

Received November 5, 2019, accepted December 28, 2019, date of publication January 1, 2020, date of current version January 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963586

Context-Aware Sliding Window for Sentiment Classification

MUHAMMAD ALI MASOOD¹, RABEEH AYAZ ABBASI¹, AND NG WEE KEONG²

¹Department of Computer Science, Quaid-i-Azam University, Islamabad 45320, Pakistan

²School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798

Corresponding author: Rabeeh Ayaz Abbasi (rabbasi@qau.edu.pk)

ABSTRACT Sentiment classification is an active area of research with applications in many domains. Many researchers in the past have proposed techniques to identify sentiments with reasonable accuracy. However, the focus is more on the syntactic and semantic features of the documents. These features are effective but they ignore the user's past sentiments. In this research, we hypothesize that the past sentiments help the classifier to effectively link the user's history along with the contents of the current tweet. Thus, allowing learning algorithms to correlate past activities in determining the current sentiments. For this sake, we propose three sliding window features to accumulate past sentiments from the time series data. In this paper, we propose seven variations of Context-aware Sliding Window (CSW) features on different machine learning and deep learning algorithms. Furthermore, we propose a temporal dataset of user tweets, which is manually labeled by nine human annotators. The proposed dataset consists of 36 users having 4,557 tweets. Results indicate significant improvements over six state-of-the-art baseline methods.

INDEX TERMS Sentiment classification, user history, temporal sliding window, deep learning, machine-learning, temporal dataset.

I. INTRODUCTION

Over the past decade, many social media platforms have gained popularity by allowing users to freely share their opinions and thoughts. These opinions and thoughts are often analyzed using automated sentiment classifiers to help various industries such as the airport services quality [1], food companies [2], many others [3], [4]. Sentiment classification is an active area of research; numerous methods have been proposed to classify sentiments [5]–[7]. Most of the past research uses the contents of the current tweet to identify sentiments. Although such kind of sentiment classification technique might be effective, it can be enhanced by utilizing sentiment clues from the past. Past experiences of a person affect his or her present sentiments. For example, if a person is angry, his or her sentiments would likely remain the same for some period of time.

To utilize the sentiment clues from the past, we propose a context-aware sliding window (CSW) features. CSW features accumulate past clues from sentiments within a given time frame to classify the current tweet. We propose two categories of features in addition to the content features (1) sliding

window and (2) ratio of sentiments features. The first category accumulates the past sentiments within a window size. The window size can be either temporal (last t minutes) or non-temporal (last k sentiments). The temporal sliding window assumes that the user's activities change over time, therefore the temporal window captures most recent activities. Generally, the temporal window accumulates sentiment clues for a specific time period, which is useful in attaining information from frequent users, who tweet regularly and normal users. However, a limited time span may not be useful for users who do not share tweets regularly, thus missing sentiment clues in the temporal sliding window feature. For this sake, the last k sentiments feature extracts past sentiments regardless of the time, which reduces the number of missing values. Both the last k sentiments and temporal sliding window features gain insights from the recent past, but they ignore insights from the overall activities of a user. Therefore, the feature *ratio of sentiments* gathers all the past sentiments of a user. The ratio of sentiments feature aims to create a user profile, which can later be combined with other features to produce an accurate classification.

In this paper, we propose seven novel variations of the features for sentiment classification. These variations learn patterns from the users who have high, normal and low

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

tweet frequencies. We analyze the accuracy of the proposed models in different window sizes, i.e., 30 minutes to seven days for the temporal sliding window and 1 to 5 for last k sentiments features. In the absence of a publicly available temporally labeled user sentiment dataset, we develop our dataset for evaluation. The dataset consists of 4,557 tweets from 36 users. The dataset is manually labeled by nine human experts. We analyze our results on a large list of window sizes to give a clear picture of the performance under different window sizes. Results suggest that the most effective time window is between 30 minutes to 24 hours, similarly, the best value of k is between 2 to 5. We compare our proposed feature-driven models with six state-of-the-art baselines from machine learning and deep learning domain. All of our proposed models have improvements over the baselines. Following are the key contributions of this research:

- 1) We propose three novel features (i.e., temporal sliding window, last k sentiment and ratio of sentiments) for sentiment classification. We also propose seven variations of these features to learn patterns regardless of the user tweet frequency (i.e., high, normal, or low).
- 2) We analyze results on various temporal and non-temporal window sizes.
- 3) We present a new user-based temporal sentiment dataset of 4,557 tweets from 36 users.

The rest of the paper is organized as follows: Section II discusses the background and related work on the sentiment classification. Sections III and IV introduce our proposed features and research questions for this study. Section V introduces the dataset and the labeling mechanism. Section VI discusses the results. Finally, Section VII presents our conclusions.

II. BACKGROUND AND RELATED WORK

Significant research has been carried out in the area of sentiment classification. We divide sentiment classifiers into four categories, i.e., lexicon, machine learning, deep learning, and temporal models.

A. LEXICON BASED SENTIMENT CLASSIFICATION

A lexicon is a dictionary of words, where each word in the lexicon has a positive, negative, or neutral polarity score. A lexicon is not limited to words, rather it includes emojis with their respective polarity scores. Therefore, some researchers apply polarity scores from emojis to determine the sentiments [8], [9]. Words in a tweet are matched with words in the lexicon. Polarities of the matched words are aggregated to measure the overall polarity of the tweet [10]–[12]. This simple matching technique does not consider the semantics or emotions of a user. Besides that, these positive and negative words are limited in a specific lexicon, so the researchers extend the lexicon by adding co-occurring words [13], [14]. Co-occurring words can effectively extend the lexicon but they may add semantically unrelated words. Therefore, Saif *et al.* [15] proposed a model SentiCircles to group similar co-occurring words using the term frequency-inverse document frequency (TF-IDF) and

prior sentiments of the words. This way, authors were able to extend the lexicon with semantically related words.

This expansion might be ineffective for social media users, who use acronyms and misspelled words. Satapathy *et al.* [11] proposed to correct the spelling mistakes with the phonetic sounds of a word using Soundex. Moreover, they also replaced the abbreviations with their full forms using a fixed abbreviation list.

Lexicon based techniques must incorporate negation words because they can flip the overall sentiment of a tweet. Mukhtar and Khan [16] proposed a total of 154 features, among them, some features obtained the occurrence and placement of the negation words to classify the sentences with negation.

In a social network, users may share tweets in multiple languages; in such a case, it might be difficult to devise a multi-lingual lexicon. Therefore, Asghar *et al.* [17] proposed some rules to classify the sentiment of a tweet. Rule-based techniques can be inaccurate because they rely on a fixed set of rules, which may become outdated after some time.

Lexicon based techniques are effective and easy to implement because they do not require any labeling. On the contrary, lexicon-based techniques rely on dictionaries, so they might not handle slang, acronyms, and misspelled words.

B. MACHINE LEARNING BASED SENTIMENT CLASSIFICATION

Sentiment depends on many features; some of them are taken from contents while others come from user's profile, images, URLs, etc. Together these features make a large search space for manual classification; therefore researchers use machine learning algorithms to effectively find patterns in high dimensional data.

Similarly, some researchers proposed to apply textual features like n-grams, bag of words (BOW), and term frequency-inverse document frequency (TF-IDF) in machine learning algorithms to classify sentiments [13], [18], [19]. On the other hand, some researchers devised a hybrid technique, which combines lexicon and machine learning together to recommend sentiments [14], [16], [20]. A tweet may consist of different opinions toward multiple entities; therefore it is important to link sentiments with entities/organizations. In such a case, these lexicon features may not be effective because they can not identify the relationship between words.

For example, a tweet saying that “*I love chocolate cookies*” is showing a positive relationship between the person and chocolate cookies. Therefore, researchers adopt POS tagger to first identify the entities and then find the sentiments towards those entities [10], [13], [21].

All the above techniques do not consider context or semantic similarity of words during classification. The context of words can be incorporated in a machine learning algorithm using word embedding [18]. This helps machine learning algorithms to learn densely distributed representation for each word. Similarly, some researchers devised a temporal

embedding of users and products, this embedding helps in learning the patterns from the history (of users and products) while recommending the current sentiment [22].

Most of the machine learning techniques rely on supervised learning, which requires manual or semi-automatic labeling of data. Labeling is a time-consuming activity, therefore Pandey *et al.* [20] proposed a hybrid learning mechanism that creates clusters from the initial samples and supplies them to Cuckoo search [23] for classification. Cuckoo search is a mechanism that takes clusters (initial samples) and then generates a new population by pairing the best samples (using a fitness function). Such an approach is effective for large unlabeled data, but it may get stuck in the local optima because of the inaccurate clusters.

C. DEEP LEARNING SENTIMENT CLASSIFICATION

Over the past, machine learning and lexicon-based techniques have been used for sentiment classification. Some researchers modeled features to enhance the functionality of the learning algorithms [13], [18], [19]. However, feature modeling requires careful shortlisting of the discriminating features; Deep learning automatically learns discriminating features from the data. The process of deep learning starts with learning word embedding to have a similar representation for similar words. However, similar words may have opposite sentiments; in such a case, word embedding will have the same representation for opposite sentiment words - thus misleading the classifier. Therefore, Tang *et al.* [24] focuses on learning a sentiment specific embedding. The idea is to separate opposite polarity words (for example, good and bad) in the neighboring vectors. In simple terms, the authors emphasize on the sense of the word while creating the embedding. In the same manner, Chen *et al.* [25] use emojis and an attention mechanism to emphasize the sense within the emojis while creating an embedding. Emoji, word, or sentiment specific embedding do not consider the variation in the sense under different topics. Therefore, Zhao and Mao [26] have applied the LDA model to learn topic embedding for sentence classification. Most of the word embedding based techniques rely on the word with no or minor spelling mistakes. Since Twitter puts a limit on the number of characters, so tweets also contain acronyms and misspelled words. Therefore, authors in [27], [28] have proposed character level embedding to learn embedding for multiple languages with spelling mistakes.

After learning word embeddings, researchers have applied it on various deep learning techniques like Convolutional Neural Network (CNN), Gated Neural Network (GNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), etc. CNN is a deep neural network with hidden layers having filtering (convolution) and pooling function to learn patterns in the data. The pooling function reduces the dimension of the input layer by extracting maximum and minimum values within bins (region of interest). CNN naïvely assumes that the layers are independent of each other. Therefore, LSTM, RNN, and GNN assume the output of the current layer is dependent on the computations in the previous layers.

With the same assumption, Tang *et al.* [29] have used Gated Neural Network (GNN) to model documents sequentially. The notion of the model is to predict the sentiment of the current sentence on the last sentence's sentiment. This idea differs from our approach in two ways: First, we hypothesize that sentiments are dependent on the recent context, which can be learned from variable window sizes rather than the last sentiment. Second, we are relying on Twitter users who use informal communication without any structure, whereas a document is formally written.

Some researchers [30], [31] have emphasized on the entity/topic of the tweet during classification. In the same manner, Shi *et al.* [32] emphasize on the hierarchies by proposing hierarchical LSTM with textual and user profile features. Each deep learning method has its positive and negative aspects, therefore, researchers have proposed ensemble techniques with CNN, RNN, and LSTM to avoid the drawbacks of each model [33], [34]. In the same fashion, Yang *et al.* [35] interpolated the output of CNN and LSTM by weighting the predictions of both the models. Generally, RNN has the problem of vanishing gradient descent (i.e., the computation in the long distant layers vanishes after a few layers), so Ding *et al.* [36] proposed to directly connect all layers to reduce the effects of gradient descent.

Deep learning methods are effective in sentiment classification but they require a large annotated dataset. Therefore, Yang *et al.* [37] have proposed a sentiment classification technique that uses both the supervised deep neural network and unsupervised probabilistic generative model. On the other hand, Xu and Tan [7] have proposed a semi-supervised learning method using a deep generative model to utilize the structure of labeled data onto a large unlabelled data.

D. TEMPORAL SENTIMENT CLASSIFICATION

Since users share their opinions over different time spans, therefore it is essential to learn the historical patterns from the temporal dimension. Temporal sentiment classification focuses on time series (temporal dimension) data to identify the sentiment of a tweet. One such research has applied user and product embedding on the temporal dimension to identify the sentiment [22]. The idea is based on the assumption that a user gives similar sentiments to similar products in the future. Users/products embedding uses the recent trends of products and users to identify the sentiment. However, the assumption of the same sentiments for similar products can be misleading because humans also give equal importance to their experiences alongside the history of the product.

Some researchers [21], [38] have focused on an analysis of user's sentiments over time.

First, they identify the sentiment of the tweets, later these sentiments are used to analyze the distribution of user's sentiments. The authors concluded that users tend to share more positive sentiments, however, on some occasions (like terrorist attacks, corruption cases, etc.) negative sentiments tend to increase [38]. Therefore, it is important to understand the context before identifying the sentiments. The authors

Sr #	Tweet Date Time	User ID	Tweet	Sentiment Clue
1	10/19/2018 12:10pm	User 1	Hey guys love the coffee	1
2	10/20/2018 01:20pm	User 1	At Eiffel tower, love the place	1
3	10/20/2018 01:30pm	User 1	Hate raining	-1
4	10/20/2018 02:10pm	User 1	hello guys I am back	0
5	10/11/2018 10:10pm	User 2	Hate this coffee	-1
6	10/11/2018 10:15pm	User 2	Even the manager was very irresponsible	-1
7	10/15/2018 11:30pm	User 2	I love to play soccer with friends	1

Aggregated Sentiments	Tweet Date Time	User ID	Tweet
Null	10/19/2018 12:10pm	User 1	Hey guys love the coffee
Null	10/20/2018 01:20pm	User 1	At Eiffel tower, love the place
1	10/20/2018 01:30pm	User 1	Hate raining
0	10/20/2018 02:10pm	User 1	hello guys I am back
Null	10/11/2018 10:10pm	User 2	Hate this coffee
-1	10/11/2018 10:15pm	User 2	Even the manager was very irresponsible
Null	10/15/2018 11:30pm	User 2	I love to play soccer with friends

FIGURE 1. An example of temporal sliding window with window size of 1 hour.

in [39] have considered the context by taking the last conversation of the current tweet. The authors extracted the conversation by grouping tweets using mentions. Authors proposed to learn the context using LSTM, which can learn long distant sentiments context. Conversation based models are only capable of learning the context within the conversation; however, they ignore the overall context of a user. These techniques are different from our research because we emphasize on the user's recent context, this recent context can help recommend accurate sentiments for both conversational and non-conversational tweets.

Few research articles have focused on the user's recent context while identifying the sentiment.

Therefore, we intend to use the temporal dimension by hypothesizing that the current sentiment of a user is dependent on his recent sentiments.

III. CONTEXT-AWARE SLIDING WINDOW (CSW)

In this paper, we propose a method that takes the past sentiments along with contents to identify the sentiment of the current tweet. We hypothesize that the sentiment of the present tweet depends on the recent moods of the user. For example, a user with a high number of negative tweets (in the recent past) will have a high probability of sharing negative content in the current tweet. Therefore, we propose to utilize the past sentiments of a user. Besides that, we emphasize to learn the profile of a user to use patterns of similar users in identifying the sentiment of the current tweet.

We propose three categories of features: the first category is the contents feature, which mainly constitutes bag of words model. The second category has sliding window features, which define the size of the window (either temporal or non-temporal) to accumulate the sentiments within the defined window size. The third category gathers the overall personality of a user by computing the ratio of positive, negative and neutral tweets. These features are applied to machine learning and deep learning algorithms to identify sentiments. The details of the categories are given below:

A. CONTENTS FEATURES

Contents features are widely used for sentiment classification [13], [18], [19]. There are many ways to model the

content features like bag of words model (BOW), n-grams, etc. In this research, for machine learning algorithms we use the BOW model because of its simplicity. For deep learning algorithms, we use GLOVE embedding [40] to obtain a lower-dimensional vector representation for each word.

Contents features are the key indicators to determine the sentiments in the textual documents. However, tweets have few words, which makes it difficult to accurately identify the sentiments. In this research, we used the following content features:

- **Bag of words:** This model is used for the machine learning models. The weights are given based on term frequency-inverse document frequency (TF-IDF), which gives high weight to rare words in the corpus.
- **GLOVE embedding:** This lower dimensional vector representation is used for all deep learning models.

B. SLIDING WINDOW FEATURES

Generally, human beings tend to share their current opinions based on their prior thoughts or opinions. These prior thoughts or opinions can be accumulated using the past sentiments of a user. Therefore, we propose a sliding window mechanism (in sentiment classification) - a mechanism well known in event detection, where the aim is to analyze words or hashtags within a time window [41], [42]. With the same approach, we based this feature on the assumption that the current tweet's sentiment is dependent on past sentiment(s). In this paper, we propose two sliding window features i.e., temporal sliding window and last k sentiment. The details of each feature are given below.

1) TEMPORAL SLIDING WINDOW

The temporal sliding window aggregates the sentiments of a user within a time span. The temporal sliding window aims to consider the most recent sentiments of a user in identifying the current sentiments. Figure 1 illustrates an example of the temporal sliding window applied for two users. The example uses the window size of one hour, therefore tweet # 2 of user # 1 is aggregated as Null (as shown in Figure 1), because the time difference is greater than one hour. Similarly, tweet # 3 of user # 1 has a value of one in its temporal window feature because only the second tweet of user # 1 was within one hour aggregated time window. Algorithm 1 shows

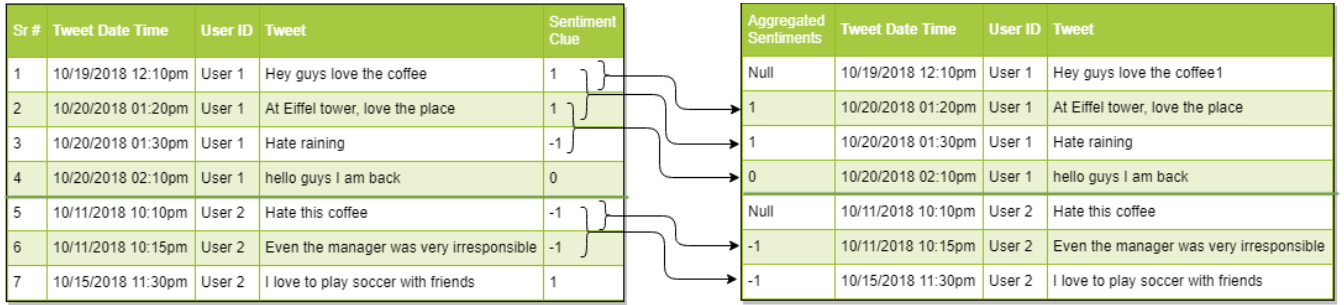


FIGURE 2. An example of last k sentiments feature with $K=2$.

Algorithm 1 Algorithm for Computing Temporal Sliding Window Feature

Input: All tweets S_i of a user U_i , Time Window: W_{size}

Output: Aggregated time window sentiments TW

Sort S_i temporally

foreach $S_{it} \in S_i$ **do**

$sum \leftarrow 0$

$count \leftarrow 0$

foreach $s_c \in S_{ip}$, where $t - 1 \geq p \geq 0$ **do**

 Compute the time difference $TD_{pc} \leftarrow$

$timedifference(s_c, S_{ip})$

if $TD_{pc} \leq W_{size}$ **then**

$sum \leftarrow sum + getSentimentClue(s_{ip})$

$count \leftarrow count + 1$

else

break

end

end

$TW_{it} \leftarrow sum/count$ if $count > 0$, null otherwise

end

return TW

recent sentiments. The goal of the learning algorithm is to use the recent sentiments of a user.

The temporal sliding window aims to learn patterns from recent past activities. We take an average of the sentiments in the sliding window. Our proposed temporal sliding window learns useful patterns like the polarity of the current sentiment with respect to the changes in the sentiments over the past.

Algorithm 2 Algorithm for Computing Last k Sentiments Feature

Input: All samples of a user U_i , K value: K

Output: Aggregated last k sentiments LS

sort S_i temporally

foreach $S_{it} \in S_i$ **do**

$sum \leftarrow 0$

$count \leftarrow 0$

foreach $s_c \in S_{ip}$, where $t - K \leq p \leq t - 1$ **do**

$sum \leftarrow sum + getSentiment(s_{ip})$

$count \leftarrow count + 1$

end

$LS_{it} \leftarrow sum/count$ if $count > 0$, null otherwise

end

return LS

the process of the temporal sliding window-based feature extraction. The process takes the samples U_i of the i^{th} user and the window size as inputs. Afterwards, the algorithm sorts all the tweets temporally. From the sorted tweets, it computes the time difference (TD_{pc}) of all the previous tweets of the i^{th} user ($S_{i(t-1)} \rightarrow S_{i(t-W_{size})}$, where W_{size} represents the window size) and stores the average of the sentiments satisfying the window size. The algorithm returns the temporal window features.

The temporal sliding window assumes that a user shares his opinion based on his present and past sentiments. For example, a person feeling angry after losing his job is likely to share more negative tweets that day. In such cases, the temporal sliding window will have more negative tweets in the recent past, so the current tweet is likely to be negative unless the content features have a strong recommendation towards positive sentiments. Similarly, a user may share more positive tweets if he is happy. Therefore, we hypothesize that the user's sentiment in the current tweet is dependent on his most

2) LAST K SENTIMENTS

The temporal sliding window is effective in gathering recent past activities. However, a tweet has *Null* value as current sentiment when there is no tweet within the specified time window. As an additional feature, we extract the last k sentiments regardless of the time. Figure 2 shows an example of aggregated sentiment with $K = 2$. Algorithm 2 shows the process of extracting this feature, it iterates over the previous K tweets ($S_{i(t-1)} \rightarrow S_{i(t-K)}$, where t is the index of the current tweet) to produce average sentiments from the previous tweets of a user. This feature helps in adding value to the users tweeting infrequently. Infrequent tweets induce many *missing* values in the temporal sliding window feature space. Therefore, the last k sentiments feature can help learn patterns from both short and long term activities of the users. The last k sentiment feature contains fewer missing values, thus allowing the learning algorithms to find patterns within a larger search space.

Sr #	Tweet Date Time	User ID	Tweet	Sentiment Clue
1	10/19/2018 12:10pm	User 1	Hey guys love the coffee	1
2	10/20/2018 01:20pm	User 1	At Eiffel tower, love the place	1
3	10/20/2018 01:30pm	User 1	Hate raining	-1
4	10/20/2018 02:10pm	User 1	hello guys I am back	0
5	10/11/2018 10:10pm	User 2	Hate this coffee	-1
6	10/11/2018 10:15pm	User 2	Even the manager was very irresponsible	-1
7	10/15/2018 11:30pm	User 2	I love to play soccer with friends	1

Ratio of Sentiments	Tweet Date Time	User ID	Tweet
Null	10/19/2018 12:10pm	User 1	Hey guys love the coffee1
{Pos:100, Neg:0, Neu:0}	10/20/2018 01:20pm	User 1	At Eiffel tower, love the place
{Pos:100, Neg:0, Neu:0}	10/20/2018 01:30pm	User 1	Hate raining
{Pos:67, Neg:33, Neu:0}	10/20/2018 02:10pm	User 1	hello guys I am back
Null	10/11/2018 10:10pm	User 2	Hate this coffee
{Pos:0, Neg:100, Neu:0}	10/11/2018 10:15pm	User 2	Even the manager was very irresponsible
{Pos:0, Neg:100, Neu:0}	10/15/2018 11:30pm	User 2	I love to play soccer with friends

FIGURE 3. A example of ratio of sentiments feature.

We hypothesize that the last k sentiment feature can help the learning algorithms to learn patterns from a large array of sentiments (of a user). Moreover, we hypothesize that this feature obtains a *broader* picture of the user by incorporating sentiments from both long and short term tweets.

Algorithm 3 Algorithm for Computing Ratio of Sentiments Feature

Input: All tweets S_i of a user U_i

Output: Ratio of sentiments RS

Sort S_i temporally

foreach $S_{it} \in S_i$ **do**

$sum_{pos} \leftarrow 0, count_{pos} \leftarrow 0$

$sum_{neg} \leftarrow 0, count_{neg} \leftarrow 0$

$sum_{neu} \leftarrow 0, count_{neu} \leftarrow 0$

foreach $s_c \in S_{ip}$, where $0 \leq p \leq t - 1$ **do**

if $getSentiment(s_{ip}) > 0.29$ **then**

$sum_{pos} \leftarrow sum_{pos} + getSentiment(s_{ip})$

$count_{pos} \leftarrow count_{pos} + 1$

else if $getSentiment(s_{ip}) < -0.29$ **then**

$sum_{neg} \leftarrow sum_{neg} + getSentiment(s_{ip})$

$count_{neg} \leftarrow count_{neg} + 1$

else

$sum_{neu} \leftarrow sum_{neu} + getSentiment(s_{ip})$

$count_{neu} \leftarrow count_{neu} + 1$

end

end

$RS_{it} \leftarrow$

$sum_{pos}/count_{pos}, sum_{neg}/count_{neg}, sum_{neu}/count_{neu}$

if $count > 0$, *null* otherwise

end

return RS

neutral and negative sentiments from all the past tweets (i.e., $S_{i0} \rightarrow S_{i(t-1)}$). Algorithm 3 shows the process of extracting the sentiment of the past tweets and then computing the percentage of all positive, neutral and negative sentiments. Figure 3 shows an example of ratio computed from all the past tweets, thus giving an overall behavior of a user. We separate the average of positive, negative, and neutral tweets in the ratio of sentiment feature because we use user's variations in the past, which might be ignored if we take the average of all the sentiments. In simple terms, we apply the ratio of positive, neutral, and negative sentiments to correlate patterns within these features with the current sentiment. For example, if a user has a high average of positive and negative sentiments in the past then, likely, his current tweet may not be neutral. To find correct sentiment for a tweet, Algorithm 3 specifies boundaries to consider a tweet as positive, negative, or neutral.

The ratio of sentiment feature builds a user profile by accumulating the overall sentiments of a user. This user profile helps in learning the general behavior of a user. For example, a user with mostly positive tweets has a high probability of sharing more positive tweets. This feature can be used with sliding window features to learn patterns from both the general and recent perspectives while classifying the current tweet.

IV. RESEARCH QUESTIONS

Figure 4 shows the framework for sentiment classification. The proposed framework first takes all the tweets of a user then extracts relevant features from it. These features are given as input to the machine learning algorithms. Finally, the input of the machine learning algorithms answer the following research questions:

- 1) **Research Question 1 (RQ1): Can the temporal sliding window help in classifying tweets effectively?**
The temporal sliding window gives insights into the user's recent activities. These activities might give recent clues (like mood, opinion, etc.) to the learning algorithms. In this research question, we test the effectiveness of the temporal sliding window during the classification.
- 2) **Research Question 2 (RQ2): Can the last k sentiments feature help in classifying tweets effectively?**

C. RATIO OF SENTIMENTS

Sliding window features are essential in learning the behavioral patterns of the users over a time span, however, they lack in handling frequent variations in the sentiments. For example, a user might change his sentiments frequently, therefore, he may have a mix of positive and negative tweets. Such behavioral patterns within a sliding window may have mixed sentiments. So, past sentiments may not give an overall picture of the user's personality or his way of thinking. Therefore, we propose to compute the ratio of positive,

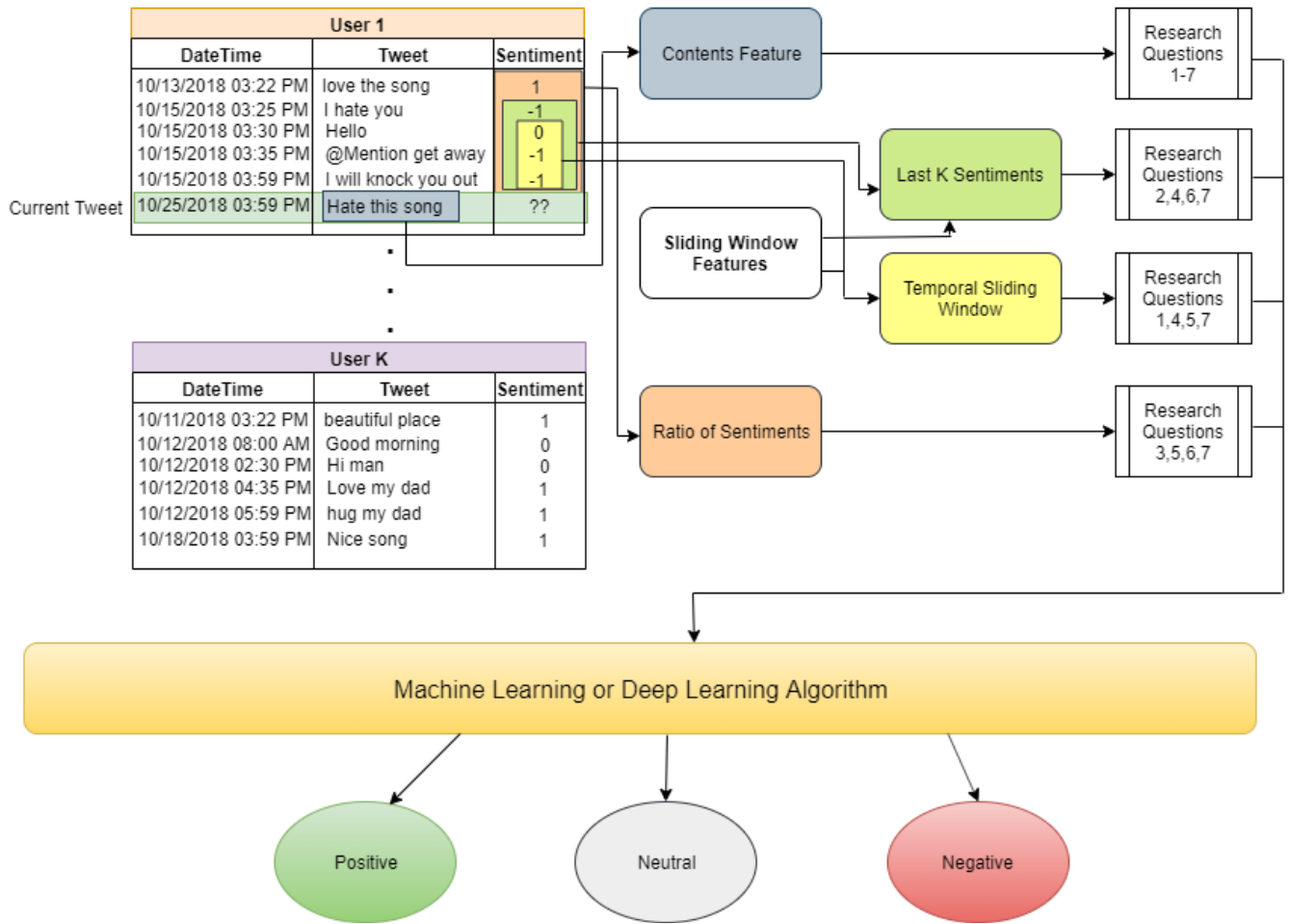


FIGURE 4. The framework of the context-aware sliding window based models. The top right part shows a snapshot of our dataset; the highlighted row shows the current tweet. Each window in the sentiment column is color-coded. The middle part shows all the features used in our framework, the details of each feature is given in Section IV). The bottom part uses the machine or deep learning algorithm to produce the final classification.

The last k sentiments feature gives an overview of the past activities regardless of time. This feature mainly focuses on obtaining patterns from a variable time span. We test the last k sentiments feature in improving the classification accuracy.

- 3) **Research Question 3 (RQ3): Can the ratio of sentiments help in classifying tweets effectively?** The ratio of sentiments provides the overall behavior of a user. This research question tests the performance of the ratio of sentiments feature in sentiment classification.
- 4) **Research Question 4 (RQ4): Can using the temporal sliding window and last k sentiments features together help in classifying tweets effectively?** As shown in Figure 4, this research question focuses on applying both the temporal sliding window and last k sentiments features in learning algorithms. The aim is to learn patterns from all kinds of users i.e., normal, frequent and infrequent users. Both these features can complement each other in different ways, like a normal user may have many missing values in the temporal sliding window, in this case, the last k sentiments

feature will give an overview for that user. We test whether these features collectively can add value to the contents while classification.

- 5) **Research Question 5 (RQ5): Can the temporal sliding window and ratio of sentiments features together help in classifying tweets effectively?** This research question emphasizes to use temporal sliding window and ratio of sentiments as shown in Figure 4. Together these features provide an overview of the personality (extracted from the ratio of sentiments) along with the recent perspective of a user. Moreover, these features can learn changes in user behavior, like a person sharing positive tweets suddenly starts posting negative tweets. So, the learning algorithm learns such patterns, later these patterns can be applied to the user with similar behavioral changes. Since the temporal window can have missing values, so the ratio of sentiments feature can give useful clues. For example, a tweet after a week will be marked null in the temporal sliding window, however, the user history has more positive tweets. In this case, the learning algorithm uses

patterns (from the ratio of sentiments) of similar users to predict the sentiment. This research question focuses on finding whether the recent perspective (from the temporal sliding window) and overall user behavior (ratio of sentiments) help in discriminating tweets.

- 6) **Research Question 6 (RQ6): Can the last k and the ratio of sentiments features together help in classifying tweets effectively?** In this research question, we use the last k and ratio of sentiments as features as shown in Figure 4. The idea is to combine the insights from both features in order to avoid any misleading information. For example, a user who posts 20 tweets in an hour and most of them are positive except the last few; assuming that we set the value of $K=5$, this means that the last k sentiment provides incomplete information. In this case, the overall view provided by the ratio of sentiments feature can give complete information about that user. Therefore, these features together might overcome incomplete information.
- 7) **Research Question 7 (RQ7): Can the temporal sliding window, last k and the ratio of sentiments features together help in classifying tweets effectively?** In this research question, we provide all the perspectives (i.e., recent, last k and the ratio of sentiments) to the learning algorithm as shown in Figure 4. These features provide learning algorithms to learn a broad set of patterns within the user’s recent, long term, and overall behavior. Since the user behavior changes over time, so these features together can correlate the patterns from recent activities to variable time span (last k sentiments) and overall sentiments (ratio of sentiment). For example, a user who has overall negative tweets may sometimes share a number of positive tweets in a short time span. We test the assumption that the correlation of behavior patterns in the recent, variable time span and overall activities can help in finding sentiments of a tweet.

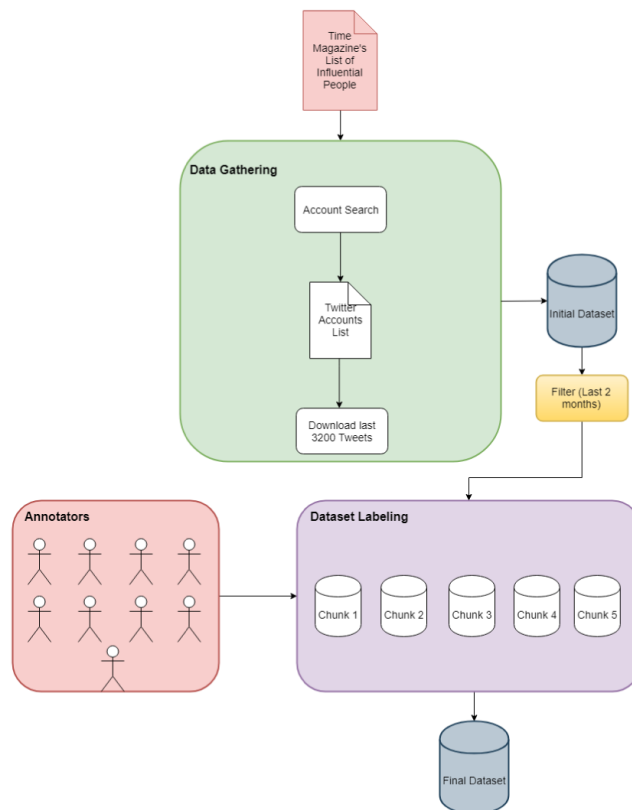


FIGURE 5. Dataset creation process.

V. DATASET

To evaluate the research questions, we require a labeled dataset of all the tweets in the user timeline. In the absence of a publicly available temporally labeled sentiment dataset, we develop our dataset using the process shown in Figure 5. The first step is to select a list of users. To gather an unbiased and diverse set of users, we use the list of top 100 influential people published by TIME magazine in 2018.¹ The list constitutes of people with diverse professions including actors, comedians, musicians, politicians, activists, and leaders. We choose influential people because they follow diplomacy and act as agenda seekers to gain more traction in the community [43], [44]. Therefore, these influential people follow diplomacy by using a specific set of vocabulary to avoid lacerating anyone in the community. On the contrary,

common people may use simple vocabulary to show their sentiments, which makes it easy for the classifier to identify sentiments. In this research, we intend to apply the proposed technique on a challenging task of learning different aspects of the diplomatic responses from influential people. Generally, the accounts of influential people stay active forever, whereas some of the regular accounts get suspended. Besides that, the accounts of influential people are active in sharing their thoughts, which allows us to obtain a large number of tweets per user. We believe that the proposed method would perform much better on datasets of common users because of their diverse vocabulary, which would help the classifiers to learn separation boundaries easily.

In the second step, we manually searched for the Twitter accounts of these users. We only considered the Twitter accounts that matched with the person’s profile in the TIME magazine. Some of the influential people did not have a Twitter account or their account privacy was set to private, therefore, we did not add those accounts. Eventually, we were left with 71 accounts and then we used Twitter API to download the recent upto 3,200 tweets of each account. In the process, we downloaded a total of 159,334 tweets.

The third step is to remove noise from the dataset. To remove noise, we applied three filters: the first filter removed all the retweets because they are not written by the users themselves. The second filter removed accounts with less than 40% of English tweets. In the third filter,

¹<https://web.archive.org/web/20181014032646/http://time.com/collection/most-influential-people-2018>

TABLE 1. Details of temporally labeled dataset.

Attribute	Value
Number of Users	36
Number of Tweets	4,557
Average Tweets per User	126
Percentage of Positive Tweets	53%
Percentage of Neutral Tweets	36%
Percentage of Negative Tweets	11%

we removed the accounts with less than 100 tweets. After applying the filters, we took the last two months of tweets to reduce the size of the dataset for human annotators. Finally, our dataset consisted of 4,557 tweets generated by 36 users as shown in Table 1. Two months duration contains sufficient tweets per user to train a machine learning model. A suitable learned model can later be used for classifying tweets in real time on a large scale. The dataset contains 53% of positive tweets, which is normal because these influential users may share more positive tweets. Furthermore, we can not use a balanced dataset for all kinds of sentiments because this research uses many tweets of a user, in which we do not have control over the distribution of sentiments.

In the last step, we ask nine human annotators to label the dataset. We gave an introductory session to each annotator to briefly explain the basic concept of sentiments.

For the labeling process, we split the data into 5 parts, each part contained all the tweets of seven users on average. Each part contained the ID of the users and temporally sorted tweets; ID helped the annotators to understand the context of the current tweet by referring to the previous tweets of the same user. Moreover, every part is labeled by two different annotators. We computed Cohen's kappa score to measure the inter-rater agreement, the kappa score was 0.62, which is considered as a substantial agreement. This kappa score shows that finding the correct sentiment for these diplomatic entities is difficult, even for humans. We only considered tweets having an agreement among the annotators. Afterwards, the dataset consisted of 3,483 tweets. To preprocess the data, we removed the stopwords and replaced URLs, emojis, user mentions with constant markers like URL, EMO_POS, USER_MENTION, etc. We used public emojis list² to classify the emojis with the sentiments. We used markers for the learning algorithm to see more useful patterns within it. For feature representation, we used the bag-of-words approach for machine learning models and 100-dimensional GloVe [40] vectors representation for the deep learning algorithms.

VI. RESULTS

This section discusses the performance of the proposed CSW techniques for the sentiment classification. We propose seven novel variations of features to incorporate context alongside the contents feature. This section is structured

²https://web.archive.org/web/20190307162715/http://kt.ijs.si/data/Emoji_sentiment_ranking/

as follows: Section VI-A discusses the experiment settings. Section VI-B introduces baselines for the comparison. Afterwards, Section VI-C presents and discusses the results.

A. EXPERIMENT SETUP

Standard methods for evaluating machine learning algorithms like k-fold cross validation are not appropriate for time series data. Therefore, we use the rolling window evaluation technique, which is commonly used to test the performance of the time series predictive systems [45]. It divides the data (of T samples) into chunks of size m , where $1 \leq m < T$. For each iteration i , the rolling window fetches sub-samples using the following equation:

$$W_i = T - m * i, \quad (1)$$

where i denotes the iteration number and W_i denotes the i^{th} window. In simple words, the first window contains 1 to m sub-samples, while the second contains samples from $m + 1$ to $2m$ and so on.

We split each window into 90% for training and 10% for testing because it allows enough tweets per user for the learning algorithm to better learn the decision boundary. We run deep learning algorithms for 150 epochs and use a dropout of 0.1. Moreover, we use default settings for machine learning models. We set the value of $m = \frac{T}{5}$ and compute the F1-score, precision, and recall for each iteration. The final score is computed by taking the average of all the scores.

For the test data, features that depend upon past sentiments do not use manually labeled sentiments. Instead, we use sentiment clues. Sentiment clues are generated using NLTK³ toolkit. Though sentiment clues are not accurate when compared to the manually annotated sentiments, they are useful in real-life scenarios where annotated data is unavailable.

For evaluating the temporal window, we use the window size ranging from 30 minutes to 7 days. We rely on the assumption that the sentiment remains the same for a short period of time, therefore we consider multiple window sizes within a day. For the last k sentiments feature, we evaluate the values of k between 1 and 5.

B. BASELINES

We use the state-of-the-art learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Stochastic Gradient Descent (SGD), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) as baselines.

All these baselines are provided with the content features, i.e., the BOW approach for the machine learning algorithms and Glove embeddings for the deep learning algorithms. Moreover, we also added an embedding layer in the deep learning models to learn new embeddings from the current dataset.

The proposed features can be applied to any kind of supervised learning algorithm. We choose widely used algorithms

³<https://web.archive.org/web/20190718130851/https://www.nltk.org/>

TABLE 2. Results of the techniques using the temporal window and contents features. The column with text *No window* shows the performance of the baseline. The cells with bold values show the best performance.

	Technique Name	No Window (baseline)	Avg Sent (30M)	Avg Sent (1H - 6H)	Avg Sent (7H - 12H)	Avg Sent (13H - 18H)	Avg Sent (19H - 24H)	Avg Sent (2 Days)	Avg Sent (7 Days)
F1 Score	LSTM	37	37.2	35.7	33.7	35	35.9	34.6	38.2
	GRU	35	31.5	33.4	32.8	34.7	34.9	32.2	36.6
	RNN	33.5	32.4	36.1	36.5	35.8	34.8	30.4	32.8
	RF	37.5	39.6	40.5	40.8	46	42.6	38.1	36.9
	SVM	36	34	38.2	37.4	37.8	36.6	34.1	31.6
	SGD	33.5	30.2	36.4	37.5	36.3	34.9	30.8	34.2
Precision	LSTM	36	36.1	34.7	33.6	34.4	35.2	34.2	37.4
	GRU	34	31	33	32.7	34.2	34.4	32.1	36.5
	RNN	33	31.8	35.5	36.1	35.6	34.5	29.7	32.3
	RF	37	40.1	41.2	41.7	48.9	41.4	38.1	37.5
	SVM	35.9	34.7	36.5	38	37.3	36	33.9	32
	SGD	34.3	29.8	36.5	38.8	35.7	35.6	33	34.8
Recall	LSTM	37.9	38.3	36.8	33.8	35.6	36.7	34.9	39.1
	GRU	36	32	33.9	32.8	35.2	35.4	32.2	36.6
	RNN	34	33	36.8	36.9	36	35	31.2	33.4
	RF	38	39.2	39.8	40	43.5	43.8	38.1	36.3
	SVM	36.1	33.3	40.1	36.8	38.4	37.3	34.3	31.2
	SGD	32.7	30.6	36.3	36.3	36.9	34.2	28.9	33.7

TABLE 3. Results of the techniques using the last k sentiments and contents features. The first row with text *No window* shows the performance of the baseline. The cells with bold values show the best performance.

Last k Sentiments	F1 Score						Precision						Recall					
	LSTM	GRU	RNN	RF	SVM	SGD	LSTM	GRU	RNN	RF	SVM	SGD	LSTM	GRU	RNN	RF	SVM	SGD
No Window (baseline)	37	35	33.5	37.5	36	33.5	36	34	33	37	35.9	34.3	37.9	36	34	38	36.1	32.7
Last 1 Sentiment	33.9	32.2	32.4	41.5	34.4	35.4	34.1	32	31.9	43.5	35.3	34.1	33.7	32.5	33	39.6	33.5	36.9
Last 2 Sentiments	33.2	33.3	37.2	42.6	36.6	37.4	33.2	33	36.9	46.8	36.2	37.9	33.3	33.6	37.6	39.1	37	36.9
Last 3 Sentiments	36.5	35.2	36.1	36.6	38.5	36.3	36.1	34.8	35.7	36.9	40.8	34.4	37	35.7	36.6	36.3	38.4	36.4
Last 4 Sentiments	40.5	32.6	33.3	43.7	32.3	32.3	39.2	32.1	32.9	44.1	32.5	32.7	41.8	33.2	33.8	43.3	32.1	31.9
Last 5 Sentiments	39.9	32.2	36.3	41.4	32.1	33.7	39.8	31.9	36.2	41	33.9	35.2	40	32.5	36.5	41.8	30.4	32.4

TABLE 4. Results of the techniques using the ratio of sentiments and contents features. The first row with text *No window* shows the performance of the baseline.

Technique Name	F1 Score						Precision						Recall					
	LSTM	GRU	RNN	RF	SVM	SGD	LSTM	GRU	RNN	RF	SVM	SGD	LSTM	GRU	RNN	RF	SVM	SGD
No Window (baseline)	37	35	33.5	37	36	33.5	36	34	33	37	35.9	34.3	37.9	36	34	38	36.1	32.7
Ratio of Sentiments	38.82	34.5	35.6	39	37.8	34	37.7	33.6	35	40.6	39.1	34.6	40	35.4	36.2	38.1	36.6	33.5

in sentiment classification research. The same algorithms are used in the baseline for the proposed features. In addition, each proposed technique is targeting a research question discussed in Section IV.

C. RESULTS & DISCUSSIONS

As seen in Tables 2 to 8, the proposed features perform significantly better than the baselines. The best performance is achieved by combining all the features (i.e., temporal sliding window, last k and ratio of sentiments). Together, these features help in learning patterns from the user's recent and overall history; moreover, they also handle all types of users, i.e., normal, frequent, and infrequent. However, the F1-score values in the results are not very high because identifying the correct sentiments for influential people is a challenging problem. The Cohen's Kappa value for annotators agreement also draws the same conclusion, i.e., the complexity in assigning sentiments for such individuals resulted in decreased agreement among the human annotators. We discuss the results of each research question in the sections below.

1) RESEARCH QUESTION 1 (RQ1): CAN THE TEMPORAL SLIDING WINDOW HELP TO CLASSIFY TWEETS EFFECTIVELY?

This research question emphasizes on utilizing the most recent sentiments to derive the sentiment of the current tweet. As shown in Table 2, the performance of the proposed models is better than the baseline. Interestingly, almost 50% precision is achieved between 13 to 18 hours window size, which clearly shows that the current sentiment is derived from the recent mood. In the same manner, the F1-score also indicates a better performance within the time range of 6-18 hours because a wider window size captures enough information for the learning algorithms to effectively recommend a sentiment. On the other hand, the shorter time span may not be effective because of many missing (null) values. For example, a user might not share tweets regularly within six hours, therefore, a short time span may have many missing values. So, it is beneficial to take insights from other features while the temporal sliding window has missing values.

The temporal sliding window outperforms all the baselines. However, in a shorter time span, it achieves a relatively

TABLE 5. F1 score of the techniques using the temporal sliding window, last k sentiments and contents features. The column with the text *No window* shows the performance of baseline. The cells with bold values show the best performance.

Last k / Window Size	Technique Name	No Window (baseline)	Avg Sent (30M)	Avg Sent(1H - 6H)	Avg Sent(7H - 12H)	Avg Sent(13H - 18H)	Avg Sent(19H - 24H)	Avg Sent (2 Days)	Avg Sent (7 Days)
1	LSTM	36.9	33.6	34.7	33.3	33.4	35.9	32.9	36
	GRU	35	31.6	32.6	33.1	32.5	34.8	33	35.5
	RNN	33.5	32.8	33.9	33.4	35.6	32.1	32.1	32.3
	RF	37.5	39.2	39.7	42.9	40.4	39.5	41.1	39.5
	SVM	36	34.1	37.1	39.2	37.8	41.5	35.5	33.7
	SGD	33.5	31.3	36.9	39.5	40.7	37.3	30.4	34.4
2	LSTM	36.9	35.7	37.8	33.4	35.8	35.9	35.9	37.4
	GRU	35	32.7	35.1	35.6	34.6	35.8	35.7	37.2
	RNN	33.5	32.6	35.5	34.1	35.9	33.7	32.8	32.6
	RF	37.5	40.4	42.1	40.5	40.6	40	41.4	38.2
	SVM	36	31.1	37.7	38	38	38.9	33.3	34.4
	SGD	33.5	36.8	39.4	34.5	35.9	37.8	35	31.9
3	LSTM	36.9	35.9	37.7	36.2	35.7	36.5	37.3	40.9
	GRU	35	35.5	35.4	35.3	35.7	35.6	35.1	34.9
	RNN	33.5	35.5	34.4	33.4	35.7	35.9	32.1	35.3
	RF	37.5	38.6	41	41.7	40.3	42	39.5	37.5
	SVM	36	35.9	36.2	39.7	36.7	39.7	37.4	33.7
	SGD	33.5	32.9	36.1	36.2	36.1	36.4	38.8	36.2
4	LSTM	36.9	35.4	38.7	34.8	33.3	36.8	34.7	38.1
	GRU	35	34.6	37.5	35.4	35.5	35.3	35.1	33.1
	RNN	33.5	36.1	36.2	35.4	35.1	35.5	34.5	36.6
	RF	37.5	36.9	41.6	39.7	40.6	41	38.9	39.3
	SVM	36	35	40	42.7	38.2	36.6	34.8	35.9
	SGD	33.5	34	38.6	37.2	38.7	37	36	34.8
5	LSTM	36.9	37	36.3	33	32.9	36	32.3	39.2
	GRU	35	38.4	36.7	35.4	35.2	36.2	37.8	31.9
	RNN	33.5	35	34.9	35.7	36.3	35.2	32.7	35.2
	RF	37.5	38.7	40.3	39.9	39.6	39.6	38.5	39.8
	SVM	36	33.1	36.8	37.7	38.3	40.6	35.3	38.1
	SGD	33.5	36.4	37.5	37.7	36.5	37.8	35.5	33.6

TABLE 6. F1-score of the techniques using the temporal sliding window, ratio of sentiments and contents features. The cell with bold values shows the best performance.

Technique Name	No Window (baseline)	Avg Sent (30M)	Avg Sent(1H - 6H)	Avg Sent(7H - 12H)	Avg Sent(13H - 18H)	Avg Sent(19H - 24H)	Avg Sent (2 Days)	Avg Sent (7 Days)
LSTM	36.9	41.5	45.3	43.1	42.4	43.9	39.6	40.6
GRU	35	35.2	39.4	37.9	40.2	38.7	39.6	35.4
RNN	33.5	35.8	33.3	34.6	35.6	35.8	34.9	33.3
RF	37.5	38.2	40.5	41.4	42	39.3	40.3	36.5
SVM	36	34.8	38	38.9	35.5	40.5	35	33.4
SGD	33.5	34.2	35.5	36.9	36.6	37.5	35.6	33.8

low performance because of missing values in the feature set. Table 2 shows that the temporal sliding window has a positive impact on the performance, i.e., between 6 to 18 hours window. In fact, the random forest-based model has achieved the highest F1-score of 46%. This performance can further be improved by adding more users in the dataset; that way, learning algorithms can learn a variety of patterns within different types of users. As per the results, the answer to this research question is yes, but we can further improve the performance by adding more data and features.

2) RESEARCH QUESTION 2 (RQ2): CAN THE LAST K SENTIMENTS FEATURE HELP TO CLASSIFY TWEETS EFFECTIVELY?

This research question focuses on avoiding missing values by taking past sentiments regardless of time. This feature helps in obtaining information from the recent tweets irrespective of time. This feature also helps in getting an overview of the

users who do not share tweets frequently. Table 3 shows that the proposed models have performed better with the last k sentiments feature. The window sizes between 2 to 4 have more discriminating results because these window sizes give a balanced view of both frequent and normal users. The performance starts deteriorating if we increase the value of k (≥ 5) because it may accumulate old sentiments for a user who rarely shares tweets. For example, the last 10 sentiments of the infrequent users may span over the last 10 years. Therefore, a value between 2 to 5 past sentiment gives a balanced view of all kinds of users.

The last k sentiments feature has significant improvements over the baseline. In fact, in some cases, the last k sentiments feature has better performance than the temporal window feature shown in Table 2. This shows that the last k sentiments feature handles the missing value problem by allowing learning algorithms to learn effective patterns for any kind of user.

TABLE 7. F1-score of the techniques using the last k , ratio of sentiments and contents features. The first row with text *No window* shows the performance of the baseline. The cell with bold values shows the best performance.

Last k Sentiments	LSTM	GRU	RNN	RF	SVM	SGD
No Window (baseline)	36.9	35	33.5	37.5	36	33.5
Last 1 Sentiments	39.8	32.1	32.8	36.7	32	33.1
Last 2 Sentiments	41.5	34.4	35.2	37.1	31	34.1
Last 3 Sentiments	44.7	36.8	32.4	37.7	38.6	36.4
Last 4 Sentiments	40.1	34.5	35.9	37.3	32.8	34.9
Last 5 Sentiments	43.1	39.3	33.2	38.6	41.5	37.3

3) RESEARCH QUESTION 3 (RQ3): CAN THE RATIO OF SENTIMENTS HELP TO CLASSIFY TWEETS EFFECTIVELY?

The ratio of sentiments gives an overview of a user. According to Table 4, this feature did not have a significant increase in performance because the overall view may add confusion. For example, a user who mostly posts negative tweets in the past occasionally shares positive tweets. In this case, the ratio of sentiments feature will recommend classifying the current tweet as negative unless the contents feature have a strong recommendation towards positive sentiment. Therefore, it is necessary to add more features to detect recent changes in activities.

Answer to this research question is considered plausible because the ratio of sentiments feature has slight improvement; however, it can be more effective if we merge both the recent and the overall perspective of a user. For brevity, in the next research questions, we only present the F1-scores.

4) RESEARCH QUESTION 4 (RQ4): CAN THE TEMPORAL SLIDING WINDOW AND LAST k SENTIMENTS FEATURE HELP TO CLASSIFY TWEETS EFFECTIVELY?

This research question focuses on handling all kinds of users. Table 5 shows the increase in performance with both these features together. Similar to results in Section VI-C.2, the proposed features have the best performance when the value of k is between 2 and 4. In the same manner, the proposed models perform better in the temporal window size within 24 hours. In this research question, machine learning algorithms have performed better than deep learning algorithms because both these features (i.e., temporal sliding window and last k sentiments) make up of four variations, i.e., both positive, both negative, one positive & one negative and vice versa. These four variations combined with the contents form a large number of permutations, thus deep learning algorithms may require more data to effectively learn the weights for each variation. The results highlight that the majority of the proposed feature driven models have obtained a significant increase in F1-score.

Answer to this research question is yes because the majority of the proposed feature driven models have significant improvements in terms of F1-score. Moreover, deep learning models have slight improvement over baseline, however we hypothesize that more data will help learning algorithms to learn weights with respect to different scenarios (particularly in which both features have different sentiments).

5) RESEARCH QUESTION 5 (RQ5): CAN THE TEMPORAL SLIDING WINDOW AND RATIO OF SENTIMENTS FEATURES HELP TO CLASSIFY TWEETS EFFECTIVELY?

The focus of this research question is to combine recent and overall sentiments to give a holistic view of a user. Since the temporal sliding window feature has missing values, in such a case, the ratio of sentiments allow learning algorithms to learn information from the overall perspective of a user. On the contrary, the ratio of sentiments feature can not extract changes in the recent activities, so the temporal sliding window complements it by adding recent activities of a user. Table 6 validates this research question, as these features lead to significant improvements over the baselines. In this research question, the ratio of sentiment feature compliments the temporal window feature. For example, when both features point to the same recommendation, the classification may have higher confidence. On the other hand, if one feature has weak recommendations then the other feature can add its weight to help in effective recommendations. Consider a user who has a similar ratio of positive or negative sentiments in the ratio of sentiments feature, whereas the temporal sliding window feature has more positive sentiments. In this case, the recommendation from the ratio of sentiments feature may have a low probability, thus combining the recommendations from the temporal sliding window feature may increase the confidence of the recommendation. Both of these features work in conjunction with the contents, so the final sentiment is based on both these features. Together these features create dependency, which can help in disambiguation.

Table 6 shows that the best performance is between 13 to 24 hours of window sizes, which means that the user's tweets are influenced by his moods within a day. Besides that, the results also show a significant increase in the recall values, which means that the proposed feature-based models have more coverage.

6) RESEARCH QUESTION 6 (RQ6): CAN THE LAST k AND RATIO OF SENTIMENTS FEATURES HELP TO CLASSIFY TWEETS EFFECTIVELY?

We use the last k and ratio of sentiments features to address this research question. The idea is to compliment the last k sentiments feature with an overall view of a user, like the last k sentiment feature gives a specific view of a user, which may not highlight user's trends, whereas the ratio of sentiments helps provide a broader perspective of a user. For example, the last 4 sentiments have equal positive and negative sentiments, now the algorithm may rely more on the ratio of sentiments and contents features to classify a tweet. Results in Table 7 confirm that the feature-driven model outperforms the baseline. The results show a significant increase in $k = 5$, this shows that the overall view from the ratio of sentiments feature has helped in further discriminated tweets.

It is evident from the answer to this research question is yes. Moreover, the feature-driven models attain a significantly higher F1-score than the baseline, which depicts that

TABLE 8. F1 score of the techniques with all the features (temporal sliding window, last k and ratio of sentiment). The column with text *No window* shows the results of the baseline. The cell with bold values shows the best performance.

Last k / Window Size	Technique Name	No Window (baseline)	Avg Sent (30M)	Avg Sent(1H - 6H)	Avg Sent(7H - 12H)	Avg Sent(13H - 18H)	Avg Sent(19H - 24H)	Avg Sent (2 Days)	Avg Sent (7 Days)
1	LSTM	36.9	44.8	41.8	42.4	40.3	39.5	37.8	44
	GRU	35	37.4	33.1	35.4	34.9	34.3	32.8	35.2
	RNN	33.5	32.4	32.8	36	35.1	34.6	31.7	33.1
	RF	37.5	40.7	40.3	39.2	39.4	39.8	41.8	38.8
	SVM	36	31.7	37.7	39.6	39.9	36.7	33.4	35.5
	SGD	33.5	37.8	40.1	38.1	37.3	39.6	32.7	32.9
2	LSTM	36.9	42.2	43	45.1	42.4	41.5	35.9	41.7
	GRU	35	31.6	37	37.7	35.6	35.7	32.7	37.8
	RNN	33.5	30.8	34.8	35.7	35.2	35.8	32.3	31.9
	RF	37.5	38.7	43.8	40.7	41.5	44.4	41.5	38.1
	SVM	36	33.7	39.5	36.3	38.1	38.1	36.6	31.2
	SGD	33.5	34.3	36.2	40.1	37.4	39.8	33.7	33.1
3	LSTM	36.9	39.8	43.8	45.1	41.7	42.5	39.8	39.7
	GRU	35	34.6	37	37.4	39.1	37.6	37.2	34.2
	RNN	33.5	31.6	35.4	35.1	35.8	35.6	31.6	33.5
	RF	37.5	42	41.6	41.5	43.1	41.3	41.4	39.3
	SVM	36	35.9	38.2	39.6	37.5	39.8	32.5	34.3
	SGD	33.5	35.2	37.9	38.2	35.8	39.4	36	38.8
4	LSTM	36.9	41.8	45.2	43.1	43.3	43	43	40.6
	GRU	34.4	38	40.4	38	37.4	35.2	38.2	38.2
	RNN	33.5	32.2	35	35.4	35.6	36.3	34.8	36.4
	RF	37.5	35.7	44.3	40.4	41.9	45.5	42.2	40.8
	SVM	36	38.9	42.3	35.9	35.3	37.1	37.2	35.7
	SGD	33.5	36.5	39.3	35.6	37.9	37.2	32.2	32.4
5	LSTM	36.9	41.5	45	44.3	44.1	42.7	43.1	45.2
	GRU	35	41.8	39.5	38.5	38.9	40.6	40.4	42.7
	RNN	33.5	33.2	37.9	38	35.2	34.8	33.8	35.5
	RF	37.5	42.1	42.1	43.7	40.6	42.4	40.7	38.6
	SVM	36	34.6	37.7	41.3	39.5	39	32	33.2
	SGD	33.5	33.4	35.3	40.6	37.8	38.2	34.9	33.4

the predictions of the proposed feature-driven models are better.

7) RESEARCH QUESTION 7 (RQ7): CAN THE TEMPORAL SLIDING WINDOW, LAST K AND THE RATIO OF SENTIMENTS FEATURES HELP TO CLASSIFY TWEETS EFFECTIVELY?

This research question focuses on using all the features to effectively classify a tweet. This research question aims to create dependencies among all the features, which allows the learning algorithms to correctly classify even if any feature has missing values. Table 8 shows that all features together gain the significantly improvement compared to the models in Sections VI-C1 to VI-C6. Together these features cover the user’s recent and overall activities. Besides that, these features play a vital role when temporal window feature has missing values, in such cases the remaining two features become useful for the classifier. This can also be seen in the results that the window sizes between 30 minutes to 24 hours achieve the significant performance even with the missing values in the temporal sliding window. Most of the feature-driven models have achieved better performance regardless of the window size, which means that together these features allow learning algorithms to learn more discriminate patterns from the data. The best F1-score is close to 50%, which also elaborates that the proposed feature-driven models perform better than the state of the art baseline.

Table 9 shows some examples of the output produced by the proposed model (with all the features); the table covers three tweets of different users. These tweets are not entirely classified on the contents rather the model has used overall user profile and previous activities to predict the sentiment. For example, the first tweet is classified as negative, although the contents did not have many negative words. In this case, the person’s most recent and overall sentiments were negative. So the learning algorithm correctly disambiguates the context of the current tweet with respect to the past tweets. The second example shows a tweet of an activist with mixed sentiments. In this case, the contents have more positive words (like Timeless rock, Timeless for smart people, etc.) and past sentiments do not portray many negative sentiments, therefore the classifier correctly classifies it as positive. In the third example, a user often criticizes the elected officials on their actions. Although the contents have some positive words but the features on the past sentiments may have a higher probability towards the negative sentiment.

The answer to this research question is yes, and it is evident that together all features have significant improvements. Moreover, these features help in disambiguating the context by incorporating the previous activities of a user. Together all these features make classification a two steps process. The first process creates patterns by correlating the values from all features. The second process ensures that the prediction relies on both the contents and past patterns.

TABLE 9. Sample predictions produced by random forest model.

Tweet	Prediction	Temporal Slide Window	Last k Sentiments	Ratio of Sentiments
I can not believe @Sony @NBCTimeless did not have the money for this fantastic show. Take 3 hours away from The Today Show and you would have enough for 8 seasons of #Timeless!	-1	-0.5	-0.2	Pos: 25%, Neg: 45%, Neu: 30%
#Timelessrocks #RenewTimeless #Timelessforsmartpeople	1	0.3	0.6	Pos: 70%, Neg: 5%, Neu: 25%
also why is it culturally acceptable for any elected official to take money from special interests in America we call it campaign donations, the rest of the world calls it a bribe	-1	-0.47	-0.6	Pos: 17%, Neg: 69%, Neu: 14%

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose seven context-aware sliding window models along with three novel features for sentiment classification. The aim is to incorporate past insights during the classification of the current tweet. The proposed models can disambiguate the context from the recent activities and history of a user. Moreover, the proposed models can effectively classify tweets of a user with the variant frequency of tweets. All the proposed models have improvements over all the six state-of-the-art baselines. Among all, the best performing model uses all the features together, which emphasizes that the past sentiments (temporal or non-temporal) and overall sentiments give effective indications to the classifier. The results also portray that the window size within 24 hours effectively captures recent mood(s) during classification. Moreover, the non-temporal window size of the past 2 to 5 sentiments provides a good balance between the insights of long and short time span. Lastly, the ratio of sentiments feature may not be effective unless combined with other features. The findings of this research will help in choosing certain window sizes for better classification.

We also present the first user-based temporally labeled dataset of 4,557 tweets labeled by nine human annotators. The labeling process achieves satisfactory agreement among the raters. The proposed work presents a new dimension by using the time series to classify sentiments; therefore, it can act as a benchmark for future researchers. In future, we intend to use the attention mechanism to signify the importance of some features. Moreover, we plan to extend the dataset to include a wide variety of contents.

REFERENCES

- [1] L. Martín-Domingo, J. C. Martín, and G. Mandsberg, "Social media as a resource for sentiment analysis of airport service quality (ASQ)," *J. Air Transp. Manage.*, vol. 78, pp. 106–115, Jul. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0969699719300079>
- [2] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, "Sentiment analysis of Twitter data," in *Proc. Int. Conf. Comput. Inf. Sci. (ICICIS)*, Trenton, NJ, USA, Apr. 2019, pp. 1–4.
- [3] D. C. Edara, L. P. Vanukuri, V. Sistla, and V. K. K. Kolli, "Sentiment analysis and text categorization of cancer medical records with LSTM," *J. Ambient Intell. Hum. Comput.*, vol. 10, no. 7, pp. 1–17, Jul. 2019, doi: [10.1007/s12652-019-01399-8](https://doi.org/10.1007/s12652-019-01399-8).
- [4] R. Liang and J.-Q. Wang, "A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in E-commerce," *Int. J. Fuzzy Syst.*, vol. 21, no. 3, pp. 963–977, Apr. 2019, doi: [10.1007/s40815-019-00606-0](https://doi.org/10.1007/s40815-019-00606-0).
- [5] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. F. Gelbukh, "Sentiment and sarcasm classification with multitask learning," 2019, *arXiv:1901.08014*. [Online]. Available: <http://arxiv.org/abs/1901.08014>
- [6] M. Yang, W. Yin, Q. Qu, W. Tu, Y. Shen, and X. Chen, "Neural attentive network for cross-domain aspect-level sentiment classification," *IEEE Trans. Affect. Comput.*, pp. 1–1, 2019.
- [7] W. Xu and Y. Tan, "Semi-supervised target-oriented sentiment classification," *Neurocomputing*, vol. 337, pp. 120–128, Apr. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231219300864>
- [8] J. Zhao, L. Dong, J. Wu, and K. Xu, "Moodlens: An emoticon-based sentiment analysis system for Chinese tweets," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2012, pp. 1528–1531. [Online]. Available: <http://doi.acm.org/10.1145/2339530.2339772>
- [9] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proc. ACL Student Res. Workshop (ACLStudent)*, Stroudsburg, PA, USA, 2005, pp. 43–48. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1628960.1628969>
- [10] M. Y. Khan, S. M. Emaduddin, and K. N. Junejo, "Harnessing English sentiment lexicons for polarity detection in urdu tweets: A baseline approach," in *Proc. IEEE 11th Int. Conf. Semantic Comput. (ICSC)*, San Diego, CA, USA, Jan. 2017, pp. 242–249.
- [11] R. Satapathy, C. Guerreiro, I. Chaturvedi, and E. Cambria, "Phonetic-based microtext normalization for Twitter sentiment analysis," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, New Orleans, LA, USA, Nov. 2017, pp. 407–413.
- [12] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2004, pp. 168–177. [Online]. Available: <http://doi.acm.org/10.1145/1014052.1014073>
- [13] V. A. Kharde and S. Sonawane, "Sentiment analysis of Twitter data: A survey of techniques," *Int. J. Comput. Appl.*, vol. 139, pp. 1–15, Apr. 2016. [Online]. Available: <http://arxiv.org/abs/1601.06971>
- [14] G. Gadek, J. Betsholtz, A. Pauchet, S. Brunessaux, N. Malandain, and L. Vercouter, "Extracting contextonyms from Twitter for stance detection," in *Proc. ICAART*, 2017, pp. 132–141.
- [15] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter," *Inf. Process. Manage.*, vol. 52, no. 1, pp. 5–19, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306457315000242>
- [16] N. Mukhtar and M. A. Khan, "Urdu sentiment analysis using supervised machine learning approach," *J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 2, Feb. 2018, Art. no. 1851001, doi: [10.1142/s0218001418510011](https://doi.org/10.1142/s0218001418510011).
- [17] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171649, doi: [10.1371/journal.pone.0171649](https://doi.org/10.1371/journal.pone.0171649).
- [18] E. A. C. Júnior, V. Q. Marinho, and L. B. dos Santos, "NILC-USP at semeval-2017 task 4: A multi-view ensemble for Twitter sentiment analysis," in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, p. 5, doi: [10.18653/v1/s17-2100](https://doi.org/10.18653/v1/s17-2100).
- [19] E. S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, R. R. Suárez, and O. S. Siordia, "A simple approach to multilingual polarity classification in Twitter," *Pattern Recognit. Lett.*, vol. 94, pp. 68–74, Jul. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865517301721>

- [20] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," *Inf. Process. Manage.*, vol. 53, no. 4, pp. 764–779, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306457316302205>
- [21] A. L. F. Alves, C. de Souza Baptista, A. A. Firmino, M. G. de Oliveira, and A. C. de Paiva, "A spatial and temporal sentiment analysis approach applied to Twitter microtexts," *J. Inf. Data Manage.*, vol. 6, no. 2, p. 118, 2015.
- [22] T. Chen, R. Xu, Y. He, Y. Xia, and X. Wang, "Learning user and product distributed representations using a sequence model for sentiment analysis," *IEEE Comput. Intell. Mag.*, vol. 11, no. 3, pp. 34–44, Aug. 2016.
- [23] X. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Coimbatore, India, Dec. 2009, pp. 210–214.
- [24] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, Stroudsburg PA, USA, vol. 1, 2014, pp. 1555–1565.
- [25] Y. Chen, J. Yuan, Q. You, and J. Luo, "Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM," in *Proc. 26th ACM Int. Conf. Multimedia (MM)*, New York, NY, USA, 2018, pp. 117–125. [Online]. Available: <http://doi.acm.org/10.1145/3240508.3240533>
- [26] R. Zhao and K. Mao, "Topic-aware deep compositional models for sentence classification," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 2, pp. 248–260, Feb. 2017.
- [27] J. Wehrmann, W. Becker, H. E. L. Cagnini, and R. C. Barros, "A character-based convolutional neural network for language-agnostic Twitter sentiment analysis," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 2384–2391.
- [28] H. Nguyen and M.-L. Nguyen, "A deep neural architecture for sentence-level sentiment classification in Twitter social networking," in *Computational Linguistics*, K. Hasida and W. P. Pa, Eds. Singapore: Springer, 2018, pp. 15–27.
- [29] 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.*, Stroudsburg, PA, USA, 2015, pp. 1422–1432.
- [30] S. T. S. and S. Tandon, "Topic based sentiment analysis using deep learning," pp. 1–8, 2017, *arXiv:1710.10498*. [Online]. Available: <http://arxiv.org/abs/1710.10498>
- [31] M. Zhang, Y. Zhang, and D.-T. Vo, "Gated neural networks for targeted sentiment analysis," in *Proc. 30th AAAI Conf. Artif. Intell.*, Sacramento, CA, USA, 2016, pp. 3087–3093.
- [32] S. Shi, M. Zhao, J. Guan, Y. Li, and H. Huang, "A hierarchical LSTM model with multiple features for sentiment analysis of sina weibo texts," in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Singapore, Dec. 2017, pp. 379–382.
- [33] D. Stojanovski, G. Strezoski, G. Madjarov, and I. Dimitrovski, "Finki at semeval-2016 task 4: Deep learning architecture for Twitter sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, San Diego, CA, USA, 2016, pp. 149–154.
- [34] A. Hassan and A. Mahmood, "Deep learning approach for sentiment analysis of short texts," in *Proc. 3rd Int. Conf. Control, Autom. Robot. (ICCAR)*, Nagoya, Japan, Apr. 2017, pp. 705–710.
- [35] T.-H. Yang, T.-H. Tseng, and C.-P. Chen, "deepSA at SemEval-2017 task 4: Interpolated deep neural networks for sentiment analysis in Twitter," in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, Vancouver, BC, Canada, Aug. 2017, pp. 616–620. [Online]. Available: <https://www.aclweb.org/anthology/S17-2101>
- [36] Z. Ding, R. Xia, J. Yu, X. Li, and J. Yang, "Densely connected bidirectional lstm with applications to sentence classification," in *Natural Language Processing and Chinese Computing*, M. Zhang, V. Ng, D. Zhao, S. Li, and H. Zan, Eds. Cham, Switzerland: Springer, 2018, pp. 278–287.
- [37] M. Yang, W. Yin, Q. Qu, W. Tu, Y. Shen, and X. Chen, "Neural attentive network for cross-domain aspect-level sentiment classification," *IEEE Trans. Affect. Comput.*, to be published.
- [38] N. Medagoda and S. Shanmuganathan, "Keywords based temporal sentiment analysis," in *Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, Zhangjiajie, China, Aug. 2015, pp. 1418–1425.
- [39] S. Feng, Y. Wang, L. Liu, D. Wang, and G. Yu, "Attention based hierarchical lstm network for context-aware microblog sentiment classification," *World Wide Web*, vol. 22, no. 1, pp. 59–81, Jan. 2019, doi: [10.1007/s11280-018-0529-6](https://doi.org/10.1007/s11280-018-0529-6).
- [40] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Vancouver, BC, Canada, 2014, pp. 1532–1543.
- [41] B. Poblete, J. Guzmán, J. Maldonado, and F. Tobar, "Robust detection of extreme events using Twitter: Worldwide earthquake monitoring," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2551–2561, Oct. 2018.
- [42] M. Hasan, M. A. Orgun, and R. Schwitter, "A survey on real-time event detection from the Twitter data stream," *J. Inf. Sci.*, vol. 44, no. 4, pp. 443–463, Aug. 2018, doi: [10.1177/0165551517698564](https://doi.org/10.1177/0165551517698564).
- [43] M. Wheeler, "Celebrity politics and cultural citizenship: Un good-will ambassadors and messengers of peace," in *Transnational Celebrity Activism in Global Politics*, L. Tsaliki, C. A. Frangonikolopoulos, and A. Huliaras, Eds. Bristol, U.K.: Intellect Ltd., 2011, ch. 2, pp. 45–62.
- [44] A. F. Cooper, H. Dobson, and M. Wheeler, "Non-western celebrity politics and diplomacy: Introduction," *Celebrity Stud.*, vol. 8, no. 2, pp. 312–317, Apr. 2017, doi: [10.1080/19392397.2017.1311509](https://doi.org/10.1080/19392397.2017.1311509).
- [45] E. Zivot and J. Wang, *Modeling Financial Time Series With S-Plus*, vol. 191. New York, NY, USA: Springer, 2007.



MUHAMMAD ALI MASOOD is currently a Ph.D. Scholar with the Department of Computer Science, Quaid-i-Azam University, Pakistan. His research interests include machine learning, sentiment classification, user classification, cybercrime, social network analysis, and big data analytics. The applications of his research include identifying and predicting outlier behavior through social media analytics and social network analysis.



RABEEH AYAZ ABBASI received the Ph.D. degree from the University of Koblenz-Landau, Germany, in 2010. Since 2011, he has been an Assistant Professor with Quaid-i-Azam University, Islamabad, Pakistan. He has a vast research experience in the fields of social media analytics and social network analysis. His research focuses on leveraging positive aspects of social media, including social media's use in saving lives, understanding events, and analyzing sentiments. He has published more than 35 articles in reputed journals, such as the *IEEE Computational Intelligence Magazine*, the *Computers in Human Behavior*, the *Telematics and Informatics*, the *Applied Soft Computing*, and *Scientometrics*, as well as the international conferences, such as the ACM Hypertext Conference, the ACM World Wide Web Conference, the Pacific Asia Conference on Knowledge Discovery and Data Mining, and the European Conference on Information Retrieval.



NG WEE KEONG is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He works in machine learning, privacy-preserving techniques, query-permissible encrypted databases, enterprise blockchain systems, data security, and blockchain security. His work is motivated by the need to harness the power of data for the betterment of stakeholders, where the data has confidentiality and privacy issues; where the data may be owned and held by different stakeholders, and where the data is so large that it must be hosted in cloud servers that are never completely secured (susceptible to insider and cyber-attacks). He contributes to companies and industries as technology consultant on projects involving data analytics, artificial intelligence, data privacy and security, and blockchain. His work is supported by various research grants. He is an Associate Editor and also a member of the Editorial Boards of five journals. Over the years, he has worked with many talented students on their Ph.D. and Master theses. He writes for online magazines and newspapers and contributes to social causes.