IJDI-ERET

INTERNATIONAL JOURNAL OF DARSHAN INSTITUTE ON ENGINEERING RESEARCH AND EMERGING TECHNOLOGIES
Vol. 12, No. 1, 2023
www.ijdieret.in

(Research Article)

A Model for Detecting Events from Twitter Data with Rough Accuracy

Mrs. K. Vasumathi[1*], Dr. S. Selvakani[2], J. Vellaiyan[3]

[1*]*Assistant Professor, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamilnadu, INDIA*
[2]*Assistant Professor and Head, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamilnadu, INDIA*
[3]*PG Scholar, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamilnadu, INDIA*

Abstract

The usage of mobile devices is growing every day, leading to a significant surge in SMS traffic. SMS, a text messaging service available on both regular and smartphones, has also resulted in an increase in spam texts. Spammers aim to send unsolicited messages to gain financial or commercial benefits, such as credit card details, market growth, lottery tickets, and more. Consequently, spam detection has become a priority, and our work focuses on detecting SMS spam using a combination of machine learning and deep learning techniques. To build our spam detection model, we utilized data from UCI, achieving an accuracy rate of 98.5%, which surpasses previous models. Our implementation was completed using Python.

Keywords: Event detection, Data models, Twitter, Computational modeling, Context modeling, Semantics.

1. Introduction

Over the course of five years, the number of smartphone users has increased from 1 billion to 3.8 billion, with China, India, and the United States ranking as the top three countries with the most mobile phone users. SMS, also known as Short Message Service, is an established text messaging service that does not require internet access and is compatible with both basic and smartphone devices. While there are several text messaging applications available on smartphones, such as WhatsApp, they are only accessible with internet connectivity. Conversely, SMS can be sent and received at any time.

The traffic for SMS service is steadily growing, and spammers send unsolicited messages to individuals. These spammers flood people with a high volume of messages to benefit their organizations or gain personal profit, which is known as spam. Although there are many SMS spam filtering solutions available, sophisticated methods are still necessary to address this issue. Spam messages on mobile devices can be frustrating, and SMS spam and email spam are two distinct types of spam. However, the terms "spam" and "SMS spam" are interchangeable and have the same meaning. Spammers employ these spam messages to promote their businesses or services.

[*]*Corresponding Author: e-mail: kulirmail@gmail.com, Tel-+91-9894726878*
ISSN 2320-7590 (Print) 2583-3863 (Online)

In this article, we conduct a multi-dimensional spatial pattern analysis of crime events that occur in San Francisco. Our analysis encompasses the influence of spatial resolution on hotspot identification, temporal effects in crime spatial patterns, and the correlation between different crime categories. We approach crime prediction as a classification problem in this study. To make predictions for a specific category, we establish a binary classification-based model, while a multiclass model is developed for analyzing all categories. The model is resilient when the parameters are modified. The results indicate that the proposed model outperforms the standard Deep Crime model in the majority of cases.

Spam emails can sometimes lead to financial losses for users. Machine Learning is a technology that enables machines to learn from past data and predict future data. Machine learning and deep learning techniques can be applied to tackle real-world problems across various fields, such as health, security, and market analysis. Supervised learning, unsupervised learning, and semi-supervised learning are some of the machine learning methods available. In supervised learning, the dataset includes output labels, while unsupervised learning deals with datasets without labels. To identify SMS spam, we employed multiple supervised learning techniques on a labeled dataset obtained from UCI.

There has been a growing body of research on machine perception of audio events, with a focus on detection and classification tasks. However, human-like perception of audio

scenes involves not only detecting and classifying sounds but also summarizing the relationship between different audio events. Although comparable research such as image captioning has been conducted, the audio field remains largely underdeveloped. Our model generates captions that are understandable and relevant to the data based on the dataset, with similar BLEU scores obtained for both languages.

The detection of events using social media data is a significant challenge. This study introduces a novel approach for detecting events by detecting patterns of herding in Twitter data. The analysis focuses solely on the metadata rather than the content of tweets. The approach is evaluated using a dataset of 3.3 million tweets collected by the authors, and the results are compared with those obtained using Twit info, a state-of-the-art method, on the same dataset. The generalizability of the method is demonstrated by testing it on a publicly available dataset of 1.28 million tweets.

The generator of an adversarial network is commonly employed to create an anomaly score based on the reconstruction loss of input for anomaly detection. Since anomalies are infrequent, training these networks can be challenging. Additionally, due to the involvement of adversarial training, this approach often exhibits instability, causing significant fluctuations in performance with each training step. Nevertheless, our model attained exceptional results on the UCSD Ped2 video dataset for anomaly detection, with a frame-level AUC of 98.1%, surpassing recent state-of-the-art methods.

Preserving the Contextual Relationship between Tweets: Collecting real-time Twitter data can lead to a vast amount of data. Despite the maximum length of a tweet being 280 characters and the most common length being 33 characters, generating the most relevant tweets remains a challenge due to the limited length. This research refers to this as uncertainty in capturing the contextual relationship among tweets. Additionally, processing Twitter data incurs a high computation cost, as the number of tweets related to a particular event can be massive.

Conducting a study on event detection techniques for social media data streams, given Twitter's growing popularity as a source of up-to-date news and information on current events. While previous approaches have demonstrated evidence of the quality of the detected events, none have correlated this performance with their run time performance in terms of processing speed, data throughput, or memory usage. To enable reproducible comparisons of run time performance, our approach utilizes a general-purpose data stream management system, and task-based performance is automatically evaluated using a series of novel measures.

In addition, we argue that the textual content authored by the message creator on social media platforms is a reliable and authentic source of information, providing valuable insights for detecting events such as natural disasters and terrorist activity in a timely manner. We aim to extend this research further by investigating this topic in detail.

To improve the efficiency of event detection, specific strategies can be employed such as integrating external knowledge bases like WordNet and utilizing unique characteristics of Twitter.

2. Review of Literature

Mengyue Wu, Heinrich Dinkel, and Kai Yu [6] have highlighted that the COVID-19 pandemic has forced us to alter our work practices since March. As people are reluctant to revert to the old ways of working, it is essential to cultivate and acquire a new set of skills that are necessary for the future of work. Over the next four weeks, I will be discussing the critical skills required for the future of work and providing tips on how to develop and implement them.

According to the research conducted by Mengyue Wu, Heinrich Dinkel, and Kai Yu [6], several advantages.

- Auditory perception were identified, including the temporal processing of information and the overlap of multiple sound events.
- To describe an audio scene, it is essential to evaluate the results on the development set. As audio-only information is insufficient for determining a sound event, video clips with sound were included.
- Each video was labeled by three raters to ensure dataset variance, and all raters received undergraduate education and were instructed to focus on the sound of the video while labeling.

In the paper, Nirmal Kumar Sivaraman, Vibhor Agarwal, and Yash Vekaria propose [7] a new approach for detecting events in social media data. The authors utilize metadata from Twitter data to detect traces of herding, without analyzing the actual content of the tweets. To evaluate their method, they use a dataset of 3.3 million tweets that they collected, and compare their results with a state-of-the-art method called Twit info. Additionally, the authors test their method on a publicly available dataset of 1.28 million tweets, which shows that their method can be applied more broadly. Overall, the authors demonstrate the effectiveness of their novel method for event detection in social media data.

According to Nirmal Kumar Sivaraman, Vibhor Agarwal, and Yash Vekaria [7], there are several benefits associated with their approach.

- In the literature, there is a significant body of research on event detection.
- Broadly speaking, event detection methods can be categorized into four types. One such type is term-interestingness-based approaches.

- In the field of event detection, there have been numerous approaches proposed, which can be classified into four main categories.
- Term-interestingness-based approaches, topic-modeling-based approaches, incremental-clustering-based approaches, and miscellaneous approaches.
- Term-interestingness-based approaches involve tracking terms that are likely to be associated with an event.
- In this section, we will focus on Herding, our specific formulation of THF, social synchrony, and the methodology we employ.

Jin-ha Lee, Marcella Astrid, and Seung-Ik Lee [4] explain that a common approach to anomaly detection involves using an adversarial network's generator to calculate an anomaly score based on the reconstruction loss of the input. However, because anomalies occur infrequently, training such networks can be a challenging task. Additionally, due to the use of adversarial training, this model may be unstable, causing performance to fluctuate significantly with each training step. Despite these challenges, the authors demonstrate that their model achieves superior performance on the UCSD Ped2 video dataset for anomaly detection, with a frame-level AUC of 98.1%, surpassing recent state-of-the-art methods.

According to Jin-ha Lee, Marcella Astrid, and Seung-Ik Lee [4], their approach offers several advantages. This paper describes an approach to anomaly detection using adversarial learning, which leverages both a generator (G) and a discriminator (D) to produce stable and robust results. When using a unified G and D model for anomaly detection, results can often be unstable due to the adversary. To align with conventional anomaly detection using generative networks, the authors propose a formulation that includes both good and bad quality reconstructions. Additionally, they introduce a pseudo-anomaly module, G, which generates fake anomaly examples from normal data.

During phase two training, the goal is to provide examples of both good and bad quality reconstructions. For good quality examples, D is provided with real data (X) and high-quality reconstructed data (Xˆ = G(X)). For bad quality examples, low-quality reconstructed data (Xˆlow) is generated using G. The pseudo-anomaly module, G, is also used to simulate reconstructed pseudo anomalies (Xˆpseudo).

In their article, Ren-De Li, Qiang Guo, Xue-Kui Zhang, and Jian-Guo Liu [8] discuss the importance of event detection in social media analysis, which often includes event detection and correlation. They then present their method for event detection.

The article describes a method for detecting sub-events and reconstructing their relationships using external expert knowledge. The process involves two main steps: using a Single Pass Clustering method to summarize social media posts and implementing a Label Propagation Algorithm to detect sub-events based on expert labeling. The resulting discussions and expressions are organized into topics that reflect the different sub-events from the user-generated content perspective. However, the classification of each post into the appropriate sub-event must be determined for effective correlation and evolution analyses. The method generates sub-event popularity curves that are verified against a null model with random labels. Overall, this approach provides insight into the popularity mechanism of unfolding events in online social media.

Ren-De Li, Qiang Guo, Xue-Kui Zhang, and Jian-Guo Liu [8] outlined several benefits.

- The WMD method is employed to establish correlations between posts within sub-events.
- The LPA results indicate that each label corresponds to a sub-event and contains multiple summarized posts. By utilizing the WMD method, pairs of summarized posts from different sub-events can be calculated.
- The WMD method measures the semantic distance between the two documents, each of which represents a summarized post.
- Our method's primary objective is to determine the prior probability and evolution probability to establish correlations and evolutionary trends.
- To evaluate the regenerated popularities, a comparison between the actual dynamic model and a random model is necessary for reference.

The analysis of social media has piqued the interest of researchers in event detection, as it has the potential to be highly significant in managing social crises. Yavari, H. Hassanpour, and M. Mahdavi [1] conducted a study that involves predicting events by analyzing Tweets. The categorization of Tweets is achieved using non-negative matrix factorization analysis and distance dependent Chinese restaurant process incremental clustering. The categorization results indicate that a cluster with a high rate of Tweets suggests the occurrence of a new event in the near future. Furthermore, a description of the event is presented through frequent words in each cluster. Notably, the content of Tweets about predictable events undergoes significant changes prior to their occurrence, which can be leveraged to predict events with a high degree of accuracy.

According to A. Yavari, H. Hassanpour, and M. Mahdavi [1], several advantages were outlined.

The following passage outlines the suggested approach for event prediction. In order to achieve precise predictions, it is necessary to identify effective features for prediction. These may include metrics such as new user sign-ups, the frequency of Tweets on the network, re-Tweets, mentions, hyperlinks, user responses, and the rate of messages sent by influential figures.

- This study utilizes variations in the frequency of message sending on social networks to demonstrate that events can be predicted by analyzing the rate of message sending prior to their occurrence. The focus is on utilizing fluctuations in the frequency of Tweet sending as a predictive feature.
- Variations in the number of Tweets sent during a given time interval or the frequency of Tweet sending can serve as the foundation for event prediction. The alterations in these metrics can be utilized to predict future events.
- There exist various incremental clustering methods that can be employed in the suggested approach. However, the reason for utilizing ddCRP clustering in the proposed method is twofold: firstly, it is founded on the well-known CRP method, and secondly, it relies on a single parameter that is simple to adjust.

Abhinay Pandya, Mourad Oussalah, and Ummul Fatima [2] assert that social media platforms play a vital role in capturing the spirit of society and serve as important social sensors. These platforms offer valuable opportunities to identify events like natural disasters and terrorist activities in a timely manner through real-time, online, and asynchronous message-sharing, supporting text, audio, video, and images. The authors contend that since the textual component of the message is authored by the sender, it is a more dependable, genuine, and credible source of information. The authors further advance this research by incorporating external knowledge bases such as WordNet and exploiting specific Twitter features for efficient event detection.

According to Abhinay Pandya, Mourad Oussalah, and Ummul Fatima [2], there are certain advantages associated with this approach.

Our system draws inspiration from current event detection techniques for identifying and tracking event-related segments. The system works by identifying significant phrases within individual tweets and constructing a temporal profile of these phrases to determine if their frequency is sufficiently high to indicate the occurrence of an event. Nevertheless, our system deviates from the current state-of-the-art in the many ways, our approach differs from the current state-of-the-art in the following ways:

- We leverage DBPedia to effectively identify noteworthy named entities.
- We use WordNet to aid in the identification of event-specific words and phrases.
- We broaden the scope of URLs embedded in tweets to locate important phrases in the titles of the web pages they reference.
- We utilize Twitter's metadata, such as 'quoted status', 'retweet', 'liked', 'in reply to', as well as the social network of the Twitter user.
- The remainder of the paper is organized as follows: Section discusses related work. Section outlines our system for detecting Twitter events. Section explains our experimental setup and findings, followed by a discussion. Lastly, Section presents our conclusions.

Ranjbar-Khadivi, Shahin Akbarpour, Mohammad-Reza [5] note that with the widespread usage of social networks, identifying the subjects being discussed within these networks has become a significant challenge. Present methods primarily rely on frequent pattern mining or semantic relations. Language structural techniques aim to discover the connection between words and how humans comprehend them. As such, this paper proposes a topic detection framework in social networks using the concept of the Imitation of the Mental Ability of Word Association, based on the Human Word Association method. The proposed approach was applied to a dataset and demonstrated superior performance in comparison to other topic detection methods.

According to Ranjbar-Khadivi, Shahin Akbarpour, Mohammad-Reza [5], some advantages of their proposed topic detection framework include:

The proposed algorithm incorporates all three methods of natural language processing, including the co-occurrence of keywords, the semantic relationship between words, and the human understanding of word association, with an emphasis on linguistic structures. The flowchart of the proposed algorithm is presented below. The data from social networks is processed for topic extraction using the following four steps:

- Receiving the social network data stream, windowing, and preprocessing.
- Keyword ranking, co-occurrence word embedding calculation, and pattern vector extraction.
- Pattern distance calculation and pattern clustering.
- Cluster ranking and topic extraction.

According to Roshni Chakraborty, Maitry Bhavsar, Sourav Dandapat, and Joydeep [9], Twitter has gained immense popularity as a means of sharing information and expressing opinions on social media platforms. The diverse range of user opinions available on Twitter can be utilized by news media outlets to gauge user reactions towards different news articles. While unsupervised techniques may struggle to handle the variability and noise in the features of these articles, the effectiveness of the summary tweets extracted with regards to specific news articles can still be validated.

According to Roshni Chakraborty, Maitry Bhavsar, Sourav Dandapat, and Joydeep [9], certain advantages were identified.

Next, we will describe the additional preprocessing steps used to filter out opinion tweets related to a news article:

- Keyword Extraction: Firstly, we eliminate duplicate tweets and subsequently remove user mentions, retweet tags, URLs, and stop words from the tweet text. We extract nouns and verbs from the remaining words using a POS tagger, which forms the representative keywords of the tweet.
- Seed Tweet Set Creation: We create an initial seed tweet set of an article that includes tweets having the collective set of seed and co-occurring hashtags, which is known as a related hashtag set of the article.
- Extended Tweet Set: We expand the seed tweet set by including relevant opinion tweets that may not have direct word overlap with the news article. However, we can often find word overlap between these tweets and other tweets that do have direct word overlap with the news article. This step enables us to extend our relevant tweet set and eliminate irrelevant tweets.
- Gaurav Hajela, Meenu Chawla, and Akhtar Rasool [3] note that Twitter's real-time capabilities and asynchronous message-sharing make it an excellent tool for timely event detection. However, existing event detection approaches have primarily centered on constructing a temporal profile.

This article presents a multidimensional approach to event detection that involves identifying named entities and detecting significant bursts in their usage as an indication of an event. By leveraging temporal, social, and Twitter-specific features, our system shows enhanced precision, recall, and DERate compared to state-of-the-art approaches, as demonstrated on the benchmarked Events2012 corpus.

According to Gaurav Hajela, Meenu Chawla, and Akhtar Rasool [3], certain advantages were identified. In this study, a model was developed using a dataset from Maggot. The training dataset comprises various attributes, each with a distinct association. The training dataset includes incidents from Cagle on San Francisco crimes. The data spans from January 2003 to May 2015. The dataset consists of nearly 12 years of criminal reports from San Francisco, classified into various crime types. The dataset was used to assess the accuracy of classification techniques with new unclassified data.

Andreas Weiler, Michael Grossniklaus, and Marc H. Scholl [10] explain that a common approach for detecting anomalies involves utilizing the generator of an adversarial network to generate an anomaly score based on the reconstruction loss of the input. However, since anomalies are infrequent, training such networks can be a challenging task. Additionally, our model outperforms recent state-of-the-art methods by achieving a frame-level AUC of 98.1% on the UCSD Ped2 video dataset for anomaly detection. According to Andreas Weiler, Michael Grossniklaus, and Marc H. Scholl [10], certain advantages were identified. The growing use of Twitter for obtaining current news and information on current events has led to the development of various research studies on event detection techniques for social media data streams. Although each proposed approach offers some indication of the detected events' quality, none establish a connection between the task-based performance and their runtime performance in terms of processing speed, data throughput, or memory usage. Additionally, this step allows for the extension of the relevant tweet set while filtering out irrelevant tweets.

This article's specific contributions are as follows:

A thorough examination of the task-based and runtime performance of established event detection techniques.

- To enable systematic performance studies of novel event detection techniques in the future, our approach is based on a platform-based approach that utilizes a general-purpose data stream management system for reproducible comparisons of runtime performance.

3. Modules

The system is divided into the following modules:

- Dataset collection
- Dataset preprocessing
- Feature extraction
- Evaluation model

This project consists of four modules: Dataset Collection, Dataset Preprocessing, Feature Extraction, and Evaluation Model.

Dataset Collection involves selecting a subset of product reviews collected from records. In Machine Learning (ML), it is essential to start with a large amount of labelled data, which is data with a known target answer. The next step is Dataset Preprocessing, which includes three common procedures: Formatting, Cleaning, and Sampling.

- Formatting is required when data is in an unsuitable format, such as a relational database.
- Cleaning involves removing or fixing missing data and anonymizing sensitive attributes.
- Sampling is necessary when there is too much data, as more data means longer.
- Running times for algorithms and greater computational and memory requirements.

Feature Extraction is an attribute reduction process that transforms the attributes. In this project, we use the classify module on the Natural Language Toolkit library in Python to train our models using a Classifier algorithm Overall, the four modules enable us to work with a representative sample of the selected data and train our models effectively. Labelled dataset gathered. The rest of our labelled data will be used to evaluate

the models. Some machine learning algorithms were used to classify pre-processed data.

The Evaluation Model is a crucial aspect of the model development process as it helps to determine the most suitable model that accurately represents the data and performs well in the future. However, evaluating model performance solely with the training data can result in over fitted and over-optimistic models, which is unacceptable in data science. To address this issue, there are two common methods for evaluating models: Hold-Out and Cross-Validation.

4. Proposed Methodology

The algorithms utilized in this project are Linear Regression, Random Forest Regression, and Decision Tree Regression.

The following steps were taken to train and predict new values:

- The training dataset containing wholesale price index was initialized.
- All rows and the first column of the dataset were selected and assigned to "x" as the independent variable.
- All rows and the second column of the dataset were selected and assigned to "y" as the dependent variable.
- DTR/SVR/LR was fitted to the dataset.
- The new value was predicted.
- The result was visualized and accuracy was checked.

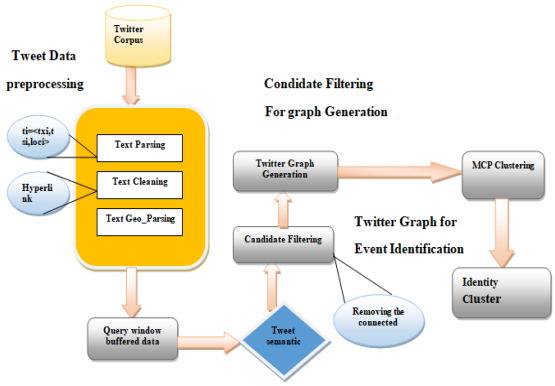

Figure 1. System architecture

4.1 Experimental results: Figure depicts a block diagram that illustrates the overall process. Initially, the tweets undergo preprocessing by means of tweet parsing and cleaning. Afterwards, JoSE is employed to generate feature representations for the tweets, and a Twitter graph is constructed in which the relationships among the tweets are modeled as uncertain and represented using similarity scores.

4.2 Dataset pre-processing: To enhance the quality of data for further processing, it's necessary to pre-process the raw Twitter data, which may have numerous impurities that can create interference during the processing stage. The following methods are employed to pre-process the dataset.

- During Tweet Parsing, the most pertinent data, such as text and creation time, is extracted from the tweets, resulting in a tuple of these features representing the Twitter corpus. Each tweet can be represented as ith tweet (ti), tweet text (tx), and timestamp (ts).
- In Tweet cleaning, all stop words, hyperlinks, recurring characters, and other non-decodable information is removed from the text of the tweets.
- The text is then analyzed using Part-of-Speech (POS) tagging and Named Entity Recognition (NER) detection to identify important information. After cleaning, the tweets are processed further to generate word embedding's. The Twitter dataset contains two main feature sets, namely tweet-level features (retweets, replies, etc.) and user-level features number of followers, number of friends, etc.

5. Feature Extraction

The accuracy of identified events through Twitter data heavily relies on the co-occurring textual information gathered from the data. Typically, the feature set vectors of textual data are extensive. This is illustrated in Figure 2.

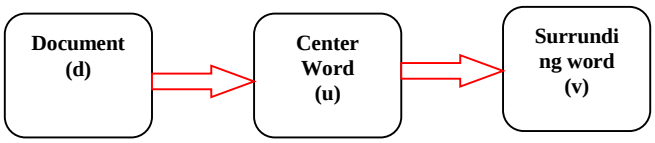

Figure 2. Twitter data

The present study revolves around identifying events from Twitter data, which consists of short-text sentences, making it more challenging to capture semantics compared to document-based datasets. Therefore, JoSE's performance is evaluated by comparing it with other state-of-the-art word embedding's models such as Word2Vec and Doc2Vec, using two labeled Twitter datasets.

5.1 Twitter graph generation: Graph-based data models are renowned for their ability to capture the relationships within data. Graphs prioritize connections over data points, meaning that strong relationships within the graph signify strong correlations or associations among data points or nodes. Events may experience bursts at different time frames depending on their scale. The proposed method for generating the Twitter graph does not employ the query window framework.

In the query window framework, both the window size and window shift width must be assigned to create a Twitter edge combination.

5.2 Component based graph filtration: The resulting Twitter graph comprises numerous disjoint components or subgraphs, as illustrated in Figure 3. As such, the Twitter graph functions as a transactional graph, where each component may represent a distinct sub-event. The size of a connected component is determined by the number of vertices in the graph. Analyzing the component sizes and respective frequencies reveals that smaller components occur more frequently, while the frequency decreases as component size increases, as shown in Figure 3.

In this section, we present the experimental results of the proposed model on a series of benchmark and real-time Twitter datasets. Initially, we provide the specifics of both the benchmark and real-time datasets.

5.3 Dataset description: The performance of the proposed model is evaluated using various types of Twitter datasets. Table 1 provides a summary of the results.

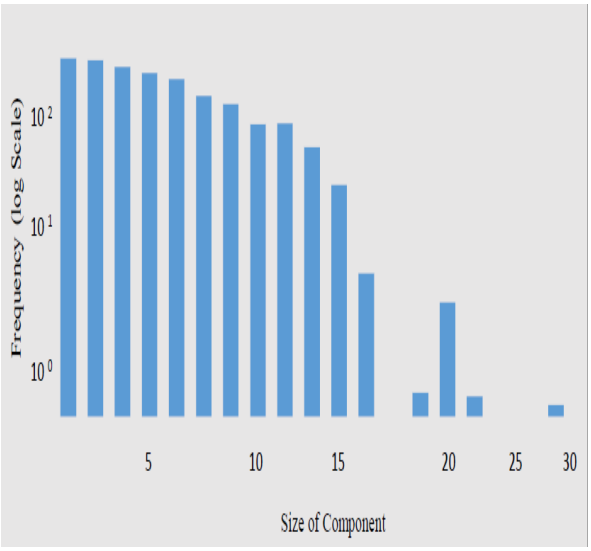

Figure 3. Frequency graph of Twitter

Table 1. Twitter datasets

Sr. No	Dataset	Number of tweets	Number of nodes	Number of components	Size of biggest component
1	RepLab 2013 Twitter data	2481	2389	785	52
2	#Auspol Twitter data	21065	12874	1601	200
3	Common disease dataset	1.085 million	195332	19434	74
4	Customer support on Twitter dataset	3.196million	601271	172742	40

Table 1 presents statistics for all the datasets used in evaluating the proposed model.

- Public Annotated Data: Twitter data from RepLab 2013
- Public Annotated Data: Twitter data containing the hashtag #Auspol
- Real-World Data: Dataset featuring common diseases
- Customer Support on Twitter Dataset is measured qualitatively using Precision, Recall, and the F1 score. Recall is computed using the equation specified. However, the model requires specifying the number of clusters to be identified, k, in advance. Additionally, since the value of k can be adjusted, the number of relevant events and retrieved events can be made identical.

$$F1 = 2 . \frac{Percision.Recall}{Percision+Recall} \quad (1)$$

6. Conclusion

The objective of the study is to predict whether a message is spam or not by utilizing efficient machine learning algorithms and technologies, achieving high accuracy. The training datasets obtained provide sufficient insights to predict the appropriate classification of messages. The system assists users in identifying whether their messages are spam or legitimate messages with a high level of accuracy.

References

[1] A.Yavari, H. Hassanpour, and M. Mahdavi, "Event Prediction in Social Network through Twitter Messages Analysis", *AAAI Conference on Weblogs and Social Media*, 2011.

[2] Abhinay Pandya, Mourad Oussalah, and Ummul Fatima, "Metadata-Assisted Twitter Event Detection System", *AAAI Conference on Weblogs and Social Media*, 2019.

[3] Gaurav Hajela, Meenu Chawla, and Akhtar Rasool, "Multi-Dimensional Crime Spatial Pattern Analysis and Prediction Model Based on Classification", 2020.

[4] Jin-ha Lee, Marcella Astrid, and Seung-Ik Lee, "Redefining the Adversarial Learned One-Class Classifier Training Paradigm", *Intelligent Systems*, 2019.

[5] Mehrdad Ranjbar-Khadivi, Shahin Akbarpour, and Mohammad-Reza, "A Human Word Association based model for topic detection in social networks", *Expert Systems with Applications*, 2019.

[6] Mengyue Wu, Heinrich Dinkel, and Kai Yu, "Audio Caption, Listen and Tell, Audio, Speech, and Language Processing", 2019.

[7] Nirmal Kumar Sivaraman, Vibhor Agarwal, Yash Vekaria, and Sakthi Balan Muthiah, "A Metadata-Based Event Detection Method Using Temporal

Herding Factor and Social Synchrony on Twitter Data", *Information Processing and Management*, 2019.

[8] Ren-De Li, Qiang Guo, and Jian-Guo Liu, "Reconstruction of Unfolding Sub-Events from Social Media Posts", *Springer*, 2018.

[9] Roshni Chakraborty, Maitry Bhavsar, Sourav Dandapat, and Joydeep Chandra, "A Network Based Stratification Approach for Summarizing Relevant Comment Tweets of news articles", *AAAI Conference on Weblogs and Social Media*, 2010.

[10] Van Quan Nguyen, Tien Nguyen Anh, and Hyung-Jeong Yang, "Real-Time Event Detection Using Recurrent Neural Network In Social Sensors", *Future Generation Computer Systems*, 2017.

Biographical notes

Mrs. K Vasumathi Working as an Assistant Professor Department of Computer Science / Computer Applications at Government Arts and Science College, Arakkonam (Formerly Thiruvalluvar University College of Arts and Science). She Completed S.E.T (Computer Science), M.Phil (Computer Science), M.Sc (university VI Rank) at Thiruvalluvar University, Vellore, Tamilnadu, India. She is a dynamic hardworking professional with rich experience of 13 years in Teaching. She is having hands on experience in Research contributions and Student Counseling. She has published more than 18 papers in the peer reviewed International Journals. Besides presented more than 35 papers in the National and International Conferences. She has published 2 books. Her interest areas are Operating Sysyem, Big Data, Machine Learning**.**

Dr. S. Selvakani Working as an Assistant Professor and Head in the Department of Computer Science / Computer Applications at Government Arts and Science College, Arakkonam (Formerly Thiruvalluvar University College of Arts and Science). She Completed Ph.D (Computer Science), M.Tech (Computer Science and Engg), M.Phil (Computer Science), MCA (university III Rank) at Manonmanium Sundaranar University, Tirunelveli, Tamilnadu, India. She is a dynamic hardworking professional with rich experience of 20 years in Teaching, Administration/ Training & Development. She is having hands on experience in general administrative activities, Research contributions and Student Counseling. Cambridge University has recognized and awarded certification of "Teachers and Trainers" for remarkable achievement in the field of Teaching. She has published more than 48 papers in the peer reviewed International Journals and more than 6 papers in National Journals. Besides presented more than 90 papers in the National and International Conferences and guiding several Ph.D Scholars in Computer Science presented more than 112 papers in National and International Journals, Conference and Symposiums. She has published 4 books and 6 Patents. Her interest areas are Data Science, Big Data, Machine Learning and Network Security.

J. Vellaiyan has received his Bachelor's Degree from Thiruvalluvar University College of Arts and Science and doing his Masters in Government Arts and Science College, Arakkonam. His research interests are Data Science, Big Data and Network Security.