

RESEARCH ARTICLE

Improving Sentiment Analysis in Social Media by Handling Lengthened Words

ASHIMA KUKKAR¹, RAJNI MOHANA², AMAN SHARMA², ANAND NAYYAR³,
AND MOHD. ASIF SHAH^{4,5}

¹Department of Computer Science, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab 140401, India

²Department of Computer Science, Jaypee University of Information Technology, Solan, Himachal Pradesh 173234, India

³Graduate School, Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam

⁴Department of Economics, Bakhtar University, Kabul 2496300, Afghanistan

⁵School of Business, Woxsen University, Kamkole, Sadasivpet, Hyderabad, Telangana 502345, India

Corresponding authors: Anand Nayyar (anandnayyar@duytan.edu.vn) and Mohd. Asif Shah (ohaasif@bakhtar.edu.af)

ABSTRACT Machines are continually being channelized in the current era of automation to deliver accurate interpretations of what people communicate on social media. The human species is today engulfed in the concept of what and how people believe, and the decisions made as a result are mostly dependent on the sway of the masses on social media platforms. The usage of internet as well as social media is booming day by day. Today, this ocean of data can be used for the fruitful purposes. Analysis of social media sentiment textual posts can supply knowledge and information that can be used in citizen opinion polling, business intelligence, social contexts, and Internet of Things (IOT)-mood triggered devices. In this manuscript, the main focus is the sentiment analysis based on Emotional Recognition (ER). The proposed system highlights the process of gaining actual sentiment or mood of a person. The key idea to this system is posed by the fact that if smile and laughter can be two different categories of being happy, then why not happpyyyyyy and happy. A novel lexicon based system is proposed that considers the lengthened word as it is, instead of being omitted or normalized. The aggregated intensified senti-scores of lengthened words are calculated using framed lexicon rules. After that, these senti-scores of lengthened words are used to calculate the level of sentiment of the person. The dataset used in this paper is the informal chats happened among different friend groups over Facebook, Tweets and personal chat. The performance of proposed system is compared with traditional systems that ignore lengthened words and proposed system outperform tradition systems by achieving 81% to 96% F-measure rate for all datasets.

INDEX TERMS Sentiment analysis, natural language processing, social media, senti-score, lengthened words, emotional recognition.

I. INTRODUCTION

Human Beings are always attracted to persons who share our values, beliefs and interests. Even studies show that individuals feel more at ease associating with other individuals who share their ideas, with people individual can trust and who can assist us in achieving our goals [1]. People have an etymological propensity to associate with like-minded communities. A community is made up of multiple clusters. Modularity is one of the most important factors to consider when deciding

the number of communities [2]. If the clusters' traits are thoroughly examined, it can aid in identifying the specific character set of individual clusters or groups of like-minded people [3], [4]. To put it another way, the presence of a common connect amongst a group of people implies that those persons share similar ideas and goals. To be more exact, four types of social media models are available as shown in Figure 1. The sort of service associated with these social media models is described in Table 1, along with several websites and applications that fall into that category. Social media is a platform that facilitates the sharing of information (like reviews of a product, debate, voting), ideas (such as eCommerce site,

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Guidi¹.

TABLE 1. Sort of service associated with these social media models.

Types	Services	Websites
Entertainment Models	Game Play and Sharing	Candy Crush, Kongregated, Miniclip, Anipang,
	Virtual Worlds	The Sims, Second Life
Collaboration Models	Q &A, Community	Askville, Yahoo! Answers, Quora, Spring.me,
	Wikis	Twiki, Pbworks, Evernote
	Reviews and Opinions	Eopinions, Kindle, amazon, City-data.com,
	Social Bookmarking	Delicious, Readwrite, Scoop.it, Digg, Diigo, Stumbleupon, Pinterest, Flipboard
Sharing Models	Magazines, Files, Books, Documents	Scribd, Google Docs, Issuu, 4shared, SlideShare,
	Music and Audio	Last.fm, Tunes, Soundcloud
	Livestreaming	Justin.tv, Ustream.tv
	Photo	Flickr, Instagram
	Video	Vine, YouTube, Vimeo
Communication Model	Video-Conferencing	Skype, Google Hangout
	Instant Messaging	Viber, WhatsApp, Kakao Talk, Line
	Event Networking	Meetup.com, Upcoming
	Social Networking	Facebook, Google+, LinkedIn, Ning, MySpace, Cyworld,
	Microblogging	Twitter, Tumblr, Me2day
	Blogging	WordPress, Blogger

advertisement, marketing) and other forms of expressions via virtual networks [5], [6], [7], [8], [9], [10]. Social networking sites such as Facebook, Twitter, WhatsApp, Instagram have become a very popular source to easily express their views with the help of text posts, status, stories, images, videos, and audios [11]. When engaging with these services users can create a highly interactive platform through which individual and organization can share, co-create, discuss and modify user generated content. If you want to buy a consumer goods, you don't have to ask your friends and family for advice anymore because there are many user reviews and debates about the product on the Internet. Because there is an abundance of publicly available information, opinion polls, surveys, and focus groups may no longer be essential for an organization to acquire public viewpoints [12], [13], [14]. We've seen how opinionated social media posts have helped restructure businesses and affect public feelings and emotions in recent years, all of which have had a significant impact on our social and political institutions [15], [16], [17]. Hence, there are a lot of raw insights and information available on social media which is collected from the individual's social media activities. This raw data can be structured,

semi-structured or unstructured. Structured data can be used for taking future decisions and performing actions. But focusing on the simple raw data points, such as likes, shares and comments are meaningless insights as they are unstructured or semi-structured [18], [19], [20]. This unstructured and semi-structured data needs to be pre-processed to be meaningful [21]. The pre-processing stage extracts the semantics, syntax, lengthened, metaphor, etc. Studies have shown that around the world 70% of the internet people use lengthened words for different purposes [22], [23], [24].

A. PROBLEM STATEMENT

In this paper, the main focus is on the lengthened words. Lengthened words are the words whose any character is repeated more than its dictionary word. For example: "Happy" is a dictionary word and one of its lengthened word is "Happyyyyy". Recently, it has been noticed that peoples how their sentiments by writing the lengthened words, mainly in social media as shown in Figure 2.

Thus, these words play an important role in the social network. It is observed that earlier these words were considered semantically wrong and were omitted from the sentence.

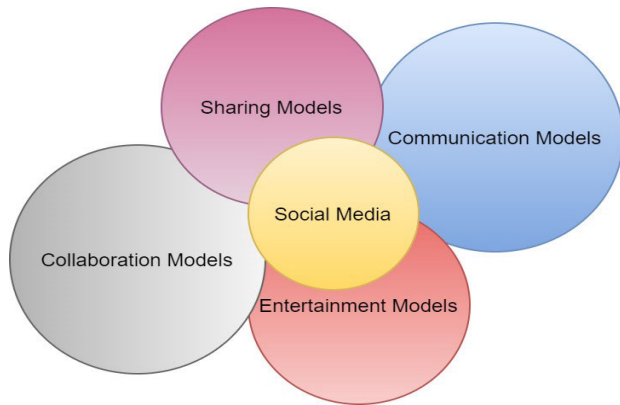


FIGURE 1. The various models of social media.

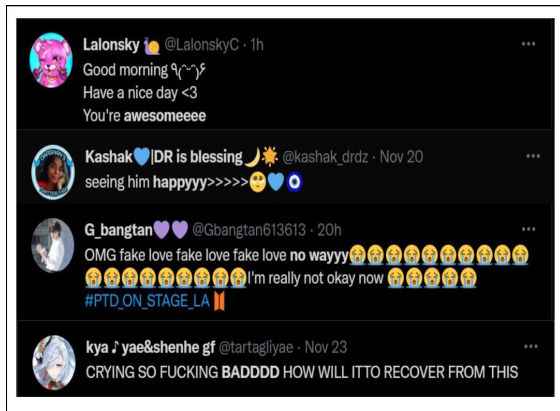


FIGURE 2. Example of twitter.

Later it is found that it changed the interpretation of what the person actually wanted to say. Thus, the researchers proposed that these should be normalized instead of being omitted as these words could be meaning bearing and could enhance the meaning of the sentence. But recently, it is being noticed that normalization is also not the efficient way of dealing with lengthened words as this ignores or nullifies the sentiment of an excitement of the person. A person generally writes a lengthened word in the situation of very happy (e.g.: Awesomeeee) or sad (e.g.: badddd) or angry (e.g.: no wayyy). These all-lengthened words are normalized to the dictionary word (Awesomeeee as Awesome and badddd as bad or no wayyy as no way) which reduces the actual excitement level, i.e., sentiment.

B. OBJECTIVES

The Objectives of the paper are:

- To understand previous sentiment analysis systems that ignore or normalize the lengthened words.
- To propose a new sentiment analysis system based on lexicon that considered the lengthened words. These lengthened words are used to calculate the level of sentiment of the person.

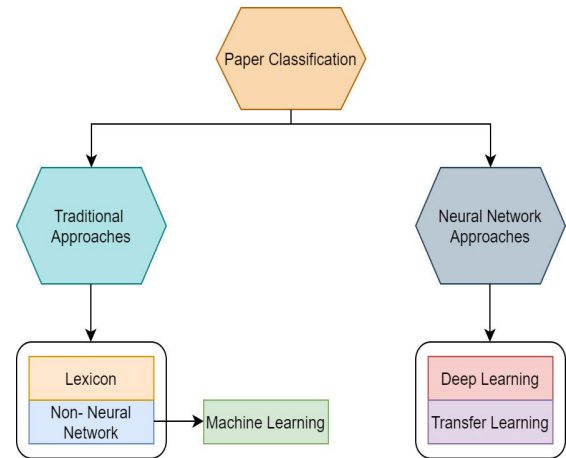


FIGURE 3. Framework of paper classification.

- Dataset used in this paper is the informal chats happened among different friend groups over Facebook, Tweets and personal chat.
- To evaluate the performance of proposed system on parameters like precision, recall and F-measure.
- To compare the proposed system with traditional systems (ignoring the lengthened words) for all datasets, to validate its performance and accuracy.

C. STRUCTURE OF PAPER

The structure of paper is organized as: related work on sentiment analysis is demonstrated in Section II. Section III represents the materials and methods. Section IV exemplified the proposed system with example to find the impact of lengthened words. Experimental results of proposed system are presented in Section V. At last, conclusion and outlook of the study is illustrated in Section VI.

II. RELATED WORK

Sentiment analysis is known as evaluation extraction, emotional polarity assessment and opinion mining, depending on the text format and application field [15], [16]. This work classifies the approaches into classes depending on the diverse concepts of sentiment analysis algorithms, including classical methods, Transfer Learning (TL) methods and Deep Learning (DL) methods in order to examine and investigate recent relevant papers.

A. TRADITIONAL APPROACHES

It includes non-neural network and lexicon-based sentiment analysis approaches.

Lexicon-based approaches: entails creating an extracting emotional value, emotional dictionary, and so on in order to judge sentiment polarity. In 2003, semantic polarity (ISA) algorithm was devised by Turney and Littman [25] to determine text sentiment tendency. Hu and Liu [26] used WordNet seed words to create a vocabulary of negative and positive sentiment terms, and then classified

sentences using the dictionary. Marquez et al. [27] proposed a technique for increasing the lexicon based on opinion. Yang et al. [28] suggested an updated SOPMI algorithm that was more successful for emotional vocabulary and computing modelling.

Brody and Diakopoulos [24] showed that there was a strong association of lengthening of words with subjectivity and sentiment for which they developed an unsupervised method for recognizing new sentiments with words that weren't in the dictionary and determining their polarity.

Birmingham and Smeaton [29] found that detecting sentiment in small corpus documents was easier than in larger form materials. They came to the conclusion that while the length of the documents influenced which feature sets and classifiers performed best, the documents' lack of information did not hinder our capability to identify them.

Bollen et al. [30] used Twitter data to detect public sentiment and anticipate future stock market swings.

Brody and Elhada [31] proposed an unsupervised approach for extracting elements from review text and calculating sentiment. The method was straightforward and adaptable in terms of language and domain, and it considers the impact of aspect on sentiment polarity, which has been mostly overlooked in earlier research.

Velikovich et al. [32] devised a method for extracting a broad sentiment lexicon from the entire internet. The lexicon that resulted had significantly more coverage than current dictionaries, and it also handled spelling problems and web-specific lingo.

Pak and Paroubek [33] demonstrated how to collect a corpus automatically for sentiment analysis and opinion mining. The authors conducted a linguistic study of the corpus and provided explanations for the phenomena they uncovered. Using the data, the author created a sentiment classifier that can tell whether a document was good, negative, or neutral.

Connor et al. [34] found a link between the frequency of sentiment words in contemporaneous Twitter messages. The findings demonstrated the value of text streams as a replacement and supplement to traditional polling.

Kivran Swaine and Naaman [35] investigated and concluded that emotional expression could explain part of the variation in users' Twitter networks, and that the use of emotion in user interactions had a substantial explanatory factor.

Diakopoulos and Shamma [36] created an analytical technique and graphic representations to aid a public or journalist affairs professional in better understanding the sentiment temporal dynamics in response to the discussion footage.

The lexicon-based method is simple to use, and it mainly focuses on the sentiment dictionary and overlooks word positional relationships. The text is extracted in a rudimentary and incomplete manner.

Non-neural network: refers to Machine Learning (ML) based approaches. ML approaches employ NLP to collect the data and information from web. It looks for polarity in texts, ranging from positive to negative. It learns how to recognize sentiments without user input by training them with the text samples of emotions. It mostly assists in understanding word data intensity, which is classified as neutral, negative, or positive [37], [39]. Several ML approaches are used for sentiment analysis like Random Forest (RF), Decision Tree (DT), Bayesian Network (BN), Logistic Regression (LR), Support Vector Machine (SVM), Maximum Entropy (ME), Ensemble Learning (EL) Naive Bayes, etc.

A model for predicting the semantic likelihood was proposed by Maas et al. [38]. It incorporated supervised and unsupervised learning. Their proposed vector-based model dealt with sentiment and semantic similarities between words. But, in the cross-domain, it did not work well. Nowadays, various social networking sites have started recommendations for the products, services, etc based on an individual's interest. This can also be used in analysis of the product through their valuable feedbacks [2].

Amrani et al. [40] found that combining SVM with other methods yielded positive results. Authors suggested a new method that combined SVM and RF algorithm. Their research revealed that the hybrid strategy beats individual algorithms.

Hasan et al. [41] used the NB algorithm to create a classifier to categorize opinions given in Bangla and English, and they achieved a substantial accuracy. They also used their classifiers to examine some random tweets and reviews, with great results in the majority of cases.

Wan and Gao [42] employed a BN as a sentiment classifier directly. They used NB, SVM, BN, DT, C4.5, and RF algorithms in an ensemble sentiment classification system. The BN surpassed all six classifiers in the individual evaluation, according to their findings.

Ngoc et al. [43] proposed a novel approach based on the C4.5 decision tree methods to do document-level sentiment classification in the data mining area.

Saleena [44] proposed a SVM, RF, NB and LR ensemble classification system. EL [42], [45], [46], NB [47], [48], [49], ME [50], [51], [52], and SVM [53], [54], [55], [56] were the most common classifiers utilized by researchers for sentiment analysis. With the growing volume of data and the requirement for a variety of tasks, more researchers were applying the Neural Network (NN) model based on DL to sentiment analysis and achieving improved results.

B. NEURAL NETWORK APPROACHES

It includes DL and TL based approaches.

DL: Many domains, including Computer Vision [57], Artificial Intelligence (AI) [58], and the Internet of Things (IOT) [59], [60], rely on DL. The application of DL to sentiment

analysis has grown in popularity as DL has progressed in the domain of NLP. Word representations of text are vital for sentiment analysis, and models such as Word Embedding and Bag of Words (BoW) are often utilized. Word2Vec [61], [60] is a well-known example of a word embedding model. In this paper, the NN approaches are divided into types that include Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) Gated Recurrent Unit (GRU), Long Term Short Memory (LSTM) and hybrid neural network models to summarize recent related researches.

Chen et al. [62] suggested a strategy that combines the benefits of ML and information retrieval approaches in their paper. Authors used semantic orientation indexes to feed a back-propagation NN and discovered that the proposed method improves sentiment categorization performance while also saving time during training.

Zhang and Wallace [63] developed a model utilizing CNN approach that also includes pre-trained word embedding's to categories texts at the sentence level, demonstrating that a basic CNN with a modest number of hyper parameters coupled with static visual words can obtain excellent outcomes on a variety of benchmarks.

Conneau et al. [64] used a deeper CNN that was applied to NLP and suggested a novel system for text processing called VDCNN. Visual Geometry Group (VGGNet) is combined with a residual network (ResNet) in this model. The authors demonstrated that having twenty-nine convolutional layers improves the results.

For text classification and prediction, Johnson and Zhang [65] employed a one-dimensional data structure. This method moved away from utilizing word vectors of low-dimensional as input and instead of it. The author has used high-dimensional text for the processing object of CNN.

It also looked at the possibility of mixing multiple convolutional layers to improve prediction accuracy. To learn vector-based document representation, Tang et al. [66] employed the bottom-up strategy. To implement single-sentence representation, firstly utilize CNN or LSTM, and then utilized a GRU network to encrypt the semantic and inherent relationships between phrases. This strategy allowed semantic information among sentences to be effortlessly recorded, resulting in superior analysis findings.

Day and Lin [67] proposed a technique for content study the consumer reviews of Chinese Google Play based on the LSTM DL model, which outperformed both the NB and SVM methods. The complicated LSTM region embedding technique was investigated by Johnson and Zhang [68]. By combining convolutional layers with LSTM region embedding, LSTM may embed text sections of different sizes and produce optimal outcomes.

Akhtar et al. [69] introduced a novel DL architecture that uses a CNN to learn the sentiment embedding vector, then a multi-objective optimization (MOO) framework was used to select a set of optimization features, and finally a SVM

classifier was utilized to classify the augment optimized vector of sentiments. This was the first time a DL model has been used to analyses sentiment in languages with limited resources. For document classification, Yang et al. [70] proposed a network of hierarchical attentional. At the word and phrase levels, a two-level attention technique was used. The model's attention layer visualization showed the precise words chosen by the model. The capacity to instinctively describe the value of specific words and sentences to the category and it has an advantage of linking the attention mechanism. CNN was fast, but it has flaws therefore new characteristics must be introduced in order to attain improved outcomes. However, the RNN model's approach has several drawbacks, such as its inability to discern the value of word context cues, necessitating the development of a hybrid framework of various mechanisms to compensate for these flaws by maximizing the model performance.

TL: is a strategy for applying knowledge learnt in one field to a new field based on similarities in data, models, data distribution, task, etc. The labelled samples can be utilized to develop the system and then utilized in the target field data to raise the target data annotation via TL.

Glorot et al. [71] investigated the field adaptable issue of sentiment classifiers, solving edge distribution differences with the Stacked Denoising Autoencoder (SDA). This approach worked well for sentiment analysis in comments, although it was more reliant on parameter setup.

Chen et al. [72] developed a marginalization SDA (mSDA) method, which addressed the SDA algorithm's flaws of computational cost and scalability of high-dimensional characteristics. This method learnt features quickly and did not need optimization strategies to train parameters. For the mSDA, Sun et al. [73] suggested an improved type of the mSDA ++ technique that can learn low-dimensional features. Text categorization accuracy can be improved by combining the mSDA and EASYADAPT algorithms. Furthermore, the mSDA ++ algorithm speeds up future calculations while reducing data storage space. Previous researches combined embedding from multiple tasks with varied levels of input.

Mccann et al. [74] used an English-German translation challenge to train the neural network model, resulting in the output "context vectors (CoVe)". Embedding's from Language Models (EMLo) was a technique for mining the word vectors of semantic qualities at a deep level, it is introduced by Ilic et al. [75]. This method introduced a richer word representation and applies it to a variety of NLP tasks, but it still trained the core task model from the start and treats pre-training embedding as a fixed parameter, limiting its impact.

Dai and Le [76] offered two strategies for improving sequence learning with RNN utilizing unlabeled input by fine-tuning the language model. The two techniques can be used as pre-training methods for post-supervised sequence machine learning methods, resulting in impressive results in a variety of classification applications. Further, in 2018,

Universal Language Model Fine-tuning (ULMFiT) method established by Howard and Ruder [77] that can be applied to NLP. Vieri et al. [78] developed a TL method based on boosting that can pass information to diverse target regions in the training process by learning training data from diverse fields and lowering the cost of labelling target areas.

In 2014, Wang and Li [79] proposed a distributed measuring approach and boosting framework for TL. In 2017, Wang et al. [80] presented a strategy of adaptive training data selection that can effectively avoid source domain training data that was fuzzy and fuzzy. Using informativeness measures, the iterative technique was utilized to integrate samples from the source field and the target field training data, evaluate the performance of the target field classifier, and update the informativeness methods for the succeeding iteration.

Zhou et al. [81] described a Hybrid Heterogeneous (HH) TL approach for solving classification of sentiment with extensive labelled source dataset and unlabeled target domain data for a heterogeneous environment.

According to the literature review, sentiment analysis is not an only issue; rather, it is a “suitcase” research challenge that necessitates the completion of a number of NLP tasks as demonstrated in Table 2. Various processes are required to extract sentiments from a given text, as well as the resolution of numerous NLP difficulties. This paper focused on the main challenge of sentiment analysis that is detecting the emotions and effects. The term “sentiment analysis” is usually used to describe the process of identifying a text’s polarity or valence. However, it can also apply to evaluating a person’s attitude toward a specific target or topic in a broader sense. Enthusiasm, grief, joy, frustration and rage and other emotional or affective attitudes are all examples of attitude. Apart from all the research, the state-of-the-art sentiment analyzers is unable to capture the implicit sentiment in the form of lengthened words. Therefore, in this manuscript, a technique is proposed based on capturing the intensification of sentiment in lengthened words.

III. MATERIALS AND METHODS

Materials

This section describes the study’s methodology. In the first section, the programming environments will be discussed. In the second section, we will examine how and where the data was collected, as well as the method for data preparation. The sentiment analysis system that considers the lengthened word will be detailed in the last section.

A. PROGRAMMING ENVIRONMENT

Python is among the most popular programming languages for Natural Language Processing (NLP), Machine Learning (ML) and data research. Python provides a vast library of NLP and ML techniques that may be used to solve a variety of problems. Python was chosen as the programming language for this investigation due to its extensive library and usability. NLTK is one of several Python packages that provide the working with human language data. It supports many

TABLE 2. List of sentiment analysis tasks [11].

Sentiment Analysis Tasks	
Tasks	Subtasks
Spam Detection	—
Sentiment Lexicon generation	—
Word sense disambiguation	—
Sarcasm detection	—
Time, Entity and opinion holder extraction	—
Sentiment search and retrieval	—
Sentiment visualization	—
Polarity classification	Cross language sentiment classification
	Sentiment rating prediction
	Cross domain sentiment classification
	Contextual polarity disambiguation
	Aspect-based sentiment classification
Sentiment summarization	—
Subjectivity classification	—

modules required namely beautiful soup, matplotlib, NumPy, pandas, and seaborn. Other strategies for feature extraction are also included.

B. DATASET

The dataset used in this is the informal chats happened among different friend groups over Facebook, tweets and chat. It is found that most of the young people express their feelings using informal text, i.e., short messages with lengthened words. The data files are created in the Comma Separated Values (CSV) format since Python can handle these types of files more easily. In the proposed algorithm, the intensification criteria are developed using three linguistic experts, with 0.68 kappa agreement.

C. DATA PREPARATION

For the preparation of the needed data, a simple Python code is built to eliminate the unnecessary features. The experimental findings for the enhancement of senti-score are determined

by combining data from three sources among young people and children. Individuals between the ages of 13 and 18 are classified as Children. Individuals between the ages of 19 and 40 are considered youth. Further, two experiments have been undertaken to assess the performance of the proposed system. In the initial trial, the children and young dataset is utilized. Since the number of chats are sufficient to obtain a decent result from proposed lexicon method. In the second trial, the impact of impact of lengthened words on sentiment analysis.

1) METHODS

In this study a new lexicon based is proposed to calculate the aggregated intensified senti-score of the words for detecting lengthened words in sentiment analysis. Lexicon method is one of the methods or strategies for semantic analysis. This method determines the sentiment orientations of a whole text or group of sentences based on the semantic orientation of lexicons. Semantic orientation can be positive, negative, or neutral. The dictionary of lexicons may be built both manually and mechanically. Many scholars make use of the WorldNet dictionary. First, lexicons are extracted from the entire document, and then WorldNet or another online thesaurus is used to find synonyms and antonyms to enlarge the dictionary.

Lexicon-based approaches utilize adjectives and adverbs to determine the text's semantic orientation. For the purpose of determining text orientation, combinations of adjectives and adverbs are extracted with their emotion orientation value. Then, these may be translated into a single score for the total value as shown in Figure 4.

With the use of a so-called valence dictionary, words in texts are categorized as positive, negative, or neutral. Consider the remark, "Good people occasionally have terrible days." A valence dictionary would classify "Good" as a positive word, "Bad" as a negative word, and the remaining words as neutral. Once each word in the text has been categorized, we can get an overall emotion score by counting the amount of positive and negative terms and summing these values. A common formula for calculating the sentiment score is as follows:

$$\text{Sentiment score} = \frac{\text{number of positive words} - \text{number of negative words}}{\text{total number of words}} \quad (1)$$

The many ways of lexicon-based approach are as follows:

2) DICTIONARY-BASED METHOD

In this method, a dictionary is compiled by selecting a few terms at first. Then, an online dictionary, thesaurus, or WordNet may be used to add synonyms and antonyms to this lexicon. The dictionary is enlarged until no other words may be added. Manual examination can be utilized to refine the vocabulary.

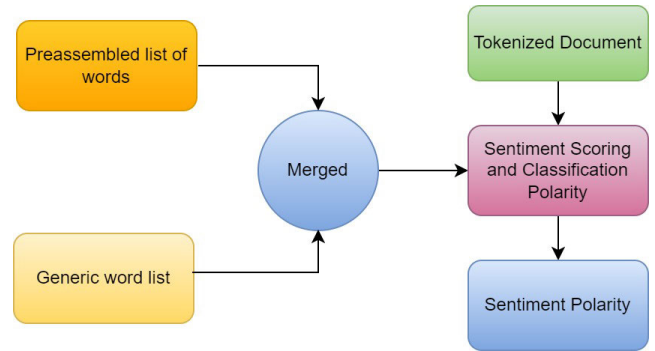


FIGURE 4. Lexicon based model.

3) CORPUS-BASED STRATEGY

This identifies the emotional valence of context-specific phrases. These are the two techniques of this strategy:

Statistical method: Positive polarity is attributed to the phrases that display unpredictable behavior in positive contexts. Negative polarity is demonstrated by negative repetition in negative writing. If the frequency of a term is the same in positive and negative literature, it is neutral polarity.

Semantic approach: This technique assigns sentiment values to words and phrases. The semantically closest words to those terms; this may be done by discovering synonyms and antonyms for that term.

IV. PROPOSED METHODOLOGY

The methodology of the proposed system involves the phases described below and in Figure 5. The raw data from the different sources, which is in the unstructured format, is taken as input to the Tokenization phase.

A. PHASE I: TOKENIZATION

In this phase, the data is divided into tokens by using some delimiters. Mainly, Space, Comma, Hash (#), @, are used as delimiters.

B. PHASE II: STOP WORD ANALYSIS

The tokens formed in Tokenization are fed into this phase in which Stop word removal will be performed. Stop word is the commonly used word that does not have any meaning in itself; such as "the". These words do not play any role in determining the sentiment. So, they are removed. Further emoji's are also removed from the text.

C. PHASE III: NORMALIZATION AND SENTI-SCORE GENERATION

The remaining tokens are sent to this phase in which the normalization of tokens will take place. In parallel, the same tokens will be sent to phase IV. Every content word has some Senti-score in the Senti WordNet. The normalized word's Senti-score will be extracted from the Senti-WordNet in the Senti-Score generation.

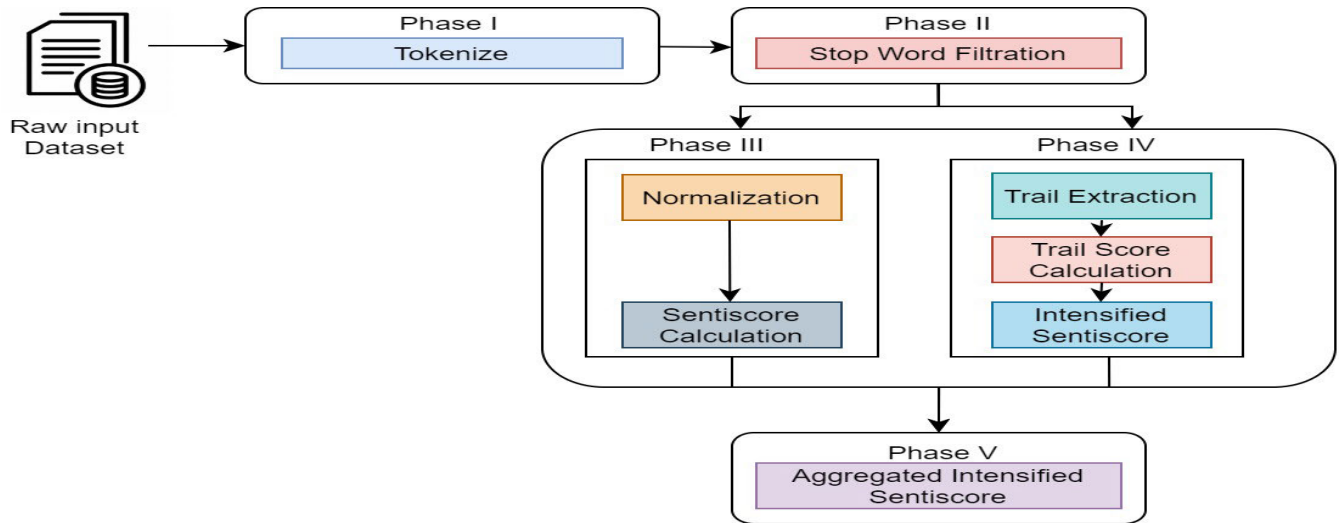


FIGURE 5. Proposed system for detecting lengthened words in sentiment analysis.

D. PHASE IV: TAIL EXTRACTION AND INTENSIFIED TAIL SCORE

Meanwhile, a separate table will be formed for each token saved in the previous phase. The first column will contain the unique characters of the respective tokens stored in the variables. For the second column of the table, it uses the respective normalized word which is the output of the normalization module. Count the frequency of each character by using the normalized word and store it corresponding to the character in the first column. In the third column, take the lengthened word which is saved in the variable in previous phase and the normalized word generated as the output of the normalization module, as the input. Compare these two words of the respective tokens and count the frequency of the character in the tail and insert it corresponding to the character in the first column. After the tables for all the tokens are generated, all the cells of the third column of the respective tables are added and stored in the Tail Score Database as each token's tail score. In this phase, the tail score of each token is retrieved from the database and multiplied with 0.01 and stored in some variables namely "intensified tail score".

E. PHASE V: AGGREGATED INTENSIFIED SENTI-SCORE

After the completion of the fifth phase, the Intensified Senti-score is calculated by the aggregation of the output of Senti-score generation and intensified tail score. Further these scores are used to make the final prediction. The calculated scores are added to the score of sentences and the score for a review is calculated. The review classification threshold value is set to 0.5. Based on a threshold value, all reviews of a dataset are evaluated to determine whether they are positive, negative, or neutral. Anything below a score of -0.05 is classified as negative, while Anything over 0.05 is classified as positive and anything between is classified as neutral.

F. ALGORITHM

Document (D) containing noisy data, where $D = d_1, d_2, d_3, \dots, d_n$

'n' is the total number of documents

'T' is list of tokens, where $T = t_1, t_2, t_3, \dots, t_m$

'm' is the total number of tokens in a document

'W' is the valid word found in WordNet

- 1) Set $i \leftarrow 1$
- 2) For each document (d_i) in D
- 3) Tokenization
- 4) Stop word of filtration
- 5) For each token (t_i)
 - a. Extract the tail based on dictionary count of each letter containing in that token.
 - b. Calculate intensified tail score of token (t_s)
6. For each token (t_i)
 - a. Normalize the word using standard WordNet
 - b. Calculate senti-score(TS_i)
7. For each token(t_i)
 - a. Calculate intensified senti-score(IS_i) = $TS_i * T_s$
8. Aggregate the intensified senti-score (IS_i)

G. EXAMPLE

After taking the input, the proposed system start working by going through the different phases. The initial input is the raw data and the final output is the intensified score of the document.

Raw data: I am feeling veryyyy happpyyyy.

Phase I: Word tokenizer is used to carry out the processing of the raw input text. During this phase, the words are extracted.

"I", "am", "feeling", "veryyyy", "happpyyyy"

Phase II: Stop words are removed in this phase.

“feeling”, “veryyyy”, “happyyyyy”

Phase III: This phase deals with the normalization of the words using standard English WordNet. Although here we normalize the word but retaining the original input is also required. For that a separate table is maintained to keep the flavor of lengthened alive. This further used in next phase for caring out the intensification.

Normalized words	Retaining original words
1. Feel	variable1=feeling
2. Very	variable2=veryyyy
3. Happy	variable3=happyyyyy

Phase IV: This phase is the backbone of the proposed system. Here, the actual intensified senti-score is being calculated.

	Unique Character	Dictionary Count	Tail
Let Senti-score for “feel” is x1	f	1	0
	e	2	0
	l	1	0
	i	1	0
	n	1	0
	g	1	0
Table for variable 1= feeling			

	Unique Character	Dictionary Count	Tail
Let Senti-score for “very” is x2	v	1	0
	e	1	0
	r	1	0
	y	1	3
Table for variable 2= veryyyy			

	Unique Character	Dictionary Count	Tail
Let Senti-score for “happy” is x3	h	1	0
	a	1	0
	p	2	1
	y	1	3
Table for variable 3= happyyyyy			

Intensified Tail Score for variable 1= 0×0.01

Intensified Tail Score for variable 2= 3×0.01

Intensified Tail Score for variable 3= 4×0.01

Intensified Senti-score for “feeling” = $x1 + \text{Intensified Tail Score for variable 1}$

Intensified Senti-score for “veryyyy” = $x2 + \text{Intensified Tail Score for variable 2}$

Intensified Senti-score for “happyyyyy” = $x3 + \text{Intensified Tail Score for variable 3}$

Phase V: Finally, aggregation of the intensified score of each token of a document give rise to the actual intensified senti-score of a document.

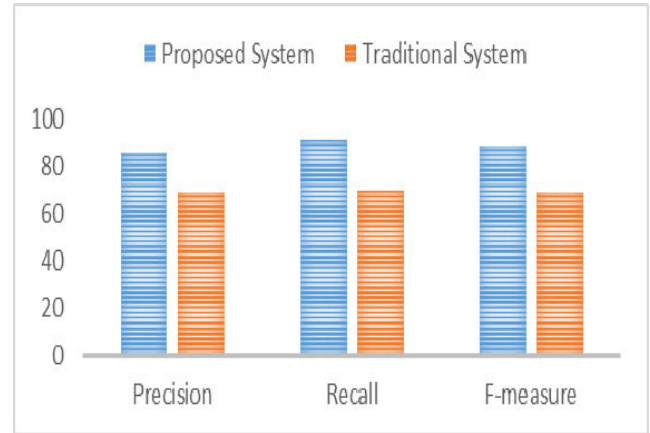


FIGURE 6. Comparison results on facebook dataset for children group.

V. EXPERIMENTAL RESULTS

The results are gathered using word level tokenization. The experiment consists the components mentioned below:

A. PERFORMANCE METRICS

Precision and recall along with accuracy is used for the evaluation [82].

Precision of an information retrieval system is defined as the proportion of the relevant documents in the retrieved set.

$$P = \text{True Positive} / (\text{True Positive} + \text{False Positive}) \quad (2)$$

Recall is defined as the proportion of the relevant documents in the collection that has actually been retrieved.

$$R = \text{True Positive} / (\text{True Positive} + \text{False Negative}) \quad (3)$$

F-measure is defined as the harmonic mean of Precision and Recall.

$$F - \text{measure} = 2 * PR / (P + R) \quad (4)$$

B. EXPERIMENTAL RESULTS

The experimental results for the intensification of senti-score is carried out on data combined by three sources Twitter, Facebook and Personal chat in young people and children. The individuals belong to 13 to 18 age group comes under Children group. The individuals belong to 19 to 40 age group comes under youngster group. The experiments are divided into two parts to evaluate the performance of proposed system and detail about these experiments are given below: -

1) EXPERIMENT 1: RESULTS

This subsection demonstrates the experimental results of proposed system using the lengthened words in the dataset for detecting the sentiment analysis. Table 3 represents the result of proposed and traditional sentimentanalysis system for children group. The traditional system ignores the lengthened words but proposed system utilized these words for detecting the sentiment analysis. For children group as shown in Figures 6 to 8, proposed system with lengthened

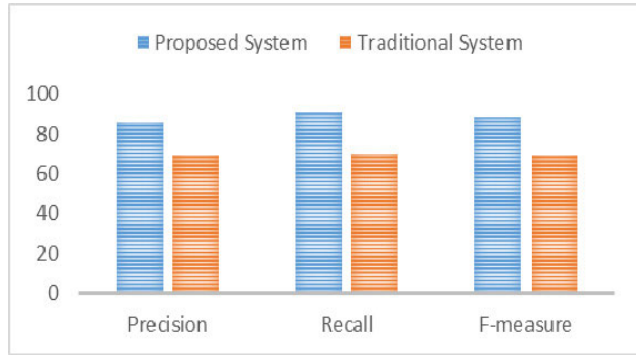


FIGURE 7. Comparison results on twitter dataset for children group.

TABLE 3. Performance measures of the Proposed and traditional system for Children Group.

Group	Dataset	Precision	Recall	F-measure
Proposed System	Facebook	89.97	91.65	90.8
	Twitter	84.56	87.04	85.78
	Personal chat	76.98	87.20	81.77
Traditional System (Ignore lengthened words)	Facebook	69.78	65.67	67.47
	Twitter	64.34	66.88	65.05
	Personal chat	58.78	61.21	59.67

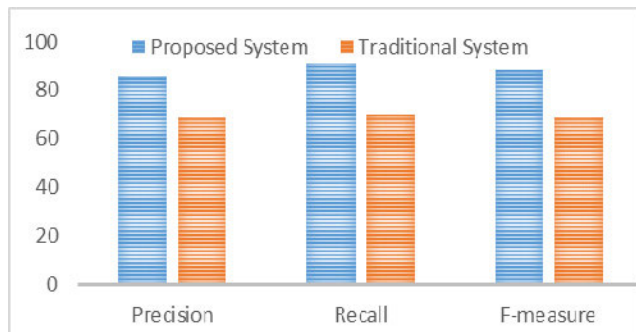


FIGURE 8. Comparison results on personal chat dataset for children group.

words obtains higher precision, recall, and F-measure rate i.e., 89.97%, 91.65%, and 90.8%, respectively on Facebook dataset as compared to traditional system. The proposed system achieves better F-measure i.e., 85.78% and 81.77 respectively for Twitter and Personal chat datasets as compare to traditional system Further, Table 4 illustrates the result of sentiment analysis system for youngster group. It can be observed from Figures 9 to 11 that proposed system outperform the traditional system by achieving better F-measure rate in the range of for all datasets. Thus, lengthened words help in improving the F-measure of sentiment analysis.



FIGURE 9. Comparison results on facebook dataset for young group.

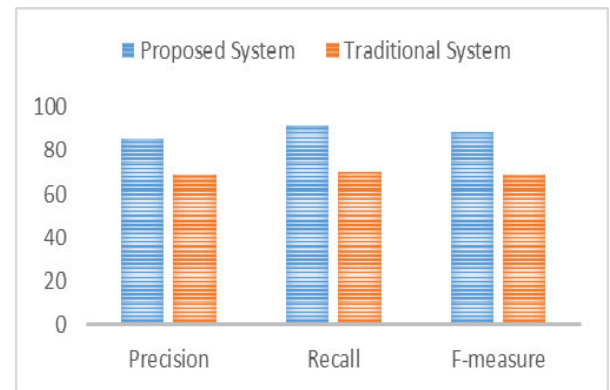


FIGURE 10. Comparison results on twitter dataset for young group.

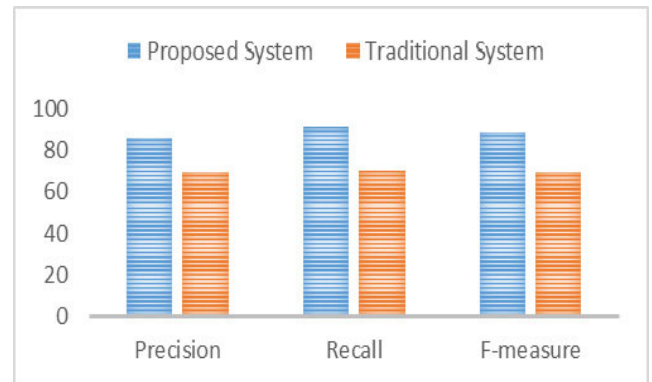


FIGURE 11. Comparison results on personal chat dataset for young group.

2) EXPERIMENT 2: IMPACT OF LENGTHENED WORDS ON SENTIMENT ANALYSIS

The experimental results of proposed system are varying due to diversity of information used in the experiment. The lengthened words like happyyy, coollll etc. are also considered to calculate the final senti-scores for detecting sentiment analysis. The averaged F-measure of proposed system is increased by 21.22% via involving lengthened words in the dataset as compared to traditional system. The lengthened words enhanced the results of proposed system. The performance of the proposed system is evaluated using the overlapping of results with the gold standards. The gold standard is being synthesized with the help of 4 graduate students. They are told to read the articles, tweets, chats and annotate them

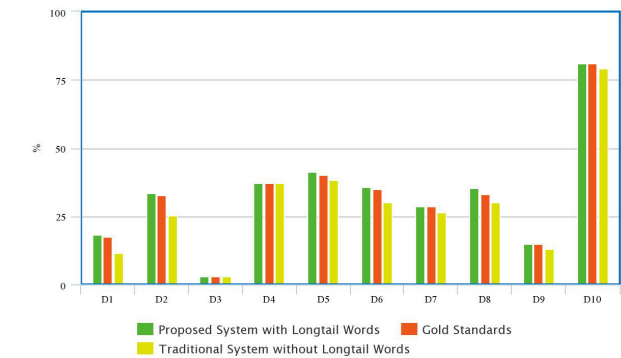


FIGURE 12. Effect over negative for young.

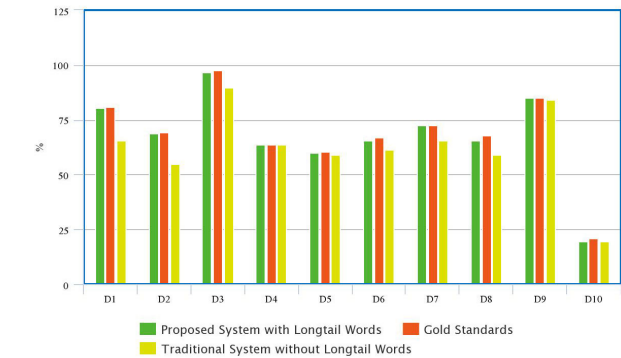


FIGURE 13. Effect over positive for young.

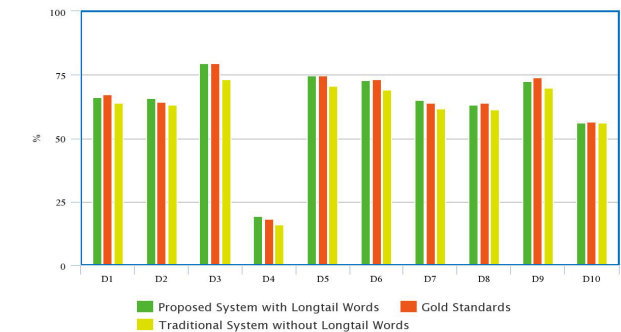


FIGURE 14. Effect over positive for children.

as if they were shareholders in the organization described in the topic statement. The graduate students are instructed to annotate the data for overall sentiment if no firm is cited. A 7-point scale, ranging from highly positive to very negative, is used for all jobs. We carried out an agreement study with all 4 students using the shared set of data in order to assess the dependability of the annotation scheme and the level of agreement between the annotators. Kappa statistic has been used to evaluate the system. The results also show that the intensified senti-score is very close to the Gold standard both for young and children. Figures 12 to 15 have shown the results of percentage of positive-ness and negative-ness in top 10 random documents. The gap of results in gold standard and proposed system is very less as compared to traditional system. In Figures 14 and 15, it is shown that the closeness of results to the gold standard is less as compared to the

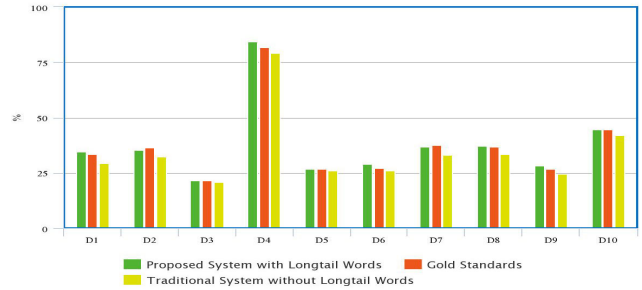


FIGURE 15. Effect over negative for children.

TABLE 4. Performance measures of the proposed and traditional system for young group.

Group	Dataset	Precision	Recall	F-measure
Proposed System	Facebook	96	94.51	95.25
	Twitter	91.45	89.43	90.43
	Personal chat	85.67	91.12	88.31
Traditional System (Ignore lengthened words)	Facebook	76	75	75
	Twitter	71	69	70
	Personal chat	69	70	69

Figures 12 and 13. The reason for this may be the less use of lengthened in children as compared to youngsters. Thus, the proposed system has shown its valuable impact over the data shared by youngsters.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, we have presented a system for extracting sentiments. It can be used in other applications as well, i.e., emotion detection. It not only important to capture actual sentiment but the actual mood of the person as well. This study confirmed that the there is a strong impact of lengthened words in sentiment analysis. It is critical to correctly detect these words in order to provide complete coverage. We also showed that lengthening is not random, and that it is frequently utilised with subjective terms to accentuate their meaning. The effect is positively related to the percentage of the internet usage as the usage is high in youngsters as compared to old age and children. As a result of this discovery, we developed an unsupervised strategy based on lengthening for detecting novel sentiment bearing words not found in the existing lexicon and determining their polarity. Such a technology is vital for up-to-date sentiment recognition in the frequently changing realm of net-speak and microblogging.

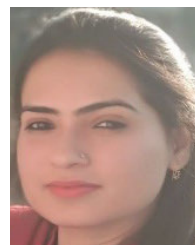
The proposed system has a limitation as well; this system assumes that the spellings are correct. The proposed system is still on the way to capture the results at par with the gold standard. This we will try to optimize as well as betterment in the performance by formulating some rules.

Only one component of the lengthening phenomena is investigated in this work. In future, other features of

REFERENCES

- 9786

- [40] Y. A. Amrani, M. Lazaar, and K. E. E. Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Proc. Comput. Sci.*, vol. 127, pp. 511–520, Jan. 2018.
- [41] K. A. Hasan, M. S. Sabuj, and Z. Afrin, "Opinion mining using naive Bayes," in *Proc. IEEE Int. WIE Conf. Elect. Comput. Eng. (WIECON-ECE)*, Dec. 2015, pp. 511–514.
- [42] Y. Wan and Q. Gao, "An ensemble sentiment classification system of Twitter data for airline services analysis," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 1318–1325.
- [43] P. V. Ngoc, C. V. T. Ngoc, T. V. T. Ngoc, and D. N. Duy, "A C4.5 algorithm for English emotional classification," *Evolving Syst.*, vol. 10, no. 3, pp. 425–451, Sep. 2019.
- [44] N. Saleena, "An ensemble classification system for Twitter sentiment analysis," *Proc. Comput. Sci.*, vol. 132, pp. 937–946, Jan. 2018.
- [45] J. J. Bird, A. Ekárt, C. D. Buckingham, and D. R. Faria, "High resolution sentiment analysis by ensemble classification," in *Proc. Intell. Comput. Comput. Conf. Cham, Switzerland: Springer*, Jul. 2019, pp. 593–606.
- [46] N. Sultana and M. M. Islam, "Meta classifier-based ensemble learning for sentiment classification," in *Proc. Int. Joint Conf. Comput. Intell.* Singapore: Springer, 2020, pp. 73–84.
- [47] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, "Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection," in *Proc. IEEE Symp. Comput. Intell. Big Data (CIBD)*, Dec. 2014, pp. 1–6.
- [48] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of roman-urdu opinions using Naïve Bayesian, decision tree and KNN classification techniques," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 28, no. 3, pp. 330–344, Jul. 2016.
- [49] L. Wikarsa and S. N. Thahir, "A text mining application of emotion classifications of Twitter's users using Naïve Bayes method," in *Proc. 1st Int. Conf. Wireless Telematics (ICWT)*, Nov. 2015, pp. 1–6.
- [50] X. Fei, H. Wang, and J. Zhu, "Sentiment word identification using the maximum entropy model," in *Proc. 6th Int. Conf. Natural Lang. Process. Knowl. Eng.*, Aug. 2010, pp. 1–4.
- [51] H. Y. Lee and H. Renganathan, "Chinese sentiment analysis using maximum entropy," in *Proc. Workshop Sentiment Anal. AI Psychol.* Nov. 2011, pp. 89–93.
- [52] M. Murata, K. Uchimoto, M. Utiyama, Q. Ma, R. Nishimura, Y. Watanabe, and K. Doi, "Using the maximum entropy method for natural language processing: Category estimation, feature extraction, and error correction," *Cogn. Comput.*, vol. 2, no. 4, pp. 272–279, 2010.
- [53] J. Smalović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Inf. Sci.*, vol. 285, pp. 181–203, Nov. 2014.
- [54] T. T. Li and D. H. Ji, "Sentiment analysis of micro-blog based on SVM and CRF using various combinations of features," *Appl. Res. Comput.*, vol. 32, pp. 978–981, Sep. 2015.
- [55] H. Jin, R. Tong, and J. Ruiquan, "Sentiment analysis in financial domain based on SVM with dependency syntax," *Comput. Eng. Appl.*, vol. 51, no. 23, pp. 230–235, 2015.
- [56] R. Liu, R. Yang, S. Li, Y. Shi, and X. Jin, "Painting completion with generative translation models," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 14375–14388, Jun. 2020.
- [57] H. Lu, D. Wang, Y. Li, J. Li, X. Li, H. Kim, S. Serikawa, and I. Humar, "CONet: A cognitive ocean network," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 90–96, 2019.
- [58] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018.
- [59] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2315–2322, Aug. 2017.
- [60] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
- [61] T. Mikolov, I. C. K. Sutskever, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–12.
- [62] L.-S. Chen, C.-H. Liu, and H.-J. Chiu, "A neural network based approach for sentiment classification in the blogosphere," *J. Informetrics*, vol. 5, no. 2, pp. 313–322, 2011.
- [63] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," 2015, *arXiv:1510.03820*.
- [64] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," 2016, *arXiv:1606.01781*.
- [65] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," 2014, *arXiv:1412.1058*.
- [66] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1422–1432.
- [67] M.-Y. Day and Y.-D. Lin, "Deep learning for sentiment analysis on Google play consumer review," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Aug. 2017, pp. 382–388.
- [68] R. Johnson and T. Zhang, "Supervised and semi-supervised text categorization using LSTM for region embeddings," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 526–534.
- [69] M. S. Akhtar, A. Kumar, A. Ekbal, and P. Bhattacharyya, "A hybrid deep learning architecture for sentiment analysis," in *Proc. COLING 26th Int. Conf. Comput. Linguistics: Tech. Papers*, Dec. 2016, pp. 482–493.
- [70] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.
- [71] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. ICML*, Jan. 2011, pp. 1–15.
- [72] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," 2012, *arXiv:1206.4683*.
- [73] M. Sun, Q. Tan, R. Ding, and H. Liu, "Cross-domain sentiment classification using deep learning approach," in *Proc. IEEE 3rd Int. Conf. Cloud Comput. Intell. Syst.*, Nov. 2014, pp. 60–64.
- [74] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in translation: Contextualized word vectors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–14.
- [75] S. Ilić, E. Marrese-Taylor, J. A. Balazs, and Y. Matsuo, "Deep contextualized word representations for detecting sarcasm and irony," 2018, *arXiv:1809.09795*.
- [76] A. M. Dai and Q. V. Le, "Semi-supervised sequence learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–15.
- [77] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," 2018, *arXiv:1801.06146*.
- [78] R. L. Vieri, A. K. Rajagopal, R. Subramanian, O. Lanz, E. Ricci, N. Sebe, and K. Ramakrishnan, "Boosting-based transfer learning for multi-view head-pose classification from surveillance videos," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 649–653.
- [79] S. Wang and Z. Li, "A new transfer learning boosting approach based on distribution measure with an application on facial expression recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 432–439.
- [80] W. Wang, L. Chen, K. Chen, K. Thirunaryan, and A. P. Sheth, "Adaptive training instance selection for cross-domain emotion identification," in *Proc. Int. Conf. Web Intell.*, Aug. 2017, pp. 525–532.
- [81] J. T. Zhou, I. W. Tsang, S. J. Pan, and M. Tan, "Heterogeneous domain adaptation for multiple classes," in *Proc. 17th Int. Conf. Artif. Intell. Statist.*, in *Proceedings of Machine Learning Research*, vol. 33, 2014, pp. 1095–1103. [Online]. Available: <https://proceedings.mlr.press/v33/zhou14.html>
- [82] A. Kukkar, R. Mohana, Y. Kumar, A. Nayyar, M. Bilal, and K.-S. Kwak, "Duplicate bug report detection and classification system based on deep learning technique," *IEEE Access*, vol. 8, pp. 200749–200763, 2020.



ASHIMA KUKKAR received the B.Tech. degree in information and technology from Rayat Bharat University, Mohali, India, in 2014, the M.Tech. degree in computer science and engineering from the Jaypee University of Information Technology, Solan, Himachal Pradesh, India, in 2016, and the Ph.D. degree in computer science from Jaypee University, in 2019, in the area of bug summarization, severity classification, and assignment for automated bug resolution process. She is currently working as an Assistant Professor with the Department of CSE, Chitkara University, Rajpura, India. Her research interests include problems relating to software engineering and natural language processing. She likes applying machine learning and deep learning in text. She has a passion for software engineering with the goal of improving the software development life cycle.



RAJNI MOHANA received the B.Tech. degree in computer science and engineering from BPUT, Rourkela, Orissa, India, in 2003, the M.Tech. degree in computer science and engineering from Punjab Technical University, Jalandhar, India, in 2009, and the Ph.D. degree in computer science and engineering from the Jaypee University of Information Technology (JUIT), Wagnaghat, Solan, Himachal Pradesh, India, in 2013, working on interoperability problem in service oriented architecture computing. She is currently working as an Associate Professor with the Department of CSE, JUIT. She has over 19 years of experience in Academia. Her research interest includes sentiment analysis cloud computing. She has published her research in various impact factor journals, such as *Sensors* (MDPI), *Proceedings of National Academy of Sciences*, and *Journal of Physical Sciences*. She has published around 40 papers in various ESCI and Scopus journals. She has guided four Ph.D. students and is currently guiding two more students. She has guided four post graduate student.



AMAN SHARMA received the B.Tech. degree in computer science and engineering from RBIEBT, Mohali, Punjab, the M.E. degree in computer science and engineering from the Thapar Institute of Engineering and Technology, Patiala, Punjab, titled "Community detection and analysis of twitter social data," and the Ph.D. degree from the Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, titled "Machine learning-based framework for drug prediction of cancerous profile." He is currently working as an Assistant Professor (SG) with the Department of Computer Science and Engineering, Jaypee University of Information Technology, Wagnaghat, Solan, Himachal Pradesh. He has around seven years of teaching and research experience. He has been a Rank Holder throughout his studies. He has also obtained A. I. R-1800 in GATE 2013.



ANAND NAYYAR received the Ph.D. degree in computer science from Desh Bhagat University, in 2017, in the area of wireless sensor networks, swarm intelligence, and network simulation. He is currently working with the School of Computer Science, Duy Tan University, Da Nang, Vietnam, as a Professor, the Scientist, the Vice Chairperson (Research), and the Director of the IoT and Intelligent Systems Laboratory. A Certified Professional with more than 125 professional certificates from CISCO, Microsoft, Amazon, EC-Council, Oracle, Google, Beingcert, EXIN, GAQM, and Cyberoam. He has published more than 150 research papers in various high-quality ISI-SCI/SCIE/SSCI impact factor journals cum Scopus/ESCI indexed journals, more than 70 papers in international conferences indexed with Springer, IEEE, and ACM digital library, more than 40 book chapters in various Scopus, Web of Science indexed books with Springer, CRC press, Wiley, IET, and Elsevier with citations: more than 8000, H-Index: 46, and I-Index: 165. He is also a member of more than 60 associations as a senior and life member, including IEEE and ACM. He has authored/coauthored cum edited more than 40 books of computer science. He has associated with more than 500 international conferences as a programme committee/the chair/an advisory board/a review board member. He has 18 Australian patents, four German patents, two Japanese patents, 11 Indian design cum utility patents, one USA patent, three Indian copyrights, and two Canadian copyrights to his credit in the area of wireless communications, artificial intelligence, cloud computing, IoT, and image processing. He awarded 38 awards for Teaching and Research—the Young Scientist, the Best Scientist, the Best Senior Scientist, the Asia Top 50 Academicians and Researchers, the Young Researcher Award, the Outstanding Researcher Award, the Excellence in Teaching, the Best Senior Scientist Award, and the DTU Best Professor and Researcher Award, in 2019, 2020, 2021, and 2022. He is listed in top 2% scientists as per Stanford University (2020, 2021, and 2022). He is acting as an Associate Editor for *Wireless Networks* (Springer), *Computer Communications* (Elsevier), *International Journal of Sensor Networks* (IJSNET) (Inderscience), *Frontiers in Computer Science*, *PeerJ Computer Science*, *Human-Centric Computing and Information Sciences* (HCIS), *IET Quantum Communications*, *IET Wireless Sensor Systems*, *IET Networks*, *IJDST*, *IJISP*, *IJCINI*, and *IJGC*. He is acting as the Editor-in-Chief of IGI-Global, USA Journal titled *International Journal of Smart Vehicles and Smart Transportation* (IJSVST). He has reviewed more than 2500 articles for diverse Web of Science and Scopus indexed journals. He is currently researching in the area of wireless sensor networks, Internet of Things, swarm intelligence, cloud computing, artificial intelligence, drones, blockchain, cyber security, network simulation, big data, and wireless communications.



MOHD. ASIF SHAH is currently working as an Associate Professor with Bakhtar University, Kabul, Afghanistan. He has more than ten years of demonstrated teaching experience in the field of research and teaching. Furthermore, he has the ability and capability to analyze both quantitative and qualitative data through valid software's like R, STATA, SPSS, and MS Excel using different statistical methods and techniques. He has published more than 30 research papers which are indexed in Scopus and Web of Science with more than 50 citations.

...