A Hybrid DBN-GRU Model for Enhanced Sentiment Analysis in Product Reviews

Shaista Khan[1], J Chandra Sekhar[2], J. Ramu[3], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[4], Dr.K.Aanandha Saravanan[5], Kuchipudi Prasanth Kumar[6], Dr. Prajakta Uday Waghe[7]

MBA Department, Datta Meghe Institute of Management Studies, Nagpur, India[1]
Professor in CSE, NRI Institute of Technology, Guntur, India[2]
Associate Professor &Head, Department of CSE, NRI Institute of Technology, Guntur, Andhra Pradesh, India[3]
Faculty of Informatics and Computing, UniSZA University, Malaysia[4]
Associate Professor, Department of ECE, VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India[5]
Assistant Professor, Dept. of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, Andhra Pradesh, India[6]
Associate Professor and Head, Department of Applied Chemistry, Yeshwantrao Chavan College of Engineering, Nagpur, India[7]

Abstract*—In an era marked by a proliferation of online reviews across various domains, navigating the extensive and diverse range of opinions can be challenging. Sentiment analysis aims to extract and interpret sentiments from these vast pools of data using computational linguistics and information retrieval techniques. This study focuses on employing deep learning methods such as Deep Belief Networks (DBN) and Gated Recurrent Units (GRU) to classify reviews into positive and negative sentiments, addressing the issue of information overload in Product Reviews. The primary objective is to develop an efficient sentiment analysis system that reliably categorizes reviews as positive or negative. The study introduces a novel sentiment analysis framework combining Deep Belief Networks and Gated Recurrent Units for online product review classification, enhancing accuracy through advanced feature extraction and classification techniques. The comprehensive preparation pipeline—comprising data splitting, stemming, stop word removal and special character separation—enhances dataset refinement for improved classification accuracy. The proposed framework consists of four main phases: pre-processing, feature extraction, classification, and evaluation. During the preparation phase, the dataset is meticulously cleaned and refined to reduce noise and enhance signal quality. Significant features are then extracted from the pre-processed data using advanced feature extraction algorithms. The DBN-GRU model leverages these features for sentiment classification, effectively distinguishing between positive and negative attitudes. The framework's performance is subsequently evaluated to assess its efficacy in accurately classifying reviews. The combination of in-depth pre-processing procedures and the DBN-GRU technique yielded promising results in sentiment categorization. The framework demonstrated a high accuracy of 98.74% in differentiating between positive and negative sentiments, thereby facilitating the effective analysis of online reviews. This study presents a robust framework for sentiment analysis, utilizing the DBN-GRU method to classify online reviews. Through extensive preprocessing and advanced classification techniques, the system addresses the challenges of noise and information overload in online reviews, providing valuable insights for both consumers and businesses.

Keywords—Sentimental analysis; product review; deep learning; DBN-GRU

I. Introduction

The modern human lifestyle includes online buying daily. There are several e-commerce platforms available in the IT industry to accommodate customer expectations. Customers may update the product review for the purchases on any e-commerce site. Nowadays, every e-commerce customer has the privilege of reading product reviews on various buying websites [1]. These are typically referred to as product ratings. A range of ratings from 1 to 5 is that entire product ratings are, after all. A higher grade indicates a higher level of product quality. In addition to evaluating the goods, customers may provide their opinions. Feeling, experience, and sentiment are the three key components of any user review [2]. These variables differ from person to person. Any consumer who wants to comprehend the review must spend some time on several websites before buying a certain product. Many customers will become disinterested in purchasing the products. Recent advancements in social media sites have provided companies and organizations with the opportunity to receive assessments from their customers and through the use of consumer comments [3]. These articles are published online through social media and websites and may be in the form of text, voice, video, or a combination of all three. Specifically, social networking text data is described by short, semi-structured, unorganized sentences that are frequently filled with ordinary language. As a result, sentimental analysis and vectors models for this text data are difficult to generate and take a lot of effort [4].

Sentiment analysis is the management of emotions, opinions, and subjectively communication. Sentiment analysis, which looks at a number of tweets and comments, gives the knowledge about what the general population thinks. It is an accurate way to forecast a number of significant events, such as the popularity of film at the blockbusters and presidential race [5]. The evaluation of a particular entity, such as a person, a product, or a location, is done using public reviews, which can be obtained on numerous websites like Yelp and Amazon. The perspectives might be categorised as good, positive, or moderate. Sentiment analysis aims to identify individual the evaluative tone of main outcomes [6]. The desire for sentiment classification has grown as a result of the growing requirement to analyse and organise the complex data that results from social establishment's secret information.

By utilising polarities and combinations, sentiments encompass a wide range of highlighted values, including tri-grams and bi-grams. As a result, using training methods for different Support vector machine (SVM), sentiments are assessed as both positive and negative components [7]. Neural Networks (NN) are employed in sentiment analysis to determine the belongingness of labels. Extracted data at the consideration of the potential is aided by the contextual connections between a number of edges and vertices of an interconnected network run by Bayesian networks. By using the best possible terms and phrases on social media platforms, learning and data integrity can be attained. The construction of the data's negative and positive properties at the word system level uses data tokenization. In order to increase the precision of social media data, methods are being used to minimise sentiment classification errors [8].

The procedure of recognising and categorising the viewpoints or opinions represented in a text span utilizing retrieval of information and computational linguistics is known as opinion mining (also known as sentiment analysis) [9]. Instead of the issue itself, the viewpoint that is voiced about it is given weight. In our information-seeking behaviour before making a decision, opinion is crucial [10]. Personal blogs and online review platforms make it easier to obtain opinions on goods or services utilising information technology. By identifying characteristics and elements of the item that have been remarked on in each document, the primary goal of opinion mining is to ascertain the polarity of comments [11].

Sentiment analysis is a method for identifying and understanding the sentiments expressed in text. Thanks to the data surge in social media platforms like Twitter, and LinkedIn, individuals now have potential approaches to express their thoughts on particular goods, persons, and regions [12]. The individual's feedback is often displayed as textual information. Each day, social media sites and online businesses send and receive millions of texts and emails. Investigating and analysing the tone of the opinion is crucial. Textual data and NLP with AI skills are used to determine whether an opinion is positive, negative, or neutral. Opinion mining and emotion research are autonomous of any one network or industry [13]. It is pervasive on all kinds of social media and in a variety of fields, including management, health insurance, finance, and many more. Additionally, it is quite beneficial for the growth of numerous organizations and companies. Furthermore, sentiment analysis provides business knowledge that can be used to decide wisely and effectively [14]. Despite the fact that they each possess their own distinctive qualities, they can sometimes be used indiscriminately. Sentiment categorization shows the sentiment polarity by assigning classifiers to the content or fragment. Sentiment perspective is a sort of text summarization that arranges text data in accordance with the consciousness – of thoughts. Sentiment orientation refers to the subjective polarity of an opinion, regardless of whether it is correct or incorrect. The subjectively or objectively nature of the presented text or review data must be distinguished using the subjective analysis approach.

Lexicon-based approaches and machine learning techniques like Deep Learning can be used to classify sentiment. Lists of words and phrases having positive and negative meanings are often the foundation of the lexicon-based sentiment analysis technique. For this strategy, a dictionary of words with assigned negative and positive emotion values is necessary. These approaches are straightforward, adaptable, and computationally effective. They are therefore mostly employed to address generic sentiment analysis issues. However, lexicon-based approaches in human-labelled texts rely on human effort and can have limited coverage. They also rely on discovering the sentiment lexicon that is used to analyses the text. The key contributions of this study can be summarized as follows:

- The DBN-GRU model outperforms other models like DBN, GRU, and LSTM, ensuring reliable sentiment analysis and a better understanding of customer feedback for businesses.
- The study implements a comprehensive data preprocessing pipeline that includes data splitting, stopword removal, stemming, and special character isolation. This thorough approach significantly enhances data quality, reducing noise and improving classification accuracy.
- The proposed DBN-GRU model achieves an outstanding accuracy in classifying sentiments from product reviews, outperforming traditional models such as DBN, GRU, and LSTM. This high level of accuracy demonstrates the model's robustness and reliability in sentiment classification tasks.
- Utilizing a large dataset of 70,000 product reviews from Amazon, spanning categories like electronics, mobile phones, and instruments, the study validates the model's effectiveness across different domains. This broad application highlights the model's versatility and scalability.
- The data pre-processing process, including splitting, stemming, stop word removal, and special character separation, enhances dataset refinement, improves feature extraction, and reduces noise for accurate real-time sentiment classification.
- The deep belief network component of the model facilitates superior feature extraction, capturing intricate patterns and sentiments in the data, which are crucial for accurate sentiment classification.
- The study evaluates the model's performance using a wide range of metrics, including accuracy, precision, recall, F-measure, specificity, and sensitivity. This thorough evaluation provides a holistic view of the model's effectiveness and reliability.

- The novelty of this work lies in the innovative integration of DBN and GRU models, leveraging their complementary strengths to enhance sentiment analysis accuracy and efficiency. Unlike traditional methods, the hybrid approach combines deep feature extraction with the ability to handle sequential data, providing a more nuanced and robust sentiment analysis.

The remainder of this paper is structured as follows: The literature review is addressed in Section II. Section III provides a brief explanation of the suggested method. With the aid of graphs and tables, Section IV illustrates the outcome of the suggested strategy. In Section V, the research is finished.

II. Related Work

Li Yang et al. [15] proposed a Sentiment Analysis for Chinese e-Commerce Reviews relying on Sentiment Words and Learning Techniques. People may now purchase and use things online more frequently than ever before thanks to the remarkable technological breakthroughs in Internet technology. Mood analysis of a large number of customer reviews on e-commerce sites can effectively boost customer pleasure. The article indicates a novel attention-based Bidirectional Gated Recurrent Unit Neural Network and Convolution neural sentiment classification method dubbed SLCABG. It focuses on a language of emotions. By fusing the advantages of sentiment words with DL technology, the SLCABG approach tackles the drawbacks of the present sentiment analytical technique of brand assessments. The SLCABG system combines the advantages of sentiment words and DL techniques. The sentiment vocabulary is used to first enhance the sentiment features in the reviews. As a result, Cns and GRU networks are used to identify the main attitude features and contextual features from the reviews, and the functional form is then used to assess them. Sort the qualities that feeling judges into categories. The article cleans and scans the authentic book evaluation of the well-known Chinese e-commerce site dangdang.com, which is totally in Chinese, for testing and training purposes. The data is beneficial for various applications in the field of Chinese attitude research because it has a level of one million orders of magnitude. The empirical findings indicate that the technique may greatly increase the efficiency of sentiment analysis of text. However, the industry research suggested the technique could only divide attitudes into positive and negative categories, rendering it useless for situations where there is a limited supply of attitude clarifications.

Using DNN and weighted word embeddings, Aytug Onan presented a sentiment analysis of customer evaluations [16]. One of the main goals of processing usual language is sentiment analysis, which involves extracting sentiments, ideas, views, or judgments about a certain issue. The internet is an unorganized, comprehensive information resource with a wide variety of paper communication, including evaluations and viewpoints. Private decision-makers, Government, and commercial groups may all benefit from understanding emotion. Researchers offer a DL-based method for sentiment analysis of invention evaluations from Twitter in this article. The suggested scheme incorporates Long Short-Term Memory Networks (LSTM)- Convolutional neural networks (CNN) architectures and IDF-TF weighted Gloves word embeddings. Five phases make up the LSTM-CNN architecture: a dense layer, a weighted embedding layer, an LSTM layer, a max-pooling layer, and a convolution layer. In the exploratory study, numerous word embedding techniques with various weighting function have indeed been compared to traditional deep neural network designs to see how well they predict the outcome. According on the empirical findings, the suggested deep learning framework operates better than the traditional deep learning techniques. However, Document length might be a limitation, and therefore no one method will necessarily produce the best prediction results including all different kinds of text categorization problems.

Text Review Sentiment Analysis Using Lexicon-Enhanced James Mutinda et al. [17] suggested using a CNN using the Bert Integration Scheme. Within the discipline of naturally occurring language processing, the study of attitude has gained importance. The technique can be applied in a variety of settings, such as politics, economics, and online review systems for businesses. To be effective, sentiment analysis involves trustworthy text summarization techniques that can convert words into precise vectors that accurately represent the text information. ML -based techniques and lexicon-based methods are two types into which text reconstruction techniques can be classified. According to studies, both methods have drawbacks. For illustration, pre-trained word embeddings produce vectors by neglecting additional factors like word sentiment orientation and instead focusing on word distance, commonalities, and appearances. The study introduces a sentiment categorization method that combines a CNN, N-grams, sentiment lexicon, and BERT, in an effort to overcome such restrictions. Words chosen from a portion of the input sentence are vectorized in the system using N-grams, sentiment lexicon, and BERT. CNN is a deep neural system classifier that maps features and assigns a sentiment category as an output. Three open datasets are used to assess the suggested approach. The model's performance measures include F-measure accuracy and precision. According to the test findings, the suggested method performs more effectively than the currently available state-of-the-art approaches. The article has certain restrictions. Only the convolutional neural network was employed in the constructed method.

Zhengjie Gao et al. [18] suggested Target-Dependent Sentiment Classification With BERT. Sentiment examination is one of the frequently used applications of machine-assisted textual analysis, which has been rapidly developing alongside online technology. Conventional sentiment analysis techniques call for intricate feature engineering, and embedded interpretations have largely dominated leader boards. Nevertheless, because of their context-independence, they have limited representational strength in rich context, which negatively affects how well they do Natural Language Processing activities. The current standard for character recognition is BERT, which outperforms previous pre-trained language models in 11 Natural language processing tasks by a significant margin. BERT has been used less frequently for sentiment categorization at the relation to the issue since it is a particularly difficult assignment. With outputs located at target words and an additional phrase with the objective built in, researchers develop three target-dependent versions of the BERT base paradigm. Studies on three types of data demonstrate that, in contrast to conventional features engineering techniques, embedding-based approaches, and older BERT implementations, the TD-BERT approach provides different state-of-the-art efficiency. The investigations aim to determine if context-aware representations of BERT may produce a comparable performance gain in aspect-based sentiment analysis given its effective delivery throughout many NLP applications. It's fascinating to see that merging it with advanced NN that traditionally conducted superbly with integrated caricatures did not always enhance efficiency above that of the basic BERT-FC version. On the contrary side, inclusion of the goal information demonstrates steady improvement in accuracy, and the experiment reveals the best strategy to use that knowledge. Yet, the classification performance of neutral situations is substantially lower than that of instances with a distinct polarity, and it is much more difficult to handle cases with mixed emotion polarities relating to the same target or various features.

The Sentiment Analysis of Explanations Texts Based on BiLSTM was developed by Guixian Xu [19]. A substantial amount of comments text is created on the Web as a result of the quick growth of social networks and online technologies. In the age of big data, it is beneficial to use AI technologies to mine the emotional tendencies of feedback in order to quickly comprehend online public sentiment. Artificial intelligence includes sentiment analysis technologies, and its study is particularly important for determining the sentiment trends of the comments. The basis of sentiment classification is the text summarization issue, and different words each participate to classification in a different way. The majority of the most current sentiment classification experiments use generalised phrase models. However, generalised sentence reconstructions only include the semantic aspects of the phrase and overlook its emotive significance. The work suggests an improved TF-IDF method that incorporates attitude data into language modeling to construct weighted word vectors. In order to effectively incorporate essential data and enhance the representation of the comments matrix, the filled expressions are input into BiLSTM. The sentiment tendency of the remarks is used by the feed-forward NN classifier to establish its categorization. In the same conditions, the proposed sentiment analysis strategy is compared to the sentiment analytical methods of NB, convolution neural network, recurrent neural networks, and Long short-term memory. The findings of the experiment demonstrate that the precision, F1 score, and recall, of the suggested sentiment analysis approach are greater. The technique has been proven to be successful with highly accurate remarks. Nevertheless, the BiLSTM-based sentiment analysis technique for comments takes a while to train.

Amlan chakrabarti and Paramita Ray made a suggestion utilizing a combination of rule-based and Deep Learning (DL) techniques; aspect-level sentiment analysis is improved [20]. The communication problems have drastically evolved as a consequence of social networking sites. Material from various social media platforms may be effectively used to analyse user opinions. Therefore, the creation of a platform that can assess consumer perceptions of their goods and services using social media would be advantageous to the companies and add value to their operations. DL has gained a lot of traction in the previous few years in fields like speech recognition and picture categorization. Nevertheless, there is little study on the application of deep learning to sentiment analysis. It has been noted that the current machine learning techniques for sentiment analysis can fail to capture certain underlying elements and may not be particularly helpful. As a result, researchers suggest a deep learning method for extracting aspects from texts and analysing user sentiment in relation to those aspects. Every component of the controversial statements is tagged using a seven layer deep convolutional neural network. Researchers have combined the DL technique with a variety of regulation methods to increase the efficacy of the feature extraction method and the emotional parameter selection. By employing a present collection of aspects classifications and the clustering approach, they also attempted to enhance the current rule-based strategy to feature extraction. Researchers then compared the suggested technique to some of the most advanced systems. The accuracy obtained from the suggested approach is higher than that of the most modern techniques but the technique only uses a limited number of datasets.

Feiran Huang developed direct memory access (DMA) fusion for image-text sentiment analysis. Sentiment analyzing of social media data sets is essential for understanding public perceptions, positions, and opinions about a specific event. This technique has various uses, including the forecasting of elections and the appraisal of products. The evaluation of multimodal social media information has received less attention than the study of a single modality. The majority of the multimodal sentiment analysis techniques now in use only integrate several data modalities, which yields unsatisfactory sentiment categorization effectiveness. In the study, researchers introduce a new image-text sentiment analysis model called DMA Fusion to take use of the racist and discriminatory characteristics and intrinsic connection between semantic and visual content with combined fusion architecture. Two distinct unimodal attention approaches are suggested to develop efficient emotion classifier for the textual and visual modalities, respectively. These models are particularly intended to automatically concentrate on exclusionary areas and crucial phrases that are most connected to the sentiment. In order to take use of the individual's interaction between textual and visual signals for joint sentiment classification, an intermediary fusion-based multimodal attention approach is then developed. The 3 attention categories are then combined for sentiment predictions using a delayed fusion approach. Manually labelled and weakly labelled datasets are used to illustrate the success of the technique in several tests. However, the accuracy of the model is not greater when compared to the other techniques [21].

III. METHODOLOGY

The deep learning techniques provided here serve as the foundation for the proposed approach for forecasting the review-related emotions. Dataset collection, data pre-processing, feature extraction and classification using the DBN-GRU model, evaluation metrics, and result analysis are the steps of the proposed system. The proposed methodology's framework, which was applied in the current investigation, is exposed in Fig. 1.

A. Data Collection

A dataset created from public datasets was used in the tests to gauge the effectiveness of the suggested strategy. Since Amazon is one of biggest e-commerce sites, a huge number of reviews may be found there. We made use of data from Amazon called item data. Here three categories from Amazon's product reviews that collectively chosen which contain about 70000 product reviews: electrical reviews, mobile phone and accessory evaluations, and instrument reviews. Whereas 5000 evaluations are for instruments, 29000 are for technology, and 36,000 are for cell devices.

B. Data Pre-Processing

The process of preparing and cleaning the texts for categorization is known as pre-processing the data. The wording of product reviews typically contains a significant amount of noise and non - informative sections. It is frequently noted that the information collected by scraping might not be suitable for inclusion in an algorithm. The data that was scraped might contain misspelt words or other information that wouldn't be helpful to the algorithm. Contrarily, the bulk of the message's sentences have very little bearing on the narrative's total perceived. By keeping such phrases, the issue becomes more complex to categorize because each phrase in the texts is treated as a one-dimensional construct. The idea behind having the data properly which was before is that doing so should improve classification performance and speed up data classification, allowing for real-time sentiment classification. The several stages of the procedure include stem, dividing, deleting stop-word, and isolating special characters are shown in the Fig. 2.

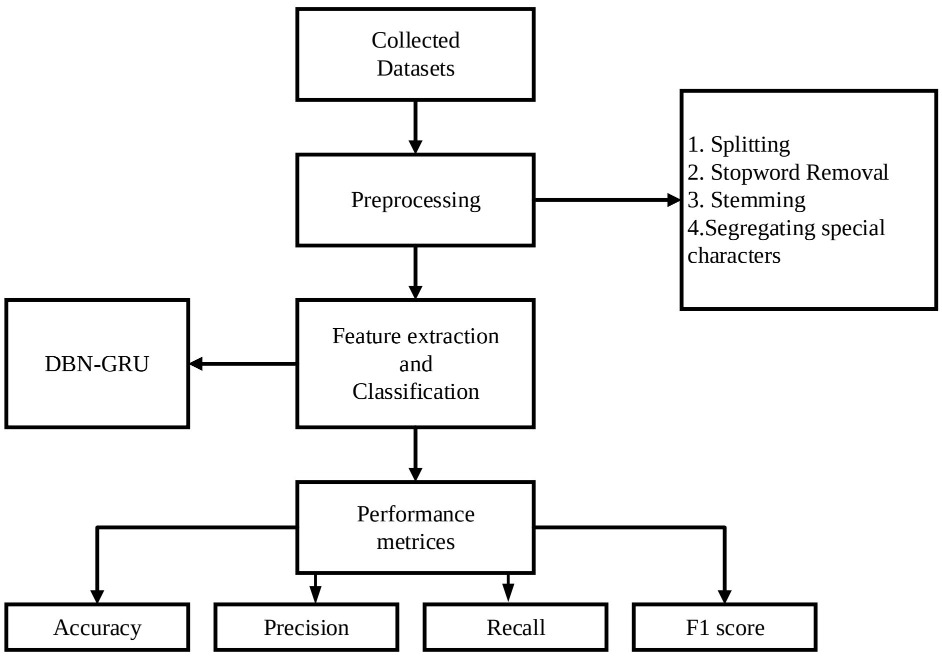

Fig. 1. Proposed model.

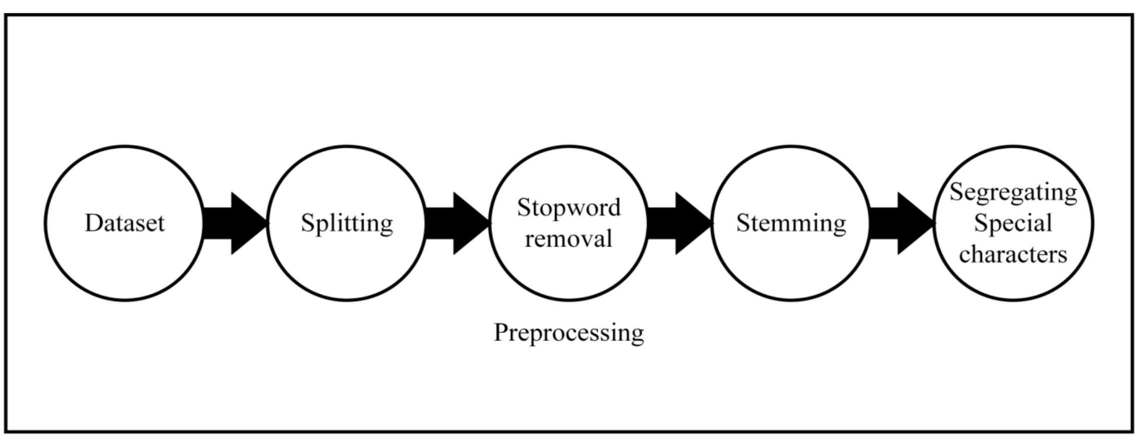

Fig. 2. Stages of Pre-processing.

1) Splitting: Dividing refers to the division of data into two or even more divisions. Evaluating the information in one portion of a two different split and training the algorithms in the second part of the divide are common practises. This method ensures the development of data structures and the activities that rely on data structures. To do this, splitting estimates a series of conditional probabilities, the sum of which is the desired outcome.

2) Stopword removal: Stop words should be eliminated to expand the enactment of the feature selection algorithm since they are frequently used and high frequency terms. The feature extraction approaches may quickly identify the remaining important words in the review corpus after the stop words removal method decreases the dimension of the data sets. High frequency stopwords include "of," "a," "she," "it," "the," "I," "he," "at," "and," and "about," among others. These words are typically referred to as "functional words" because they don't convey any sentimental content. In this experiment, we eliminate stop words to shrink the file index without affecting the accuracy of the user.

3) Stemming: Stemming is an important element in the pre-processing phase of extracting features. Each of the text's words is transformed into their stem or root form throughout this procedure. Stemming is a quick and easy method that simplifies the feature extraction process. The fundamental stemming procedure converts the words "automatic," "automate," and "automation" into the stem "automat." The prominent English language stemming algorithm is Porter's stemmer. The fundamental stemming procedure can change the words in the manner described in Table I in the following manner.

TABLE I. STEMMING

List of words	Stem form
Playing, Plays, Played,	Play
Argues, Argue, arguing, argued,	argu

4) Segregating special character: In general, special characters like the hyphen (-) and the slash (/) are separated since they don't provide any value. According to the use case, characters are eliminated. Usually, we eliminate the $ or any other currency sign if we are carrying out a task in which the currency is irrelevant (such as sentiment analysis).

C. Feature Extraction and Classification

The relevant features are extracted and classified using DBN-GRU mechanism. The proposed method, DBN-GRU approach of emotion recognition, reconstructs the positive and negative comments to relieve the emotions from the product review data to generate product recommendation which not only intervenes in the emotional direction of subjects in a targeted way, but also recommends the product based on their reviews.

1) Deep belief networks (DBN): The deep belief network is a neural system composed of many Restricted Boltzmann Machine layers, in which the outputs of one RBM serves as the inputs for the next, and the hidden layer of the preceding Restricted Boltzmann Machine serves as its visible layer. The present layer's RBM may only be trained throughout the training procedure after the final layer's RBM has been fully trained. It may be viewed as a discriminating model as well as produces better results. Unsupervised learning's goal is to minimise the dimensionality of characteristics while maintaining as many of the initial characteristics' properties as feasible. Its goal is to minimise the categorization error rate from the standpoint of supervised learning. The method of extracting features, or how to achieve a more accurate feature expression, is the core of the DBN algorithm regardless of whether unsupervised learning or supervised learning is being used. Fig. 3 depicts the unique DBN network topology.

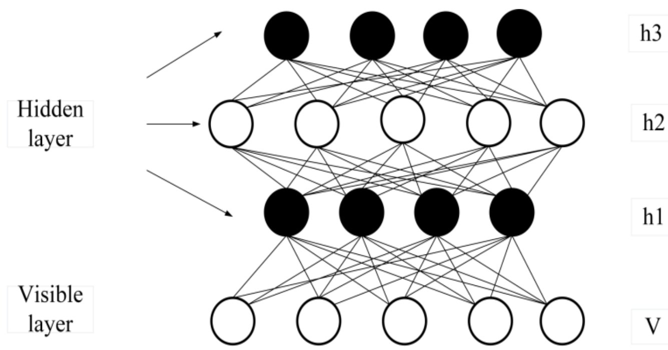

Fig. 3. DBN network structure.

2) Gated recurrent unit (GRU): In recurrent neural networks, the GRU model was most commonly employed to address the gradient vanishing problem (RNN). GRU contains three major gates and an inner cell state, making it more efficient than LSTM. Within the GRU, the data is held in a safe location. The reset gate just provides prior knowledge, but the update gate provides both previous and future information. The current memory gate utilizes the reset gate to maintain and save the necessary data from the system's previous state. The inputs modulation gate concurrently gives the input zero-mean qualities and permits the insertion of nonlinearity. The following Eq. (1) and Eq. (2) are the definitions of the fundamental GRU of rest and updated gates' mathematical formulation:

$$U_t = \sigma \left(X_t . Z_{xu} + F_{t-1} . Z_{hu} + d_u \right) \tag{1}$$

$$V_t = \sigma \left(X_t . Z_{xv} + F_{t-1} . Z_{hv} + d_v \right) \tag{2}$$

where Z_{xu} and Z_{xv} present weight parameters, while the d_v , $d_u$are biased. Fig. 4 represents the fundamental design of the GRU model.

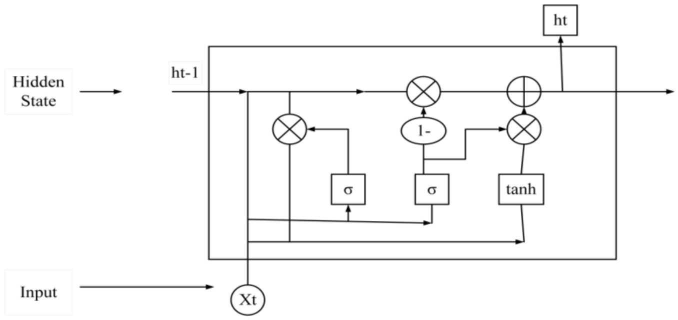

Fig. 4. The fundamental design of the GRU model.

IV. Results and Discussion

The findings of numerous tests are presented to evaluate the effectiveness of the classifier in this section. Based on accuracy, we evaluate the classifier on each of the feature metrics produced by every information extraction and compare the findings to the performance obtained by executing the classifier on unprocessed data. Utilizing the product review datasets, the sentimental analysis is done. The datasets undergo the pre-processing stage and then the pre-processed data is used for the feature extraction and classification. The feature extraction and classification are carried out by the DBN-GRU classifiers to classify the product reviews as positive, and negative.

A. *Performance Evaluation*

Evaluation metrics are crucial for gauging categorization performance. The most frequently employed tool for this is an accuracy measure. The proportion of a test dataset that is properly categorised by a classifier indicates the classifier's accuracy for that dataset. We also used some other measures to assess classifier performance because the accuracy metric alone is insufficient to provide appropriate decision-making. Measures of accuracy, precision, recall, and F1-score were used to evaluate the efficacy of the suggested technique. Also evaluated are accuracy metrics like false negative and false positive rates. The Matthews Correlation Coefficient and Negative Predictive Value are also assessed because these are the indicators that are most frequently used for classifying effectiveness. The following are descriptions of each metric's definitions:

- TP (True Positive) denotes the quantity of correctly categorised data.
- FP (False Positive) refers to the number of accurate data that was incorrectly categorised.
- The term "False Negative" (FN) refers to instances when wrong data have been classed as valid.
- TN (True Negative) refers to the classification of inaccurate data values.

1) Accuracy: The classifier's accuracy indicates how frequently it makes the right guess. The percentage of accurate forecasts to all other guesses is known as accuracy. It is shown in Eq. (3).

$$Accuracy = \frac{Tpos+Tneg}{Tpos+Tneg+Fpos+Fneg} \tag{3}$$

2) Precision: The amount of correctly classified returns is measured by a classifier's precision, or how accurate it is. Reduced false positives result from higher accuracy, whereas more false positives result from lower precision. The ratio of correctly categorised instances to all instances is known as precision. It is characterised by Eq. (4).

$$P = \frac{Tpos}{Tpos+Fpos} \tag{4}$$

3) Recall: The amount of effective data a classification produces, or its sensitivity, is determined by recall. Greater recall reduces the number of FN. Recall is the proportion of correctly categorised instances to all of the expected instances. This is demonstrable by Eq. (5).

$$R = \frac{Tpos}{Tpos+Fneg} \tag{5}$$

4) F-measure: The unified metrics known as F-measure, which is the weighted mean of accuracy and recall, is created by combining precision and recall. It is characterised by Eq. (6).

$$F\ measure = \frac{2\times precision \times recall}{precision \times recall} \tag{6}$$

5) Specificity: The percentage of favourable events that were predicted. In Eq. (7), the expression is provided.

$$Specificity = \frac{T_N}{T_N+F_P} \tag{7}$$

6) Sensitivity: The percentage of circumstances when a negative outcome was predicted. Eq. (8) presents the formula.

$$Sensitivity = \frac{T_P}{T_P+F_N} \tag{8}$$

7) False positive rate: The percentage of situations when a positive outcome was expected but turned out to be untrue. Eq. (9) presents the formula.

$$FPR = \frac{F_P}{T_N+F_P} \tag{9}$$

8) False negative rate: The percentage of cases that is positive even though they were expected to be negative. Eq. (10) presents the formula.

$$FNR = \frac{F_N}{T_P+F_N} \tag{10}$$

9) Matthews correlation coefficient and negative predictive value: One of the most often used indicators of categorization effectiveness is the Matthews Correlation Coefficient (MCC). It is widely accepted as a trustworthy approximation that can be applied although when class sizes vary significantly. The equation for the Matthews correlation coefficient is found in Eq. (11).

$$MCC = \frac{T_P T_N - F_P F_N}{\sqrt{(T_P+F_P)-(T_P+F_N)(T_N+F_P)(T_N+F_N)}} \tag{11}$$

The subject-to-outcome ratio is defined as the proportion of subjects with genuinely negative findings to all subjects with unsatisfactory results. The percentage of times that every forecast was completely wrong is known as the negative predictive value. In Eq. (12), the formula is provided.

$$NPV = \frac{T_N}{T_N+F_N} \quad (12)$$

The value of the suggested strategy is compared to that of other classification approaches in this portion of the study.

Fig. 5 shows schematically the specificity, sensitivity, and reliability of the proposed DBN-GRU in contrast to previous methods. When compared to DBN, GRU and LSTM approaches, improved DBN-GRU offers superior characteristics. Table II compares the suggested approach's Precision, Recall, and F-measure values to those of previous approaches.

TABLE II. COMPARISON OF PRECISION, RECALL AND F-MEASURE

Methods	Precision	Recall	F-Measure
Proposed DBN-GRU	0.968109	0.968109	0.968109
DBN	0.924829	0.924829	0.924829
GRU	0.920273	0.920273	0.920273
LSTM	0.835991	0.835991	0.835991

Fig. 6 displays the proposed improved DBN- GRU's FPR and FNR graphs along with the results of earlier technologies. It demonstrates that the FPR and FNR ratios of the proposed enhanced DBN-GRU are lower than those of the already employed approaches, such as DBN, GRU, and LSTM.

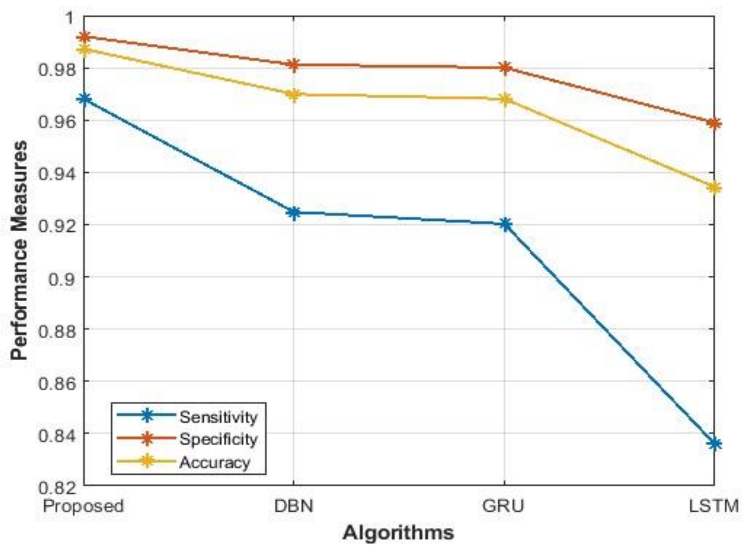

Fig. 5. Comparison graph of accuracy, sensitivity, and specificity.

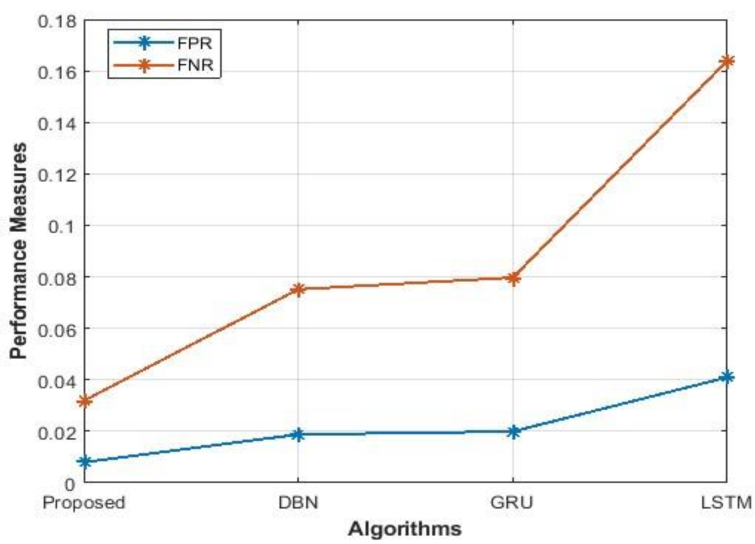

Fig. 6. Comparison graph of FPR and FNR.

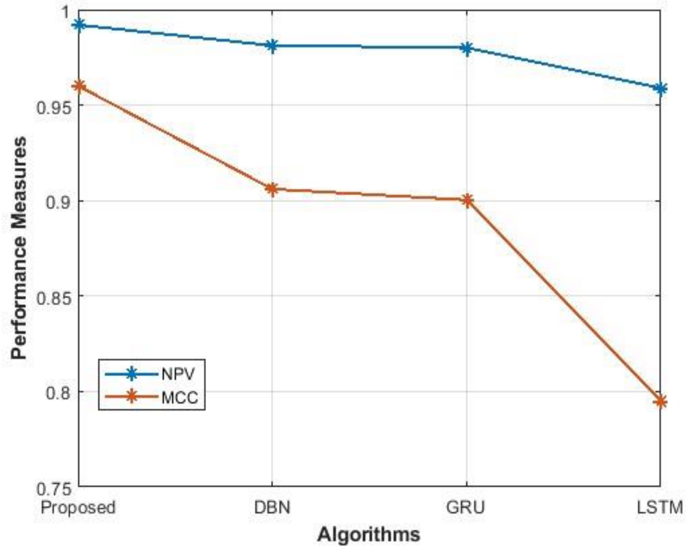

Fig. 7. Comparison graph of MCC and NPV.

The Matthews Correlation Coefficient and Negative Predictive Value graphs of the proposed augmented DBN-GRU created using conventional techniques are shown in Fig. 7. It demonstrates that the recommended enhanced DBN-GRU has a higher MCC and NPV than the current DBN, GRU, and LSTM.

Graphical representations of the recall and accuracy of the recommended technique are shown in Fig. 8 and Fig. 9. The Figures show that the proposed DBN-GRU is superior to that of the existing techniques proposed earlier.

In contrast to past methods, the recommended technique was constructed utilising an improved version of the sentimental analysis in product review data. Additionally, it leverages augmentation and has a 98.72 percent accuracy rate for sentimental analysis of product reviews.

B. Discussion

Previous sentiment analysis models, such as DBN, GRU, and LSTM, faced limitations in handling extensive and diverse datasets due to their inability to effectively address information overload and noise in data preprocessing. These models struggled with balancing precision, recall, and overall accuracy, often resulting in suboptimal sentiment classification [22]. The proposed DBN-GRU model overcomes these limitations by incorporating a comprehensive pre-processing pipeline that includes data splitting, stemming, stop word removal, and special character separation, ensuring refined and clean datasets for improved feature extraction and classification. This approach enhances the classifier's ability to accurately distinguish between positive and negative sentiments, achieving a high accuracy rate of 98.74%. However, the proposed study still has limitations, such as dependency on the quality of pre-processed data and potential overfitting due to the model's complexity. Additionally, while the model excels in the specified product review categories, its performance may vary across different domains and types of reviews, requiring further validation and potential adjustments for broader applicability. Despite these limitations, the DBN-GRU model represents a significant advancement in sentiment analysis, offering robust and reliable classification of online reviews.

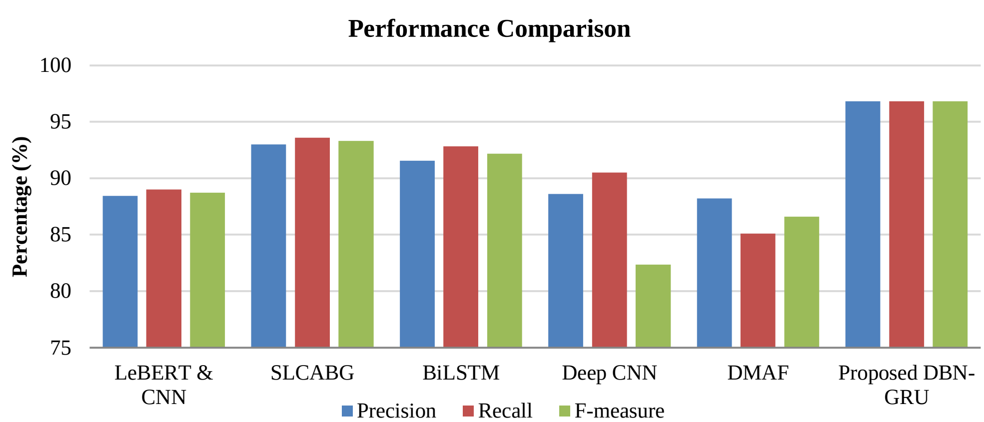

Fig. 8. Effectiveness of suggested technique over other methods.

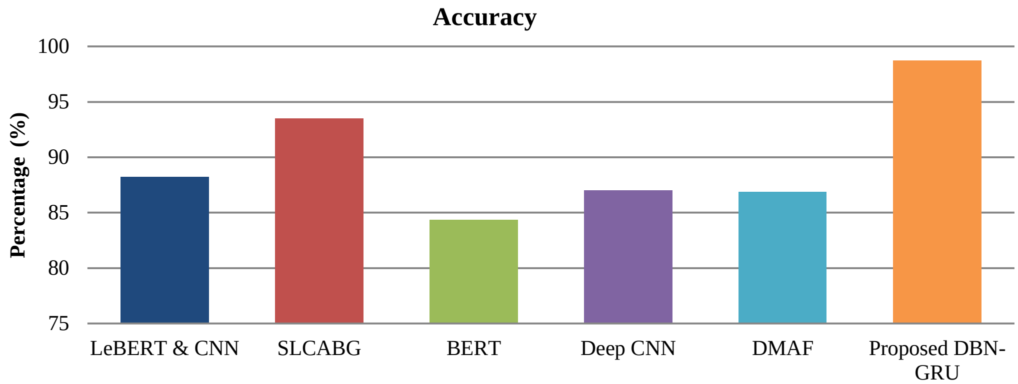

Fig. 9. Accuracy comparison of suggested technique over other methods.

V. Conclusion and Future Work

The study successfully demonstrates the effectiveness of the DBN-GRU hybrid model for sentiment analysis of product reviews. By leveraging a large dataset comprising 70,000 reviews across three categories (electronics, mobile phones, and instruments), the proposed method achieves an impressive accuracy of 98.74%. The robust preprocessing pipeline, which includes data splitting, stopword removal, stemming, and special character isolation, significantly enhances the quality of the input data, contributing to the model's high performance. The DBN-GRU model excels in extracting and classifying relevant features, outperforming traditional methods such as DBN, GRU, and LSTM in terms of accuracy, precision, recall, and F-measure. This study provides a comprehensive framework for sentiment analysis, addressing the challenges of noise and information overload in online reviews and delivering valuable insights for consumers and businesses alike.

Future research can focus on refining the DBN-GRU model to further enhance its applicability and efficiency. One area of improvement could be the expansion of the dataset to include a broader range of product categories and review languages, thereby increasing the model's generalizability. Additionally, optimizing the model for real-time sentiment analysis could be explored, enabling immediate feedback for users and businesses. Reducing the computational requirements of the model will be crucial for its deployment on edge devices, making it accessible in resource-constrained environments. Finally, integrating advanced natural language processing techniques and exploring unsupervised learning approaches could further refine the feature extraction process and improve the model's overall performance.

References

[1] H. Zhao, Z. Liu, X. Yao, and Q. Yang, “A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach,” Information Processing & Management, vol. 58, no. 5, p. 102656, 2021.

[2] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, “Sentiment analysis on product reviews using machine learning techniques,” in Cognitive Informatics and Soft Computing: Proceeding of CISC 2017, Springer, 2019, pp. 639–647.

[3] Y. Yao, M. Yang, J. Wang, and M. Xie, “Multivariate time-series prediction in industrial processes via a deep hybrid network under data uncertainty,” IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1977–1987, 2022.

[4] G. N. Tikhe and P. S. Patheja, “A Wrapper Feature Selection Based Hybrid Deep Learning Model for DDoS Detection in a Network with NFV Behaviors,” Wireless Personal Communications, vol. 133, no. 1, pp. 481–506, 2023.

[5] S. Smetanin and M. Komarov, “Sentiment analysis of product reviews in Russian using convolutional neural networks,” in 2019 IEEE 21st conference on business informatics (CBI), IEEE, 2019, pp. 482–486.

[6] S. Kausar, X. Huahu, W. Ahmad, and M. Y. Shabir, “A sentiment polarity categorization technique for online product reviews,” IEEE Access, vol. 8, pp. 3594–3605, 2019.

[7] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The impact of features extraction on the sentiment analysis,” Procedia Computer Science, vol. 152, pp. 341–348, 2019.

[8] F. Ali et al., “Transportation sentiment analysis using word embedding and ontology-based topic modeling,” Knowledge-Based Systems, vol. 174, pp. 27–42, 2019.

[9] Q. Li, C. Yu, and G. Yan, “A new multipredictor ensemble decision framework based on deep reinforcement learning for regional gdp prediction,” IEEE Access, vol. 10, pp. 45266–45279, 2022.

[10] F. Xu, Z. Pan, and R. Xia, “E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework,” Information Processing & Management, vol. 57, no. 5, p. 102221, 2020.

[11] D. A. K. Khotimah and R. Sarno, “Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM,” International Journal of Intelligent Engineering and Systems, vol. 12, no. 4, pp. 275–290, 2019.

[12] J. Jabbar, I. Urooj, W. JunSheng, and N. Azeem, “Real-time sentiment analysis on E-commerce application,” in 2019 IEEE 16th international conference on networking, sensing and control (ICNSC), IEEE, 2019, pp. 391–396.

[13] G. R. Devi et al., “COOT-Optimized Real-Time Drowsiness Detection using GRU and Enhanced Deep Belief Networks for Advanced Driver Safety.,” International Journal of Advanced Computer Science & Applications, vol. 15, no. 4, 2024.

[14] N. Shrestha and F. Nasoz, “Deep learning sentiment analysis of amazon. com reviews and ratings,” arXiv preprint arXiv:1904.04096, 2019.

[15] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, “Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning,” IEEE Access, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[16] A. Onan, “Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks,” Concurrency and Computation: Practice and Experience, vol. 33, Jun. 2020, doi: 10.1002/cpe.5909.

[17] J. Mutinda, W. Mwangi, and G. Okeyo, "Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network," Applied Sciences, vol. 13, no. 3, p. 1445, 2023.

[18] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-Dependent Sentiment Classification With BERT," IEEE Access, vol. 7, pp. 154290–154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[19] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," IEEE Access, vol. 7, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[20] P. Ray and A. Chakrabarti, "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis," Applied Computing and Informatics, vol. 18, no. 1/2, pp. 163–178, Jan. 2020, doi: 10.1016/j.aci.2019.02.002.

[21] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, "Image–text sentiment analysis via deep multimodal attentive fusion," Knowledge-Based Systems, vol. 167, pp. 26–37, Mar. 2019, doi: 10.1016/j.knosys.2019.01.019.

[22] T. Lu, Y. Du, L. Ouyang, Q. Chen, and X. Wang, "Android malware detection based on a hybrid deep learning model," Security and Communication Networks, vol. 2020, no. 1, p. 8863617, 2020.