positive reviews (CNN): 441		
Number of negative reviews (LSTM): 719	Number of negative reviews (CNN): 559		
Number of total reviews: 3000	Number of total reviews: 3000		
Number of positive reviews (LSTM): 1239	Number of positive reviews (CNN): 1596		
Number of negative reviews (LSTM): 1761	Number of negative reviews (CNN): 1404		

Table 4: Experimental results of CNN and LSTM algorithms

After the model training process, we decided to choose CNN's model for the following four reasons: Firstly, in terms of objective assessment of Accuracy and F1-score scales, CNN (0.93) compared with the other two models is LSTM (0.88). Secondly, although LSTM has advanced in natural language processing compared to CNN. However, LSTM needs to have a large amount of input data to learn better. Thirdly, during the training process, the processing time for the CNN model only takes 30 s to 1–2 min. Meanwhile, the model of LSTM takes more than 7 min to process. From there, it shows that the effectiveness of the CNN model is not only in accuracy but also in processing time. Finally, in the classification in practice, CNN gives better results than LSTM. With CNN emotional analysis model, the input will be the user's reviews, the output will be the number of positive (>=0.5) and negative (>=0.5) reviews. Applying the model into practice will help businesses in general easily access the needs and thoughts of customers about their products. Understand clearly whether the products/services we provide meet the expectations of customers, and whether customers' reactions to them are positive or negative so that manufacturers can adjust. Adjusted to optimize functions and designs to bring the best product and experience to users. From there, businesses can easily achieve high profits, firmly grasp as well as expand their market share.

Implications

+ Trends of applying Deep Learning to market research for production and business

Deep Learning has been put into practice to help computers train themselves with large amounts of data. Therefore, in recent years, all industries have tended to convert some jobs to computers to save time and human costs. Market analysis by Deep Learning has always been of interest to businesses, which shows that many scientific articles have been researched on this issue, especially Facebook and Amazon have been at the forefront of applying intelligence. Artificially in analyzing the market, and understanding customers to serve to build the best production and business strategy.

+ The need for emotional analysis from customer product reviews

The demand for emotional analysis of customer reviews is also a small area in the large block of market analysis. Businesses always want to know if customers are targeting a product/service with what functions and whether the product they provide has met that customer's need or not. Of course, with hundreds to thousands of customer feedback about the product, we must have a staff to be able to read all and write down how many positive reviews and how many negative reviews in total. From the user. This is simple, but it takes time and wastes company resources. Therefore, the emotional analysis model has been researched and applied so that businesses can optimize their working performance.

+ The advantages that this method brings to manufacturers and businesses

Currently, the method of emotional analysis to evaluate and respond to customers brings the opinions, suggestions, complaints, satisfaction, and desires of customers given to businesses about the products and services that the business provides. Grant. Therefore, feedback analysis and evaluation by Deep Learning, more specifically by modeling of convolutional neural network (CNN) with accuracy up to 95%. In the experimental results, with 1000 reviews, including 500 positive and 500 negative, the model gave classification results: 504 positive reviews, and 496 negative reviews. Analyzing customer reviews and feedback brings the following benefits and advantages. To enhance business products and services, understanding customer needs is paramount. By analyzing these reviews, businesses can refine product features, and tailor offerings to customer preferences. This data-driven approach not only boosts customer retention but also attracts new customers. For consumers, these reviews serve as a reliable source of information to make informed purchasing decisions. Furthermore, this marks the initial step for businesses to leverage artificial intelligence in analyzing market trends and user behavior, paving the way for future innovation and growth.

5 Conclusions

Deep Learning have revolutionized numerous industries and aspects of daily life. Deep Learning can be applied to market analysis and consumer trend analysis has been adopted by leading companies like Facebook and Amazon. Our research aimed to delve deeper into the workings of an emotional analysis model, to understand its potential to evaluate users, predict trends, and inform strategic business decisions. Through our experiments, we gained valuable insights into the capabilities of deep learning models. We observed how these models can significantly enhance the precision of sentiment analysis, enabling businesses to accurately gauge customer emotions and preferences. Furthermore, we recognized the power of deep learning to personalize marketing strategies by tailoring campaigns to individual consumer behavior and preferences. While our research yielded promising results, we encountered certain challenges, including data quality issues and computational constraints. To address these limitations and further advance the field, future research could explore more sophisticated deep-learning architectures and incorporate multimodal data into the analysis.

Besides the benefits that this approach brings, there are still several limitations that the team found still exist in this model. Because the training input data is small. For Deep Learning, the input data source plays a very important role because it works according to the simulation like the brain. So it takes hundreds of thousands of data inputs for the machine to learn.

Future research can be developed in two directions: First, expand the data. For the project to be practical, diversify industries, and bring higher accuracy, it is necessary to always update the data and expand the vocabulary for the model. Second, perform Aspect-Based Sentiment Analysis (ABSA) implemented by enhancing short-term memory (LSTM) with hierarchical attention mechanisms. In addition, to tightly integrate conventional knowledge into the repeater encoder, the authors could use the Sentic LSTM extension of the LSTM.

Acknowledgement: This research is supported by Ho Chi Minh City University of Technology and Education.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Phan-Anh-Huy Nguyen, Luu-Luyen Than; analysis and interpretation of results: Phan-Anh-Huy Nguyen, Luu-Luyen Than; draft manuscript preparation: Phan-Anh-Huy Nguyen, Luu-Luyen Than. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available at https://www.kaggle.com/datasets/rishikumarrajvansh/womens-clothing-reviews-data (accesed on 27 February 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- 1. Jha I, Bharti N, Malik S. The impact of online reviews on product perception and purchase intention. In: 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS); 2023; Bangalore, India. p. 111–5. doi:10.1109/ICAECIS58353.2023.10169949.
- 2. Raymond R. When word-of-mouth goes online: evaluating the characteristics and effects of Ewom communication. Int J Arts Sci. 2015;8(5):499–508.
- 3. Yaylı A, Bayram M. E-WOM: the effects of online consumer reviews on purchasing decisions. Int J Inter Market Advert. 2012;7(1):51–64. doi:10.1504/IJIMA.2012.044958.
- Shih H-P, Lai K-H, Cheng T. Constraint-based and dedication-based mechanisms for encouraging online selfdisclosure: is personalization the only thing that matters? Eur J Inform Syst. 2017;26(4):432–50. doi:10.1057/s41303-016-0031-0.
- 5. Nawaz Z, Zhao C, Nawaz F, Safeer AA, Irshad W. Role of artificial neural networks techniques in development of market intelligence: a study of sentiment analysis of ewom of a women's clothing company. J Theor Appl Elect Commerce Res. 2021;16(5):1862–76. doi:10.3390/jtaer16050104.
- 6. Agarap AF. Statistical analysis on e-commerce reviews, with sentiment classification using bidirectional recurrent neural network (RNN). arXiv:180503687. 2018.
- 7. Hsu P-Y, Lei H-T, Huang S-H, Liao TH, Lo Y-C, Lo C-C. Effects of sentiment on recommendations in social network. Electr Mark. 2019;29(2):253–62. doi:10.1007/s12525-018-0314-5.
- 8. Salinca A. Convolutional neural networks for sentiment classification on business reviews. arXiv:171005978. 2017.
- 9. Kalaivaani P, Thangarajan R. Enhancing the classification accuracy in sentiment analysis with computational intelligence using joint sentiment topic detection with MEDLDA. Intell Automat Soft Comput. 2020;26(1):71–9. doi:10.31209/2019.100000152.
- 10. Shirani-Mehr H. Applications of deep learning to sentiment analysis of movie reviews. In: Technical report. Stanford University; 2014.
- 11. Kaur G, Sharma A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data. 2023;10(1):5. doi:10.1186/s40537-022-00680-6.
- 12. Holzhaider JC, Sibley MD, Taylor AH, Singh PJ, Gray RD, Hunt GR. The social structure of New Caledonian crows. Anim Behav. 2011;81(1):83–92. doi:10.1016/j.anbehav.2010.09.015.
- 13. Smith T, Rana RS, Missiaen P, Rose KD, Sahni A, Singh H, et al. High bat (*Chiroptera*) diversity in the Early Eocene of India. Naturwissenschaften. 2007;94(12):1003–9. doi:10.1007/s00114-007-0280-9.
- 14. Yu LC, Wang J, Lai KR, Zhang X. Refining word embeddings for sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; 2017; Copenhagen, Denmark: Association for Computational Linguistics. p. 534–9.
- 15. Wang Y, Huang G, Li J, Li H, Zhou Y, Jiang H. Refined global word embeddings based on sentiment concept for sentiment analysis. IEEE Access. 2021;9:37075–85. doi:10.1109/ACCESS.2021.3062654.
- Sharma P, Kumar R, Gupta M. Impacts of customer feedback for online-offline shopping using machine learning. In: 2nd International Conference on Smart Electronics and Communication (ICOSEC); 2021; Trichy, India. p. 696–1703. doi:10.1109/ICOSEC51865.2021.9591939.

- Lin N, Shi Z, Gao R, and Zheng S. A focus on female consumer review-based recommender systems. In: 2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA); 2024; Shenyang, China. p. 656–9. doi:10.1109/ICPECA60615.2024.10470983.
- Xiao L, Xue Y, Wang H, Hu X, Gu D, Zhu Y. Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. Neurocomputing. 2022;471:48–59. doi:10.1016/j.neucom. 2021.10.091.
- Li R, Chen H, Feng F, Ma Z, Wang X, Hovy E. Dual graph convolutional networks for aspect-based sentiment analysis. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2021; Association for Computational Linguistics (ACL).
- 20. Dai J, Yan H, Sun T, Liu P, Qiu X. Does syntax matter? A strong baseline for aspect-based sentiment analysis with roberta. arXiv:210404986. 2021.
- Mao R, Liu Q, He K, Li W, Cambria E. The biases of pre-trained language models: an empirical study on promptbased sentiment analysis and emotion detection. IEEE Transact Affect Comput. 2022;14(3):1743–53. doi:10.1109/ TAFFC.2022.3204972.
- 22. Keeling R, Chhatwal R, Huber-Fliflet N, Zhang J, Wei F, Zhao H, et al. Empirical comparisons of CNN with other learning algorithms for text classification in legal document review. In: 2019 IEEE International Conference on Big Data; 2019; Los Angeles, CA, USA. p. 2038–42. doi:10.1109/BigData47090.2019.9006248.