International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Subjectivity Detection and Semantic Analysis for Opinion Mining using SBNDNN

Mekala Susmitha[1*], Dr. Shaik Razia[2]

Submitted: 26/12/2023 **Revised**: 06/02/2024 **Accepted**: 15/02/2024

Abstract: Given the increased accessibility and popularity of resources with a diversity of opinions, such as personal blogs and online review sites, employ technological advances effectively which learns about the beliefs of others. A sentiment analysis (SA) which defines a text as opinionated or non-opinionated is known as subjectivity. In this paper, we proposed the subjectivity detection and semantic analysis for opinion mining using SBNDNN (Sentiment based normalized deep neural network). Initially under goes the process of pre-processing, and then attains the process of subjectivity detection and feature extraction. Finally classification takes place by means of Sentiment based normalized deep neural network used for the sentiment classification of opinion Analysis to predict whether it's positive, negative or neutral002E.

Keywords: *Subjectivity Detection, Data pre-processing, Semantic Analysis, feature extraction, support value, Sentiment based normalized deep neural network*

Introduction

One of the main task in natural language processing (NLP) is sentiment analysis, which is also known as opinion mining [1] and involves extracting attitudes, evaluations, thoughts, views, or judgements about a specific subject. A significant amount of user-generated, unstructured content that includes opinions and thoughts is available online as a result of social media advancements. For private individuals, corporate entities, and governmental bodies making decisions, the ability to recognise emotion might be crucial. Organisations, decision support systems, and individuals can all benefit greatly from understanding public opinions about policies, goods, and organisations [2]. For the purpose of comprehending and producing human language, NLP, a branch of Artificial Intelligence (AI), blends linguistic and computational language foundations. To demonstrate language and extract insightful knowledge from it, NLP is made up of a variety of tasks that are each focused on a different component of language, including information extraction, text summarization, machine translation, and argument mining [3].

Online forums have proliferated dramatically during the last few decades. Additionally, the emergence of social media (such as Facebook, YouTube, and Twitter) has allowed people to often express their ideas. Social media encourages conversation and allows users to express their opinions on a variety of problems and issues. An online medium offers a forum for idea exchange and entices the general public to participate in group conversations. Social media also enables businesses and organisations to get product reviews in the form of texts, photographs, and videos [4–7].In NLP-based techniques, it is estimated what Parts of Speech (PoS) are used in each sentence and how they relate to one another syntactically. The utilisation of domain-neutral lexical resources, such as SentiWordNet, WordNet, and General Inquirer, is crucial in lexicon-based techniques. Additionally, systems that use lexical-based techniques typically use pre-made glossaries that contain the phrases connected to the polarity score. A database of positive and negative words, each tagged with a determined previous polarity score, is referred to as a semantic lexical resource. Senti- WordNet (SWN) is the most popular lexicon currently in use since it converges all three labels—Objective, Subjective, and Negative—assigned to each synset [8].

One of these areas is Subjective Sentiment Analysis (SSA), also known as Opinion Mining (OM), which is a challenge for NLP that tries to identify, extract information from texts, categorise them into groups, and identify the subjectivity and sentiments present in opinions. Numerous contemporary NLP studies have centred on emotion recognition in recent years. Arabic Subjective Sentiment Analysis (ASSA) is a study area that has lately attracted a lot of interest because the majority of published studies in SA were in English and some in Arabic [9]. The extraction of usable information from a vast volume of data, however, necessitates the application of sophisticated and effective processing tools. In many sentiment analysis systems

[1] Research Scholar, Department of CSE, KoneruLakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh, India
[1]Working as Assistant Professor, Department of Information Technology, VNRVJIET,Bachupally,Hyderabad,Telangana
ORCID ID : 0000-3343-7165-777X
[2] Associate Professor, Department of CSE, KoneruLakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh, India
ORCID ID : 0000-3343-7165-777X
** Corresponding Author Email: 183030081@kluniversity.in*

nowadays, Deep Learning (DL) techniques, a branch of Machine Learning (ML), play a significant role [10] and numerous types of research have begun to examine them to enhance the process of manipulating data [11,12].

Another crucial area of research in SA is the identification of subjectivity and objectivity, which entails determining whether a given textual source is objective or subjective. The task of subjectivity/objectivity identification is more difficult than the task of polarity categorization. This problem arises because the subjectivity of words and phrases may rely on their context, allowing subjective statements to be included in an objective document and vice versa. Additionally, the concept of subjectivity used while annotating texts substantially influences the outcomes [13].

A subjective sentence reveals one's own sentiments or beliefs, whereas an objective sentence conveys some real information about the outside world. For instance, the statement "This past Saturday, I bought a Nokia phone and my girlfriend bought a Motorola phone" is objective because it doesn't include any subjective information. However, the statement "The voice on my phone was not so clear, worse than my previous phone" is subjective. There are many different types of subjective expressions, including judgements, accusations, aspirations, convictions, beliefs, suspicions, and conjectures. As a result, an opinion may not be present in a subjective sentence [14, 15].This paper's primary contribution is:

- The pre-processing step using the methods of Tokenization, stop word removal, removal of special symbol and user name removal.
- After subjectivity detection is used to predict subjective or objective.
- Feature extraction includes the senti word net, unigram, polarity and lexicon.
- Finally, the sentiment classification using SBNDNN and indicates positive polarity or negative polarity or neutral.

The following is about the structure of the paper : In section 2 gives the Related Work, Section 3 explains the proposed Subjectivity Detection and Semantic Analysis technique, The experimental results are discussed in section 4, and finally the conclusion is presented in part 5.

1. Related Work

To anticipate the elements mentioned in a phrase and the emotions connected to each one is a more sophisticated and general problem that Bo Wang et al [16] have proposed. Aspect-based sentiment analysis (ABSA) is the name of this all-encompassing endeavour. Since 2014, an ABSA challenge has been added to the SemEval competition. The majority of successful models among submissions from the previous two years employ support vector machines (SVM). This work uses deep neural networks to complete this task, riding the recent deep learning trends. We create a hybrid model that combines sentiment and aspect prediction. Using deep learning models, we surpass or come close to state-of-the-art performance for both predictions.

A new method with a set of guidelines for merging social network reviews (Twitter reviews, Facebook reviews), blog reviews (Amazon reviews), and opinion mining has been created by Uma Maheswari et al. [17]. Then, the reviews are analysed to give new clients and organisations clear information about reviewer opinions on products. The updated information will help customers choose wisely when purchasing goods. By providing customers with popular products, businesses can enhance their operations by reading their customers' minds through the information they have retrieved.

Subjectivity detection is the process of identifying subjective assertions in data, as given by Sindhu et al. [18]. Sentiment analysis is employed to automate the analysis of such data. Finding opinionated data and classifying it according to its polarity, such as positive, negative, or neutral feedback, is the goal. This process is known as sentiment classification, and it is followed by SA. However, multiple pre-processing techniques are applied to the data before SA, which ultimately produces the desired optimised output. This enables us to learn more about the public's attitude towards or opinion on a certain subject. This summary aids relevant organisation or the general public in improving their good or service in light of the comments received.

Serena Y. Kim et.al [19] used data from Twitter, a microblogging site where users may send messages known as tweets, to deliver the general mood towards solar energy in the United States of America. We performed a classification challenge using robustly optimised Bidirectional Encoder Representations from Transformers (RoBERTa) after filtering tweets that were specifically about solar energy. Our RoBERTa-based sentiment classification algorithm achieves 80.2% accuracy for ternary (positive, neutral, or negative) classification using 6300 manually annotated tweets that are specifically about solar energy. We found that public opinion varied significantly among states after analysing 266,686 tweets from January to December 2020 (Coefficient of Variation = 164.66%). The Northeast U.S. Region has a more favourable attitude towards solar energy during the study period than did the South U.S. Region. In places where there will be a higher proportion of Democratic votes in the 2020 presidential election, public opinion on solar energy is more favourable.

In addition to Sarojini Yarramsetti [20] a brand-new technique dubbed Intensive Deep Learning based Voice Estimation Principle (IDLVEP) has been developed to help recognise the content of voice messages and extract features using NLP principles. An effective method for developing a potent data processing model to recognise the emotive elements from the social networking medium is provided by the association of such DL and NLP. Both text-based and voice-based tweet emotional feature estimations are supported by this hybrid logic. The principles of NLP help the suggested approach of IDLVEP to extract the voice content from the input message and gives a raw text content; on the basis of this, the principles of deep learning classify the messages with respect to the estimation of dangerous or normal tweets. Voice tweets and text tweets are the initial subcategories into which the user's tweets are separated. As a result, the IDLVEP technique recognises the hazardous materials from user tweets and removes them deftly utilising the suggested approach categorization strategies.

2. Proposed System

In this work, we proposed a subjectivity detection and semantic analysis for opinion mining using SBNDNN. Initially Pre-processing entails four fundamental procedures. They are, tokenization, removal of special symbols, removal of stop words and removal of username. After subjectivity detection is to classify the data into subjective and objective and remove the data based on their subjectivity scorekeeping only the data having score higher than the specified threshold using support value. Then employing feature extraction based on Senti Word Net, Unigram, polarity and lexicon. Finally Sentiment based normalized deep neural network (SBNDNN) is used to predict the result of sentiment classification opinion Analysis. Experiment results will demonstrate that our suggested strategy outperforms existing methods in terms of information. Figure 1 shows a proposed process flow of subjectivity detection and semantic analysis.

3.1 Pre-processing

Initially twitter dataset is taken and split the dataset undergoes the process of training and testing. Since Twitter data is largely unstructured, it must first be cleaned and processed before analysis. The pre-processing [21] of the data entails numerous steps. In this instance, we are solely concerned in text. As Twitter product evaluations and the data we use to measure sentiment are unstructured, so are the reviews. Links, hashtags, and other special symbols and characters that machines cannot comprehend are among the content types in this data. The sentiment analysis process makes sense of all the content in product reviews. We simply needed text from product reviews and data to analyse sentiment. All of this is done during pre-processing, where links, hashtags, special characters, and other formatting elements are taken out of product reviews and the text that is left is then utilised to evaluate sentiment. Additionally, some capitalised and repeated words can be found in product reviews. The repetition of words is eliminated during pre-processing, capitalised words are changed to lower case, and spelling errors are also fixed in product reviews. For instance, "Gm" stands for "Good Morning" and there are spelling errors like "goood" to "good" that make it simpler to understand the sentiment of product reviews. Product reviews are used to analyse user sentiment after processing is complete.

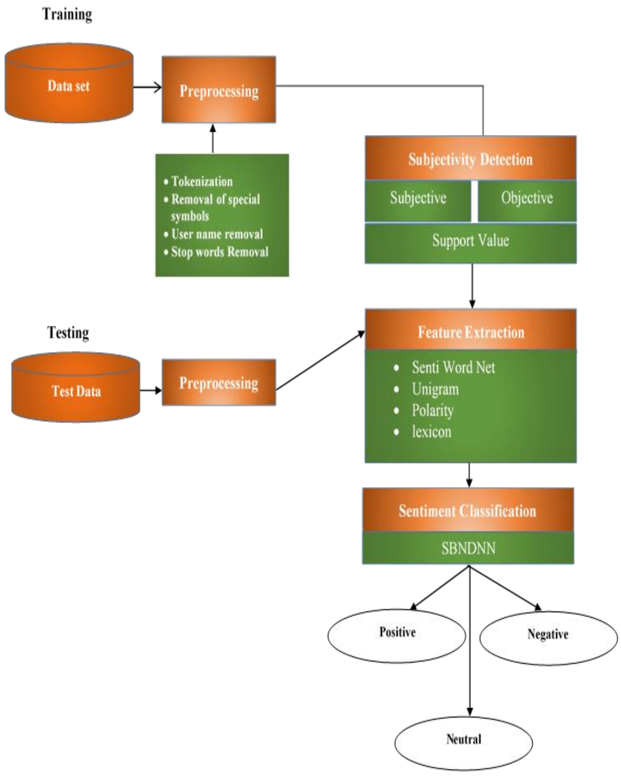

Fig. 1. Proposed process flow of subjectivity detection and semantic analysis.

3.1.1 Pre-processing steps includes below procedures:

a) Tokenisation

The product reviews that have been collected so far have been categorised into different words. Spaces are deleted from product reviews that contain them, and capitalised words are transformed to lowercase terms [21].

b) Removal of Special Symbols

Removal of special symbols [22] phase deals with getting rid of certain symbols like full stops (.) and exclamation points (!) that don't actually convey anything.

c) Username Removal

The username elimination [22] phase removes any usernames that begin with @ and do not affect sentiment analysis, such as @username.

d) Stop word removal

Stop words are eliminated [22] step, which also converts emoticons to words that are equivalent to them because they have no bearing on sentiment analysis.

3.2 Support Value based Subjectivity Detection:

Using a subjectivity classifier, the pre-processed data is divided into subjective and objective statements. The classifier is trained using a remaining product reviews after filtering out all reviews with subjectivity scores below the defined threshold. It has been shown that a sizable portion of product reviews are filtered out as the subjectivity level is raised [22].

Following is a taxonomy of subjectivity: Let D = {d1,…dn} represent a collection of texts in dataset D. The issue with subjectivity categorization is that it is difficult to distinguish between sentences that present opinions and other forms of subjectivity and sentences that present factual information objectively. Subjective and objective sentences are set as Ss and So), where Ss U So= S [24].N →total classes (N=2: i.e. Ss, So);

- M → total count of different words in the corpus in D
- R → total count of observed sequences in the training process
- $Tr_i = \{Tr_1, Tr_2, Tr_3......Tr_i\}$ →sentences in the training dataset, where Tr_i → length of i^{th} sentence, $i = 1,2,3,..., R$
- $\vartheta_{k,j}$ → association between k^{th} term and the j^{th} class $k = 1,2,3,...,M$; $j = 1,2,.,N$
- $M_{k,j}$ →number of times k^{th} term occurred in the j^{th} class
- $T_k = \sum M_{k,j}$ →occurrence times of the k^{th} term in the corpus
- Frequency (1) of the k^{th} term in the j^{th} class.

$$F_{k,j} = {(M_{k,j})}/{T_k} \tag{1}$$

- Based threshold value is define by support value estimation (2)

$$\tilde{S} = 1/\delta * \mathrm{N} \tag{2}$$

- Where $\tilde{S}$ → support value, □□is identified statistically with □□□□1.4 being best for the corpus investigated.
- Subjectivity Detection used an analytical formula to determine the degree to which each phrase was associated as, $F_{k,j} > \tilde{S}$ true Ss, otherwise so [24].

3.3 Feature Extraction

For the dataset P that contains the collection of sentences D in the product reviews, feature extraction is a crucial step. Since the data has been pre-processed for classification purposes, the useful aspects of the information have been removed. It can be challenging to pull useful information out of data. The properties of the datasets that are particularly helpful in recognising sentiments are identified in this step. The procedures are described below,

3.3.1 Senti Word Net

A sentiment dictionary that includes the polarity score of opinion words is called Senti Word Net. Senti Word Net has roughly 2 million nouns with part-of-speech tags, adjectives, adverbs, and verbs [25].

3.3.2 Unigrams

The existence or absence of a feature is a unigram feature. Words are merely a collection of words that have been retrieved from text using noise characters like spaces. For instance, the words "this," "is," "an," "awesome and so on are all different unigram features in the statement "This is an awesome movie" [26].

3.3.3 Polarity

The polarity of the opinion word is also established in the manner described below. If the word is present, the ambiguous polarity list is first used to get the polarity of the word. If the opinion word is not ambiguous, the unambiguous polarity lexicon (a combination of SenticNet, SentiWordNet, and General Inquirer) is used to determine the polarity value. The polarity (height of ontology) that the opinion word contributes to the overall polarity is considered. By adding together each term's contribution, the document's overall polarity is finally established [25].

3.3.4 Lexicon

Opinion words that convey either positive or negative attitudes, are the foundation of the lexicon-based approach. Words which denote a favoured state (such as "great", "good") have a positive polarity, whereas words denoting an unfavoured state (such as "bad", "awful") have a negative polarity. Although adjectives and adverbs typically fall under the umbrella of opinion polarity, verbs and nouns can also express opinions [27].

3.4 Sentiment based Normalized Deep Neural Network (SBNDNN)

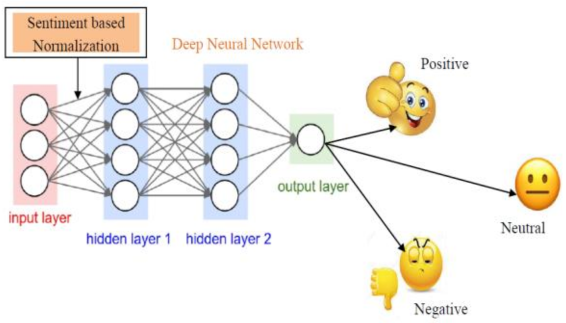

Fig. 2. Architecture of the SBNDNN of Semantic Analysis for opinion mining.

The Neural Network must continuously learn to handle problems with more expertise or to use a range of techniques to get better results. When it receives fresh knowledge from system, it learns to react to a replacement scenario. Many of the previous studies just employ DNN, however in our study, we also use recently developed sentiment-based normalisation that is integrated with DNN. The architecture of the SBNDNN of Semantic Analysis for opinion mining is depicted in Figure 2. This is one of the cutting-edge methods we employed in our study. The input is first sent to the DNN convolutional layer, which performs sentiment-based normalisation before moving on to the max pooling layer and convolutional layer, which includes a fully connected layer to the Softmax regressor. A proposed SBNDNN is used to categorise sentiment. Sentiment-based normalisation (SBN) is the layer where input features are moved in DNN [28]. Based on the Sentiment features that were collected and provided to the DNN layer, SBN begins its procedure. The output of a feature is forwarded to the max pooling layer once it has finished processing in order to minimise the depiction of boundaries and calculations in the system. The sentiment features are processed by max pooling, which then moves to con1 where it retrieves the features in a clear manner using a matrix or kernel. The sentiment features are then transmitted back to SBN, where they are processed by max pooling and moved to con2. In order to forecast the outcome, con2 finally forwards all of the emotion data to softmax regressor. The SBNDNN is a classifier whose final judgement is based on the weights and biases of the prior structural layers. It has a variety of strata, as shown by the following:

Step 1: Convolutional layer: Using a condition, this layer completes the convolution of the input data with the kernel (5).

$$C_k = \sum_{m=0}^{M-1} y_n \hat{h}_{k-n} \tag{3}$$

Where y_n →Sentiments, $\hat{h}$ →filter, M →number of components in y , C_k → output vector. Subsequently, the model is upgraded with conditions (4) and (5) each layer independently.

$$\Delta W_l = -\frac{x\lambda}{r} W_n - \frac{x}{N_t} \frac{\partial C}{\partial W_n} + m\Delta W_n(t) \tag{4}$$

$$\Delta B_n = -\frac{x}{n} \frac{\partial C}{\partial B_n} + m\Delta B_n(t) \tag{5}$$

Where W_n →weight, B_n → bias, n →layer number, λ →regularization parameter, x →learning rate, N_t →total amount of training sets, m →momentum, t → upgrading phase and C →cost function word sentiment polarity features vector of the tweet.

Step 2: Sentiment-based normalization layer

The linear alteration of data to fit into a specific range is known as normalisation. For data standardisation, the Z-score normalization method is used, which transforms data linearly. Equation (6) describes how to normalise Z-scores:

$$Z_{norm} = \frac{se - \mu}{\sigma} \tag{6}$$

Here, Z_{norm} →normalized output, se → senti WorldNet value, μ →mean weight (W_n), σ →means bias (B_n). The convolutional layer output data is normalized using the sentiment feature by using the equation (6). This layer leads to the Sentiment-based normalized data and is given as input to the pooling layer. This layer brings about the sentiment based normalized feature and is anticipated to contribute to the pooling layer.

Step 3: Pooling layer: Another name for this layer is down-sampling. The pooling method shrinks the convolution layer's output neurons to lessen computational load and prevent over fitting. Fewer output neurons are produced by the max-pooling algorithm since it only chooses the highest value in each feature map. Convolution layers are frequently followed by pooling layers in order to clean up the data in their output.

Step 4: A completely connected layer: The activation determines the classes' probability distribution. The output layer then searches for a previous layer result that matches the top positive, negative, or neutral value using the softmax algorithm.

$$p_i = \frac{e^{y_i}}{\sum_1^k e^{y_i}} \tag{7}$$

Where y → resultant data. SBNDNN is adapted with the Sentiment-based normalization to direct the over-fitting in layers and conclusions is the important classification of Sentiment Analysis to Predict the Positive, Negative or Neutral.

4. Result and Discussion

The suggested subjectivity detection and semantic analysis for opinion mining using SBNDNN classification was implemented in the working platform of python. In this section, the experimental results accomplished for the proposed techniques and comparing with the existing methods LSTM, DNN [32], RoBERT [19] and SVM [33].

4.1 Dataset description:

a) Amazon dataset:

The dataset displays information about the products that were evaluated on US Amazon sites between 2005 and 2015. The Amazon product review dataset under analysis has 6.93 million records and 15 characteristics. The entire file weighs 3.63 GB, an amount that traditional computers cannot handle or that makes computing the prediction time-consuming. This dataset consists of links, product metadata and reviews. The reviewer's profile name, the review summary, and the review content have been taken out of the original data [29].

b) Yelp dataset:

Yelp.com is a popular online resource for reviews and a directory. Data was gathered through the Yelp Dataset Challenge, which can be accessed from the Yelp1 and Kaggle2 websites and is open to the public. The dataset for this project was obtained from the Kaggle repository. Five CSV files make up the Yelp dataset: business, users, reviews, check-in, and tips. Yelp_business, Yelp_business_attributes, and Yelp_Business_Hours are the three business-related files. The yelp_business and yelp review files are used only for the research's focus on businesses and reviews. The most popular category for businesses is restaurants, hence this study only looks at this category. The Yelp review dataset was also filtered to only display restaurant-related businesses. The amount of reviews varies greatly between eateries. Due to this circumstance, further filtering was done to keep only restaurants with over 500 reviews, as this criterion results in an appropriate amount of reviews that can be utilised for comparison analysis [30]. The lowest number of reviews is 2, and the largest number is 10,323 reviews. Table 1 lists the preparation and testing test measurements for each of the two datasets of Amazon and yelp.

Table 2. Preparing and Testing Samples in the values of two datasets

Dataset		Total	Positive	Negative	Neutral
Amazon	Training	2608	1170	709	222
	Testing	1100	602	180	189
yelp	Training	2300	1003	879	450
	Testing	838	451	102	152

4.2 Performance Analysis:

4.2.1 Accuracy:

It is determined by taking the ratio of the total number of correct predictions to total number of observations in the dataset. It measures the level of data classification accuracy.

$$\text{Accuracy} = \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{TP} + \text{FN} + \text{FP})} \tag{8}$$

Where, TN → true negative, TP → true positive, FP → false positive, and FN →false negative.

4.2.2 Specificity:

Specificity refers to the ability of the precise identification of the system and it is calculated as,

$$Specificity = TN/(TN + FP) \tag{9}$$

4.2.3 F-measure:

It evaluates how accurate a test is. It has its largest value at 1 and its lowest value at 0. It's stated as,

$$F = \frac{2TP}{2TP + FP + FN} \tag{10}$$

4.2.4 Recall:

Recall is the ratio of the amount of common information that is recognised to the total amount of information that is generally available in the dataset. It is determined by utilising,

$$R = \frac{TP}{TP + FN} \tag{11}$$

4.2.5 Precision:

It is the proximity of the two measured values to each other which is measured as,

$$P = \frac{TP}{TP + FP} \tag{12}$$

4.2.6 RMSE (Root Mean Square Error):

$$\text{RMSE} = \sqrt{\frac{\sum_{x=1}^{N} \left(Z_x - \hat{Z}_x\right)^2}{N}} \tag{13}$$

We generate polarity scores that range from +1 to -1. If the polarity score >0, the product review is considered positive; otherwise as neutral. We compute the polarity sentiment analysis to visualise the general public perceptions or emotions regarding the product reviews

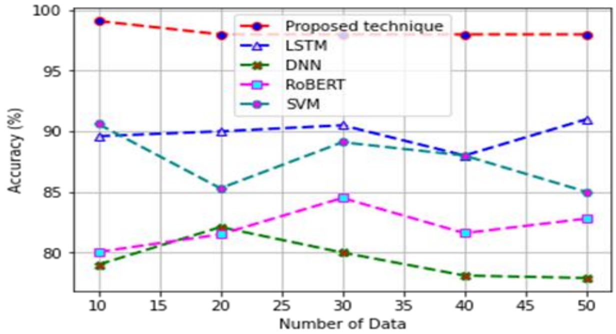

Fig 3 (a) Accuracy Amazon Dataset.

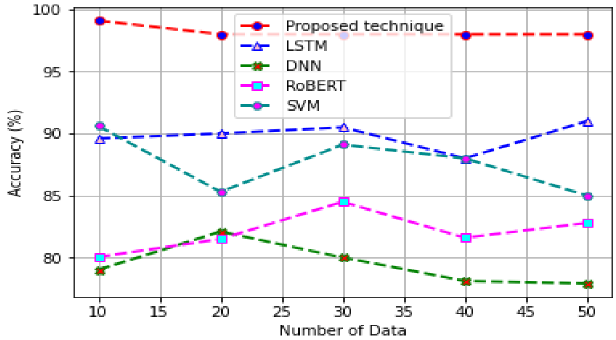

Fig 3 (b) Accuracy Amazon Dataset.

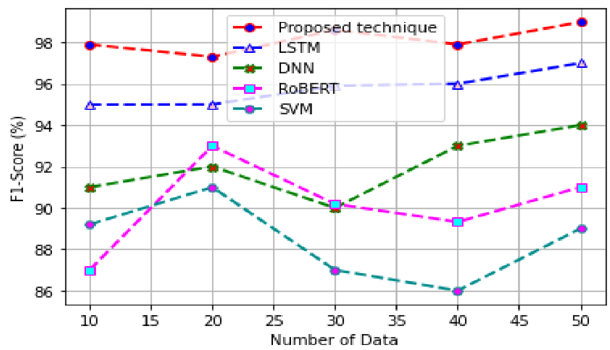

Fig 3 (c) f1-score Amazon Dataset.

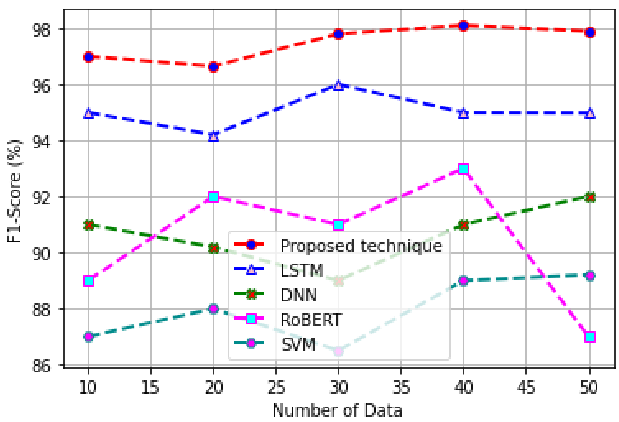

Fig 3 (d) Accuracy Amazon Dataset.

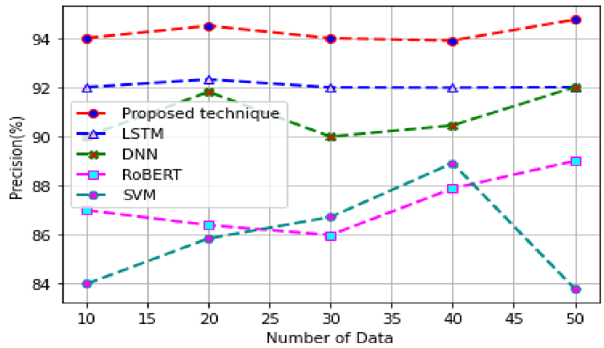

Fig 3 (e) Precision Amazon Dataset.

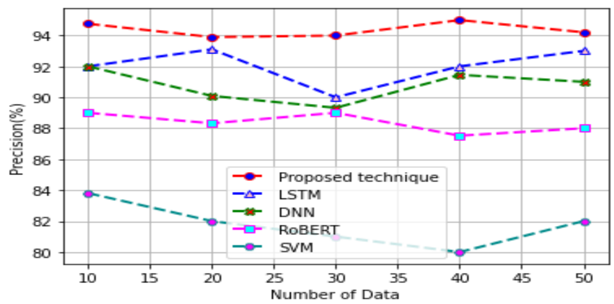

Fig 3 (f) Precision yelp Dataset.

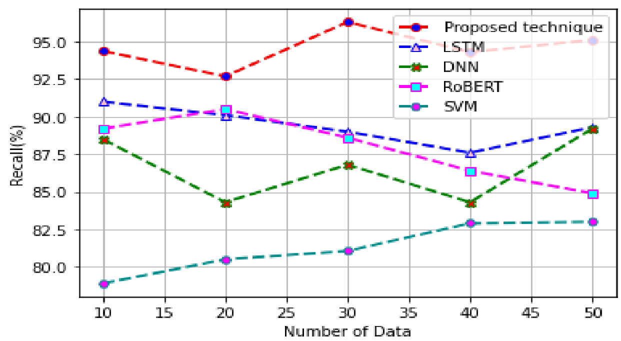

Fig 3 (g) Recall Amazon Dataset.

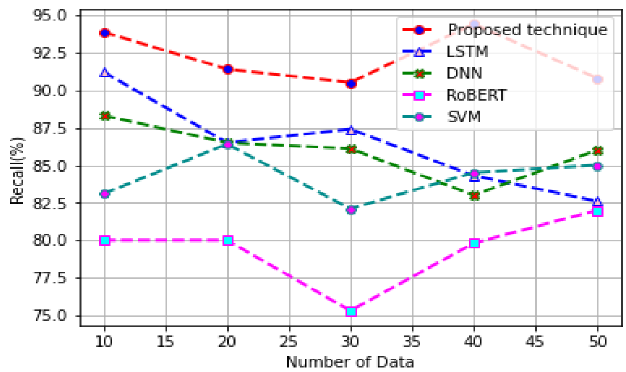

Fig 3 (h) Recall yelp Dataset.

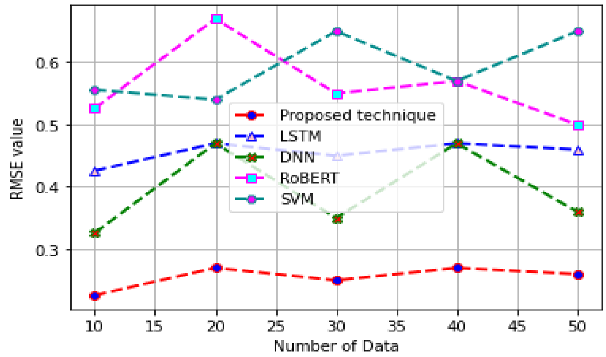

Fig 3 (i) RMSE Amazon Dataset.

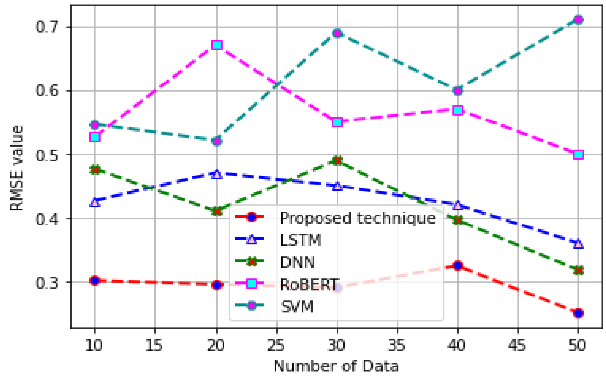

Fig 3 (j) RMSE yelp Dataset.

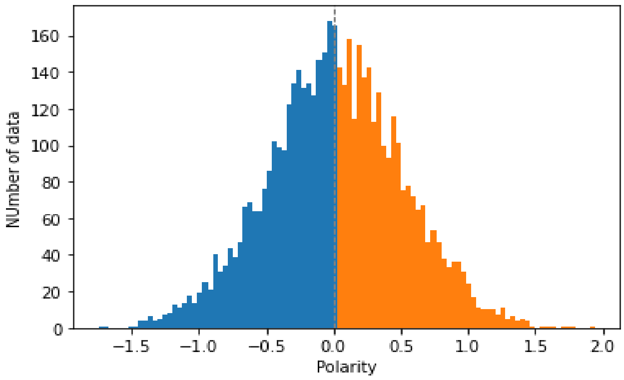

Fig 3 (k) PSA Amazon Dataset.

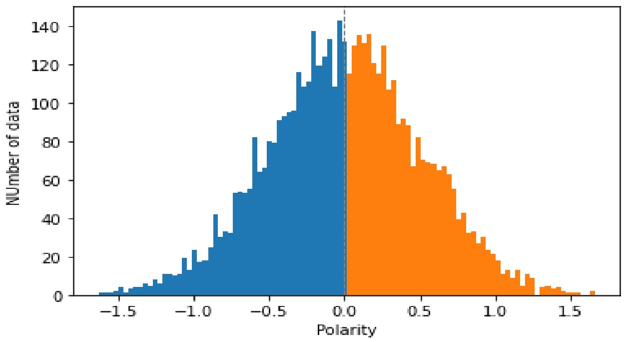

Fig 3 (l) Accuracy Amazon Dataset.

Fig 3: Representation of Performance Analysis

Figure 3(a) denotes the Representation of Performance Analysis accuracy for the Amazon dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better accuracy 96.9% is reached when compared to the existing techniques. Figure 3(b) denotes the Graphical Representation of Performance Analysis accuracy for the yelp dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better accuracy 97% is reached when compared to the existing techniques.

Figure 3(c) denotes the Graphical Representation of Performance Analysis f1-score for the Amazon dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better f1-score 99.7% is reached when compared to the existing techniques. Figure 3(d) denotes the Graphical Representation of Performance Analysis f1-score for the yelp dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better f1-score 98% is reached when compared to the existing techniques.

Figure 3(e) denotes the Graphical Representation of Performance Analysis precision for the Amazon dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better precision 94.8% is reached when compared to the existing techniques. Figure 3(f) denotes the Graphical Representation of Performance Analysis precision for the yelp dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better precision 94% is reached when compared to the existing techniques.

Figure 3(g) denotes the Graphical Representation of Performance Analysis recall for the Amazon dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better recall 95% is reached when compared to the existing techniques. Figure 3(h) denotes the Graphical Representation of Performance Analysis recall for the yelp dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Better recall 91% is reached when compared to the existing techniques.

Figure 3(i) denotes the Graphical Representation of Performance Analysis RMSE for the Amazon dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Very low RMSE 0.2% is reached when compared to the existing techniques. Figure 3(j) denotes the Graphical Representation of Performance Analysis RMSE for the yelp dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique. Very low RMSE 0.1% is reached when compared to the existing techniques.

Figure 3(k) and figure3(l) denotes the Graphical Representation of Performance Analysis of Polarity sentiment analysis for the Amazon dataset and yelp dataset compared with existing methods LSTM, DNN [32], RoBERT [19] and SVM [33] with our proposed technique.

Table 2 shows the datasets for both movie reviews and product reviews, we achieve higher results that are compared to other methods.

Table 2. Subjectivity detection tasks in sentiment analysis

Comparison	Data Set	Algorithms	Features	Accuracy
[34]	Movie reviews	MaxEnt	Dependency forest-based features	91.6%
[35]	Senti Word Net based features and feature selection	SVM, NB	Movie reviews	87.15
[36]	Movie reviews	SVM	Stylistic and syntactic features	87.9
[30]	Yelp Dataset, SemEval-2014 Dataset	CNN,BERT, ALBERT	BOW(bag of word)	83.04%
Proposed	Amazon, Yelp	SBNDNN	Senti Word Net, unigram, lexicon, polarity,	96.9% 97%

5. Conclusion:

According to current study, determining sentiment analysis is one of the key responsibilities in opinion mining. Sentiment estimation offers standardised techniques for identifying traits or qualities that are frequently used and well-known. In this work, Subjectivity Detection and Semantic Analysis for Opinion Mining is clearly portrayed with steps involving data pre-processing, subjectivity detection and feature extraction. And the sentiment classification is done using SBNDNN. Dataset of two different categories are taken Amazon review and yelp review datasets are taken respectively. The results are compared with the existing methods LSTM, DNN [32], RoBERT [19] and SVM [33]. Performance metrics are calculated and listed in the above table specifies the proposed work reaches the good Accuracy 96.9% and 97% when compared with existing ones.

Declarations

Funding : In this research article has not been funded by anyone.

Data availability: There is no data availability in this research.

Conflict of interest : All authors do not have any conflict of interest.

Ethical Approval : This article does not contain any studies with human participants or animals performed by any of the authors.

Reference:

[1] Younis, Suhaib Bin. "Opinion mining on web-based communities using optimised clustering algorithms." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 9 (2021): 438-447.

[2] Onan, Aytuğ. "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks." Concurrency and Computation: Practice and Experience 33, no. 23 (2021): e5909.

[3] Zuheros, Cristina, Eugenio Martínez-Cámara, Enrique Herrera-Viedma, and Francisco Herrera. "Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews." Information Fusion 68 (2021): 22-36.

[4] Dashtipour, Kia, Mandar Gogate, Erik Cambria, and Amir Hussain. "A novel context-aware multimodal framework for persian sentiment analysis." arXiv preprint arXiv:2103.02636 (2021).

[5] G. Rabby, S. Azad, M. Mahmud, K. Z. Zamli, M. M. Rahman, Teket: a tree-based unsupervised keyphrase extraction technique, Cognitive Computation (2020) 1–23.

[6] X. Zhong, E. Cambria, A. Hussain, Extracting time expressions and namedentities with constituent-based tagging schemes, Cognitive Computation12 (4) (2020) 844–862.

[7] R. Satapathy, E. Cambria, A. Nanetti, A. Hussain, A review of shorthandsystems: From brachygraphy to microtext and beyond, Cognitive Computation12 (4) (2020) 778–792.

[8] Husnain, Mujtaba, Malik Muhammad Saad Missen, Nadeem Akhtar, Mickaël Coustaty, Shahzad Mumtaz, and VB Surya Prasath. "A systematic study on the

role of SentiWordNet in opinion mining." Frontiers of Computer Science 15, no. 4 (2021): 1-19.

[9] M. Al-Ayyoub, A.A. Khamaiseh, Y. Jararweh, M.N. Al-Kabi, A comprehensive survey of arabic sentiment analysis, Inf. Process. Manag. (2019). doi:10.1016/j.ipm.2018.07.006.

[10] A.B. Nassif, I. Shahin, I. Attili, M. Azzeh, K. Shaalan, Speech Recognition Using Deep Neural Networks: A Systematic Review, IEEE Access. 7 (2019) 19143–19165. doi:10.1109/ACCESS.2019.2896880.

[11] J. Xu, F. Huang, X. Zhang, S. Wang, C. Li, Z. Li, Y. He, Sentiment analysis of social images via hierarchical deep fusion of content and links, Appl. Soft Comput. 80 (2019) 387–399. doi:10.1016/J.ASOC.2019.04.010.

[12] Nassif, Ali Bou, Ashraf Elnagar, Ismail Shahin, and Safaa Henno. "Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities." Applied Soft Computing (2020): 106836.

[13] Tedmori, Sara, and Arafat Awajan. "Sentiment analysis main tasks and applications: a survey." Journal of Information Processing Systems 15, no. 3 (2019): 500-519.

[14] Onyenwe, Ikechukwu, Samuel Nwagbo, Njideka Mbeledogu, and Ebele Onyedinma. "The impact of political party/candidate on the election results from a sentiment analysis perspective using# AnambraDecides2017 tweets." Social Network Analysis and Mining 10, no. 1 (2020): 1-17.

[15] Sindhu, C., Binoy Sasmal, Rahul Gupta, and J. Prathipa. "Subjectivity detection for sentiment analysis on Twitter data." In Artificial Intelligence Techniques for Advanced Computing Applications, pp. 467-476. Springer, Singapore, 2021.

[16] Wang, Jie, Bingxin Xu, and Yujie Zu. "Deep learning for aspect-based sentiment analysis." In 2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), pp. 267-271. IEEE, 2021.

[17] Uma Maheswari, S., and S. S. Dhenakaran. "Opinion Mining on Integrated Social Networks and E-Commerce Blog." IETE Journal of Research (2021): 1-9.

[18] Sindhu, C., Binoy Sasmal, Rahul Gupta, and J. Prathipa. "Subjectivity detection for sentiment analysis on Twitter data." In Artificial Intelligence Techniques for Advanced Computing Applications, pp. 467-476. Springer, Singapore, 2021.

[19] Kim, Serena Y., Koushik Ganesan, Princess Dickens, and Soumya Panda. "Public sentiment toward solar energy—opinion mining of twitter using a transformer-based language model." Sustainability 13, no. 5 (2021): 2673.

[20] Yarramsetti, Sarojini, J. Anvar Shathik, and P. S. Renisha. "Intelligent Estimation of Social Media Sentimental Features using Deep Learning with Natural Language Processing Strategies."

[21] Suryawanshi, Rakesh, Akshay Rajput, Parikshit Kokale, and Subodh S. Karve. "Sentiment Analyzer using Machine Learning." International Research Journal of Modernization in Engineering Technology and Science 2, no. 6 (2020): 1-12.

[22] Sindhu, C., Binoy Sasmal, Rahul Gupta, and J. Prathipa. "Subjectivity detection for sentiment analysis on Twitter data." In Artificial intelligence techniques for advanced computing applications, pp. 467-476. Springer, Singapore, 2021.

[23] Kumar, Akshi, and Mary Sebastian Teeja. "Sentiment analysis: A perspective on its past, present and future." International Journal of Intelligent Systems and Applications 4, no. 10 (2012): 1.

[24] Tang H, Tan S, and Cheng X. A survey on sentiment detection of reviews. Expert Systems with Applications: An International Journal, September 2009, 36(7):10760–10773.

[25] Agarwal, Basant, Namita Mittal, Pooja Bansal, and Sonal Garg. "Sentiment analysis using common-sense and context information." Computational intelligence and neuroscience 2015 (2015).

[26] Agarwal, Basant, Soujanya Poria, Namita Mittal, Alexander Gelbukh, and Amir Hussain. "Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach." Cognitive Computation 7, no. 4 (2015): 487-499.

[27] Zhang, Lei, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and Bing Liu. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis." HP Laboratories, Technical Report HPL-2011 89 (2011): 1-8.

[28] Jianqiang, Zhao, Gui Xiaolin, and Zhang Xuejun. "Deep convolution neural networks for twitter sentiment analysis." IEEE Access 6 (2018): 23253-23260.

[29] Woo, Jongwook, and Monika Mishra. "Predicting the ratings of Amazon products using Big Data." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11, no. 3 (2021): e1400.

[30] Alamoudi, Eman Saeed, and Norah Saleh Alghamdi. "Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings." Journal of Decision Systems 30, no. 2-3 (2021): 259-281.

[31] Huang, Xin, Wenbin Zhang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liu, and Ji Zhang. "Lstm based sentiment analysis for

cryptocurrency prediction." In International Conference on Database Systems for Advanced Applications, pp. 617-621. Springer, Cham, 2021.

[32] Umer, Muhammad, Imran Ashraf, Arif Mehmood, Saru Kumari, Saleem Ullah, and Gyu Sang Choi. "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model." Computational Intelligence 37, no. 1 (2021): 409-434.

[33] Obiedat, Ruba, Raneem Qaddoura, Al-Zoubi Ala'M, Laila Al-Qaisi, Osama Harfoushi, Mo'ath Alrefai, and Hossam Faris. "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution." IEEE Access 10 (2022): 22260-22273.

[34] Tu Z, Jiang W, Liu Q, Lin S. Dependency forest for sentiment analysis. In: First CCF conference, natural language processing and Chinese computing; 2012. p 69–77.

[35] O'Keefe T, Koprinska I. Feature selection and weighting methods in sentiment analysis. In: Proceedings of the 14th Australasian document computing symposium, Sydney, Australia, ACL; 2009.

[36] Abbasi A, Chen H, Salem A. Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums. ACM Trans Inf Syst. 2008;26(3):12.