AVE Trends in Intelligent Computer Letters

Semi-Deep Structural Neural Network for Sentimental Analysis in Twitter Text: A Computational and Linguistic Approach

G. Kalpana[1,*], A. S. Aathira[2], K. Shantichitra[3], B. Jaiganesh[4]

[1]Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.
[2]Department of English, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.
[3]Department of English, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.
[4]Department of Tamil, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.
kalpanag@srmist.edu.in[1], aathiraa2@srmist.edu.in[2], shanthik@srmist.edu.in[3], jaiganeb@srmist.edu.in[4]

*Corresponding author

Abstract: Twitter is the ultimate micro-blogging social networking tool, generating 6000 Tweets per second, or 500 million per day. Twitter generates and stores a lot of data because businesses and politicians market their brands there. Classifying and identifying tweet sentiments is sentiment analysis. To represent public opinion, sentiment analysis analyses available data to capture feelings. Twitter sentiment analysis is harder than wide sentiment analysis due to misspellings, slang, and repeated words/characters. Word emotion must be determined. SDSNN ML techniques provide sentiment analysis for current feedback/reviews on impending news-based data, making it more accurate and efficient. Region-level sentiment analysis shows how domain information impacts classification. Sentiment analysis involves data collection, processing, RST feature selection, categorisation, and prediction. An innovative feature model for tweet classification into neutral, positive, and negative categories, and for news-related public opinion extraction, is advised to identify the best Sentiment Analysis approach for large datasets. This study analyses sentiment analysis as a classification and computational linguistics problem. In the model's interpretative framework, Twitter's morphological noise, slang, and pragmatic intricacy of sarcasm are examined. It makes the SDSNN model human-centric in communication.

Keywords: Sentiment Analysis; Feature Selection; Rough Set Theory; Semi-Deep Structural Neural Network (SDSNN); Classification and Prediction; Support Vector Machine (SVM).

Cite as: G. Kalpana, A. S. Aathira, K. Shantichitra, and B. Jaiganesh, “Semi-Deep Structural Neural Network for Sentimental Analysis in Twitter Text: A Computational and Linguistic Approach,” *AVE Trends in Intelligent Computer Letters*, vol. 2, no. 1, pp. 1–12, 2026.

Journal Homepage: https://avepubs.com/user/journals/details/ATICL

Received on: 12/11/2024, **Revised on:** 02/01/2025, **Accepted on:** 01/04/2025, **Published on:** 03/01/2026

DOI: https://doi.org/10.64091/ATICL.2026.000251

1. Introduction

Data mining techniques involve extracting hidden information, analysing it, and thereby deriving valuable patterns and relationships from data gathered in centralised systems such as data warehouses to support effective analysis. Sentiment analysis can be considered the contextual mining of text, in which subjective information is detected and extracted from the source material, thereby assisting the organisation in comprehending the social emotions/sentiments associated with their brand/service by supervising online interactions. Twitter stands out as the ultimate microblogging social networking site, with around 271 million active users and a record of approximately. 500 million tweets were generated in a single day, creating a large volume of data. Sentiment analysis (opinion mining) basically works by classifying and identifying sentiments or emotions concerning every tweet. Sentiment analysis can be divided into 2 parts: feature-based or aspect-based sentiment analysis and objectivity-based sentiment analysis. Unlike general sentiment analysis, Twitter sentiment analysis tends to be more complicated, as it involves misspellings, slang, and repeated words/characters. Hence, the right sentiment for every word must be identified. The research proposes a superior and precise approach to sentiment analysis of the most recent feedback/reviews for upcoming News-based data by adopting ML (machine learning) techniques based on SDSNN (Semi-Deep Structural Neural Network). The various levels of the sentiment analysis system include data collection, data processing, feature selection via RST (Rough Set Theory), and classification and prediction.

To model relations among features, ML (Machine learning) techniques are quite significant for predictive analysis. By adopting the ML techniques, software applications can anticipate the results more accurately. In the proposed sentimental analytic system, dataset collection is performed via the Tweedy API using Java. The dataset consists of tweets in the form of news content, presented as text or emojis. Emojis are pictorial symbols that represent facial expressions, food, places, and other things. The assembled datasets undergo pre-processing, which is essential for data preparation and for eliminating unwanted Twitter data. Then comes the process of feature selection, which utilises RST (rough set theory) to determine a subset of features that are significant and useful for classification. To minimise computational complexity without affecting classification accuracy, the SDSNN-based machine learning model uses training and testing. The proposed model classifies the tweets as positive, negative, or neutral. The model is evaluated using Precision, Accuracy, F-score, and recall. Tweets were evaluated using multiple sentiment analysers to determine their accuracy. With the implementation of research on a large dataset, many features yield better results than the combination of RF (Random Forest), SVM (Support vector machine), NB (Naive Bayes), and SDSNN (Semi Deep Structure neural network).

2. Related Work

Zhao et al. [1] proposed a model, GloVe-DCNN, and an algorithm for Twitter sentiment analysis that trains a deep neural network to improve the speed and accuracy of sentiment analysis. Various works analyse lexical and syntactic features to explicitly derive sentiment features from sentiment words, exclamation marks, emoticons, and related features. The proposed work uses an unsupervised learning method to derive word embedding. The proposed model, the GloVe-DCNN, is presented for binary classification of tweets into positive or negative sentiment. The results show that the proposed model achieves a high classification accuracy. Zhang et al. [2] proposed an algorithm, ASCF (Aspect Sentiment Collaborative Filtering), that combines sentiment analysis with the fuzzy Kano model within item-based collaborative filtering (Item CF). Item CF uses a neighbourhood-based method and is a well-established collaborative filtering model, though it ignores the influence of sentiment on various aspects. ASCF identifies user attitudes toward product features by performing a precise sentiment analysis of user purchase records. Then, the user’s expectations for each feature are evaluated using the fuzzy Kano model. Experiments conducted on Amazon datasets show that ASCF significantly improves the accuracy of Item CF. Moreover, the accuracy of opinion-aware collaborative filtering is high.

Vo et al. [3] have proposed a method for extracting and summarising information about products, including corresponding opinions, within a particular domain. It is difficult to form a general opinion about products sold online. Aspect-based evaluation is an important part of opinion mining. The method for extracting the various aspects of a product follows two main stages: knowledge extraction and sentiment analysis. This method automatically extracts general syntactic knowledge and opinion-aspect relationships using NLP tools such as EDP, CR, and NER. The system has achieved an F1-score of 0.714 for camera reviews and 0.774 for laptop reviews. Yu et al. [4] propose a word-vector-refinement model for refining already available pre-trained word vectors by making use of real-valued sentiment intensity scores given by sentiment lexicons. The prevailing word embeddings (that are context-based), like Word2vec and GloVe, are generally incapable of capturing adequate sentiment information. This gives rise to words with the same vector representation but opposite sentiment polarity (e.g., good and bad), which minimises the performance of sentiment analysis. The refinement model aims to enrich each word vector to make it closer to the lexicon, towards semantically and sentimentally equivalent words. The above method offers the benefit in the sense that it can be applied to any word embeddings (that are pre-trained). It Its revealed from the experimental results that the proposed refinement model enhances the performance of the conventional word embeddings.

Qiu et al. [5] proposed a sentiment classification method for Tibetan microblogs based on multi-feature fusion. It’s observed that the sentiment analysis method struggles to analyse the sentiments/emotions of Tibetan microblogs, as they have a distinct format, particularly with features such as emoticons, speech, and grammatical relations. According to the proposed system, the theme of Weibo texts is initially determined, and the smart campus’s theme is selected to analyse the influence of each feature on the microblog's sentiment. It uses a JSON-formatted file and Tibetan POS tagging. The results show that the feature-fusion-based sentiment classification algorithm achieves higher accuracy in microblog sentiment classification. Xia et al. [6] propose a remotely monitored lifelong learning model for large-scale social media sentiment analysis. Sentiment analysis of social media texts is a complex and demanding research approach. The present model focuses on continuous sentiment learning in social media. It successively learns from past tasks, preserves the knowledge gained, and utilises it to support future learning. Firstly, various single-task classification algorithms are compared in the Entire Learning Setting. The PMI-SO Lexicon shows the worst performance, followed by NB (naive Bayes), and logistic regression performs best. Secondly, the Entire Learning and lifelong learning are being compared. With the results, it’s clear that Lifelong Bagging yields remarkable and improved performance compared to Entire Learning.

The results show that the lifelong sentiment learning approach is efficient and feasible for addressing the challenge of continuously updated dynamic-topic texts in social media. Bouazizi and Ohtsuki [7] propose a conventional multi-class classification model trained on a dataset gathered from Twitter. The vast amount of information available on social media has a deep influence on an individual’s behaviour. Multi-class sentiment analysis, ‘quantification’ signifies, determining all the prevailing sentiments related to an online post. The proposed model automatically assigns separate scores to each sentiment in a tweet, then selects the sentiments with the highest scores, as estimated and expressed in the text. To execute this task and achieve the goal, essential components are being added to the SENTA tool. The experimental outcome reveals the task's feasibility and yields an F1 score of 45.9%. Fang et al. [8] propose a multi-strategy sentiment analysis method using semantic fuzziness. The process of Sentiment analysis involves extracting opinions at the word, sentence, and document levels, resulting in article strengths and sentiment polarities. Consumers' reviews or opinions are conveyed as Chinese phrase sentiments. And, in general, Chinese characters are difficult to comprehend, so conventional machine learning models are unable to extract the article’s opinion effectively.

Multi-strategy sentiment analysis methods are being proposed that rely on SVM (support vector machine) and NB (Naïve Bayes) for document sentiment analysis. NB considers adversative conjunctions. The method is assured to be feasible and effective. Chen et al. [9] propose the WS-MDL (Weakly Supervised Multi-modal Deep Learning) scheme for scalable, powerful sentiment prediction. However, collecting an adequate quantity of training labels for training a discriminative model (for multi-modal prediction) remains a challenging issue. To train a discriminative model from low-cost existing emoticon labels for multi-modal prediction, a novel strategy, WS-DML (weakly supervised multi-modal deep learning), is being adopted. The capabilities of the proposed approach are being proven by measuring its superior performance against various state-of-the-art and other techniques. Chaitanya et al. [10] outline various steps for automated assumption analysis of Twitter data. Considering the sentiments of a tweet or a Facebook post can help reveal users’ opinions. The growth and popularity of user-generated blogs, reviews, and posts in social media and online retail demand an understanding of this data, which catalyses the development of Recommender systems and guides business plans. Sentiment Analysis uses ML (Machine Learning) and NLP (Natural Language Processing) to extract, classify, and analyse tweets for emotions or sentiments. Rani and Kumar [11] proposed a sentiment analysis approach that improves learning and teaching by analysing emotions and temporal sentiment in multilingual students’ opinions using ML (machine learning).

The proposed approach divides sentiments into two sections, i.e., positive and negative. Using the above approach, both learning and teaching at the university level can be made more effective by examining emotions, sentiments, and satisfaction factors in students’ feedback, thereby further assisting teachers and administrators in understanding challenging issues and taking appropriate measures. Designing SRSs must be effective enough to ensure teacher engagement. The proposed system output is compared with direct class performance evaluations, indicating that the system is reliable. Chelliah et al. [12] propose a method for multidimensional sentiment classification based on microblog emotion classification of Twitter data via CNN (Convolutional Neural Networks). Also, the system builds relationships among users and ranks highly popular topics on social media and microblogging platforms. People nowadays publicly post their views on services, products, political issues, social events, situations, and any other topics that strongly influence society. N-gram features on words by making use of the word-sentiment polarity score feature to form a set of tweets. The final output reveals that the Glove model using a CNN is efficient and yields high accuracy, with an astounding Figure of 87%. Gupta et al. [13] propose two ML algorithms in a hybrid format: KNN (K-Nearest Neighbours) and SVM (Support Vector Machines). Sentiment Analysis is a subfield of NLP (Natural Language Processing) that helps identify emotions/opinion/ sentiment in text. The present research work aims to identify sentiments in Twitter data, which is quite tough due to its limited size, unstructured nature, and use of slang, abbreviations, misspellings, etc. Using the proposed approach, the tweets are classified into positive, negative, and neutral sentiments. The various stages include preprocessing, feature generation, and classifier learning. The assessment reveals that the proposed hybrid scheme yields higher accuracy and F-measure than individual classifiers.

Hegde et al. [14] propose machine learning techniques and algorithms for polarity classification. Twitter has a record of nearly 6000 Tweets per second. 500 million Tweets generated in a single day. It handles and produces an abundance of data, since various organisations and political leaders use this blogging site to enhance their brand popularity and growth. Algorithms such as NB (Naive Bayes), SVM (Support Vector Machines), and Logistic Regression are used for execution. A comparison is being performed to identify the best algorithm for the existing dataset, considering the parameters of accuracy, precision, recall and F1 Score. The comparison shows that SVM accuracy is quite high. Jiang et al. [15] propose a novel approach for sentiment analysis of news events. WEAN (Word Emotion Association Network) is being established to represent the co-occurrence of semantics and emotions. In the social media environment, which generates abundant data, sentiment analysis of news events is particularly significant. Various researchers are gaining its attention to assist real-world-based applications. Still, the prevailing sentiment-computing methods mostly rely on supervised methods or standard emotion thesauri, which are non-scalable. A word emotion computation algorithm (WEAN) is recommended to acquire initial word emotions, which are then successively refined using a standard emotion thesaurus. The proposed approach yields remarkable performance, as demonstrated by results on real-world datasets for news event emotion computing.

Zhao et al. [16] propose an innovative deep learning framework, referred to as 'Weakly-supervised Deep Embedding', for performing review sentence sentiment classification. Product reviews are very significant for subsequent buyers, helping them make decisions based on the nature of the review (i.e., positive, negative, or neutral) as their primary factor. WDE-CNN and WDE-LSTM are proposed frameworks compared to WDE-LSTM, which possesses fewer model parameters. The output shows that the proposed approach achieves high efficiency and outperforms baselines. Dragoni and Petrucci [17] propose a NeuroSent tool for multi-domain sentiment analysis, utilising linguistic overlaps across domains to infer document polarity. The approaches presented in the study address the shortcomings of being implemented in domains that differ from those used to construct the opinion model. The Dranziera protocol is employed to validate the proposed technique, enabling hassle-free repetition of experiments and easy comparison of results. The outcome shows that the proposed approach is effective and provides a reasonable starting point for future work. Tago and Jin [18] examine the impact of emotional behaviours on user relationships using two dictionaries of emotional words in Twitter data. Using the keyword matching, the emotion scores are computed. Additionally, three experiments are formulated with varying settings: computing the user's average emotion score via random sampling; computing the average emotion score from all emotional tweets; and computing the average emotion score from emotional tweets, excluding users with no emotional tweets.

The Brunner–Munzel test is carried out. The output shows that a positive user tends to be highly active in building user relationships under a particular condition, whereas a negative user is less active. Chen et al. [19] employ a machine learning (ML) approach to propose a RESOLVE context-aware emotion synonym suggestion system for educational purposes. Inadequate educational materials and incompetent tools offer limited assistance to learners in mastering emotion words. With the help of the proposed system, synonymous emotion words are being suggested that are well-suited to learners' contexts. Moreover, usage information for all emotion words, definitions, scenario descriptions, and example sentences is provided, which aids language learners in building vocabulary and simplifying their use of words. Parlar and Ozel [20] propose a novel feature selection method, Query Expansion Ranking, that relies on query expansion term weighting. Given the abundance of review documents, researchers have developed other feature selection methods to remove irrelevant or insignificant features. Query Expansion Ranking is being compared with feature selection methods such as the Chi-Square method and Document Frequency. The results indicate that the new feature selection method yields better classification accuracy than the Document Frequency Difference and Chi-Square methods.

3. Proposed Work

3.1. Overview

Data mining techniques involve extracting patterns and relationships from large datasets to resolve issues through data analysis. The most popular one is predictive data mining, which has the greatest direct business applications. Figure 1 shows that SA (Sentiment Analysis) is a sub-process of NLP (Natural Language Processing) and a very popular text classification tool that identifies the emotion/sentiment/opinion hidden within a text. It examines the incoming message and determines whether its sentiment is positive, negative, or neutral. Firstly, the dataset is collected via the Tweedy API. Tweets may be in text or emoji form. The assembled datasets undergo pre-processing, which is essential for data preparation and for eliminating unwanted Twitter data. Then comes the process of feature selection, which utilises RST (rough set theory) to determine a significant subset of features and discard the rest. To minimise computational complexity without affecting classification accuracy, the SDSNN-based machine learning model uses training and testing. A mathematical model of the sample data (referred to as training data) is being generated using ML algorithms. Using the training data, the classifiers are trained, and a test dataset is fed to the model to achieve the required output for decision-making, without explicitly programming. In SA, target-based sentiment analysis is a basic process that analyses the sentiment polarity of various targets within a single sentence. The tweet sentence is divided into words to compute its polarity, which can be positive, negative, or neutral. This leads to the prediction of the tweet polarity.

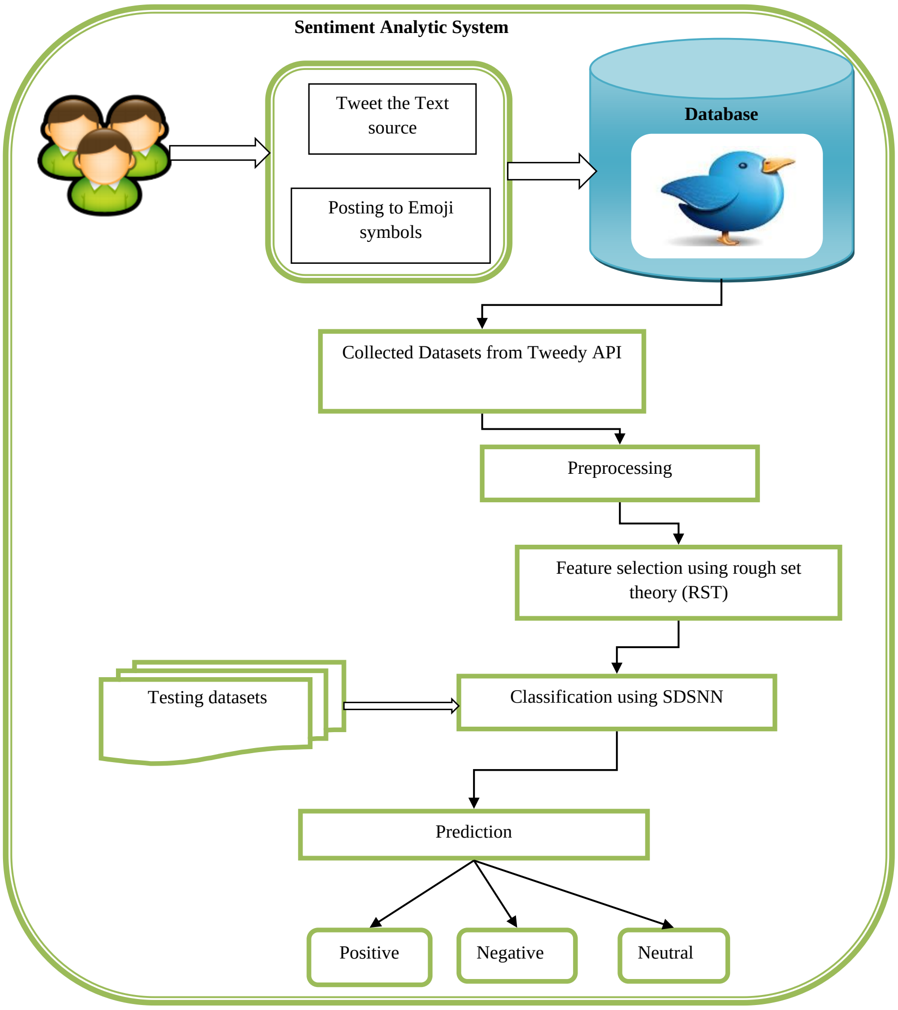

Figure 1: Proposed system architecture

3.2. Datasets Collection

Twitter is used to assemble a stream of tweets using a Twitter Programming interface. The tweet data can be structured, unstructured, Semi-Structured, or a Combination. Dataset collection is done via the Tweedy API using Java. The dataset consists of tweets in the form of news content, presented as text or emojis. An emoji is a pictorial symbol (like smileys and ideograms) used on webpages and in e-mails.

3.3. Pre-Processing

Pre-processing signifies eliminating insignificant data. There may exist irrelevant words in the assembled tweet data. This unwanted data can burden the analysis process. So, it's important to eliminate such data to improve the value of the information. The following are the steps involved:

- **Case Conversion:** Converting the entire set of words either in upper-case or lower-case to avoid the difference between say "Content" and "content" for further preparation.
- **Stop-Words Evacuation:** All the insignificant common words, which don't aid in determining any opinion, such as an, the, like, a, had, have, etc., are eliminated from the input content.
- **Accentuation Evacuation:** Accentuation symbols such as commas (,) or colons (;) are also inexpressive for the analysis; hence, they too need to be removed from the content.
- **Stemming:** Stemming represents trimming off the start or end of the word, considering a set of generally used prefixes and suffixes existing in the modified word.

- **Lemmatisation:** In contrast, lemmatisation considers the morphological analysis of words. Lemmatisation substitutes words for their base forms, called lemmas.
- **Spelling Adjustment:** For the erroneous words, the spellings can be ended via robotised determination of more probable words.

3.4. Feature Selection Using Rough Set Theory

Feature selection remains a challenging concern for any pattern classification task. Feature selection selects an optimal subset of features to improve performance and minimise data dimensionality. These features can be wrapper-, filter-, hybrid-, or embedded-based. The research uses filter-based techniques to rank and select the top features. Since these techniques are scalable and fast, they are appropriate for tweet sentiment classification. As sentiment classification is performed on large-scale datasets, scalability and speed are vital. There are thousands of features derived from Twitter data, resulting in an enlarged feature set. Therefore, a suitable feature selection method is required to minimise dimensionality. Feature selection uses RST (rough set theory) to determine a subset of features that are significant and useful for classification, thereby minimising computational complexity without affecting classification accuracy. RST effectively addresses and resolves vagueness and indefiniteness. Also, RST helps minimise attribute redundancy. With numerous attributes, their combination can lead to a combinatorial explosion, making the analysis process impossible. Hence, RST is utilised to eliminate unwanted attributes.

3.5. Rough Set Theory

RST effectively addresses and resolves vagueness and indefiniteness. It doesn't require any additional parameters or information to analyse data. A decision or information Table structures the rough set.

3.5.1. Rough Set Algorithm

Input: A decision Table ($DT = (U, C \cup D)$)
Output: A significant reduction of attributes (R)

1: $R \leftarrow \{\}$
2: repeat
3: $T \leftarrow R$
4: for $x \in (C - R)$ do
5: if $\gamma R \cup (D) > \gamma T\ (D)$ then
6: $T \leftarrow R \cup x$
7: end if
8: $R \leftarrow T$
9: end for
10: until $(D) == \gamma C\ (D)$

3.6. Classifications

It involves a training and a testing set. Datasets assembled from various sources, whose values and behaviour are well known, are referred to as training datasets. Where else are the testing datasets, the ones with unknown values or behaviour? Training data is used to train the classifiers, and thereafter, test/unknown data is fed into the model to achieve the expected outcome. SDSNN (a machine learning technique) is used in the classification process to classify sentences into layers.

3.7. Semi-Deep Structural Neural Network (SDSNN)

Machine learning (ML) is one of the AI (artificial intelligence) applications that enable systems to learn and improve automatically from experience without being explicitly programmed. SDSNN (a machine learning technique) utilises multiple hidden layers. Every layer learns features at various abstraction levels, which are used to improve the output. Every tweet/sentence traverses multiple layers to determine its meaning. NNs (neural networks) effectively resolve classification problems with complex data. Using this model, sentiments/responses (positive or negative) can be predicted.SA (sentiment analysis) identifies targets across similar sentences, extracting sentiment features from the entire sentence and monitoring information transmission via separate weight matrices to comprehend the sentiment polarities of these targets. The tweets can be classified as either positive, negative or neutral. SA analyses the sentiment polarity of various targets in similar sentences. Every word is assigned a positive or negative polarity. Each tweet is divided into words, and the polarity of each word is computed. Next, the sum of all positive and negative word polarities is computed. Subsequently, a comparison between positive and negative is performed.

3.8. Polarity Calculation and Sentiment Analysis

Sentiment analysis provides valuable insights from social media by identifying opinions and emotions from large volumes of unstructured data. The three polarity classes offered by SA include: neutral, positive and negative. A tweet's polarity is decided by assigning it a score from -1 to 1, depending on the usage of words. A positive score indicates positive sentiment, a negative score indicates negative sentiment, and a zero indicates neutral sentiment.

3.9. Prediction

Sentiment polarity classification signifies labelling an opinionated text and classifying it depending on neutral, positive and negative classes. SDSNN (a machine learning technique) classifies sentences or tweets as positive, negative, or neutral based on their polarity. Eventually, the polarity of news reviews is expected to depend on the sentence weights. Altogether, 800 tweets were assembled from the database, of which 500 were positive, 200 were negative, and 100 were neutral.

3.10. Performance Evaluation

The dataset is evaluated based on the following criteria:

- **Accuracy:** The ratio of correctly predicted observations to the total observations, and is the most intuitive performance metric.
- **Precision:** Represents the ratio of correctly anticipated positive observations against the overall anticipated positive observations.
- **Recall:** Represents the ratio of correctly anticipated positive observations against the overall observations in the actual class.
- **F1 Score:** Represents the weighted average related to Precision and Recall. The F1 score tends to be more significant than accuracy in data with irregular distributions.

The equations related to accuracy, precision, recall, and F1 Score are depicted in equations 1, 2, 3, and 4, respectively:

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (1)$$

$$\text{Precision} = \frac{t_p}{t_p + f_p} \quad (2)$$

$$\text{Recall} = \frac{t_p}{t_p + f_n} \quad (3)$$

$$\text{F1-score} = \frac{2 * recall * precision}{recall + precision} \quad (4)$$

Here, tp signifies true positive, which precisely predicts positive values; tn represents true negative, which precisely predicts negative values; fp denotes false positive, which signifies a falsely predicted positive class; and fn denotes false negative, which signifies a falsely predicted negative class.

4. Result and Discussion

There has been a significant rise in opinion mining and sentiment analysis, which focus on discovering the sentiments/opinions expressed in texts available across multiple social media sites, using ML (machine learning) techniques that compute polarity scores and perform sentiment or subjective analysis. Although numerous ML (machine learning) approaches and algorithms exist for performing sentiment analysis during elections, there is an urgent need for a state-of-the-art approach. To address these concerns, the research adopts the SDNN (Semi Deep Structure Neural Network) approach, which incorporates a sentiment analyser. Carrying out SA (sentiment analysis) helps identify a user's opinion through text analysis. Relying upon the sentiment score for a particular tweet, the tweets are classified as positive and negative. As only a particular domain is chosen, subjective and objective tweets need not be analysed separately. The existing section presents the experimental details of the research, along with the results discussion. SA (sentiment analysis) focuses on evaluating the accuracy of sentiment analysers by examining tweets from multiple analysers. The existing research work employs a large dataset and numerous features, yielding better results than SVM (Support Vector Machine), NB (Naive Bayes), RF (Random Forest), and SDSNN (Semi Deep Structure neural network).

Table 1: Comparison algorithms

No.	Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Time (MS)
1	Support Vector Machine	93.02	88.	95	88.90	1.48
2	Naïve Bayes	86.34	83.76	92	77	1.80
3	Random Forest	72	84.56	89	83	2.65
4	Semi-Deep Structural Neural Network	96.20	93	95	91	0.53

Table 1 compares the performance of the proposed techniques with SVM (Support Vector Machine), NB (Naive Bayes), and RF (Random Forest). The proposed SDSNN achieves better performance than the other existing techniques.

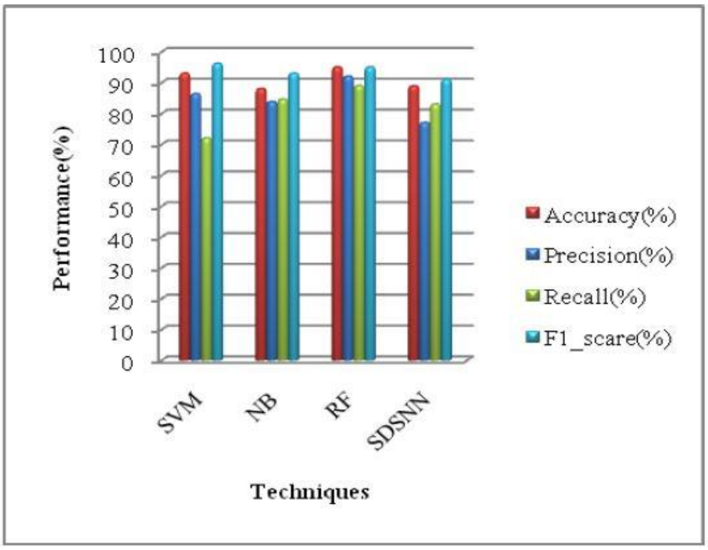

Figure 2: Comparison of sentiment analysis in overall performance

Processing time is also being evaluated. Figures 2 and 3 compare the performance of SDSNN (Semi Deep Structure neural network) with SVM (Support vector machine), NB (Naive Bayes), and RF (Random Forest). The SDSNN approach achieves better performance than the other existing techniques. Also, the work reports the F1-score, precision, recall, and time values.

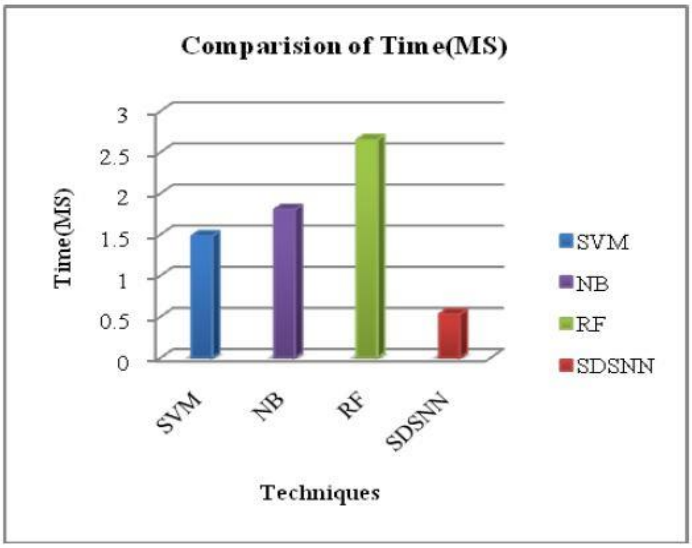

Figure 3: Comparison of processing time

4.1. Linguistic and Contextual Interpretation Framework

The Semi-Deep Structural Neural Network (SDSNN) provides a competitive computational engine for classification; its effectiveness depends on a deep understanding of the linguistic features prevalent in microblogging environments like Twitter. The section delves into how the qualitative nature of the data and the language tricks the model uses to achieve sentiment accuracy. It outlines a framework that unravels the challenges posed by the unique way people use language on Twitter and demonstrates how the SDSNN model's structure silently resolves these challenges.

4.2. Linguistic Analysis of Tweets: The Nature of "Twitter-speak"

The need for brevity, from 140 to 280 characters, has led to the development of a unique, condensed language, termed "Twitter-speak." This language is casual, unique, and distinguished by anomalies from traditional writing. Linguistic analysis focuses on three key elements:

4.2.1. Morphological Variation and Noise

Individuals use unconventional spellings to add emphasis, as in "stooopid" instead of "stupid," and for convenience, as in "u" for "you" and "gr8" for "great." Preprocessing techniques, such as stemming and lemmatisation, are critical in addressing these anomalies. These techniques minimise the effect of morphological noise by reducing words to their root, as in "happier," "happiest," and "happyyy" being mapped to "happy." This allows the model to identify the word and its sentiment, reducing the effect of unconventional spellings:

- **Syntactic Sparsity and Ellipsis:** The emphasis on conciseness results in syntactic sparsity, with tweets lacking a subject, auxiliary verb, or an article. The tweet "Absolutely loved the event. Best day ever!" lacks a subject. Conventional parsers are not effective with such sparsity. However, the SDSNN's use of several hidden layers is effective for learning non-traditional syntactic patterns. The SDSNN does not need a grammatically correct sentence. Instead, it uses its ability to learn to connect a sequence of words to capture the meaning hidden within such sentences.
- **Orthographic Signals:** In addition to words, Twitter users use orthographic signals to express their sentiment. Emoticons such as :) or :((), emojis such as 😂 or 😠, and repetition of punctuation marks to express emphasis, such as "This is amazing!!!," are common. The SDSNN's input layer includes these as valid words to be integrated into sentiment.

4.3. Understanding Sarcasm, Slang, and Contextual Sentiment

The challenges in Twitter sentiment analysis include effectively handling sarcasm and keeping up with the ever-evolving slang. These features can completely invert or obscure the meaning of a sentence at the literal level:

- **The Sarcasm Challenge as a Sentiment Flipper:** Sarcasm as a literary and linguistic tool is employed where the intended meaning is the opposite of the literal meaning that is communicated. For example, when a tweet "Great, another rainy day. Just what I needed." uses the positive word "great" to express a negative sentiment. The SDSNN addresses this challenge by giving greater weight to contextual analysis. By considering the complete sequence of words and their associated Parts of Speech tags, the model detects incompatibilities. The model adapts to identify patterns involving negative and contrast patterns. If a highly positive word (like "great" or "love") appears in proximity to a negation ("not," "don't") or within a context of negative words ("rainy," "another"), the SDSNN's features effectively imply inverting the polarity score of that phrase, classifying it as negative.
- **Slang and the Dynamic Lexicon:** These words are temporal and unique to a certain group. For example, words such as "lit," "sick," or "fire" may be associated with positive attributes in an informal setting. This is a significant deviation from their dictionary definitions. The effectiveness of SDSNN lies in its ability to utilise semantic embeddings. Instead of relying on a dictionary definition that classifies words like "fire" as dangerous, it learns from its vast training data. The word "fire" appears in many tweets with positive emoticons (🔥) alongside words such as "amazing" and "awesome." This causes its representation to shift towards positive words in its embedding space.

4.4. Manual Annotation and Validation of Sentiment Labels: Establishing Ground Truth

The success of any supervised learning model depends on the quality of the training data it uses. When it comes to something as nuanced as sentiment analysis, it's not possible to get by on algorithmic labels alone. "Human Intuition Remains Essential." To ensure that the SDSNN was trained on a credible dataset, the data used for training was carefully selected:

- **Human-in-the-Loop for Data Quality:** From the total dataset, 800 tweets were selected for human labelling. The dataset was carefully curated, comprising 500 positive, 200 negative, and 100 neutral tweets. This is an important part of the process because it ultimately determines the ground truth, meaning labels that are based on human interpretation and are the standard against which all other labels are judged.
- **Validation and Teaching Distinction:** The human validation process serves two purposes. It helps validate the system-generated labels against potential human error and bias in the initial data collection process. It's also the main way in which the SDSNN is "taught" the non-literal, pragmatic aspects of language that are discussed in Section 4.2. So, in the case of the sentence "Great, another rainy day," the human would label it as negative, emphasising that the real meaning is not contained in the word "great."

4.5. Discourse and Pragmatics Analysis: Meaning Beyond the Sentence

On the one hand, pragmatics is the study of how context influences what a speaker or writer means. When analysing tweets related to the news, this is more important than ever, as words can take on a vastly different meaning based on what's surrounding them:

- **News-Specific Sentiment Shift:** This paper focuses on "upcoming News-based data." In journalism, negative words can signal high engagement or interest, indicating positive sentiment toward a news organisation. An example would be the use of the word "shocking" in the tweet "Shocking new development in the election." The word itself has a negative connotation, yet the ultimate goal of this tweet is to inform and pique the reader's interest. This is the sentiment this paper seeks to analyse. The SDSNN's ability to analyse sentiment for a given target enables this. The SDSNN will learn to distinguish between the event's sentiment and the sentiment of reporting on the event.
- **Target-Based Sentiment Resolution:** Tweets can convey multiple sentiments depending on the target. An example would be the tweet "I love the way this reporter is covering this tragic event." The SDSNN's ability to analyse this tweet will allow it to identify positive sentiment for the word "reporter" and negative sentiment for the word "tragic event," demonstrating a sophisticated understanding of discourse.

4.6. Interpretation of Results from a Language Perspective

When analysing the results from a linguistic perspective, the SDSNN's high accuracy, such as the 96.2% noted in Table 1, demonstrates not only a high accuracy rate but also a sophisticated understanding of the complex linguistic landscape on Twitter:

- The high accuracy rate shows that the SDSNN has learned to generalise across the lexicon and syntax described earlier.
- The 3.8% Margin of Error: The boundary of semantic fuzziness. This remaining sliver of error is no mistake; it's a reflection of where current computational linguistics is today. Linguistically, this error marks the boundary of what current models can process, not a failure to process. This is probably an effect of extreme ambiguity, situations so dependent on context, shared world knowledge, or interpretation that even state-of-the-art neural networks are unable to identify patterns in.

This includes:

- The faint irony of a sarcasm so subtle that only knowledge of a certain event or person will allow you to pick up on it.
- The culture-specific reference to a certain slang term or meme that's only popular in a very small online community.
- The high-stakes pragmatics of a tweet that's dependent on the connection to a live event in real-time news, a connection that only comes from being aware of the world in a certain way.

5. Conclusion

The architectural design of the SDSNN shows remarkable potential to address the major linguistic challenges in Twitter sentiment analysis. The low yet steady rate of errors points to the fundamental characteristics of human language, such as creativity, ambiguity, and contextual complexity, which still pose fundamental challenges to the field of artificial intelligence.
The process of classifying opinions and sentiments in the inherently noisy and dynamic environment of Twitter data poses a two-fold challenge: both computational and linguistic. The results suggest that to resolve this challenge, a synergistic interplay between machine learning architectures and a sophisticated understanding of the linguistic characteristics of human communication is a necessity. The use of the Semi-Deep Structural Neural Network (SDSNN) allows the research to transcend

the complexities of “Twitter-speak” and interpret sentiment within its appropriate linguistic context. The SDSNN model demonstrated high proficiency in handling morphological variation, syntactic sparsity, and orthographic cues, such as emojis and slang, typical of microblogging discourse. The SDSNN model's ability to learn distributed representations of meaning was instrumental in its capacity to respond to the changing nature of slang dynamically and to begin making inroads into the difficult problem of sarcasm, where the literal meaning of the text is frequently inverted. The inclusion of the manually annotated "gold standard" dataset was seen as instrumental to the process, as it provided the vital human-in-the-loop validation needed for the SDSNN model to learn in the context of pragmatic discourse.

The framework provided a more sophisticated target-based analysis of sentiment, an essential prerequisite for understanding the complex interplay of emotion in news-related microblog discourse. The quantitative results demonstrate the effectiveness of the SDSNN framework for sentiment analysis. The SDSNN methodology achieved a high accuracy of 96.23%, well above that of conventional machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF). The high accuracy of the SDSNN is directly attributable to the model's ability to generalise across linguistic noise and extract useful patterns from the text. The small margin of error is seen as an indicator of the frontier of the problem of semantic fuzziness, where even the most advanced models have difficulty with the extreme irony, culture-specific references, and deep contextual dependencies of human language. The SDSNN framework proves highly appropriate and effective for predictive sentiment analysis on Twitter, thanks to its ability to implicitly account for the platform's linguistic environment. The direction for future research includes not just the extension of the model to carry out a prediction on news topic classification, but also its linguistic capabilities. The future direction of research will include integrating more explicit pragmatic reasoning and extending the dataset to include a broader range of cultural references, to close further the gap between computational pattern recognition and true human understanding.

Acknowledgement: The authors express their sincere gratitude to SRM Institute of Science and Technology, Kattankulathur, for providing the necessary support, facilities, and academic environment for this work. The authors also acknowledge the guidance of faculty members and the encouragement of family and peers in the successful completion of this study.

Data Availability Statement: The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request. Supporting data are included in the manuscript.

Funding Statement: None.

Conflicts of Interest Statement: The authors declare no conflicts of interest that could potentially influence the research findings.

Ethics and Consent Statement: The authors of this work consent to its availability to all interested readers for their learning and enjoyment.

References

1. J. Zhao, X. Gui, and Z. Xuejun, “Deep convolution neural networks for Twitter sentiment analysis,” *IEEE Access*, vol. 6, no. 1, pp. 23253-23260, 2017.
2. J. Zhang, D. Chen, and M. Lu, “Combining sentiment analysis with a fuzzy Kano model for product aspect preference recommendation,” *IEEE Access*, vol. 6, no. 10, pp. 59163-59172, 2018.
3. A. D. Vo, Q. P. Nguyen, and C. Y. Ock, “Opinion aspect relations in cognizing customer feelings via reviews,” *IEEE Access*, vol. 6, no. 1, pp. 5415–5426, 2018.
4. L. C. Yu, J. Wang, K. R. Lai, and X. Zhang, “Refining word embeddings using intensity scores,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 671–681, 2017.
5. L. Qiu, Q. Lei, and Z. Zhang, “Advanced sentiment classification of Tibetan microblogs on smart campuses based on multi-feature fusion,” *IEEE Access*, vol. 6, no. 1, pp. 5415–5426, 2018.
6. R. Xia, J. Jiang, and H. He, “Distantly supervised lifelong learning for large-scale social media sentiment analysis,” *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 480–491, 2017.
7. M. Bouazizi and T. Ohtsuki, “Multi-class sentiment analysis in Twitter: What if classification is not the answer,” *IEEE Access*, vol. 6, no. 10, pp. 64486-64502, 2018.
8. Y. Fang, H. Tan, and J. Zhang, “Multi-strategy sentiment analysis of consumer reviews based on semantic fuzziness,” *IEEE Access*, vol. 6, no. 4, pp. 20625-20631, 2018.
9. F. Chen, R. Ji, J. Su, D. Cao, and Y. Gao, “Predicting microblog sentiments via weakly supervised multi-modal deep learning,” *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 997-1007, 2017.

10. G. K. Chaitanya, D. R. Meka, V. S. Vamsi, and M. V. S. R. Karthik, “A survey on Twitter sentiment analysis with machine learning techniques,” *International Journal of Engineering and Technology*, vol. 7, no. 2.S32, pp. 462–465, 2018.
11. S. Rani and P. Kumar, “A sentiment analysis system to improve teaching and learning,” *Computer*, vol. 50, no. 5, pp. 36–43, 2017.
12. B. J. Chelliah, D. Lathia, S. Yadav, M. Trivedi, and S. S. Soni, “Sentiment analysis of Twitter data using CNN,” *International Journal of Emerging Technologies in Engineering Research (IJETER)*, vol. 6, no. 4, pp. 97–103, 2018.
13. A. Gupta, J. Pruthi, and N. Sahu, “Sentiment analysis of tweets using machine learning approach,” *International Journal of Computer Science and Mobile Computing*, vol. 6, no. 4, pp. 444–458, 2017.
14. B. Hegde, H. S. Nagashree, and M. Prakash, “Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 6, pp. 1874–1880, 2018.
15. D. Jiang, X. Luo, J. Xuan, and Z. Xu, “Sentiment computing for the news event based on the social media big data,” *IEEE Access*, vol. 5, no. 9, pp. 2373–2382, 2016.
16. W. Zhao, Z. Guan, L. Chen, X. He, D. Cai, and B. Wang, “Weakly-supervised deep embedding for product review sentiment analysis,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 185-197, 2017.
17. M. Dragoni and G. Petrucci, “A neural word embeddings approach for multi-domain sentiment analysis,” *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 457-470, 2017.
18. K. Tago and Q. Jin, “Influence analysis of emotional behaviors and user relationships based on Twitter data,” *Tsinghua Science and Technology*, vol. 23, no. 1, pp. 104–113, 2018.
19. M. H. Chen, W. F. Chen, and L. W. Ku, “Application of sentiment analysis to language learning,” *IEEE Access*, vol. 6, no. 5, pp. 24433-24442, 2018.
20. T. Parlar and S. A. Ozel, “A new feature selection method for sentiment analysis of Turkish reviews,” in *Proc. 2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, Sinaia, Romania, 2016.