

Sentiment Drift Detection and Analysis in Real Time Twitter Data Streams

E. Susi* and A. P. Shanthi

Department of Computer Science & Engineering, Anna University, Chennai, 600025, Tamil Nadu, India

*Corresponding Author: E. Susi. Email: susi.e.ceg.94@gmail.com

Received: 06 May 2022; Accepted: 05 July 2022

Abstract: Handling sentiment drifts in real time twitter data streams are a challenging task while performing sentiment classifications, because of the changes that occur in the sentiments of twitter users, with respect to time. The growing volume of tweets with sentiment drifts has led to the need for devising an adaptive approach to detect and handle this drift in real time. This work proposes an adaptive learning algorithm-based framework, Twitter Sentiment Drift Analysis-Bidirectional Encoder Representations from Transformers (TSDA-BERT), which introduces a sentiment drift measure to detect drifts and a domain impact score to adaptively retrain the classification model with domain relevant data in real time. The framework also works on static data by converting them to data streams using the Kafka tool. The experiments conducted on real time and simulated tweets of sports, health care and financial topics show that the proposed system is able to detect sentiment drifts and maintain the performance of the classification model, with accuracies of 91%, 87% and 90%, respectively. Though the results have been provided only for a few topics, as a proof of concept, this framework can be applied to detect sentiment drifts and perform sentiment classification on real time data streams of any topic.

Keywords: Sentiment drift; sentiment classification; big data; BERT; real time data streams; twitter

1 Introduction

In today's digital world, people rely more on social media, not only to stay in contact with friends, but also to find information on any topic. From the obtained information, users tend to form opinions on their own, which plays a vital role in the formation of opinions in different areas including finance, health care and sports. Sentiment is an essential part in analyzing social behavior, which constitutes attitude, views, feelings on a topic. Sentiments or opinions of a tweet can be projected as positive, negative and neutral, based on the user's view. The process of classifying a tweet as sentiments or opinions are commonly referred to as sentiment analysis or opinion mining [1]. Sentiment analysis can be performed by lexicon-based approaches, machine learning based approaches and hybrid approaches.

A commonly used social media platform is twitter, which is a gold mine of data in the form of tweets. A tweet or post is a short message with a maximum of 280 characters, consisting of text, image, video, graphics



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

interchange format (GIFs), Uniform Resource Locator (URLs), user mentions and hashtags [2]. Huge amounts of tweets in the form of temporal data streams are generated at rapid speed in real time. These tweet streams, due to their inherent dynamic characteristics, make it difficult to perform sentiment analysis in real time. The volume and features of the tweets tends to change over time, driven by the user's perspective in real-world [3]. This sudden change in opinion of an underlying topic, based on public perspective over time, in an evolving data stream is often called sentiment drift [4]. The sentiment analysis approaches are affected by these sentiment drifts, and adequate measures have to be devised to address these in real time. This provided the motivation to perform sentiment classification in the presence of sentiment drifts on real time twitter data streams.

This paper proposes a Twitter Sentiment Drift Analysis-Bidirectional Encoder Representations from Transformers (TSDA-BERT) learning system to identify sentiment drifts in twitter data streams and handle it with adaptive learning in real time. This adaptive framework consists of various phases to process simulated tweet data streams and live tweet data streams. The Kafka tool is used to simulate the static tweets data as streams and a live processing system is developed using Twitter developer Application Programming Interface (API). The live or simulated data streams are processed in a window-based system, incorporating Spark, where the learning system gets adaptively retrained with recent knowledge. The learning system is designed using transformer-based learning i.e., fine-tuned Bidirectional Encoder Representations from Transformers (BERT) [5]. The proposed framework utilizes the recent data from the tweet data streams and plays a vital role in avoiding the performance degradation of the model, whenever a sentiment drift occurs in a time window.

Our main contributions can be summarized as follows:

- Development of a framework using Spark to handle and detect sentiment drifts using sentiment drift measure in real time sentiment analysis of twitter data streams
- Handling of static historical data by simulation using Kafka
- Adaptive learning with relevant tweets based on domain impact score using the sliding window approach when sentiment drift occurs
- Evaluation of the proposed framework for different domains with and without sentiment drifts.

The rest of the paper is organized as follows: Section 2 discusses the related works in twitter sentiment analysis. Section 3 describes the proposed architecture along with a detailed description of modules and algorithms. Section 4 describes the experimental setup. Discussion on results and their analysis is given in Section 5, and finally, the conclusion is given in Section 6.

2 Related Works

In recent years, sentiment analysis using twitter data streams has been an active research area and finds various application areas in domains such as health, sports, finance and politics [6]. Several researchers have addressed the issues related to sentiment analysis using twitter data.

Alharbi et al. [7] have developed a convolution neural network model, which uses the behavior information within the tweet document to perform sentiment analysis. 700,000 tweets posted by more than 3500 users have been collected and annotated using the SentiStrength algorithm. The algorithm has been implemented on the static collected data set and proves that adding user behavior features with the content of a tweet is beneficial in sentiment classification, with an accuracy of 88.71% using Convolution Neural Network (CNN). Hassonah et al. [8] have introduced a hybrid machine learning approach to improve sentiment analysis using Support Vector Machines (SVM) with feature selection methods such as relief and multiverse optimizer algorithms. The framework produces an accuracy of 96.85% by reducing the features.

In the health care sector, considering the recent pandemic era, Chakraborty et al. [9] have designed a Gaussian membership based fuzzy rule with SVM to correctly identify the sentiments from the tweets related to COVID-19 and claim that the work yields an accuracy of 79%.

Garcia et al. [10] have developed a system which detects topics related to COVID-19 and performs sentiment analysis. Tweets in two different languages, namely Portuguese and English have been collected for four months. Initial labeling of training data is performed by annotators, and the study is analyzed using different word embedding techniques with different classification methods. This work shows that the negative emotions are dominant in almost all the topics identified during the COVID-19 pandemic using unigrams with Multilingual Universal Sentence Encoder for semantic retrieval (mUSE).

Ghani et al. [11] have developed a system using Flume, to track dengue epidemic using hybrid filtration polarity. It calculates the sentiment analysis value of the twitter data to extract the useful information from the tweet content. This work detected the dengue epidemic based on the negative polarity score of each tweet content.

Sentiment analysis can also be used to identify the fanatic and anti-fanatic emotions on sports related tweets. Alqmase et al. [12] have proposed a classification tool that automatically evaluates and classifies Arabic tweets against sports fanaticism. This model automatically filters the messages focusing on fanaticism to avoid negative effects on society. Similarly, Aloufi et al. [13] have developed a lexicon-based approach to classify the sentiments of football related tweets. The Fédération Internationale de Football Association (FIFA) world cup and championship league events related tweets have been annotated manually. The results show that the developed football lexicon dataset has performed well on finding sentiments of tweets with an accuracy of 56% rather than using opinion lexicons.

Daniel et al. [14] have performed sentiment analysis in the financial field by considering the various events of selected companies. The finance related tweets have been filtered from the day-to-day tweets using a dictionary related to finance. Sentiments are assigned using state-of-the-art sentiment lexicons. Valle-Cruz et al. [15] have developed a lexicon-based approach to find the correlation matrix, which denotes the relationship among the twitter posts and the behavior of financial markets during the pandemic period of COVID-19. The developed senti-computing model has better performance in finding the impact of twitter indices.

Another major field where sentiment analysis plays a vital role is in areas like prediction of election results [16,17]. Xia et al. [18] have collected 2.6 million tweets related to the U.S presidential election and have developed a multilayer perceptron with feature extraction methods to predict the polarity of the election candidates, with an accuracy of 81.53%. Rodríguez-Ibáñez et al. [19] have designed an algorithm to analyze the temporal evolution of sentiment in critical scenarios involving political elections. They have also proved the usefulness of indices with moderate complexity to obtain information on sentiments in politics with temporal dynamics and interpretability.

Several researchers have introduced and proved that sentiment analysis can also be performed in real time [20]. Gupta et al. [21] have developed a novel approach based on hybrid hashtags for effective tweet classification with limited number of data information. They also have incorporated the Apache storm, a distribution framework, to achieve classification in real time big data streams. Combining both, they have proved that the proposed framework achieves a better performance with an accuracy of 97.5% with Naïve Bayes, 80.2% with Hoeffding trees and 94.5% with Hoeffding adaptive trees.

Chatterjee et al. [22] have developed a real time incremental learning framework for sentiment classification, concentrating on the Indian general elections 2019. Their motivation was to find authentic tweets for sentiment classification in real time. They have devised a machine learning based generic framework and Representational State Transfer (REST) plugin component. From the results they have proved that the proposed framework can be used for any sentiment analysis on any major public event, with an accuracy of 72%.

Kılınç [23] has claimed that data produced by fake personalities affects the sentiment analysis model. In order to overcome this, a spark based big data framework, which incorporates a sentiment analysis method and a fake detection method has been proposed. Their framework has produced an accuracy of 86.77% and 80.93% on static and real time datasets, respectively. Ahmed et al. [24] have developed a framework to identify the user's sentiment changes for top trending subtopics related to COVID-19. They also have designed an algorithm to find the top-involved users at various times using the user's involvement score. Similarly, Rowe et al. [25] have analyzed the tweets using topic modeling and Valence Aware Dictionary and sEntiment Reasoner (VADER) with respect to immigrants. They have observed that there is a steady increase of negative attitudes towards the immigrants during the pandemic era.

In real time text data streams, whenever there is a sudden change in data distribution over time, the devised model tends to perform poorly. Several researchers coined this phenomenon as concept drift [26]. Costa et al. [27] have devised a distinctive, window-based ensemble incremental approach, to handle these concept drifts. The results proved that the proposed incremental model performs well even with these concept drifts, with an average accuracy of 76.39%. Zhang et al. [28] addressed this concept drift using a three-layer approach based on label space, feature space and relationship among labels and features. Moreover, unlike other concept drift detection methods, the proposed method is independent of the classifier and increases the efficiency layer by layer, while detecting the different types of concept drifts.

Müller et al. [29] have explored the concept drifts, by focusing on vaccine sentiments during the pandemic period. A sentiment classifier has been devised and trained on 57.4 million vaccine related tweets. The drifts are identified in classifiers by measuring the performance changes of the devised model. From the F1-scores, it is clearly understood that the concept drift is impactful on the negative class, indicating anti-vaccine concepts are changing faster than pro-vaccine concepts.

Gözüaçık et al. [30] have presented an unsupervised, one class drift detector to detect concept drift. They have performed a comprehensive evaluation on various drift detection methods using thirteen datasets. The adaptive retraining is performed by dropping the old samples only when data distribution changes, thus maintaining the model performance.

To the best of our knowledge, only one study has directly addressed the concept of sentiment drifts in twitter sentiment classification. Zhang et al. [31] have addressed the concept of sentiment drift in public sentiments, where the information is hidden in a temporal sequence. The model devised using historical data tends to perform poorly on new tweets. To handle this, a novel Hierarchical Variational Auto-Encoder (HVAE), to measure the drift among the historical and new tweet data streams has been proposed. From experimental results, they show that the proposed framework achieves an accuracy of 83.3%.

To adaptively learn the tweets data in real time, several researchers have used BERT. BERT is a pre-trained neural network model, which is based on bidirectional transformer architecture. BERT, which is trained on general book corpus, underperforms on domain specific tasks. In order to overcome this, Qudar et al. [32] have introduced domain specific BERT, TweetBERT and tested on various benchmark datasets. Based on their experimental results, TweetBERT outperforms BERT, BioBERT [33], SciBERT [34] and FinBERT [35] on each twitter data set.

In most of the literatures presented above, the researchers have focused on sentiment analysis, without considering the impact of sentiment drifts. Also, most of the researchers have used static data sets rather than real time data streams in their experiments. In a real time, scenario, the traditional sentiment classification frameworks do not scale up to the speed of incoming twitter data. It is also observed that the drift in real time sentiment classification affects the performance of the devised model. The performance of the model can be improved by the concept of adaptive learning. This work, taking cue from the literature survey done, proposes an adaptive transformer based TSDA-BERT using Spark, to detect and handle sentiment drifts in real time twitter data streams for sentiment classification. The details of the proposed work are presented below.

3 System Architecture

An adaptive framework for identifying the sentiments and any changes (drifts) in sentiments is proposed in this research work. Fig. 1 shows the overall architecture of the proposed framework to process and perform sentiment analysis, both, in the real time and simulated scenarios. The framework consists of the following phases:

- Data collection and splitting
- Tweets pre-processing
- Automated training data generation
- Static tweets simulation
- Sentiment classification and drift detection using TSDA-BERT.

All these phases are explained in this section.

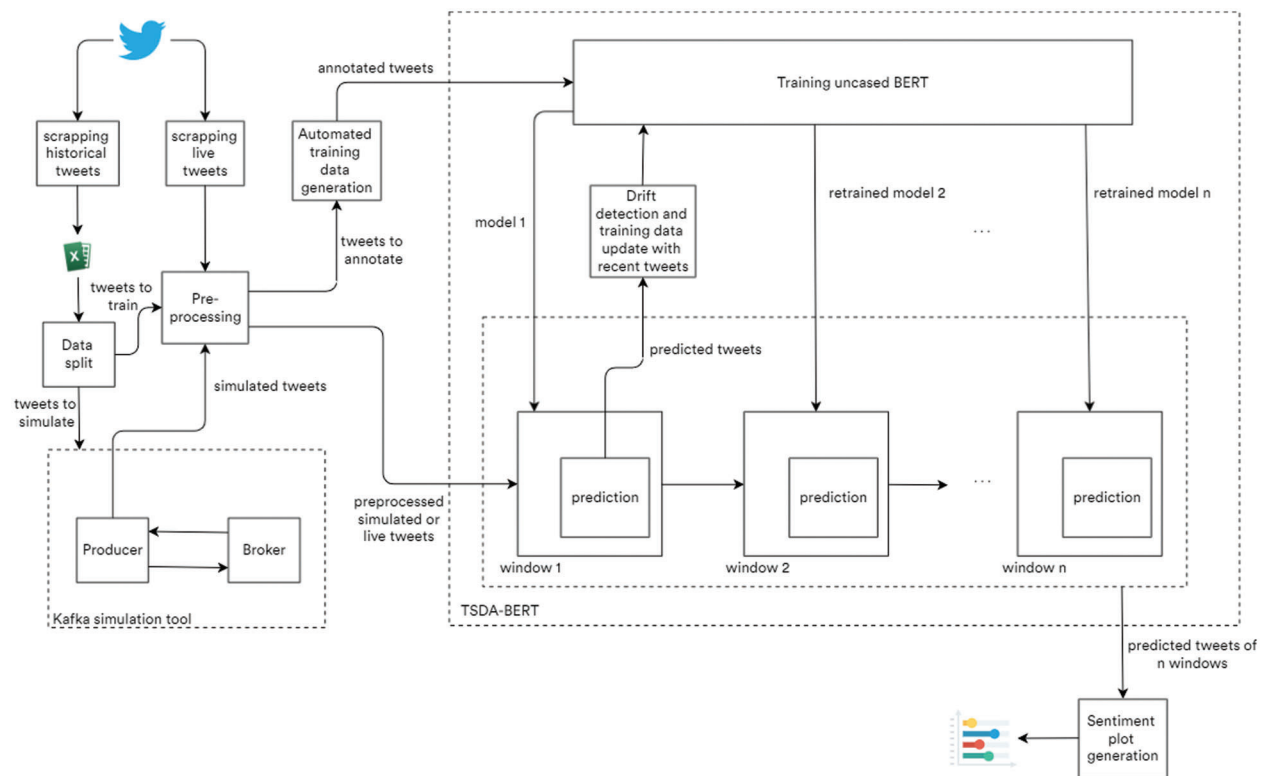


Figure 1: Overall architecture of the proposed system

3.1 Data Collection and Splitting

The data collection and splitting phase collects tweets from Twitter as follows:

- The static historical data, given the time frame, which is scrapped using ‘snsrape’
 - The historical tweets data are split into a trainable tweets data set and simulatable tweets data set based on the given time frame
- The dynamic live tweet data is tracked using the Twitter developer API, ‘tweepy’ with hashtags (#).

3.2 Tweets Preprocessing

The preprocessing phase in the proposed framework performs lower casing, removes URLs, user mentions, stop words, punctuations, number and hashtags and decodes emoticons. This process is performed to make sure that the input tweets are in proper format for the TSDA-BERT phase.

3.3 Automated Training Data Generation

The tweets, which are used for training are passed to this phase. This phase performs the automatic annotation of tweets using the state of art Hugging face pretrained ‘sentiment-roberta-large-english’ [36]. This model is fine-tuned using ‘RoBERTa-large’ [37], which performs binary sentiment analysis on English language text.

3.4 Static Tweets Simulation

The tweets, which have been split for simulation, are passed to this phase. This component uses the publish-subscribe distributed messaging tool, Kafka. These tweets are read one by one in asynchronous fashion, a Kafka topic is created in the producer and gets published. The intermediate broker phase ensures the equal distribution of tweet message topics among the partitions. Then the producer produces the tweets at regular intervals with respect to tweet scrapped date and time. These simulated tweets are then passed to TSDA-BERT for sentiment analysis.

3.5 Sentiment Classification and Drift Detection Using TSDA-BERT

The sentiment drift analysis of the incoming tweets is performed using TSDA-BERT. TSDA-BERT performs the following operations:

3.5.1 Fine Tuning of Uncased BERT for Sentiment Classification of Different Domains in Real Time

Fine-tuning is an approach to transformer learning, where the pretrained model’s output is fitted with a new task with a specific dataset. The initial training dataset is used to fine tune the BERT uncased model. The sentiment classifier consists of three dense layers with a dropout layer. The first two dense layers are connected with the fine-tuned BERT encoding layer and the output layer generates the sentiment of the encoded tweets.

3.5.2 Implementation of Sliding Window with Spark for Parallel Sentiment Analysis

Sliding window along with spark is used here for parallel processing of tweet data streams. With the sliding window approach, the latest observed data is added to the real time solution and raises the efficiency of the devised classifier model. For example, a specific time frame is selected for the sliding window, say four days. The initial dataset is sorted in ascending order based on date. The new tweets coming in for the latest four days are added and its count is calculated, while the corresponding number of oldest tweets are dropped from the initial dataset.

3.5.3 Sentiment Classification and Drift Detection

In the first window w_0 , the sentiments are classified using the initially trained model. The tweets are classified as positive or negative, depending on whether the positive score or negative score is higher. The positive and negative scores are computed by Eqs. (1) and (2),

$$\text{positive score} = \frac{\text{number of positive tweets}}{\text{total number of tweets}} \quad (1)$$

$$\text{negative score} = \frac{\text{number of negative tweets}}{\text{total number of tweets}} \quad (2)$$

These classified tweets are passed to the system for sentiment drift analysis, where the drifts are identified by the sentiment drift measure, which is calculated by Eq. (3),

$$\text{sentiment drift measure} = \text{positive score}_t - \text{negative score}_t \quad (3)$$

The sentiment drift is identified from the sentiment drift measure, whenever there is a change in the score from positive to negative and vice versa in the current and previous time periods.

3.5.4 Retraining of BERT with Recent Tweets Based on Sentiment Drift

In order to keep the transformer-based model up to date, the learning data needs to be updated. This update of learning data takes place under two circumstances after the sentiment classification and drift detection phase. Under normal circumstances, where there is no sentiment drift in a window, the training set gets updated depending on the number of live tweets that get added, the same number of old tweets are removed from the dataset. In other circumstances, the update of training data depends upon the domain impact score. The domain impact score defines the impact percentage of the tweet, based on the domain corpus. The domain impact score is defined by Eq. (4),

$$\text{domain impact score} = \frac{\text{Number of } (W \cap D)}{\text{Total number of words in a tweet}} \quad (4)$$

where, W is the set of words of a tweet, $\{w_1, w_2, \dots, w_n\}$ and D is the set of domain specific impact words $\{d_1, d_2, \dots, d_n\}$. Based upon the domain, a threshold value is selected, say 0.5. The tweets which pass the threshold value are selected and appended to the training data set and its count is calculated. The old number of tweets corresponding to the selected domain impact tweets are dropped from the training data set and the model gets retrained.

The overall working principle of the TSDA-BERT is explained in Algorithm 1. The initial model (M_0) is built with static historical tweets (T_0) collected over a specific time period. In real time or in simulation tweet data streams (T_i), a sliding window (SW) is designed to process a chunk of tweets (Tw_i). Sentiment drifts are detected in a window using the sentiment drift measure. The model gets retrained with the updated training data set, based on the number of tweets (l_i) processed in the window, both with and without sentiment drifts. A maximum of 324685 tweets (T_0) have been used to build the initial model (M_0).

Algorithm 1: TSDA-BERT

Input: Tweets T_i , Model M_i

Output: Sentiment classified tweets s_i

train data $T_0 \leftarrow$ annotated historical tweets data

model $M_0 \leftarrow$ train (T_0)

for **each** window SW_i in data stream T_i **do**

$Tw_i \leftarrow$ chunk of tweets in SW_i

$df_i \leftarrow$ spark data frame (Tw_i)

$s_i \leftarrow M_0.\text{transform}(df_i)$

$s_i.\text{save}()$

calculate positive score, negative score

(Continued)

Algorithm 1 (continued)

```

calculate sentiment drift measure
if drifted do
   $S_{domain\_impact} \leftarrow \text{calculate domain\_impact}(s_i)$ 
   $S_{domain\_impact} \leftarrow \text{drop tweet if domain impact} < \text{threshold}$ 
   $l_i \leftarrow \text{count}(S_{domain\_impact})$ 
   $T_0 \leftarrow \text{drop old } l_i \text{ tweets from } T_0$ 
   $T_0 \leftarrow T_0 + S_{domain\_impact}$ 
   $M_0 \leftarrow \text{train}(T_0)$ 
end
else do
   $l_i \leftarrow \text{count}(s_i)$ 
   $T_0 \leftarrow \text{drop old } l_i \text{ tweets from } T_0$ 
   $T_0 \leftarrow T_0 + s_i$ 
   $M_0 \leftarrow \text{train}(T_0)$ 
end
end

```

4 Experimental Setup**4.1 System Configuration**

All the live processing of the proposed framework is run on a machine with Graphics Processing Unit (GPU) cores with 8 workers with a Random Access Memory (RAM) size of 25 GB for parallel processing, using Spark enabled Natural Language Processing (NLP) frameworks. Distribution strategy was incorporated in TensorFlow to achieve faster and parallel execution of training and re-training of TSDA-BERT. The Kafka simulation system is run on a Mac system with 8 core GPU and 16 core neural engine with a RAM size of 16 GB for faster execution.

4.2 Dataset

The dataset to test and train the proposed framework is collected based on the twitter trending topics such as finance, sports and politics. These datasets consist of two classes, positive and negative defining the sentiment of a tweet. These historical tweet datasets are scraped using the Twitter based python library (snsrcape), which forms the ground truth. [Tab. 1](#) shows the basic statistics of the collected tweets. These tweets are annotated as explained in Section 3.3.

Table 1: Datasets

| Topic | Domain | Number of tweets | Positive | Negative |
|--------------|-------------|------------------|----------|----------|
| Stock market | Finance | 209634 | 107518 | 102116 |
| Manutd | Sports | 324685 | 164117 | 160568 |
| COVAXIN WHO | Health care | 59866 | 25931 | 33935 |

4.3 Evaluation Metrics

The metric chosen for evaluating the proposed framework using the real time twitter tweets is ‘accuracy’ by Eq. (5). This paper deals with the online opinion-based sentiment dataset, which comes at high velocity and might face some sentiment drifts that cause performance degradation of the devised model.

$$accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (5)$$

True positive (TP) is an outcome, where the model correctly classifies the positive sentiment. Similarly, true negative (TN) is an outcome, where the model correctly classifies the negative sentiment. False positive (FP) is an outcome, where the model incorrectly classifies the positive sentiment and false negative (FN) is an outcome, where the model incorrectly classifies the negative sentiment.

5 Results and Analysis

Experiments have been carried out on the twitter dataset to evaluate the performance of the TSDA-BERT framework. The results obtained are discussed for the following:

- Real time tweets from twitter data stream
- Simulated data streams from stored tweets
- Effects of sentiment drift on sentiment analysis
 - Comparison of accuracy of the model with and without domain impact score

Before the processing of live and simulated tweets in a window as in Algorithm 1, a transformer-based model is trained on the domain related historical tweets and fine-tuned with a learning rate of $5e-5$ for 10 epochs with a batch size of 64. Tab. 2 gives the accuracy information and computation time of the initial trained model. From the computation time in Tab. 2 we could infer that the computation time is directly related to the number of tweets in the initial training dataset.

Table 2: Accuracy of the initial trained model

| Topic | Accuracy | Computation time (s) |
|--------------|------------|----------------------|
| Manutd | 0.8060566 | 2399 |
| Stock market | 0.8056331 | 2450 |
| COVAXIN WHO | 0.79088724 | 240 |

5.1 Real Time Tweets from Twitter Data Stream

The proposed TSDA-BERT framework works on a sliding window approach when analyzing the live tweet data streams. The TSDA-BERT performs sentiment classification, whenever the assigned time frame of a window is reached. The collected live tweets are transformed into a spark data frame, where the fine-tuned model classifies sentiments. The number of days considered, the count and the distribution of positive and negative tweets are shown in Tab. 3 for each hashtag.

The sentiment drift detection is performed on each window once classifications get completed. From Tab. 4, it can be observed that there were several opinion changes based on the performance of the Manchester United team. On considering the time period from March 6–March 11, the opinion drifted suddenly from positive to negative from March 3–March 5 due to poor performance against the Manchester city game. The overall sentiment drift of #manutd can be visually represented in Fig. 2. From this we could infer that there were 4 drifts based on the team’s performance during the pre and post live

match. The computation time of the proposed frame work for the first two sliding windows are 1911, 2102 s respectively.

Table 3: Number of tweets collected based on hashtag at each window and their sentiment distribution

| Hashtag | Window | Days | Count | Positive | Negative |
|--------------|--------|-----------------------|-------|----------|----------|
| Manutd | 1 | 01/03/2022–03/03/2022 | 15746 | 8388 | 7358 |
| | 2 | 04/03/2022–06/03/2022 | 16663 | 9739 | 6924 |
| | 3 | 07/03/2022–09/03/2022 | 67687 | 36148 | 31539 |
| | 4 | 10/03/2022–13/03/2022 | 23472 | 12574 | 10898 |
| | 5 | 14/03/2022–16/03/2022 | 57541 | 33903 | 23638 |
| Stock market | 1 | 01/03/2022–04/03/2022 | 32712 | 14851 | 17556 |
| | 2 | 05/03/2022–08/03/2022 | 29802 | 13797 | 15627 |
| | 3 | 09/03/2022–12/03/2022 | 29361 | 14479 | 14624 |
| | 4 | 13/03/2022–16/03/2022 | 29156 | 14508 | 14298 |
| | 5 | 17/03/2022–20/03/2022 | 21246 | 11337 | 9624 |

Table 4: Sentiment drift measure and detection

| Hashtag | Window | Time period | Positive score | Negative score | Sentiment drift measure | Drift detected? |
|---------|--------|-------------|----------------|----------------|-------------------------|-----------------|
| Manutd | 1 | 01/03/2022 | 53.99 | 46.01 | 7.98 | N |
| | | 02/03/2022 | 48.3 | 51.7 | -3.4 | Y |
| | | 03/03/2022 | 58.75 | 41.25 | 17.5 | Y |
| | 2 | 04/03/2022 | 62.34 | 37.66 | 24.68 | N |
| | | 05/03/2022 | 63.63 | 36.37 | 27.26 | N |
| | | 06/03/2022 | 43.39 | 56.61 | -13.22 | Y |
| | 3 | 07/03/2022 | 40.46 | 59.54 | -19.08 | N |
| | | 08/03/2022 | 42.78 | 57.22 | -14.44 | N |
| | | 09/03/2022 | 43.24 | 56.76 | -13.52 | N |
| | 4 | 10/03/2022 | 45.41 | 54.59 | -9.18 | N |
| | | 11/03/2022 | 48.99 | 51.01 | -2.02 | N |
| | | 12/03/2022 | 58.15 | 41.85 | 16.3 | Y |
| | 5 | 13/03/2022 | 55.5 | 44.5 | 11 | N |
| | | 14/03/2022 | 59.74 | 40.26 | 19.48 | N |
| | | 15/03/2022 | 46.32 | 53.68 | -7.36 | Y |

(Continued)

Table 4 (continued)

| Hashtag | Window | Time period | Positive score | Negative score | Sentiment drift measure | Drift detected? |
|--------------|--------|-------------|----------------|----------------|-------------------------|-----------------|
| Stock market | 1 | 01/03/2022 | 48.75 | 51.25 | -2.5 | N |
| | | 02/03/2022 | 49.23 | 50.77 | -1.54 | N |
| | | 03/03/2022 | 50.04 | 49.96 | 0.08 | Y |
| | | 04/03/2022 | 45.37 | 54.63 | -9.26 | Y |
| | 2 | 05/03/2022 | 45.56 | 54.44 | -8.88 | N |
| | | 06/03/2022 | 52.17 | 47.83 | 4.34 | Y |
| | | 07/03/2022 | 50.9 | 49.1 | 1.8 | N |
| | | 08/03/2022 | 45.45 | 54.55 | -9.1 | Y |
| | 3 | 09/03/2022 | 52.14 | 47.86 | 4.28 | Y |
| | | 10/03/2022 | 55.43 | 44.57 | 10.86 | N |
| | | 11/03/2022 | 53.11 | 46.89 | 6.22 | N |
| | | 12/03/2022 | 44.25 | 55.75 | -11.5 | Y |
| | 4 | 13/03/2022 | 47.84 | 52.16 | -4.32 | N |
| | | 14/03/2022 | 52.04 | 47.96 | 4.08 | Y |
| | | 15/03/2022 | 51.7 | 48.3 | 3.4 | N |
| | | 16/03/2022 | 55.66 | 44.34 | 11.32 | N |
| | 5 | 17/03/2022 | 59.07 | 40.93 | 18.14 | N |
| | | 18/03/2022 | 57.28 | 42.72 | 14.56 | N |
| | | 19/03/2022 | 57.32 | 42.68 | 14.64 | N |
| | | 20/03/2022 | 60.52 | 39.48 | 21.04 | N |

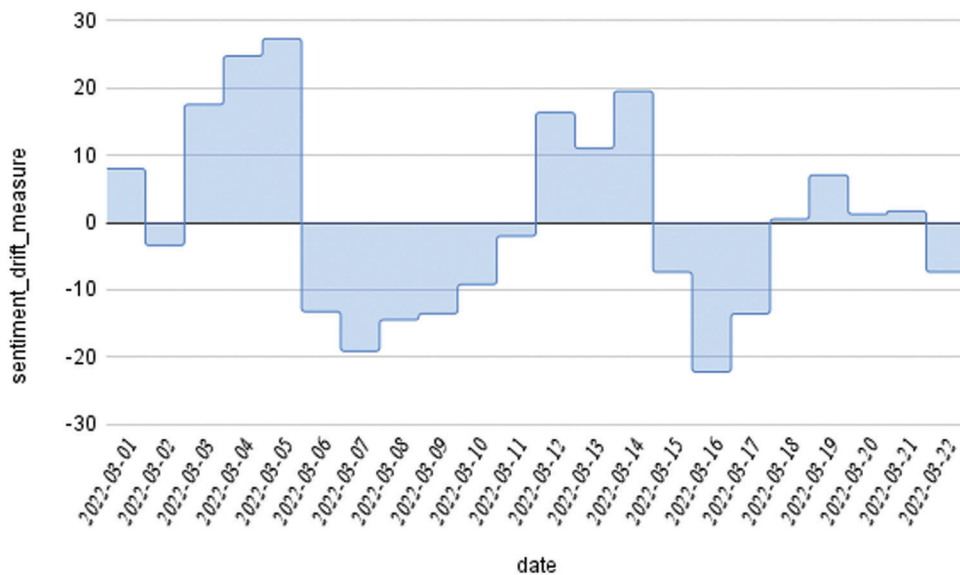


Figure 2: Overall sentiment change of #manutd over the period March 03–March 22

Likewise, for the stock market related tweets, considering the time window of March 1–March 14, there were several oscillations of sentiments among the twitter users from the sentiment drift measure of Tab. 4, due to the unstable stock market. Fig. 3 shows the prevalence of sentiment drifts during March 1–March 14, but later on, that is after March 15, it can be seen that there prevailed a positive opinion towards the stock market, since there was a bull in the stock market. The computation time of the proposed frame work for the first two sliding windows are 5256, 5100 s respectively. The time taken for processing the tweets in a window depends on the number of tweets that are adaptively included and processed in a window.

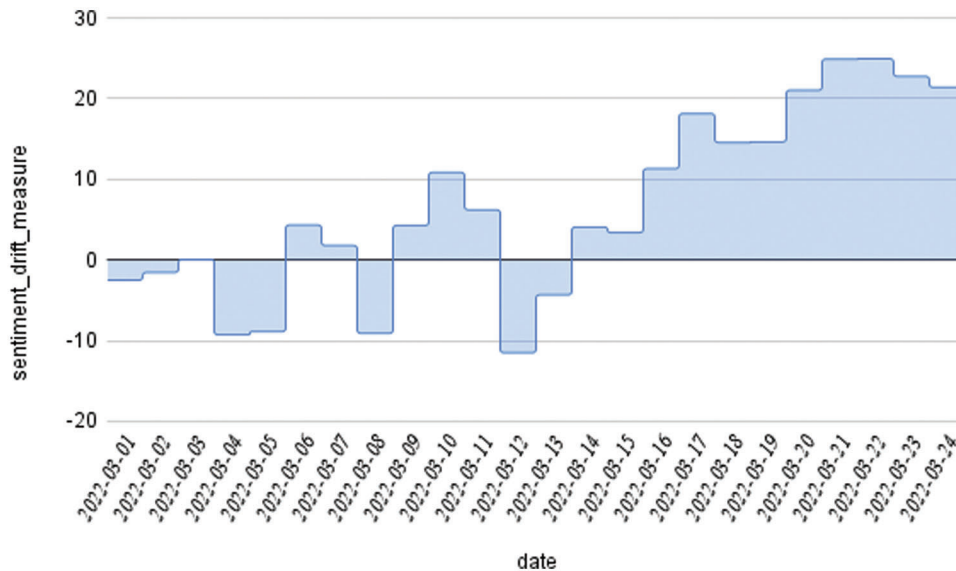


Figure 3: Overall sentiment change of #stockmarket over the period March 03–March 22

5.2 Simulated Data Streams from Stored Tweets

In health care related tweets, the view of World Health Organization (WHO) against COVAXIN related tweets were considered to measure the sentiment changes of online users during the timeline Jan 2022. The collected tweets were simulated using Kafka and classification was performed. Tab. 5 shows a sample of the sentiment drift measure obtained for various time periods. For example, the first entry for window 1 (7 days time period) shows that the sentiment drift measure obtained is -41 , indicating that most of the tweets during that window had a negative sentiment. The computation time of the proposed frame work for the first two sliding windows are 892, 551 s respectively.

Table 5: Sample of sentiment drift measure values of simulated #covaxinwho tweets

| Window | Time period | Sentiment drift measure |
|--------|-----------------------|-------------------------|
| 1 | 01/03/2022–07/03/2022 | -41 |
| 2 | 08/03/2022–15/03/2022 | -65 |
| 3 | 16/03/2022–22/03/2022 | -48.38 |
| 4 | 23/03/2022–26/03/2022 | -34.78 |
| 5 | 27/03/2022–31/03/2022 | -37.94 |

Fig. 4 shows the overall sentiment change of '#covaxinwho'. From this plot, it is observed that the negative sentiment predominantly dominates and sentiment drift was minimal with a smaller number of temporary spikes.

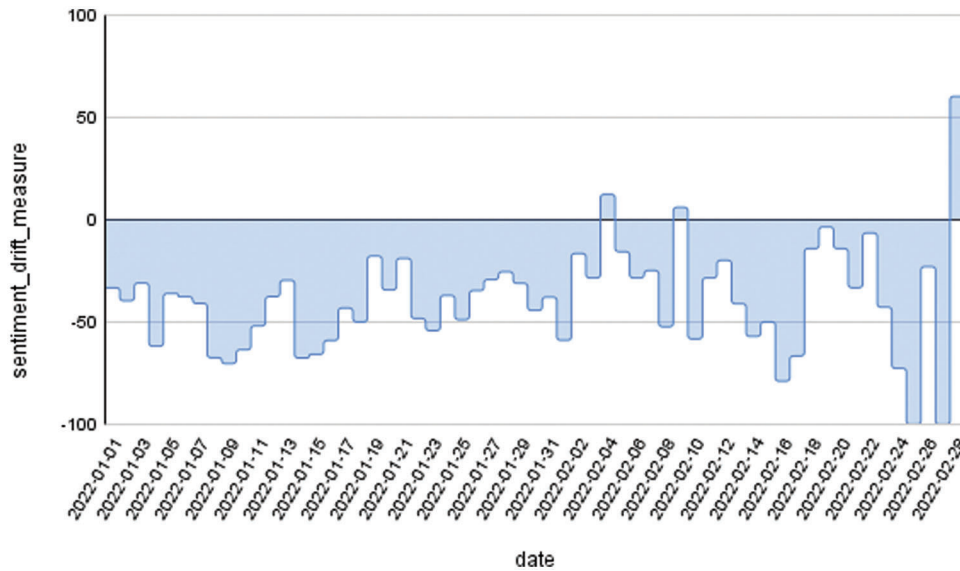


Figure 4: Sentiment change of simulated #covaxinwho tweets

5.3 Effects of Sentiment Drift on Sentiment Analysis

The TSDA-BERT uses the concept of domain impact score to overcome the performance degradation of the devised model due to sentiment drift. Tab. 6 gives the accuracy of the TSDA-BERT model with and without considering the domain impact score. It can be seen that there is a maximum performance improvement of 7.6% and 5.4% respectively, for the '#manutd' and '#stockmarket' at the end of window 5. In both cases, sentiment drifts were observed and the model is able to perform well, in spite of the sentiment drifts. In the case of '#covaxinwho', the performance of the model remains the same, with or without domain impact score. This is because, as stated earlier, the tweets for this hashtag showed only negative sentiments throughout and no sentiment drifts were detected.

Table 6: Comparison of accuracy with and without domain impact score

| Hashtag | Window | Accuracy (without domain impact score) | Accuracy (with domain impact score) | Improvement in performance (%) |
|--------------|--------|--|-------------------------------------|--------------------------------|
| Manutd | 1 | 0.8202144 | 0.8224896 | 0.2 |
| | 2 | 0.81905264 | 0.8361177 | 1.7 |
| | 3 | 0.81541765 | 0.8589702 | 4.3 |
| | 4 | 0.8341456 | 0.87429476 | 4.0 |
| | 5 | 0.8398827 | 0.91592425 | 7.6 |
| Stock market | 1 | 0.8401542 | 0.84028305 | 0 |
| | 2 | 0.8378009 | 0.8444307 | 0.6 |
| | 3 | 0.8244166 | 0.8688822 | 4.4 |
| | 4 | 0.82316025 | 0.8949854 | 7.1 |
| | 5 | 0.84981385 | 0.9044058 | 5.4 |

(Continued)

Table 6 (continued)

| Hashtag | Window | Accuracy (without domain impact score) | Accuracy (with domain impact score) | Improvement in performance (%) |
|-------------|--------|--|-------------------------------------|--------------------------------|
| COVAXIN WHO | 1 | 0.79597515 | 0.80409294 | 0.8 |
| | 2 | 0.79829293 | 0.8054671 | 0.7 |
| | 3 | 0.79856026 | 0.8063644 | 0.7 |
| | 4 | 0.7986512 | 0.8066369 | 0.7 |
| | 5 | 0.7986311 | 0.80787736 | 0.9 |

Fig. 5 shows the accuracy variation of the model with domain impact score and without domain impact score of the experimented hashtags.

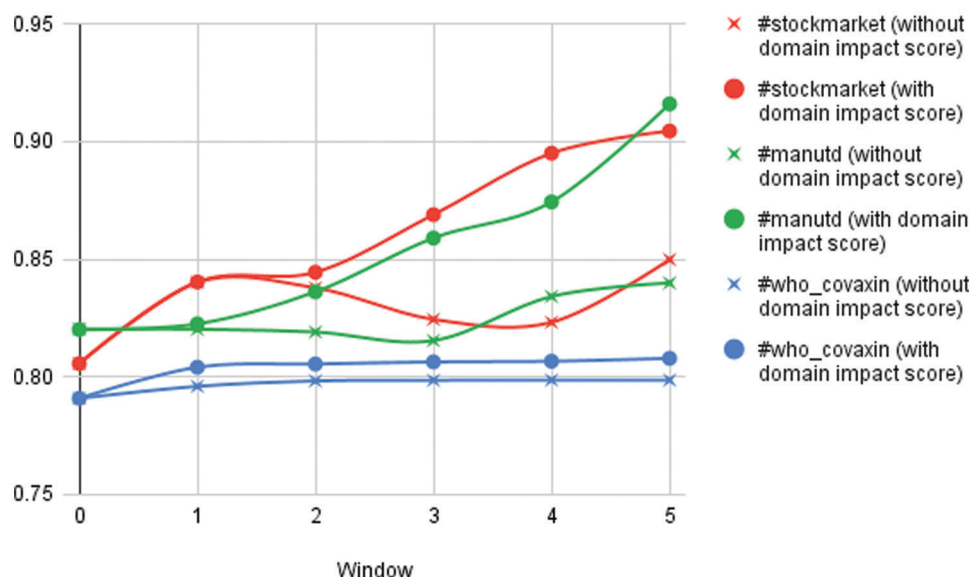


Figure 5: Accuracy variation of #stockmarket, #manutd and #covaxinwho over window

The experimental results performed on the TSDA-BERT framework thus show that it is able to perform well, in spite of sentiment drifts. This is because of the inclusion of relevant tweets, based on the domain impact score and adaptive learning that is performed on the framework.

6 Conclusion

This paper has presented a transformer based adaptive learning mechanism to identify sentiment drifts and perform sentiment classification on real time and simulated twitter data streams. The proposed TSDA-BERT framework on sentiment classification of various domains like sports, finance and health care shows that the framework has improved performance even in the presence of sentiment drifts. This is mainly attributed to the adaptive learning performed on the framework, based on the domain impact score. The adaptive learning of the model achieves considerably better accuracy even with sentiment drifts. The future work will focus on exploring the possibility of extending the TSDA-BERT framework to

automatically handle any trending topic in real time and also to filter out the tweets from bot and fake accounts and improve the performance of the framework.

Acknowledgement: Thanks to our families, friends & colleagues who supported us morally.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] K. Chakraborty, B. Siddhartha and R. Bag, “A survey of sentiment analysis from social media data,” *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 450–464, 2020.
- [2] D. Antonakaki, F. Paraskevi and I. Sotiris, “A survey of twitter research: Data model, graph structure, sentiment analysis and attacks,” *Expert Systems with Applications*, vol. 164, pp. 114006, 2021.
- [3] V. Iosifidis and E. Ntoutsi, “Sentiment analysis on big sparse data streams with limited labels,” *Knowledge and Information Systems*, vol. 62, no. 4, pp. 1393–1432, 2020.
- [4] A. Bechini, A. Bondielli, P. Ducange, F. Marcelloni and A. Renda, “Addressing event-driven concept drift in twitter stream: A stance detection application,” *IEEE Access*, vol. 9, pp. 77758–77770, 2021.
- [5] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv: 1810.04805, 2018.
- [6] M. Ebrahimi, A. H. Yazdavar and A. Sheth, “Challenges of sentiment analysis for dynamic events,” *IEEE Intelligent Systems*, vol. 32, no. 5, pp. 70–75, 2017.
- [7] A. S. M. Alharbi and E. de Doncker, “Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information,” *Cognitive Systems Research*, vol. 54, pp. 50–61, 2019.
- [8] M. A. Hassonah, R. Al-Sayyed, A. Rodan, A. M. Al-Zoubi, I. Aljarah *et al.*, “An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter,” *Knowledge-Based Systems*, vol. 192, pp. 105353, 2020.
- [9] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag *et al.*, “Sentiment analysis of covid-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media,” *Applied Soft Computing*, vol. 97, pp. 106754, 2020.
- [10] K. Garcia and L. Berton, “Topic detection and sentiment analysis in twitter content related to covid-19 from Brazil and the USA,” *Applied Soft Computing*, vol. 101, pp. 107057, 2021.
- [11] N. B. A. Ghani, S. Hamid, M. Ahmad, Y. Saadi, N. Jhanjhim *et al.*, “Tracking dengue on twitter using hybrid filtration-polarity and apache flume,” *Computer Systems Science and Engineering*, vol. 40, no. 3, pp. 913–926, 2022.
- [12] M. Alqmase, H. Al-Muhtaseb and H. Rabaan, “Sports-fanaticism formalism for sentiment analysis in arabic text,” *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1–24, 2021.
- [13] S. Aloufi and A. E. Saddik, “Sentiment identification in football-specific tweets,” *IEEE Access*, vol. 6, pp. 78609–78621, 2018.
- [14] M. Daniel, R. F. Neves and N. Horta, “Company event popularity for financial markets using twitter and sentiment analysis,” *Expert Systems with Applications*, vol. 71, pp. 111–124, 2017.
- [15] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau and R. Sandoval-Almazán, “Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods,” *Cognitive Computation*, vol. 14, no. 1, pp. 372–387, 2022.
- [16] L. Belcastro, R. Cantini, F. Marozzo, D. Talia and P. Trunfio, “Learning political polarization on social media using neural networks,” *IEEE Access*, vol. 8, pp. 47177–47187, 2020.
- [17] K. D. S. Brito, R. L. C. S. Filho and P. J. L. Adeodato, “A systematic review of predicting elections based on social media data: Research challenges and future directions,” *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 819–843, 2021.

- [18] E. Xia, H. Yue and H. Liu, "Tweet sentiment analysis of the 2020 US presidential election," in *Companion Proc. of the Web Conf., WWW '21: The Web Conf.*, Ljubljana, Slovenia, Balkans, pp. 367–371, 2021.
- [19] M. Rodríguez-Ibáñez, F. J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez, "Sentiment analysis of political tweets from the 2019 spanish elections," *IEEE Access*, vol. 9, pp. 101847–101862, 2021.
- [20] A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using naive Bayes," in *2nd Int. Conf. on Next Generation Computing Technologies (NGCT)*, Dehradun, India, IEEE, pp. 257–261, 2016.
- [21] V. Gupta and R. Hewett, "Real-time tweet analytics using hybrid hashtags on twitter big data streams," *Information*, vol. 11, no. 7, pp. 1–23, 2020.
- [22] S. Chatterjee and S. Gupta, "Incremental real-time learning framework for sentiment classification: Indian general election 2019, a case study," in *IEEE 6th Int. Conf. on Big Data Analytics (ICBDA)*, Xiamen, China, pp. 198–203, 2021.
- [23] D. Kılınç, "A spark based big data analysis framework for real time sentiment prediction on streaming data," *Software: Practice and Experience*, vol. 49, no. 9, pp. 1352–1364, 2019.
- [24] M. S. Ahmed, T. T. Aurpa and M. M. Anwar, "Detecting sentiment dynamics and clusters of twitter users for trending topics in covid-19 pandemic," *PLoS One*, vol. 16, no. 8, pp. 1–20, 2021.
- [25] F. Rowe, M. Mahony, E. Graells-Garrido, M. Rango and N. Sievers, "Using twitter to track immigration sentiment during early stages of the covid-19 pandemic," *Data & Policy*, vol. 3, pp. E36, 2021.
- [26] R. Mohawesh, S. Tran, R. Ollington and S. Xu, "Analysis of concept drift in fake reviews detection," *Expert Systems with Applications*, vol. 169, pp. 1–20, 2021.
- [27] J. Costa, C. Silva, M. Antunes and B. Ribeiro, "Concept drift awareness in twitter streams," in *2014 13th Int. Conf. on Machine Learning and Applications*, 1730 Massachusetts Ave., NW Washington DC, United States, IEEE, pp. 294–299, 2014.
- [28] Y. Zhang, G. Chu, P. Li, X. Hu and X. Wu, "Three-layer concept drifting detection in text data streams," *Neurocomputing*, vol. 260, pp. 393–403, 2017.
- [29] M. Müller and M. Salathé, "Addressing machine learning concept drift reveals declining vaccine sentiment during the covid-19 pandemic," arXiv preprint arXiv: 2012.02197, 2020.
- [30] Ö. Gözüağık and F. Can, "Concept learning using one-class classifiers for implicit drift detection in evolving data streams," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3725–3747, 2021.
- [31] W. Zhang, X. Li, Y. Li, S. Wang, D. Li *et al.*, "Public sentiment drift analysis based on hierarchical variational auto-encoder," in *Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3762–3767, 2020. <http://dx.doi.org/10.18653/v1/2020.emnlp-main.307>.
- [32] M. M. A. Qudar and V. Mago, "TweetBERT: A pretrained language representation model for twitter text analysis," arXiv preprint arXiv: 2010.11091, 2020.
- [33] I. Beltagy, K. Lo and A. Cohan, "Scibert: A pre-trained language model for scientific text," arXiv preprint arXiv: 1903.10676, 2019.
- [34] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim *et al.*, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.
- [35] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," arXiv preprint arXiv: 1908.10063, 2019.
- [36] J. Hartmann, M. Heitmann, C. Siebert and C. Schamp, "More than a feeling: Accuracy and application of sentiment analysis," *International Journal of Research in Marketing*, in press, 2022. <https://doi.org/10.1016/j.ijresmar.2022.05.005>.
- [37] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi *et al.*, "Roberta: A robustly optimized bert pretraining approach," arXiv preprint arXiv: 1907.11692, 2019.