I.J. Modern Education and Computer Science, **2023, 1, 58-72**
Published Online on February 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2023.01.05

Automatic Cyberstalking Detection on Twitter in Real-Time using Hybrid Approach

Arvind Kumar Gautam*
Department of Computer Science, Indira Gandhi National Tribal University, Amarkantak, MP, 484886, India
Email: analyst.igntu@gmail.com
ORCID ID: https://orcid.org/0000-0001-6057-1006
*Corresponding Author

Abhishek Bansal
Department of Computer Science, Indira Gandhi National Tribal University, Amarkantak, MP, 484886, India
Email: abhishek.bansal@igntu.ac.in
ORCID ID: https://orcid.org/0000-0001-5968-3625

Received: 03 January, 2022; Revised: 19 March, 2022; Accepted: 19 June, 2022; Published: 08 February, 2023

Abstract: Many people are using Twitter for thought expression and information sharing in real-time. Twitter is one of the trendiest social media applications that cybercriminals also widely use to harass the victim in the form of cyberstalking. Cyberstalkers target the victim through sexism, racism, offensive language, hate language, trolling, and fake accounts on Twitter. This paper proposed a framework for automatic cyberstalking detection on Twitter in real-time using the hybrid approach. Initially, experimental works were performed on recent unlabeled tweets collected through Twitter API using three different methods: lexicon-based, machine learning, and hybrid approach. The TF-IDF feature extraction method was used with all the applied methods to obtain the feature vectors from the tweets. The lexicon-based process produced maximum accuracy of 91.1%, and the machine learning approach achieved maximum accuracy of 92.4%. In comparison, the hybrid approach achieved the highest accuracy of 95.8% for classifying unlabeled tweets fetched through Twitter API. The machine learning approach performed better than the lexicon-based, while the performance of the proposed hybrid approach was outstanding. The hybrid method with a different approach was again applied to classify and label the live tweets collected by Twitter Streaming in real-time. Once again, the hybrid approach provided the outstanding result as expected, with an accuracy of 94.2%, recall of 94.1%, the precision of 94.6%, f-score of 94.1%, and the best AUC of 98%. The performance of machine learning classifiers was measured in each dataset labeled by all three methods. Experimental results in this study show that the proposed hybrid approach performed better than other implemented approaches in both recent and live tweets classification. The performance of SVM was better than other machine learning algorithms with all applied approaches.

Index Terms: Cyberstalking Detection, Cyberbullying, Machine Learning, Lexicon, TF-IDF, Support Vector Machine, Naive Bayes, Sentiment Analysis, Feature Extraction, Twitter.

1. Introduction

Twitter is a real-time social media application that has gained global popularity in the virtual world. As per statistics [1], More than 300 million worldwide users use Twitter, and more than 500 million daily posts are tweeted on Twitter. Twitter is a great way to remain socially connected to family, friends, and colleagues to share the tweets for individual, official, and business reasons [2]. The use of Twitter also raises challenging issues in the form of cyberstalking, cyberbullying, and other cyber harassment. Cyberstalking is a dangerous and convoluted cybercrime that affects and targets numerous people, communities, and organizations [3]. Cyberstalkers and gangs of cyberstalkers are active on Twitter with pre-defined plans and agendas to insults, profanity, harassing the victim through repeated activities of sexism, racism, offensive, abuse, hate, trolling, fake news, and fake accounts [4, 5, 6]. Impressive cyberstalking detection, controlling, and counteraction arrangements are required to handle this troublesome cyberstalking circumstance on Twitter.

Researchers widely use lexicon-based and machine learning techniques for cyberstalking detection with sentiment analysis support [7-9]. Sentiment analysis performs an imperative task in text analysis and deciding the score of

words classified as positive or negative comments [10]. The lexicon-based approach [11, 12] uses a pre-defined and pre-trained rule-based dictionary of good and bad words to determine the score of any word and assign a positive or negative sentiment. The main limitation of the lexicon-based approach is that sentiment polarity scores can not be specified to those words which are not in the dictionary. In other cases, machine learning techniques for sentiment analysis are not dependent on any pre-defined dictionary [13]. In the machine learning methodology, the detection model is initially trained using the labeled dataset to predict the probability of words for positive or negative sentiment [14].

A more improved detection model is required to enhance the performance of cyberstalking detection on Twitter in real-time. There is still much scope for comparative analysis of lexicon-based cyberstalking detection and machine learning-based cyberstalking detection to determine and design a better approach. The main research objective of this paper is to analyze and compare the different methods of automated cyberstalking detection on Twitter and propose a better approach to enhance the performance. Initially, this paper applied both lexicon-based and machine learning approaches separately. Finally, to combine the benefits of both methods, this paper implemented an automated hybrid approach to detect cyberstalking tweets on Twitter in real-time. The significant contributions from this study are as follows.

- We performed the comparative analysis of lexicon-based, machine learning, and hybrid approaches for cyberstalking detection in recent tweets collected directly through Twitter API.
- We proposed a hybrid approach for automatic cyberstalking detection on live tweets directly fetched through Twitter streaming in real-time. The proposed hybrid method can classify and label live tweets in real-time with high accuracy.
- The proposed approach can also be used in other social media platforms that provide live comments through API.

The proposed hybrid approach was applied with recent tweets (collected through Twitter API in real-time) and live tweets (collected through Twitter Streaming in real-time). Initially, the proposed hybrid approach was trained with a labeled dataset and then auto-trained through classified tweets. With both recent and live tweets, the proposed hybrid approach performed better than traditional lexicon-based and machine learning techniques. The subsequent part of the research study is structured section-wise. In section 2, the notable and recent contribution of researchers in the related field is presented in the form of a literature review. In section 3, applied materials and the proposed methodology used in this paper are described. The experimental setup, results, and detailed discussion are mentioned in section 4. Finally, the conclusion and future works are finalized in section 5.

2. Review of Literature

In the literature survey, some related research papers were chosen to observe the contributions of past work performed by researchers to the automatic detection of cyberbullying, cyberstalking, and other cyberharassment. Ghasem et al. [15] suggested a model for automatically detecting and controlling cyberbullying and cyberstalking using machine learning techniques. This approach was generally focused on automatic email-based cyber-stalking detection as well as evidence documentation to combat cybercriminals. Frommholz et al. [16] suggested a textual analysis-based cyberstalking detection model using machine learning algorithms. The proposed method of authors was mainly focused on author identification, text classification, personalization, and digital text forensics. Saravanaraj et al. [17] implemented an automated model for detecting cyberbullying tweets on Twitter using supervised machine learning techniques. The authors used Random Forest and Naïve Bayes algorithms to classify tweets and found adequate results with their experiment. Another machine learning-based automated cyberbullying detection model was developed by Zhang et al. [18] to detect the bully tweets on Twitter. The authors performed the experimental work using various machine learning models using multiple textual features and found maximum accuracy of 90%. Liew et al. [19] suggested an automated security alert model using supervised machine learning techniques to detect and control phishing tweets in real-time on Twitter. The authors implemented their proposed model using random forest and found better accuracy. Balakrishnan et al. [20] utilize the user's psychological personalities, sentiments, and emotions to design a cyberbullying detection model on Twitter. The author used the machine learning technique to filter and categorize the tweets into bully tweets, aggressor tweets, spammer tweets, and regular tweets. Shah et al. [21] have also designed a machine learning-based framework for automatically detecting cyberbullying tweets on Twitter. The author implemented their proposed approach using several machine learning algorithms and found the maximum accuracy of 93% for logistic regression.

Kazim Raza et al. [22] applied a lexicon-based methodology to detect cyberbullying tweets automatically tweeted in Roman-Urdu language on Twitter. With their proposed approach, the authors found better results than previous work of researchers. Another model using text analysis features with lexicon-based offered by Geetha et al. [23] for automatic detection of offensive language on Twitter. The authors used LIWC, POS, and Twitter Tag Scores (TTS) for lexicon-based text analysis and implemented the model with deep learning and machine learning. The authors achieved 91.72% accuracy for the C-LSTM method while 90.8% accuracy for logistic regression and SVM. Another machine-learning-

based methodology was suggested by Bandi Yoshna et al. [24] to detect cyberbullying on Twitter. The authors tested their model using random forest and SVM algorithms and successfully obtained an accuracy of 71.2% for the support vector machine. Real-time cyberbullying detection on Twitter for Hindi-English mixed tweets was suggested by Kumar Akshi et al. [25] with the support of transfer learning and deep neural networks. The author's model converted the tweets in Hindi and mixed language into English and automatically classified the tweets. Yuvaraj et al. [26] applied deep decision tree classification with multi-feature-based AI for their proposed automatic cyberbullying detection model on Twitter. In experimental work, authors classified and labeled the 30,384 tweets using the deep decision tree classification method. Another detection model based on deep neural networks was implemented by Sadiq et al. [27] for the automatic detection of aggression tweets on Twitter. The authors performed the experimental work with multilayer perceptron methods using CNN-LSTM and CNN-BiLSTM methods to classify aggression tweets and found expected results with an accuracy of 92%.

Sangwan et al. [28] designed a filter-wrapper-based hybrid model for automatic detection of cyberbullying on Twitter and Instagram. After implementing the hybrid detection model using the lexicon-based method and machine learning model, the authors found better results. Lepe-faúndez et al. [29] proposed a model for automatic detection of cyberbullying in the Spanish language on Twitter using a hybrid method. The authors evaluated their hybrid model using lexicon-based and machine learning methods and found a maximum of 89.2% of accuracy. Madan et al. [30] suggested a real-time sentiment analysis model using lexicon-based, machine learning-based, and hybrid methods for tweets in the Hindi language on Twitter. Another hybrid model was proposed by Almutairi et al. [31] for the automatic detection of cyberbullying in tweets in the Arabic language. The authors implemented their proposed approach using lexicon-based with machine learning and obtained 82% accuracy. Arora et al. [32] proposed a novel methodology for automatically detecting cyber harassment on Twitter using a mixed-methods approach. Authors performed the experimental work using lexicon-based and SVM to classify cyber harassment into spam, hateful, abusive, and neutral tweets. Ayo et al. [33] successfully implemented a clustering model for automatic hate speech detection on Twitter. The authors applied the rule-based clustering method and fuzzy logic for automatically classifying tweets and hate speech detection, respectively in real-time. The authors achieved 96.4 % of AUC and 94.5% f-score.

In the literature, authors at [22, 23] applied a lexicon-based approach, while authors at [15-21, 24-27] have implemented the detection model using machine learning techniques. Authors at [29-33] have also suggested some hybrid approaches, including lexicon-based and machine learning techniques for automatic detection. The majority of researchers applied machine learning techniques for automatic tweets classification. Automatic cyberstalking detection on Twitter and other social media networks in a real-time manner is still a challenging task. There is still a lack of automated cyberstalking detection approaches in real-time, with an impressive performance.

3. Material and Methodology

This section describes the detailed algorithms used for designing the proposed model. In Fig. 1, the basic functioning layout of the proposed automatic detection model is explained. The proposed automated model consists of the following main phases for real-time cyberstalking detection on Twitter.

1. Tweets Collection and Making the Dataset
2. Tweets pre-processing
3. Features extraction
4. Classification and Labeling of the Tweets
5. Real-Time Cyberstalking Detection on Live Tweets.
6. Measuring the Performance of Model

The proposed methodology was implemented on recent and live tweets both. After fetching the recent tweets using Twitter API, a lexicon-based approach was initially applied for tweets classification. Machine learning classifiers were trained using the pre-defined dataset, and after that, machine learning and hybrid approaches were both applied separately to the same recent tweets. Finally, the hybrid approach is applied again for tweets classification on live tweets in a real-time manner. The detailed procedure for each applied approach shown in Fig. 1 is explained as follows.

3.1. Tweets Collection and Making the Dataset

In the initial stage, this paper used Twitter API to collect the recent tweets and make the dataset while live tweets were fetched during the real-time cyberstalking detection. Several hashtags keywords regarding cyberstalking, cyberbullying, cyber harassment, and cybercrimes were used to collect the recent tweets from Twitter. The following steps were used for collecting tweets and making the dataset.

Procedure for Tweets Collection and Making the Dataset

Step:1. Logged in to a Twitter developer account, registered Twitter API, and obtained the required authentication keys and tokens by creating a new application or existing application on a Twitter developer account. Twitter generally provides four authentication keys and tokens, namely "consumer-key," "consumer-secret," "access-token," and "access-token-secret," for fetching the tweets from Twitter.

Step:2. Required libraries (Tweepy in python) were imported, and the Twitter API key was authenticated. After that, several related hashtags keywords were defined to fetch the tweets. Such as #harassment, #cyberstalking, #cyberbullying, #stalker, #stalking, #cyberharassment, #revengeporn, #sexy, #hate, #troll, #hate speech, #sexism, #racism, #cybercrime, #hacking, #abuse, #victim, #love, #onlinesafety, #bullyingsucks, #thebullyexposed, #internetsafety etc.

Step:3. Fetched the recent tweets from Twitter based on hashtags, time intervals, and user profiles and finally saved them to text and CSV file. This paper collected tweets with the user name, user id, tweets location, retweet count, follower count, and tweets date. Some tweets were also collected from the timeline of the suspicious user profile as per a pre-defined small dataset.

Step:4. Step 3 was repeated until the collection of a sufficient number of tweets. In the first phase, more than 8000 tweets were collected on several attempts. All collected tweets were saved into dataset D2.

Step:5. A mixed labeled training dataset D1 (classified as cyberstalking and non-cyberstalking text) containing 35734 unique records was prepared separately to train the machine learning classifiers. So that a trained machine learning model can predict the probability of the collected tweets. This pre-defined training dataset contains tweets and comments from different sources of the internet world. Further, this labeled dataset was automatically updated through classified live tweets using the proposed model.

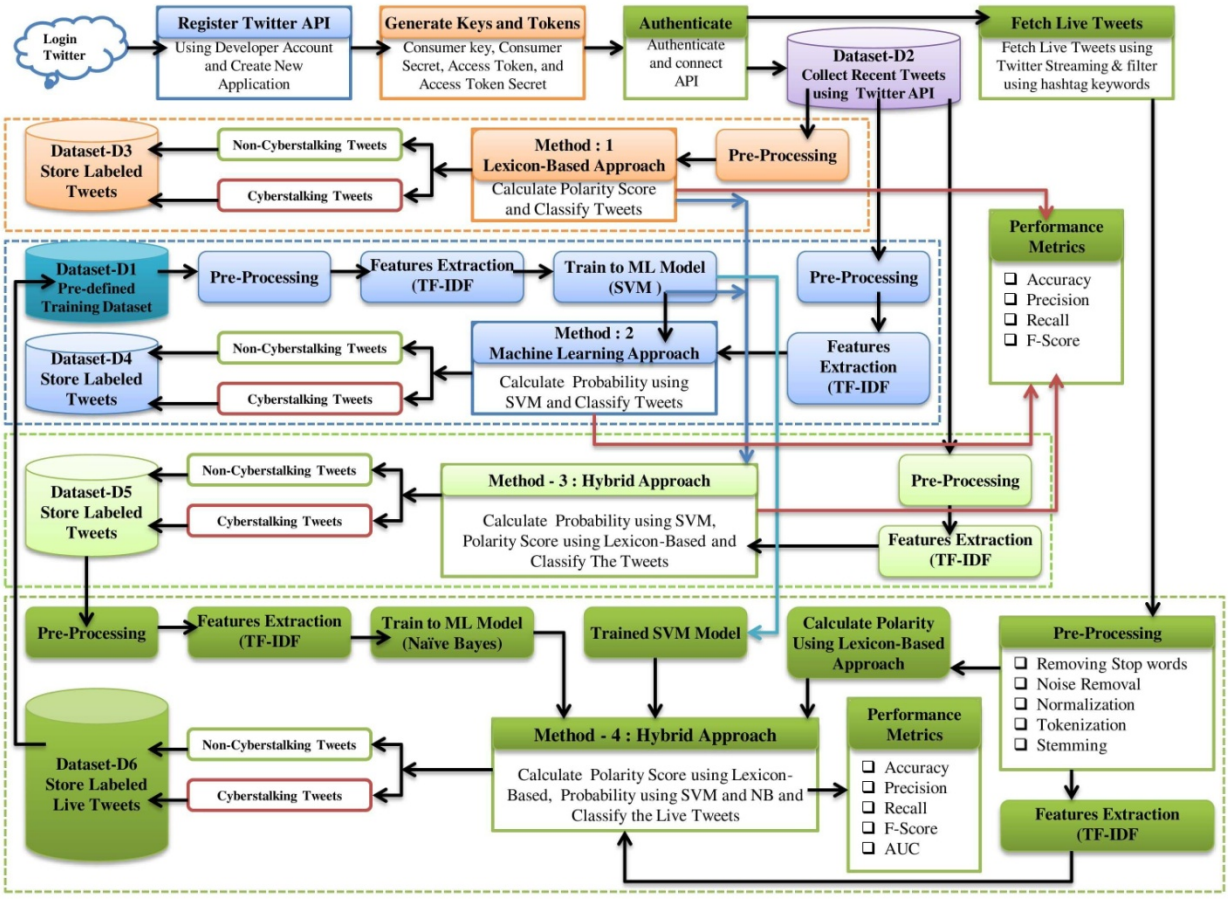

Fig. 1. The basic layout of the proposed automatic model for real-time cyberstalking detection on Twitter

3.2 Tweets Pre-processing

The collected tweets from Twitter API contain raw text with unnecessary characters, blank spaces, blank lines, meaningless characters, and different symbols. Properly cleaning the tweets is highly required before feature extraction and classification of tweets. In this phase, collected tweets were cleaned, filtered, and normalized into proper format. This paper performed several pre-processing tasks: Removing stop words, Noise removal, Tokenization, Normalization, and Stemming. In the first step of pre-processing, all stop words were removed from the tweets. Meaningless words such as articles, prepositions, and pronouns that are not useful for sentiment analysis and tweet classification are called

stop words [34]. Collected tweets from Twitter also contain different noise data, which were removed. In tweets, repeated words, symbols (such as @,#, etc.), blank lines, blank spaces, special characters (such as RT, etc.), URLs, punctuation marks, and any useless digits are called noise data [35]. After removing the noise data and stop words, the texts of the tweets were divided into individual words and added to a separate list. This process for splitting the sentence into words is called tokenization [35]. Further, tokenized tweets were converted to lower case letters using normalization [35] to make the uniformity. After that, tokenized words are required to be restored to their original form using the lemmatization [35] and stemming [36] methods. Lemmatization may be used instead of stemming for proper morphological analysis of the words. Lemmatization is a method to combine the synonyms relation words into a single word and remove all other concerned synonyms words from the list [37]. In this paper, the stemming method was used.

3.3 Feature Extraction

After performing the pre-processing tasks, the tweets dataset was ready for classification and labeling using the lexicon-based approach. In contrast, the machine learning model uses feature vectors to estimate the predicted probability of cleaned tweets. Feature extraction is essential in the machine learning-based process before classifying tweets because the machine learning algorithms work on feature vectors and can not understand tweets as text forms. Feature extraction computes the weights of tweet words and creates a feature vector in numerical form. Feature extractions play a crucial role in improving the performance of classifiers [38]. Several traditional-based, word embedding-based and language model-based feature extraction methods are available for feature extraction in the word-level, sentence-level, and n-gram levels [38]. TF-IDF, Word2Vec, BOW, BERT, FastText, GloVe, XL-NET ,ELECTRA, InferSent, GPT-2, and Universal Sentence Encoder are some widely used examples of feature extraction methods [39-43]. The proposed detection model of this study applied TF-IDF methods for feature extractions. TF-IDF is an efficient calculation-based feature extraction method that measures the weight of any word of documents in a collection of documents [44]. TF-IDF finds most occurring words and assigns more consequences because regularly occurring words are more important for the classification [45]. Equation (1) is used to calculate the feature vector in the TF-IDF.

$$TF-IDF\left(T,D\right)=\frac{\sum T\,in\,D}{\sum W\,in\,D}\times Log\left(\frac{N}{\left(\sum T\,in\,N\right)+1}\right) \tag{1}$$

Where:

$$\left.\begin{array}{c}\sum T\ in\ D = Number\ of\ \ times\ word\ T\ appears \\ in\ a\ document\ "D \\ \sum W\ in\ D = Total\ number\ of\ words\ in\ \ the\ document\ "D"\end{array}\right\} \rightarrow R\ epresents\ the\ Term\ Frequency$$

$\sum T\,in\,N$ = {Total occurrence of Word “T” in total documents} → Represents the Document Frequency

N= Total Documents

3.4 Classification and Labeling of the Tweets

In this phase, collected tweets through Twitter API were classified into cyberstalking tweets and non-cyberstalking tweets using different methods, as explained in Fig. 1. Recent tweets directly collected through Twitter API were classified in the primary detection phase. The lexicon-based method was applied in the first approach, and labeled tweets were saved in a separate dataset. After that, in the second approach, a machine learning technique with a trained SVM model was applied to classify the same tweets, and labeled tweets were saved in a separate dataset. In the third approach, a hybrid approach was implemented using the lexicon-based polarity and SVM-based probability to classify the same tweets, and labeled tweets were saved in another dataset. Finally, another hybrid approach using polarity score through lexicon-based, probability score through trained SVM, and Naïve Bayes was applied for automated cyberstalking detection on Twitter in a real-time manner during the fetching of live tweets. The detailed procedure of each approach is explained in the subsection as follows.

3.4.1 Lexicon-based approach for classification and labeling of the tweets

Lexicon represents the vocabulary of any word, person, and language. The lexicon-based is an admired method for sentiment analysis that uses a dictionary and rules to assign a positive or negative score to a word. The lexicon-based process uses pre-papered sentiment to give a score to the words. The lexicon-based method uses different techniques, namely dictionary-based and context-based lexicon, to produce the polarity score [46, 47]. The dictionary-based lexicon [48] uses a pre-defined word dictionary of good and bad words updated using synonyms and antonyms. Context-based lexicon [49] uses semantics and statistical methods to find the context-specific sentiment. The semantic approach finds the synonyms and antonyms of the word and semantically closer words for assigning the sentiment value. The statistical

technique of the lexicon-based process finds positive and negative words in a positive and negative context. Suppose words behave irregularly in a positive context. In that case, positive polarity is assigned, while in other cases, if word behavior returns negative in a negative context, then negative polarity is assigned. Neutral polarity is given in case of equal occurrence of a positive and negative word. Several pre-defined and pre-trained lexicon-based libraries are widely used, namely TextBlob, Vader, SentiWordNet, and AFINN [50]. This paper used the TextBlob library as a Lexicon-based approach for classifying and labeling tweets. TextBlob computes the sentiment and returns polarity within the range of [-1.0 to 1.0] and subjectivity within the range of [0.0 to 1.0]. Equation (2) is used to calculate the polarity of the tweet.

$$Polarity(tweet) = \frac{\sum_{k=1}^{n} PS_k}{n} \tag{2}$$

Where: 'n' is a total word in a tweet and PS_k is the polarity score of words of tweet available in the dictionary.

This approach used the following stepwise procedure to classify and label the tweets.

Method 1 Lexicon-based approach for classification and labeling of the tweets

Step:1. The polarity of the unlabeled tweet (denoted by PT) from dataset D2 was calculated using equation (2).
Step:2. If PT >= 0, then the tweet was classified as non-cyberstalking and assigned a label (value=0, positive tweet) to a tweet of dataset D2.
Step:3. If PT < 0, the tweet was classified as a highly suspicious tweet. In this case, the tweet was very near to cyberstalking tweet, but before taking the final decision, tweets on the user timeline and retweets count were checked to confirm.
Step:4. The average polarity of the tweets (denoted by UPT) from the user timeline and retweet count (represented by RT) were calculated (at least three recent tweets were considered from the user timeline).
Step:5. If PT < 0 AND (RT>0 or UPT < 0), then the suspicious tweet was classified as cyberstalking, assigned label as cyberstalking tweets (value=1, negative tweets), otherwise classified as a non-cyberstalking tweet.
Step:6. After classification, the labeled tweet was stored in a separate dataset D3.
Step:7. Steps 1 to step 6 were repeated until the classification of all tweets of Dataset D2.

3.4.2 Machine Learning-based approach for classification and labeling of the tweets

Machine learning is broadly used to classify and label tweets with sentiment analysis support. In this approach, this paper used Support Vector Machine (SVM) for classification and labeling the tweets. Support vector machine is an efficient, versatile, and trendy supervised machine learning broadly used to classify tweets with more accurate results [51]. SVM creates hyperplanes and computes the distance between the line and support vector to classify the text. The SVM offered several kernels (polynomial, sigmoid, Radial Basis Function, linear, and nonlinear kernels) with different mathematical functions [52]. Although, as per its native nature, SVM use prediction and does not support probability directly but using Platt scaling and isotonic regression methods, SVM determines the probability of any text for the target class. This paper used the probability calibration classifier method for SVM to calculate the prediction probability of tweets. The mathematical expression (3) is used to calculate the prediction probability of tweets in the SVM model.

$$P(y|tweet) = \frac{1}{1 + exp(Af(tweet) + B)} \tag{3}$$

Where 'A' and 'B' are scalar parameters learned by the algorithm during the training, 'y' is target class(y=1 for cyberstalking and y=0 for non-cyberstalking) ***f***(tweet) is a real-valued function.

In this approach, the following stepwise procedure was used for classifying and labeling the tweets.

Method 2 Machine Learning-based approach for classification and labeling of the tweets

Step:1. In the first step, a pre-defined training dataset (D1) containing 35734 unique records (as discussed in step 5 of section 3.1) with cyberstalking and non-cyberstalking texts were cleaned using pre-processing tasks.
Step:2. After getting the feature vectors from dataset D1 using the TF-IDF feature extraction, the SVM model was trained through dataset D1. A trained SVM model can predict the probability of any unlabeled tweets for positive or negative sentiment.
Step:3. The trained SVM model was applied to the unlabelled tweet of dataset D2 (collected from Twitter API, cleaned through pre-processing tasks, and obtained feature vectors using TF-IDF, as discussed in sections 3.2 and 3.3), and the predicted probability of tweet (represented by PPT) was estimated using equation(3).

Step:4. If predicted probability (PPT) <=0.5, then the tweet was classified as a non-cyberstalking tweet and assigned a label (value=0, positive tweet) to the tweet of dataset D2.
Step:5. If PPT >0.5, the tweet was classified as a suspicious tweet. In this case, tweets from the concerned user timeline were checked, and retweets (denoted by RT) were counted.
Step:6. The average predicted probability of tweets from the user timeline (denoted by UPPT) was calculated (at least three recent tweets were considered from the user timeline)
Step:7. If PPT >0.5 AND (RT>0 or UPPT >0.5), then the suspicious tweet was classified as cyberstalking tweet and assigned a label (value=1, negative tweets) otherwise classified as a non-cyberstalking tweet.
Step:8. The classified tweet was saved into a separate Dataset D4.
Step:9. Steps 3 to step 8 were repeated until the classification of all tweets of Dataset D2.

3.4.3 Hybrid approach for classification and labeling of the Tweets

The first segment of a hybrid approach used lexicon-based polarity scores and machine learning-based probability scores to classify and label the tweets. In this approach, the following main stepwise procedure was used.

Method 3 Hybrid approach for classification and labeling of the Tweets

Step:1. The polarity of the unlabeled tweet (denoted by PT) from dataset D2 was calculated using (as discussed in section 3.4.1) using lexicon-based sentiment analysis.
Step:2. The predicted probability of unlabelled tweet (denoted by PPT) from dataset D2 was calculated (as discussed in section 3.4.2) using the trained SVM model through Dataset D1.
Step:3. If PT >= 0 AND PPT <= 0.5 then tweet was classified as non-cyberstalking and assigned label (value=0, positive tweet) to tweet of dataset D2. In this case, both lexicons-based and machine learning methods produced the same sentiment (non-cyberstalking) for the tweet.
Step:4. If PT < 0 AND PPT > 0.5, the tweet was classified as highly suspicious. In this case, the tweet was very near the cyberstalking tweet, and both lexicons-based and machine learning methods produced the same sentiment (cyberstalking). In this case, tweets on the user timeline and retweets were checked to confirm before making the final decision.
Step:5. The average predicted probability of tweets (denoted by UPPT) and average polarity of the tweet (represented by UPT) from the user timeline and retweet (RT) were calculated (at least three recent tweets were considered from the user timeline).
Step:6. If (PT < 0 AND PPT > 0.5) AND (RT > 0 or UPPT > 0.5 or UPT < 0), then high suspicious tweet was classified as cyberstalking, assigned label as cyberstalking tweets (value=1, negative tweets) otherwise classified as non-cyberstalking tweet. Again Dataset D2 was updated and saved.
Step:7. Labeled tweet, classified by using this approach, was saved into separate dataset D5.
Step:8. Steps 1 to step 7 were repeated until the successful classification of all tweets of Dataset D2.

After classification and labeling the tweets of Dataset D2, several ML classifiers, specifically SVM, Logistic Regression (LR), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbor (KNB), were trained and tested on datasets D3, D4, and D5. Performances were measured for all applied methods of classifications and labeling: lexicon-based, machine learning, and hybrid approach.

3.5 Real-Time Cyberstalking Detection on Live Tweets

Tweets collected through Twitter's search API (as discussed in section 3.1) contained tweets that already happened and were not in real-time. In this section, live tweets in real-time were fetched using Twitter's Streaming API and Twitter's Firehose. Further, using a hybrid approach, tweets were automatically classified and labeled as cyberstalking or non-cyberstalking tweets in real-time while fetching the live tweets. At this time, the proposed hybrid approach used the lexicon-based method, trained SVM model, and trained Naïve Bayes model. Naïve Bayes (NB) is an efficient and straightforward supervised machine learning algorithm. The functioning of NB is according to the Bayes Theorem and derived from conditional probability [53]. In this paper, the multinomial NB model was used, while other models offered by NB are Gaussian NB and Bernoulli NB. In Naïve Bayes, the following equation calculates the predicted probability of tweets for the target class (cyberstalking or non-cyberstalking tweets).

$$P(y|tweet) = \frac{P(y)\prod_{i=1}^{n} P(x_i \mid y)}{P(x1)\times P(x2)\times \ldots \times p(x_n)} \tag{4}$$

Where 'y' is the target class (y=1 for cyberstalking and y=0 for non-cyberstalking). P(y|tweet) represents the posterior probability of tweet for target class 'y'. P(tweet)=P(x1)P(x2)….P(x_n) is the preceding probability of predictor tweet. P(y) is the preceding probability of the target class. P(x_i|y) is the likelihood conditional probability of predictor tweet for target class (y).

SVM model was trained by dataset D1 while NB was trained by recently created labeled dataset D5 (contain recent tweets classified by hybrid approach as discussed in section 3.4.3). The following stepwise procedure was used for real-time automated cyberstalking detection on live tweets.

Method 4 Hybrid approach for Real-Time Cyberstalking Detection on Live Tweets

Step:1. The live tweet was fetched in real-time through Twitter's streaming API and filtered through various hashtags keywords.
Step:2. The polarity of the fetched unlabeled tweet PT was calculated (as discussed in section 3.4.1) using the lexicon-based method.
Step:3. The predicted probabilities of fetched unlabelled tweet PPT_SVM and PPT_NB were calculated using the trained SVM and NB model, respectively.
Step:4. The average predicted probability of fetched unlabelled tweet APPT_ML was calculated using PPT_SVM and PPT_NB
Step:5. If PT>=0 OR APPT_ML <= 0.5, then tweet was classified as non-cyberstalking and assigned label (value=0, positive tweet).
Step:6. IF PT <0 AND APPT_ML >0.5, the tweet was classified as highly suspicious. In this case, tweets of the user timeline were checked to confirm before making the final decision.
Step:7. The average predicted probabilities UAPPT_ML (UAPPT_ML= (UAPPT_SVM+ UAPPT_NB)/2) and average polarity of tweets UAPT were calculated (at least three recent tweets were considered from the user timeline).
Step:8. If (PT <0 AND APPT_ML >0.5) AND (UAPT < 0 AND UAPPT_ML >0.5), then the highly suspicious tweet was classified as cyberstalking and assigned a label (value=1, negative tweets) otherwise classified as a non-cyberstalking tweet.
Step:9. The live labeled tweet (cyberstalking and non-cyberstalking tweet along with user id, username, location, and date) was stored in dataset D6.
Step:10. Dataset D1 was updated from the labeled tweet of Dataset D6 for further use.
Step:11. Steps 1 to step 10 were repeated until fetching a sufficient number of live tweets (more than 10000 live tweets).

3.6 Measuring the Performance of Model

Performance of classifiers with each applied method (lexicon-based, machine learning, and hybrid approach) for classification and labeling of recent tweets (fetched through Twitter API) and live tweets (fetched through Twitter Streaming) were measured separately. Performance metrics are several factors used to measure a model's performance during training and testing time [54]. The performance parameters are usually determined through the confusion matrix. In this study, the confusion matrix is a 2x2 truth table matrix containing the total value of True_Pos, True_Neg, False_Neg, and False_Pos. True_Pos (True Positive) is a successful hit showing the total number of correctly detected cyberstalking tweets, while True_Neg (True Negative) explains the total number of correctly detected non-cyberstalking tweets. In contrast, False_Pos (False Positive) is miss-hit, illustrating the total number of incorrectly detected cyberstalking tweets, while False_Neg (False Negative) is the failure count representing the total number of wrongly detected non-cyberstalking tweets. In this paper, broadly used parameters such as accuracy, precision, f-score, and recall were calculated to measure the performance of cyberstalking detection method. AUC (Area Under the Curve) was also calculated during the automatic detection of live tweets in real-time.

3.6.1 Accuracy

Accuracy addresses the complete number of rights predictions anticipated by the classifier. Accuracy can be calculated using equation (5).

$$Accuracy = \frac{True\_Pos + True\_Neg}{True\_Pos + False\_Pos + False\_Neg + True\_Neg} \tag{5}$$

3.6.2 Precision

Precision shows the proportion between the true positives and the wide range of various positives. Precision can be calculated using equation (6).

$$Precision = \frac{True\_Pos}{True\_Pos + False\_Pos} \tag{6}$$

3.6.3 Recall

Recall describes the sensitivity and measures the proportion of true positive prediction to total positive. Recall can be determined using equation (7).

$$Recall = \frac{True\_Pos}{True\_Pos + False\_Neg} \tag{7}$$

3.6.4 F-Score

F-Score measures test accuracy and describe the harmonic average between precision and recall. F-score can be determined using the equation (8).

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

3.6.5 AUC (Area Under the Curve)

AUC estimates the capacity of the classifier to separate among classes correctly. ROC (Receiver Operator Characteristic) is a likelihood curve that plots the True Positive Rate (TPR) against the False Positive Rate (FPR). Equation (9) can be used to calculate the AUC.

$$AUC = \frac{1}{2}\left(\frac{True\_Pos}{True\_Pos + False\_Neg} + \frac{True\_Neg}{True\_Neg + False\_Pos}\right) \tag{9}$$

4. Experimental Setup, Results, and Discussion

This section will discuss the experimental setup and results for automatically detecting cyberstalking tweets in real-time. The experiments used python language with Scikit Learn, Tweepy, Twitter Streaming, TextBlob, NLTK, and other library packages to implement the proposed model. To train the machine learning classifiers in the initial phase (in machine learning and hybrid approach for classification and labeling of the collected tweets), a mixed labeled dataset D1 was prepared [55-59]. Training dataset D1 contains 35734 unique records classified as cyberstalking and non-cyberstalking text. Fig. 2 shows the distribution of tweets/comments in the training dataset D1.

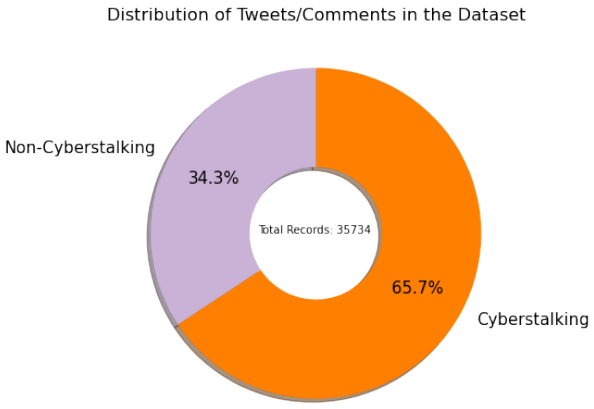

Fig. 2. Distribution of Tweets/Comments in the training dataset D1

In the first stage of the experiment, recent tweets were collected using the Twitter API. A total of 24178 tweets were collected using several attempts. After removing the duplicate tweets and blank lines, a total of 8066 unique tweets were selected and saved to dataset D2 for classification and labeling. After that, separate experiments separately classified tweets using different methods (as discussed in the methodology section). In the second experiment, collected recent tweets were classified and labeled using the lexicon-based method with the support of TextBlob sentiment analysis. Experimental work was also performed using other pre-trained and pre-defined lexicon-based methods such as Vader, SentiWordNet, and AFINN and found almost similar results. The classified tweets were stored in a separate dataset (D3), and the model was tested using different machine learning classifiers. The performance of different classifiers with lexicon-based labeling is explained in Table 1. As per experimental results, using a lexicon-based approach, 24.2% of recent tweets were classified as cyberstalking tweets, while 75.8% of recent tweets were classified as non-cyberstalking tweets. The lexicon-based approach provided maximum accuracy of 91.1%, a precision of 91.4%,

a recall of 81%, an f-score of 80.9%, and an AUC of 90.9% in the classification and labeling of the recent tweets. SVM achieved the maximum accuracy and AUC, Logistic Regression achieved maximum precision, while the Decision Tree achieved maximum recall and f-score.

Table 1. Performance of Classifiers with Lexicon-Based Classification and labeling of Tweets

Dataset (D2): 8066 unique recent tweets collected through Twitter API Tweets classified and labeled by: Method1 - Lexicon-based sentiment Cyberstalking tweets found: 24.2 %, Non-Cyberstalking tweets found: 75.8%					
S. No	ML Algorithm	Accuracy	Precision	Recall	F-Score
1	Support Vector Machine (SVM)	0.911254	0.891509	0.739726	0.808556
2	Decision Tree	0.903322	0.808594	0.810176	0.809384
3	Random Forest	0.896381	0.881313	0.682975	0.769570
4	Logistic Regression	0.865642	0.913793	0.518591	0.661673
5	Naive Bayes	0.861675	0.953333	0.679843	0.632678
6	K-Nearest Neighbor	0.836886	0.760000	0.520548	0.617886

In the third stage of the experiment, a trained SVM model as a machine learning (as discussed in the methodology section 3.4.2) was used to classify and label the recently collected tweets. The classified tweets were again stored in a separate dataset (D4), and the model was tested using the different machine learning classifiers. The performance of the machine learning approach for classification and labeling the tweets is described in Table 2. As per experimental results, 23.3% of recent tweets were classified as cyberstalking tweets, while 76.7% of recent tweets were classified as non-cyberstalking tweets using a machine learning approach. The machine learning approach for tweets classification provided maximum accuracy of 92.7%, precision of 90.5%, recall of 89.3%, f-score of 89.9%, and AUC of 96.6%. SVM performed better than other classifiers.

Table 2. Performance of Classifiers with Machine learning approach for Labeling of Tweets

Dataset (D2): 8066 unique recent tweets collected through Twitter API Tweets classified and labeled by: Method 2- Machine Learning Approach Cyberstalking tweets found: 11.7 %, Non-Cyberstalking tweets found: 88.3%					
S. No	ML Algorithm	Accuracy	Precision	Recall	F-Score
1	Support Vector Machine	0.923649	0.834646	0.443515	0.579235
2	Random Forest	0.908775	0.747748	0.347280	0.474286
3	Logistic Regression	0.902826	0.957447	0.188285	0.314685
4	K-Nearest Neighbor	0.898364	0.621429	0.364017	0.459103
5	Naive Bayes	0.892910	0.727536	0.096234	0.175573
6	Decision Tree	0.869608	0.453488	0.489540	0.470825

In the fourth stage of the experiment, a hybrid approach was used (as discussed in the methodology section 3.4.3) to classify and label the recently collected tweets. After classification, the labeled tweets were saved in a separate dataset (D5), and the hybrid approach was tested using the different machine learning classifiers. The performance of the hybrid approach for the classification and labeling of the tweets is exposed in Table 3. As per experimental results, 5.1% of recent tweets were classified as cyberstalking tweets, while 94.9% of recent tweets were classified as non-cyberstalking tweets using a hybrid approach. The hybrid approach for tweets classification achieved the highest accuracy of 95.8%, precision of 98.2%, recall of 38.8%, and an f-score of 40.6%. SVM again performed better than other classifiers.

Table 3. Performance of Classifiers with Hybrid Approach for Labeling of Tweets

Dataset (D2): 8066 unique recent tweets collected through Twitter API Tweets classified and labeled by: Hybrid Approach Cyberstalking tweets found: 5.1 %, Non-Cyberstalking tweets found: 94.9%					
S. No	ML Algorithm	Accuracy	Precision	Recall	F-Score
1	Support Vector Machine	**0.958004**	0.813725	0.247761	0.379863
2	Decision Tree	0.941113	0.426230	0.388060	0.406250
3	Random Forest	0.957229	0.953846	0.185075	0.310000
4	Naive Bayes	0.956454	0.965517	0.167164	0.284987
5	K-Nearest Neighbor	0.952735	0.636364	0.208955	0.314607
6	Logistic Regression	0.956609	0.982456	0.167164	0.285714

The comparative performance of all three applied approaches is presented in Fig. 3. As per experimental results shown in Table 1, Table 2, Table 3, and Fig. 3 show that the performance of the machine learning approach is better than the lexicon-based approach. In contrast, the performance of the hybrid approach is outstanding. SVM outperformed and achieved the highest accuracy of 91.1%, 92.5%, and 95.8% for lexicon-based, machine learning, and hybrid approach.

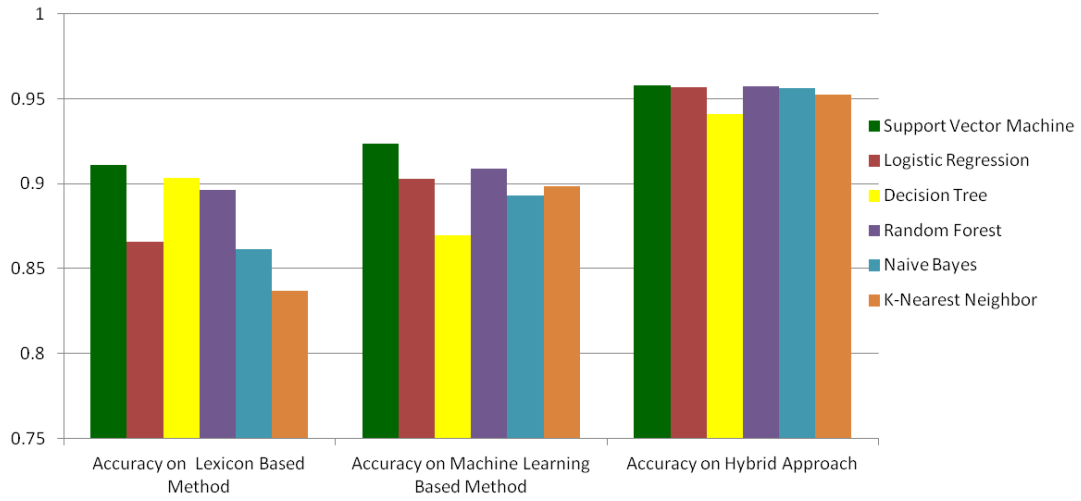

Fig. 3. Performances of Classifiers with all classification Methods

In the final experiment, an enhanced hybrid approach was applied again for automatic cyberstalking detection on live tweets in real-time due to its best performance. This time, the live tweets were fetched through Twitter Streaming, and the tweets were classified in real-time using a hybrid approach during the fetching of live tweets. During the fetching and classification, live labeled tweets were recorded into a separate dataset (D6), and the model was tested using several machine learning algorithms. The performance of the hybrid approach for automatic classification and labeling of the live tweets in real-time is shown in Table 4 and Fig. 4. AUC score and ROC curve are shown in Fig. 5, while the distribution of classified live tweets is described in Fig. 6. As per experimental results, 48.1% of tweets were labeled as cyberstalking, while 51.9% were labeled as non-cyberstalking during the fetching and classification of live tweets using a hybrid approach. Results mentioned in Table 4 and Fig. 4 show that the hybrid approach accomplished the results with notable performance. Accuracy of 94.2%, recall and f-score of 94.1%, the precision of 94.6%, and AUC of 98 % were achieved by the hybrid approach for automatic cyberstalking detection of live tweets in real-time. SVM again accomplished the highest accuracy, recall, and f-score, while random forest obtained the highest precision. AUC score and ROC curve are plotted in Fig. 4, indicating that SVM and random forest achieved the highest AUC of 98%.

Table 4. Performance of Classifiers with Hybrid Approach for Labeling of Live Tweets in Real-Time

Dataset size: 13294 unique live tweets collected through Twitter Streaming Live Tweets classified and labeled by: Hybrid Approach Cyberstalking tweets found: 48.1 %, Non-Cyberstalking tweets found: 51.9%						
S. No	ML Algorithm	Accuracy	Precision	Recall	F-Score	AUC
1	Support Vector Machine	**0.941937**	0.940564	0.941140	0.940852	0.980237
2	Random Forest	0.937425	0.946082	0.925199	0.935524	0.980135
3	Decision Tree	0.928700	0.929716	0.924586	0.927144	0.927545
4	Logistic Regression	0.923887	0.924784	0.919681	0.922226	0.971913
5	Naive Bayes	0.867329	0.872340	0.854690	0.863425	0.946197
6	K-Nearest Neighbor	0.726233	0.943419	0.470264	0.627660	0.792647

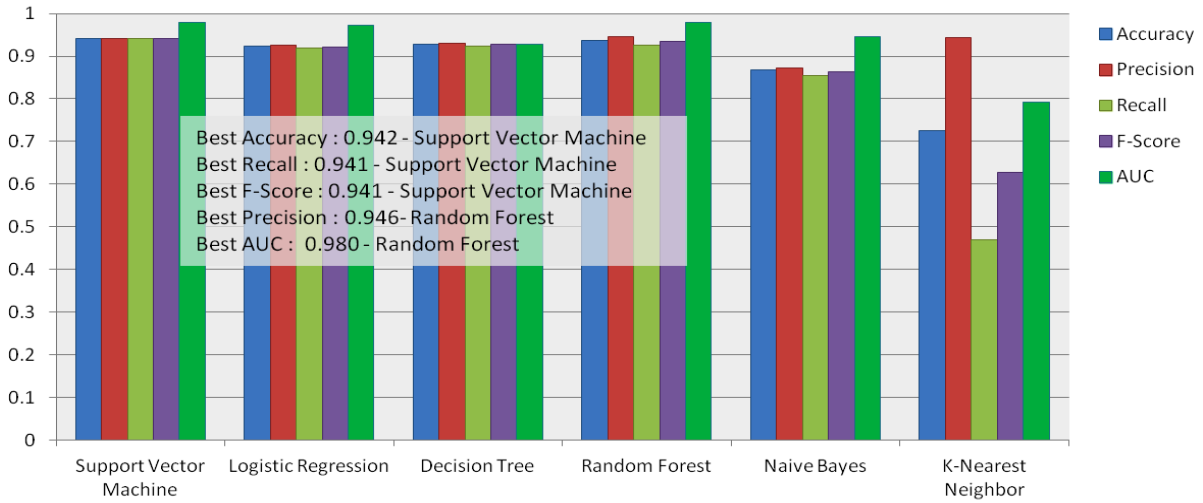

Fig. 4. Performances of Classifiers with Hybrid Approach for Labeling of Live Tweets in Real-Time

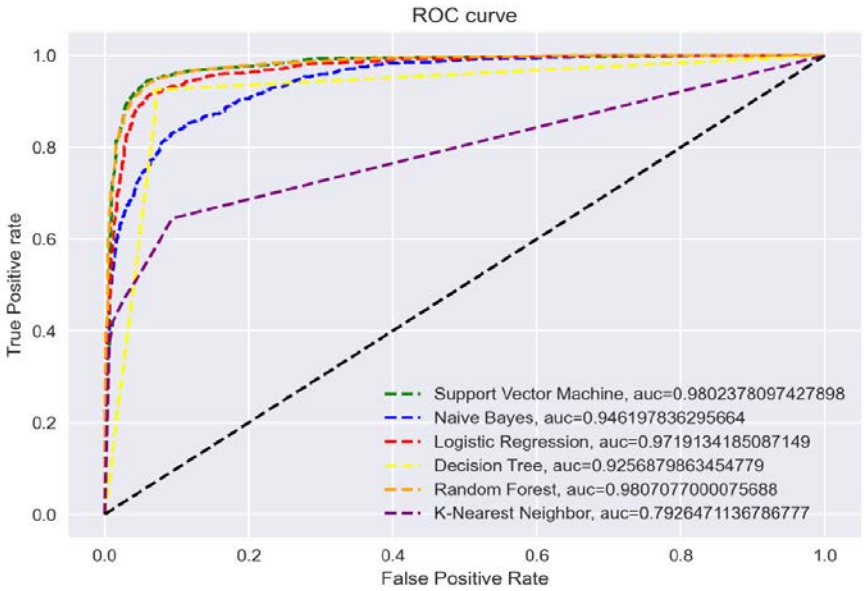

Fig. 5. AUC score and ROC curve for Live Tweets Classification in Real-Time

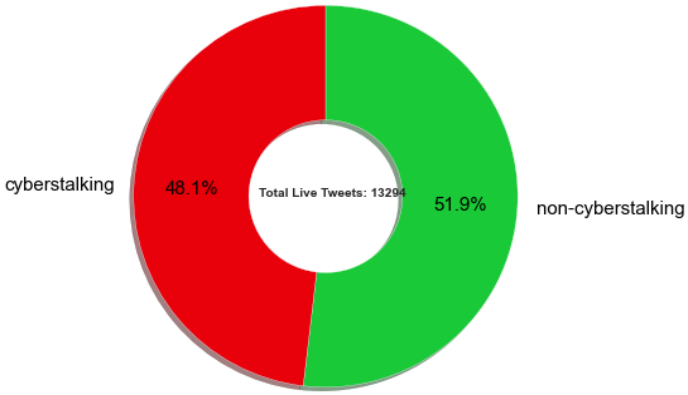

Fig. 6. Distribution of Live Tweets Classification in Real-Time

5. Conclusion and Future Work

Cyberstalkers are making a negative and fearful face of Twitter, and it is a challenging task to combat cyberstalking in real-time automatically. This paper proposed a hybrid approach using lexicon-based and machine learning-based models using different manners on separate segments for automatically cyberstalking detection on live tweets on Twitter in real-time. Using the Twitter API total of 24178 recent tweets were collected. In the initial stage, separate experiments were performed using lexicon-based, machine learning-based, and hybrid approaches on recent 8066 tweets (unique tweets out of 24178). The machine learning-based and hybrid approach used a pre-defined dataset containing 35734 individual tweets and comments to train the machine learning model. The performance of each method was measured using several parameters. The lexicon-based model obtained a maximum accuracy of 91.1%, while the machine learning-based model achieved 92.4% accuracy. The proposed hybrid approach successfully achieved the highest accuracy of 95.8%. Experimental results show that the performance of the machine learning-based model was better than the lexicon-based model, while the hybrid approach outperformed during cyberstalking detection on Twitter.

Due to the better performance of the hybrid approach, once again, another hybrid approach was applied for cyberstalking detection on live tweets directly fetching through Twitter streaming in real-time. Cyberstalking detection was successfully performed on 13294 live tweets using the hybrid approach in real-time. 48.1% of live tweets were classified as cyberstalking tweets, while 51.9% were classified as non-cyberstalking tweets during real-time cyberstalking detection. This time, the hybrid approach again outperformed for cyberstalking detection on live tweets on Twitter. The proposed hybrid approach successfully achieved the maximum accuracy of 94.4%, highest recall of 94.1%, highest precision of 94.6%, maximum f-score of 94.1%, and impressive AUC of 98% during cyberstalking detection on live tweets in real-time. In all approaches, the support vector machine outperformed other classifiers. Experimental results show that the hybrid approach is much better than other methods to combat the cyberstalking in

real-time. Lexicon-based models are often dependent on rules and dictionaries, while machine learning models require labeled datasets for training before the prediction. The proposed hybrid approach utilized the benefits of both approach lexicon-based and machine learning. The training dataset was automatically updated through the proposed detection model to improve the performance of each subsequent execution of the model. The performance of the proposed model can be enhanced through a more accurate and large dataset. Future work includes designing a more efficient hybrid model with lexicon-based, machine learning, deep learning, and fuzzy logic for cyberstalking detection in real-time.

References

[1] (2021) The Blacklinko Website. How Many People Use Twitter in 2021? [New Twitter Stats] [Online]. Available: https://backlinko.com/twitter-users

[2] Arkaitz Zubiaga, Alex Voss, Rob Procter, Maria Liakata, Bo Wang, Adam Tsakalidis, "Towards Real-Time, Country-Level Location Classification of Worldwide Tweets," IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 29(9), 2017.

[3] M. Baer. "Cyberstalking and the Internet Landscape We Have Constructed," Virginia Journal of Law & Technology, 154(15), 2020, pp. 153-227.

[4] Gautam, Arvind Kumar, and Abhishek Bansal. "Email-Based Cyberstalking Detection On Textual Data Using Multi-Model Soft Voting Technique Of Machine Learning Approach." Journal of Computer Information Systems (2023): 1-20. doi: 10.1080/08874417.2022.2155267

[5] Tarmizi, Nursyahirah, Suhaila Saee, and Dayang Hanani Abanag Ibrahim, "Detecting the usage of vulgar words in cyberbully activities from Twitter," International Journal on Advanced Science, Engineering and Information Technology 10(3), 2020, pp. 1117-1122.

[6] S. Lal, L. Tiwari, R. Ranjan, A. Verma, N. Sardana, & R. Mourya, Analysis and classification of crime tweets. Procedia Computer Science, 167, 2020, pp. 1911-1919.

[7] Arvind Kumar Gautam, and Abhishek Bansal, "A Review on Cyberstalking Detection Using Machine Learning Techniques: Current Trends and Future Direction." International Journal of Engineering Trends and Technology, 70(3), 2022, pp. 95-107. Crossref, https://doi.org/10.14445/22315381/IJETT-V70I3P211

[8] Salawu, Semiu, Yulan He, and Joanna Lumsden, "Approaches to automated detection of cyberbullying: A survey," IEEE Transactions on Affective Computing, 11(1), 2017, pp. 3-24.

[9] Abdur Rahman, Mobashir Sadat, Saeed Siddik, "Sentiment Analysis on Twitter Data: Comparative Study on Different Approaches," International Journal of Intelligent Systems and Applications, 13(4), 2021, pp.1-13.

[10] K. Rakshitha, H. M. Ramalingam, M. Pavithra, H.D. Advi, & M. Hegde, "Sentimental analysis of Indian regional languages on social media," Global Transitions Proceedings, 2(2), 2021, pp. 414-420.

[11] Khoo, Christopher SG, and Sathik Basha Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," Journal of Information Science 44(4), 2018, pp. 491-511.

[12] Norah AL-Harbi, Amirrudin Bin Kamsin, "An Effective Text Classifier using Machine Learning for Identifying Tweets' Polarity Concerning Terrorist Connotation," International Journal of Information Technology and Computer Science, 13(5), 2021, pp.19-29.

[13] A. Hasan, S. Moin, A. Karim, & S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," Mathematical and Computational Applications, 23(1), 2018, pp. 11.

[14] Golam Mostafa, Ikhtiar Ahmed, Masum Shah Junayed, "Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data," International Journal of Information Technology and Computer Science, 13(2), 2021, pp.38-48.

[15] Z. Ghasem, I. Frommholz, and C. Maple, "Machine learning solutions for controlling cyberbullying and cyberstalking," International Journal of Information Security, 6(2), 2015, pp. 55-64.

[16] Ingo Frommholz, Haider M. al-Khateeb, Martin Potthast, Zinnar Ghasem, Mitul Shukla , Emma Short, "On Textual Analysis and Machine Learning for Cyberstalking Detection," Datenbank Spektrum 16, 2016, pp. 127–135.

[17] Saravanaraj, A., J. I. Sheeba, and S. Pradeep Devaneyan, "Automatic detection of cyberbullying from twitter," International Journal of Computer Science and Information Technology & Security (IJCSITS), 2016.

[18] J. Zhang, T. Otomo, L. Li, & S. Nakajima, "Cyberbullying Detection on Twitter using Multiple Textual Features," In 2019 IEEE 10th International Conference on Awareness Science and Technology (CAST), IEEE, 2019, pp. 1-6.

[19] S. W. Liew, N. F. M. Sani, M. T. Abdullah, R. Yaakob & M. Y. Sharum, "An effective security alert mechanism for real-time phishing tweet detection on Twitter," Computers & Security, 83, 2019, pp. 201-207.

[20] V. Balakrishnan, S. Khan, H.R. Arabnia, "Improving cyberbullying detection using Twitter users' psychological features and machine learning," Science Direct, ELSEVIER, Computer & Security, 90, 2020.

[21] R. Shah, S. Aparajit, R. Chopdekar, & R. Patil, "Machine Learning based Approach for Detection of Cyberbullying Tweets," International Journal of Computer Applications, 175(37), 2020

[22] Kazim Raza Talpur, Siti Sophiayati Yuhaniz, Nilam Nur binti Amir Sjarif, Bandeh Ali, "Cyberbullying Detection In Roman Urdu Language Using Lexicon Based Approach," JOURNAL OF CRITICAL REVIEWS, 16, 2020, pp. 834-848. doi: 10.31838/jcr.07.16.109

[23] R. Geetha, S. Karthika, C. J. Sowmika, & B. M. Janani, "Auto-Off ID: Automatic Detection of Offensive Language in Social Media," In Journal of Physics: Conference Series, 1911(1), 2021.

[24] Bandi Yoshna, A. K. Jaithunbi, G. Lavanya, D.V. Smitha, "Detecting Twitter Cyberbullying Using Machine Learning," Annals of the Romanian Society for Cell Biology, 2021, pp. 16307–16315.

[25] Kumar, Akshi, and Nitin Sachdeva, "Multi-input integrative learning using deep neural networks and transfer learning for cyberbullying detection in real-time code-mix data," Multimedia systems, 2020, pp. 1-15.

[26] N. Yuvaraj, V. Chang, B. Gobinathan, A. Pinagapani, S. Kannan, G. Dhiman, & A. R. Rajan, "Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification," Computers & Electrical Engineering, 92, 2021.
[27] S. Sadiq, A. Mehmood, S. Ullah, M. Ahmad, G.S. Choi, "Aggression detection through deep neural model on twitter," Future Generation Computer Systems, 114, 2021, pp. 120-129.
[28] Sangwan, Saurabh Raj, and M. P. S. Bhatia, "D-BullyRumbler: a safety rumble strip to resolve online denigration bullying using a hybrid filter-wrapper approach," Multimedia Systems, 2020, pp. 1-17.
[29] Lepe-Faúndez M, Segura-Navarrete A, Vidal-Castro C, Martínez-Araneda C, Rubio-Manzano C, "Detecting Aggressiveness in Tweets: A Hybrid Model for Detecting Cyberbullying in the Spanish Language," Applied Sciences, 22(11), 2021. https://doi.org/10.3390/app112210706
[30] Madan, Anjum, and Udayan Ghose. "Sentiment Analysis for Twitter Data in the Hindi Language," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021.
[31] Almutairi, Amjad Rasmi, and Muhammad Abdullah Al-Hagery, "Cyberbullying Detection by Sentiment Analysis of Tweets' Contents Written in Arabic in Saudi Arabia Society," International Journal of Computer Science & Network Security 21(3), 2021, pp. 112-119.
[32] I. Arora, J. Guo, S. L. Levitan, S. McGregor, & J. Hirschberg, "A novel methodology for developing automatic harassment classifiers for Twitter," In Proceedings of the Fourth Workshop on Online Abuse and Harms, 2020, pp. 7-15.
[33] F.E. Ayo, O. Folorunso, F.T. Ibharalu, I.A. Osinuga, & A. Abayomi-Alli, "A probabilistic clustering model for hate speech classification in twitter," Expert Systems with Applications, 173, 2021.
[34] S. Vijayarani, J. Ilamathi, and Nithya, "Pre-processing techniques for text mining-an overview," International Journal of Computer Science & Communication Networks, 5(1), 2015, pp. 7-16..
[35] (2020) Towardsdatascience website. All you need to know about text pre-processing for NLP and Machine Learning. [Online]. Available: https://towardsdatascience.com/all-you-need-to-know-about-text-preprocessing-for-nlp-and-machine-learning-bc1c5765ff67.
[36] Kadhim, Ammar Ismael, "An evaluation of pre-processing techniques for text classification," International Journal of Computer Science and Information Security (IJCSIS), 16(6), 2018, pp. 22-32.
[37] Dimple Tiwari, Nanhay Singh, "Ensemble Approach for Twitter Sentiment Analysis", International Journal of Information Technology and Computer Science, 11(8), 2019, pp. 20-26.
[38] Gautam, Arvind Kumar, and Abhishek Bansal, " Effect of Features Extraction Techniques on Cyberstalking Detection using Machine Learning Framework," Journal of Advances in Information Technology, 13(5), 2022.
[39] Rui, Weikang, Kai Xing, and Yawei Jia. "BOWL: Bag of word clusters text representation using word embeddings." International Conference on Knowledge Science, Engineering and Management. Springer, Cham, 2016.
[40] (2020) Medium Website. All about Embeddings. [Online]. Available: https://medium.com/@kashyapkathrani/all-about-embeddings-829c8ff0bf5b
[41] T. Mikolov, K. Chen, G. Corrado, J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013. https://arxiv.org/pdf/1301.3781.pdf
[42] Pennington, Jeffrey, Richard Socher, and D. Christopher, "Glove: Global vectors for word representation," Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014.
[43] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, "Bag of tricks for efficient text classification," arXiv preprint arXiv:1607.01759, 2016. https://arxiv.org/pdf/1607.01759.pdf
[44] C. Raj, A. Agarwal, G. Bharathy, B. Narayan, M. Prasad, "Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques," Electronics, 22(10), 2021.
[45] B. Das, S. Chakraborty, "An improved text sentiment classification model using TF-IDF and next word negation," arXiv preprint arXiv:1806.06407, 2018.
[46] S. Alashri, S. Alzahrani, M. Alhoshan, I. Alkhanen, S. Alghunaim, & M. Alhassoun, "Lexi-Augmenter: Lexicon-Based Model for Tweets Sentiment Analysis," In 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), IEEE, 2019, pp. 7-10.
[47] Gupta, Neha, and Rashmi Agrawal, "Application and techniques of opinion mining," Hybrid Computational Intelligence. Academic Press, 2020, pp. 1-23.
[48] Osman, Aida, Said Ahmad. "Current trends and research directions in the dictionary-based approach for sentiment lexicon generation: a survey," Journal of theoretical and applied information technology 97(2), 2019.
[49] Kumar, Akshi, and Geetanjali Garg, "Systematic literature review on context-based sentiment analysis in social multimedia," Multimedia tools and Applications, 2020, pp. 15349-15380.
[50] Sazzed, Salim, and Sampath Jayarathna. "Ssentia: a self-supervised sentiment analyzer for classification from unlabeled data," Machine Learning with Applications, 4, 2021.
[51] Gautam, Arvind Kumar, and Abhishek Bansal, "Performance Analysis of Supervised Machine Learning Techniques For Cyberstalking Detection In Social Media," Journal of Theoretical and Applied Information Technology, 100(2), 2022.
[52] (2017) Data Flair website. Kernel Functions-Introduction to SVM Kernel & Examples. [Online]. Available: https://data-flair.training/blogs/svm-kernel-functions/
[53] Rish, "An empirical study of the naive bayes classifier", IJCAI 2001 workshop on empirical methods in artificial intelligence, 3(22), 2001, pp. 41–46
[54] Eman Bashir, Mohamed Bouguessa, "Data Mining for Cyberbullying and Harassment Detection in Arabic Texts," International Journal of Information Technology and Computer Science, 13(5),2021, pp.41-50.
[55] (2020) Mendeley Cyberbullying datasets. [Online]. Available:https://data.mendeley.com/datasets/jf4pzyvnpj/1
[56] (2020) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/mrmorj/hate-speech-and-offensive-language-dataset
[57] (2022) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/andrewmvd/cyberbullying-classification
[58] (2021) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/sanamps/toxiccommentclassification

[59] (2014) The Kaggle website-dataset. [Online]. Available: https://www.kaggle.com/c/detecting-insults-in-social-commentary/data

Authors' Profiles

Arvind Kumar Gautam was born in Rewa, Madhya Pradesh, India. He received his Master of Philosophy degree in Computer Science in 2009 from APS University, Rewa, Madhya Pradesh, India. He is a Ph.D. research scholar in the Department of Computer Science, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India. He is also working as a System Analyst for 9 years at Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh. He has more than 10 years of working experience in server administration, networking, cyber security, web programming, and teaching. He has published several research papers in international journals and conferences. His academic research interests mainly include Cyber Security, Machine Learning, and Web Engineering.

Abhishek Bansal received the MCA degree from Dr. B. R. Ambedkar University, Agra, Uttar Pradesh, India, in 2004, and the Ph.D. degree from Delhi University, Delhi, India. He is currently working as a Senior Assistant Professor with the Department of Computer Science, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India. He has more than 12 years of teaching and research experience and supervised several Ph.D., research scholars. He has also published several papers in reputed journals and conferences.

How to cite this paper: Arvind Kumar Gautam, Abhishek Bansal, "Automatic Cyberstalking Detection on Twitter in Real-Time using Hybrid Approach", International Journal of Modern Education and Computer Science(IJMECS), Vol.15, No.1, pp. 58-72, 2023. DOI:10.5815/ijmecs.2023.01.05