Journal of Autonomous Intelligence (2024) Volume 7 Issue 4
doi: 10.32629/jai.v7i4.971

ORIGINAL RESEARCH ARTICLE

Hybrid model of unsupervised and supervised learning for multiclass sentiment analysis based on users' reviews on healthcare web forums

Anuj Kumar*, Shashi Shekhar

Department of Computer Engineering & Application, GLA University, Mathura, Uttar Pradesh 281406, India

*** Corresponding author:** Anuj Kumar, anujkumar.gla@gla.ac.in

ABSTRACT

Twitter has become a popular platform for sharing health information, including diabetes-related content. Recent research studies have shown that Twitter data can be used for various purposes such as monitoring illnesses, promoting health, analyzing sentiment, and potentially aiding in medical directing. However, detecting fitness-related tweets in the vast amount of data on Twitter can be difficult. This pilot study, therefore, aimed to classify patient text about drugs and disease-associated tweets into meaningful health-related segments. The unlabeled dataset is divided into several groups using an unsupervised learning technique called K-Means Clustering, using this first label the text and followed by a combination of neural networks and machine learning classifiers, they classified 32046 diabetes-related tweets and 161290 drug text lines into five groups. Approximately 66.38% of drug line text was classified as health-related, with 55.14% "treatment and medication", 7.10% "prevention" and 4.14% "symptoms and causes". Over 33% were categorized as "Other and News". If we talk about the tweets as a dataset then the tweet was classified as health-related, with 44.30% "treatment and medication", 7% "prevention" and 5.3% "symptoms and causes". Over 56.10% were categorized as "Other and News. After this multiclass classification, we applied three machine learning and two deep learning models to find accuracy, precision, recall, and F1 scores. Drug review was used as a dataset then SVM and LR models provided an accuracy of 98% and when tweets were used as a dataset then LR models provided an accuracy of 97%. This research shows the importance of social media data in the decision-making system in the healthcare domain.

Keywords: k-means labeling; CNN-LSTM; tweets; sentiment analysis; feature extraction

ARTICLE INFO

Received: 13 July 2023
Accepted: 21 December 2023
Available online: 25 January 2024

1. Introduction

One of the four main chronic illnesses with a sizable worldwide impact, diabetes mellitus (DM), has lately been given the "pandemic" moniker. By 2030, 578 million people will have diabetes worldwide, up from 450 million currently. Diabetes prevalence increased in Singapore from 8.3% in 2010 to 8.6% in 2017. If not adequately managed, diabetes patients will become sicker, which is a concerning trend. The ability for stakeholders and consumers to exchange and obtain information online has considerably improved access to medical information. The phrase "information dissemination" is used to characterize the exchange of information among Twitter users, demonstrating the growing significance of social media platforms like Facebook and Twitter in the information dissemination process. This activity motivates people to do internet searches for health-related information and creates a substantial volume of health-related data. Healthcare professionals, careers, and patients should utilize social media often as Internet connection becomes more widespread and inexpensive. Patients can get answers to their medical issues and have an impact on medical decisions by doing web searches for health information. To enhance patient care, broaden medical education, and progress medical research, healthcare organizations have started integrating social media into their strategy. Khalifa et al.[1] and Wahi et al.[2] said the study aimed to group tweets about diabetes into useful categories that healthcare stakeholders might refer to while making choices and investigate patterns. Twitter has become the most significant microblog service in the past ten years, and academics are paying a lot of attention to Twitter's data source. Twitter, in contrast to many other social network platforms, makes the majority of user data publicly available. Twitter also includes a huge number of tweets, most of which are opinions on a wide range of subjects. These tweets may contain insightful comments and viewpoints from patients on a particular illness or set of medical interventions.

Unlike many other social networking sites, Twitter makes the larger part of client data accessible to the open. These tweets, which offer commentary on a extend of subjects, may also offer advantageous exhortation and reflect smart quiet conclusions on certain sicknesses or therapeutic medicines. Twitter has developed into a considerable social media stage, with key characteristics counting communication, community improvement, and collective activity organizations, claims O'Leary[3]. Assist investigation in this region is essential, especially in content opinion examination, since social media posts are ordinarily composed in casual dialect and it can be troublesome to get it and survey these writings in a therapeutic (epidemiological) setting. Estimation investigation particularly looks at how opinion is communicated in writings. Estimation frameworks may be utilized to extricate estimation categories from writings Gridach et al.[4] and Tariyal et al.[5]. Past thinks have portrayed opinion investigation as a free strategy for analyzing composed or verbal explanations, to extricate subjective information such as evaluations, appraisals, and estimations by Kaur and Kautish[6] and Bansal and Kaur[7]. Rathi and Tedmori[8] assumption investigation positions suppositions communicated through tweets to get users' sees on subjects. is for Opinion Examination which has been proposed as a portion of different advances by Yasen et al.[9]. One of the proposals is to overlook the foremost self-assertive parts of the content, or to naturally propose online advertisements for items custom-made to the viewer's tastes and to exclude others, hence expanding the data. Dictionary-based or unsupervised learning approaches apply rules, regularly found through etymological examination of a dialect, to estimation examination or machine learning or administered utilization known as programmed machine learning calculations to illuminate estimation analysis as a classification assignment. One or the other learning method is used in sentiment analysis. In addition, related research shows that Twitter has different types of sentiment analysis techniques, such as dictionary learning, machine learning, associative (using dictionaries and machine learning), and conceptual (using text messages) learning or context[10].

Information and communications technology (ICT) has been utilized for an assortment of exercises over the past decades, including healthcare and therapeutic improvement. This final knowledge-intensive range continually produces and devours huge sums of information and data. Furthermore, we found that about 80% of look motor questions are related to health-related subjects, indications, and medications. Patients and the common open are progressively turning to the Web for health-related back and data. Online health communities are utilized by chronic patients, in particular by Kamakshi[11]. According to surveys Kaoud[12] found that these patients gain a lot by talking to other people and exchanging ideas and experiences with them. In the world where we now reside, sentiment analysis might assist the healthcare industry in using reliable data to develop by taking the necessary steps. Algorithms are used in sentiment analysis to look through patient tweets about their contacts with hospitals, physicians, and other healthcare professionals. Many healthcare businesses benefit from this since users may use it to comprehend the thoughts of their customers and take the necessary steps to address any gaps. Since it enables them to research the negative effects of medications, complementary therapies, the spread of pandemics (COVID-19), as well as the environment's quality and pollution, content on social media platforms like Twitter and Facebook is also of significant interest to academics and professionals. The contributions made by this study are as follows:

This study divides the text of two different datasets (medicine reviews and Twitter tweets) into multiclass classifications like treatment and medication, prevention, symptoms causes, news, and others.

A hybrid framework is proposed for multiclass classification sentiment analysis in the healthcare domain. The developed hybrid model is based on unsupervised and supervised learning and uses both machine learning and deep learning model algorithms for classification.

K means approach is used for labeling, and clustering and is used to annotate medicine reviews to five groups (treatment and medication, prevention, symptoms causes, news, and others). The efficacy of TF-IDF feature engineering approaches is evaluated on the Medicine reviews and Twitter tweets datasets.

To evaluate the performance, three machine learning models and two deep learning models are used: Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM), and Deep learning models are Convolutional Neural Network (CNN) and Hybrid of CNN-LSTM.

Several experiments are run to see how well the suggested technique performs on two publicly accessible datasets. To confirm its effectiveness, Outcomes of this study demonstrate the significance of social media data in the decision-making mechanism in the healthcare services area.

2. Literature survey

The extensive research that has been done in the fields of text categorization models and sentiment analysis is covered in this section. It is one of the most extensively studied subjects that frequently combines machine learning (ML) and natural language processing (NLP) sciences. Text categorization in literature is one of the key challenges of text mining. It is a supervised ML issue as well. Text categorization has been the subject of extensive research.

Abuelenin et al.[13] This paper focused on sentiment analysis for Arabic tweets. The authors propose a methodology for sentiment analysis using a combination of machine learning and natural language processing techniques. They evaluate their approach using Arabic tweets and report promising results. Al-Hadhrami et al.[14], the authors compared supervised and unsupervised approaches for sentiment analysis of English tweets. They evaluate various classifiers and clustering algorithms and analyze their performance in terms of accuracy and efficiency. The findings provide insights into the effectiveness of different techniques for sentiment analysis. Al-Khasawneh[15] This paper mentioned a method for classifying diabetes using data mining techniques. The author applies various algorithms to a healthcare information system dataset and evaluates their performance in diabetes classification. The study provides valuable insights into the application of data mining in healthcare. Alhanjouri[16] focused on pre-processing techniques for clustering Arabic documents. The author explores various pre-processing steps such as tokenization, stop-word removal, and stemming, and evaluates their impact on the clustering results. The findings provide insights into effective pre-processing techniques for Arabic document clustering. Alsaeedi and Zubair[17] investigated sentiment analysis techniques for Twitter data in this study. They compare different approaches, including lexicon-based and machine learning-based methods, and evaluate their performance using Twitter datasets. The findings offer valuable insights into sentiment analysis techniques for Twitter data. Anitha and Asha[18] presented a hybrid machine-learning algorithm for improving prediction accuracy on medical datasets. The authors propose a combination of different machine learning techniques and evaluate their performance on medical datasets. The research contributes to improving prediction accuracy in medical data analysis. Ariestya et al.[19] focused on decision-tree learning for determining the graduation path of students. The authors apply decision tree algorithms to student data and analyze the accuracy and performance of the models. The research provides insights into using decision tree learning for student progression prediction. Solanki et al.[20] investigated supposition mining utilizing machine learning procedures. They center on estimation investigation and utilize machine learning calculations to classify conclusions. The think gives experiences into the application of machine learning for opinion investigation assignments. Gautam and Yadav[21] conducted estimation examination of Twitter information utilizing machine learning procedures. They investigate different machine learning calculations and compare their execution in classifying estimations. The think about highlights the potential of machine learning for assumption investigation errands on social media information. Rajurkar[22] gave an outline of assumption investigation methods for social media information. They examine different approaches, counting machine learning and lexicon-based strategies, and analyze their qualities and restrictions. The study serves as a comprehensive direct-to-opinion examination method within the setting of social media. Poongothai and Vijayalakshmi[23] performed assumption examination of social media information utilizing machine learning strategies. They investigate distinctive machine learning calculations and assess their viability in classifying estimations in social media posts. The consideration contributes to understanding assumption investigation strategies for social media information. Priyanka and Rekha[24] conducted estimation examination of Twitter information utilizing machine learning methods. They apply different machine learning calculations to classify estimations in tweets and compare their execution. The consideration gives bits of knowledge into the opinion examination of social media information and the viability of distinctive machine-learning approaches. Rajendran and Kousalya[25] performed estimation examination on Twitter information utilizing machine learning procedures. They utilize distinctive machine learning calculations to classify opinions in tweets and assess their execution. The consideration contributes to the application of machine learning for opinion investigation on social media stages.

Sedhain et al.[26] proposed hierarchical attention networks for document classification tasks. They introduce a hierarchical structure to capture the context at different levels of granularity and apply attention mechanisms to focus on informative parts of the input. The study presents an innovative approach to document classification using attention networks. Social media is playing an increasingly important role in information sharing and searching. More than two billion people are currently active users of Facebook (2414 million) and Twitter (336 million) globally by Pershad et al.[27]. Approximately 500 million tweets are created every day by Statista[28]. Recent studies have demonstrated the value of Twitter in health research. Using Twitter data for network conversation analysis, sentiment analysis, illness surveillance, health promotion, and medical treatments has been done on purpose Finfgeld-Connett et al.[29–33]. Twitter could be a prevalent stage for quiet support, enlistment, and collaboration in biomedical inquiry[30,34]. Twitter is becoming a basic source of health-related information for open well-being analysts due to the real-time nature of the substance and the ease Chen et al.[33]. Agreeing to Finfgeld-Connett[29], Twitter clients with therapeutic accounts are more likely to tweet "common data for open well-being or modern inquiries about almost medicines and innovation," such as from the Mayo Clinic. Thus, user-generated wellbeing substance on Twitter has the potential to provide valuable and germane data. This will give healthcare decision-makers more data almost the persistent, which can be utilized to make strides in clinical decision-making and patient-physician communication. Understanding how to help and oversee the growing persistent populace is pivotal given the predominance of diabetes around the world. Ali et al.[34] and focused on multiclass event classification from the text. They present a methodology for classifying events based on textual data. The paper discusses the challenges and provides insights into the classification of events using text-based features. Divya[35] presented an opinion-based learning model in the medical sector. The paper focuses on utilizing opinion-based information for learning and decision-making in medical contexts. It explores the integration of opinion mining techniques with medical data analysis, offering a unique perspective on knowledge extraction in the medical domain. Emadi and Rahgozar[36] proposed a fuzzy integral classifier fusion approach for sentiment analysis of Twitter data. Vijayaraghavan et al.[37-40] focused on combine multiple classifiers using a fuzzy integral to enhance sentiment classification accuracy. The paper demonstrates the effectiveness of the proposed approach in capturing the sentiment expressed in Twitter posts. Go et al.[41] proposed a Twitter sentiment classification method using distant supervision. They utilize a large-scale dataset with labeled emoticons as distant supervision to train a sentiment classifier. The paper offers an approach to sentiment classification in Twitter data using distant supervision. Gohil et al.[42] reviewed the methods used for sentiment analysis of healthcare tweets. They discuss the different approaches and techniques employed to analyze sentiments in healthcare-related tweets. The paper provides an overview of sentiment analysis methods in the context of health care. Gridach et al.[4] conducted an empirical evaluation of word representations on Arabic sentiment analysis. They compare different word representation models and assess their performance in sentiment analysis tasks specific to the Arabic language. The paper contributes to understanding the effectiveness of word representations in sentiment analysis for Arabic text.

Sentiment analysis and classification, according to the literature, are two extremely tough jobs. domains of study that necessitate varied tasks. The most attention has been paid to the tasks of sentiment classification and aspect sentiment classification. Furthermore, as a result of medical concerns and health-related incidents, Because of the history that social media, notably tweets, provide for practitioners and patients, it was important to develop sentiment analysis tools for application in medical fields. It is later discovered that ML algorithms may extract emotions connected to domain-specific data sources. To provide relevant insights for making informed decisions, the execution of administered learning calculations for multi-class assumption classification on tweets approximately healthcare and the study of disease transmission must be assessed.

3. Proposed methodology

We integrated various background concepts and research efforts for automated multi-class sentiment categorization of healthcare tweets according to our goals and objectives. The entire process of contributing may be broken down into three steps: gathering tweets that might be relevant to healthcare, pre-processing the data, and categorization.

To improve its efficiency, a suggested hybrid framework for multiclass sentiment analysis includes several methodologies. Several major components of the framework include data preprocessing, feature extraction, sentiment lexicon utilization, and sentiment clustering and classification. **Figure 1** depicts the general flow of the sentiment analysis algorithm design. The first phase is data preparation, which includes five critical processes: removing punctuation, converting text to lowercase, deleting stop words, managing null variables, and removing single-letter words. An English list of opinion words combined with a TF-IDF term weighting model is employed as the default technique for feature extraction. For multiclass sentiment clustering, k-means is also employed. SVM, Logistic Regression, Random Forest, CNN, and CNN-LSTM are among the sentiment classifiers included in the system. As previously stated, studies on sentiment analysis provide insufficient information on the best medical data solutions. As indicated in **Figure 1**, tests are performed on the comparison using one feature selection approach, one clustering strategy, and five classificational algorithms to support the proposed framework's default choice. Among the feature extraction approaches provided are BoWs, term presence, and TF-IDF models. One method for sentiment clustering is K-means labeling. Finally, sentiment classification alternatives such as SVM, RF, SVM, CNN, and CNN-LSTM exist.

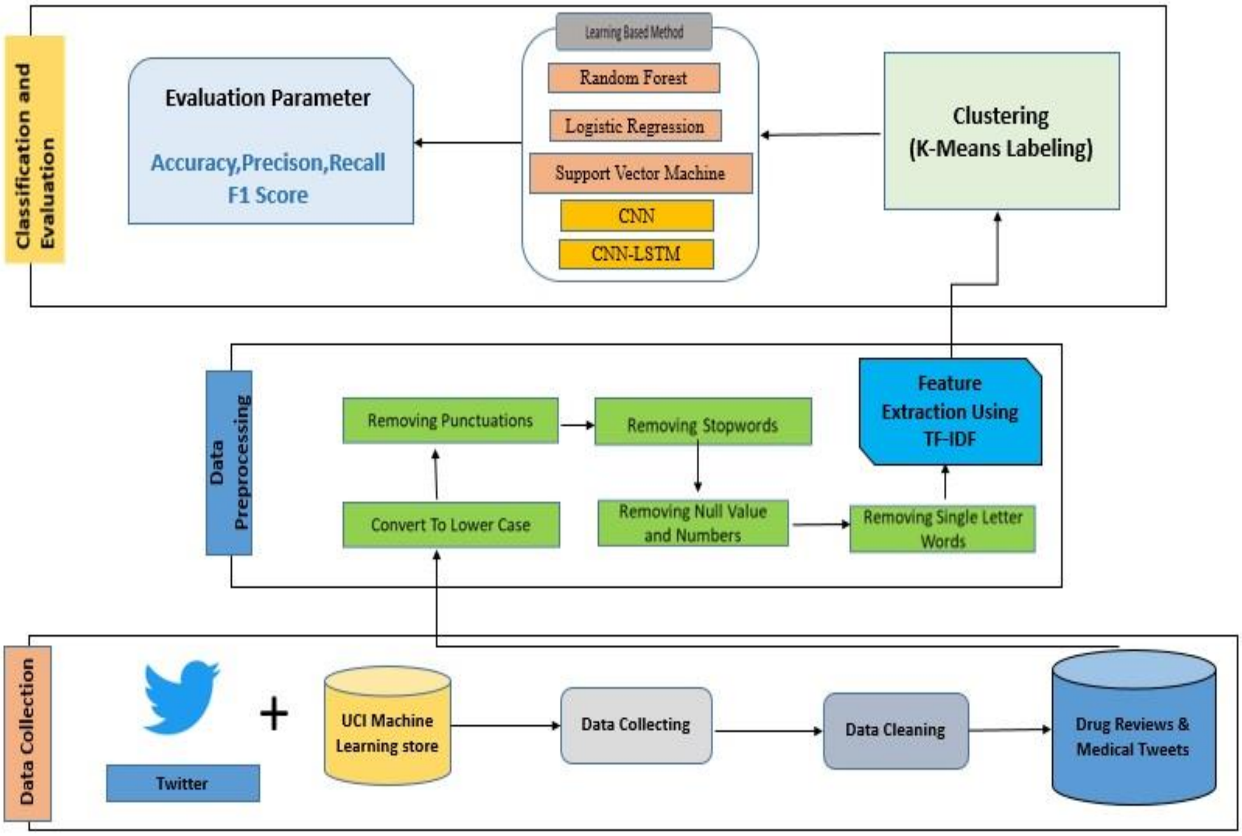


Figure 1. Architecture of the proposed approach.

3.1. Collecting health tweets dataset

Two types of datasets were employed for this research project. The first dataset, medicate surveys, was obtained from the UCI machine learning repository, a well-known source of benchmark datasets, and the second was tweets obtained from Twitter.

(I) Medicine Review Dataset: The content in this dataset contains patient reviews of pharmaceuticals and patients' underlying diseases at the time of drug use. There are 161,291 data items in the dataset, each having 7 waverings. The sole variable utilized in trials is examined. The evaluations include user testimonies of bad effects, side effects, and a drug's good conclusion, which demonstrates the user's satisfaction.

(II) Diabetes and Type 2 diabetes tweets are utilized as datasets for sentiment analysis. 32046 tweets concerning diabetes were retrieved from various Twitter accounts using Python scripts.

3.2. Data preprocessing

- Removing Punctuation: Punctuation signs such as periods, commas, exclamation points, and question marks are often meaningless in many NLP tasks. As a result, deleting them reduces the dimensionality of the data and concentrates on the important material.

- Converting Text to Lowercase: By converting all text to lowercase, you ensure that the identical words, regardless of cause, are treated uniformly. This step prevents the model from treating the same word differently depending on how it is capitalized.

- Deleting Stop Words: Stop words are regularly used words like "a," "an," "the," "in," and so on that do not add anything to the overall meaning of the text. Getting rid of them can assist in minimizing noise and boost the efficiency of the following analysis.

- Managing Null Variables: Datasets frequently contain null variables or missing data. Depending on the job and dataset, you can deal with null values by either eliminating the related data points, replacing them with a placeholder value, or employing more complex techniques like imputation to estimate missing values based on existing data.
- Removing Single-Letter Words: Single-letter words such as “a,” “I,” or “s” are generally uninformative and should be avoided to focus on more significant terms in the text.

3.3. Feature extraction

A feature selection strategy is used to discover essential qualities from pre-processed data to improve the performance of prediction models employing unobserved data. A variety of strategies may be used to feature engineer textual data. In this paper, we use TF-IDF to extract essential features from data for model training.

TF-IDF is the product of the Term Frequency (TF) and Inverse Document Frequency (IDF) for a specific term in a document. The formula for TF-IDF of a term (t) in document (d) within a document collection (D):

$$\text{TF-IDF}\ (t, d, D) = \text{TF}\ (t, d) \times \text{IDF}\ (t, D) \qquad (1)$$

The resulting TF-IDF score provides a measure of how relevant a term is to a specific document in the context of the entire document collection. Higher scores indicate that the term is more important or distinctive to the document, while lower scores suggest less significance.

By calculating TF-IDF scores for all terms in a document collection, you can identify important terms and perform various tasks such as keyword extraction, document similarity calculations, or text classification based on term importance.

3.4. Clustering and labeling

It is difficult to find health-related text in huge databases. It is critical to categorize text into categories or topics by identifying text that may include relevant health-related information. It is also critical to classify text that contains medical phrases but is otherwise unrelated to health. An unsupervised learning approach, K – means clustering used for labeling our text of the dataset. We established five text categories that describe our notion of what constitutes a health-related and non-health-related text. Preventive Measures (PM), Symptoms and causes (SC), and Treatment and medication (TM) are three categories that are strongly connected with health-related information that is commonly presented during a patient visit. The remaining two, News (NW) and Others (OT), are used to classify tweets that are irrelevant to our context.

3.4.1. K-means clustering

The unlabeled dataset is divided into several groups using an unsupervised learning technique called K-Means Clustering. In this instance, K denotes the number of pre-established clusters that have to develop during the process; for instance, if $K = 2$, two clusters will form; if $K = 3$, three clusters will form, and so on. **Figure 2** focuses on the difference between before and after K-means Clustering effects.Using an iterative procedure, the unlabeled dataset is split into k unique clusters, each dataset belonging to a single group with similar qualities. It allows us to partition the data into discrete groups and offers a useful way to autonomously detect the group categories in the unlabeled dataset without training. This method, which is based on centroid theory, associates each cluster with a centroid.

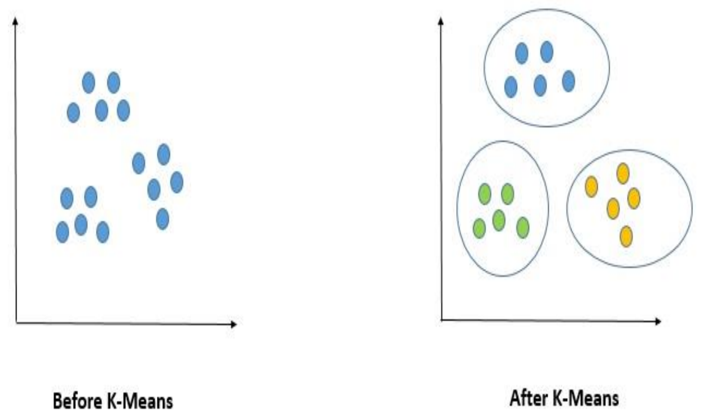


Figure 2. K Means clustering.

3.4.2. K-means algorithm work?

The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.
Step-2: Select random K points or centroids. (It can be other from the input dataset).
Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.
Step-4: Calculate the variance and place a new centroid of each cluster.
Step-5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.
Step-7: The model is ready.

3.5. Brief description of classifiers

Although machine learning has been used in data analysis and related fields for decades. Machine Learning on a known dataset is the process of solving issues in the medical sector utilizing various tools, methodologies, and strategies. This article made use of a variety of machine learning and Deep Learning methods to get the highest possible performance on medical datasets and enable effective text categorization. Random Forest (RM), Logistic Regression (LR), Support Vector Machine (SVM), Convolutional Neural Networks (CNN), and a hybrid of CNN and Long Short Term Memory (LSTM) were chosen to improve classification results in the domain of sentiment analysis and, on the other hand, to increase the efficiency and reliability of doctors' decisions (prediction).

(i) **Random Forest** is an ensemble learning approach that creates predictions by combining many decision trees. The base of each decision tree is a randomly selected set of variables that are then merged to produce the final forecast.

(ii) **Logistic Regression** is a linear framework that is used to solve issues involving binary categorization. A logistic equation predicts whether or not an instance belongs to a certain class.

$$P\ (\text{sentiment} \mid \text{features}) = 1 / (1 + \exp(-z)) \tag{2}$$

where z is the result of the linear combination of the characteristic weights.

(iii) **Support Vector Machine** is a binary classification model that attempts to find the best hyperplane for categorizing data points. A feature that reflects the text input can be used to train SVM for sentiment classification. The margin maximization problem, which requires optimizing the decision boundary, lies at the heart of SVM.

(iv) **Convolutional Neural Network (CNN)** is a form of deep learning algorithm that excels at picture detection and processing. It has several layers, including convolutional layers, pooling layers, and fully linked layers.

(v) **CNN-LSTM (Convolutional Neural Network Long Short-Term Memory)** is an advanced deep learning technique that combines the benefits of convolutional neural networks (CNNs) with long short-term memory (LSTM). This combination enables the algorithm to process geographical and temporal data at the same time. It can find patterns in time series data like audio, text, or pictures. It's been utilized in a variety of applications, including natural language processing (NLP), picture recognition, and speech recognition. It may also be used to produce simulated training data for supervised learning applications. CNN-LSTM achieves cutting-edge performance on a wide range of tasks across several domains by merging these two strong algorithms.

4. Result and evaluation

4.1. Drug reviews used as a dataset

In our framework, k-means labeling is chosen as the labeling process method above polarity labeling and SWN labeling. For many years, K-Means labeling has been one of the most versatile clustering techniques. **Table 1** describes when we applied K means labeling approach to the drug review dataset we found 161290 reviews in five clusters 55.14% "treatment and medication", 7.10% "prevention", 4.14% "symptoms and causes",17.70% as News and 15.83% were categorized as Other. **Figure 3** describes the Pie chart representation of the Number of instances for each label and In **Table 2** represents the comparative performance analysis of classification models in accuracy, precision, recall, and F1 Score.

Table 1. Description of the Number of instances for each label.

NEWS	Treatment	Prevention	Symptoms	Others	Total
28651	88942	11458	6692	25547	161290
17.70%	55.14%	7.10%	4.14%	15.83%	-

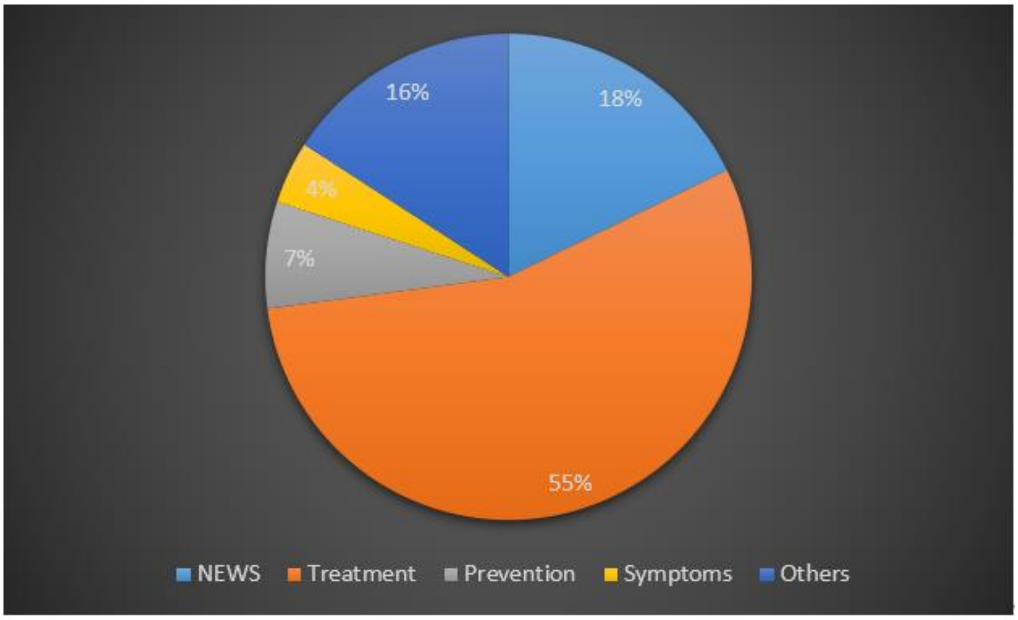


Figure 3. Pie chart representation of the Number of instances for each label.

Table 2. Comparative performance analysis of classification models in Accuracy, precision, recall, and F1 Score.

Models	**Accuracy**	**Precision**	**Recall**	**F1 Score**
RF	0.94	0.94	0.94	0.94
LR	0.98	0.98	0.98	0.98
SVM	0.98	0.98	0.98	0.98
CNN	0.96	0.95	0.97	0.96
CNN-LSTM	0.96	0.96	0.96	0.96

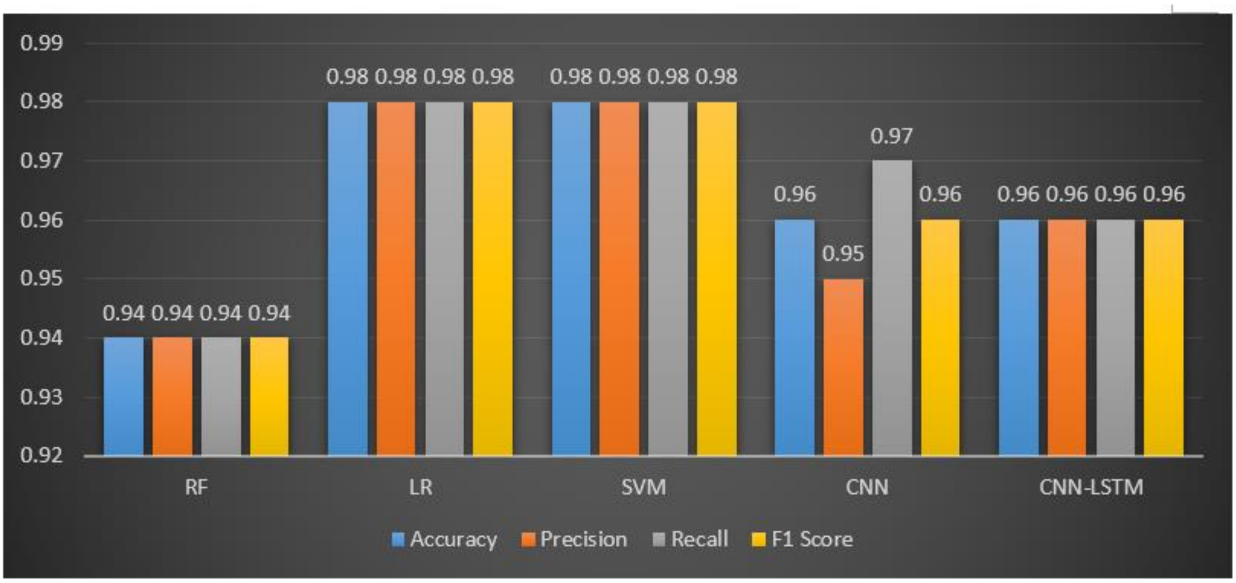
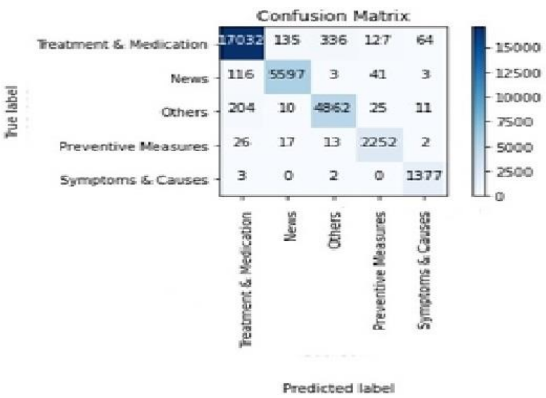
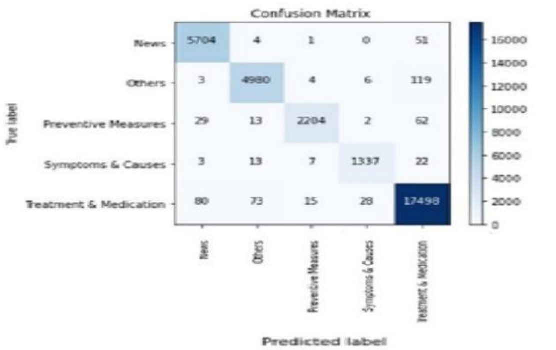


Figure 4. Accuracy, Precision, Recall, and F1 score comparison of classification models.

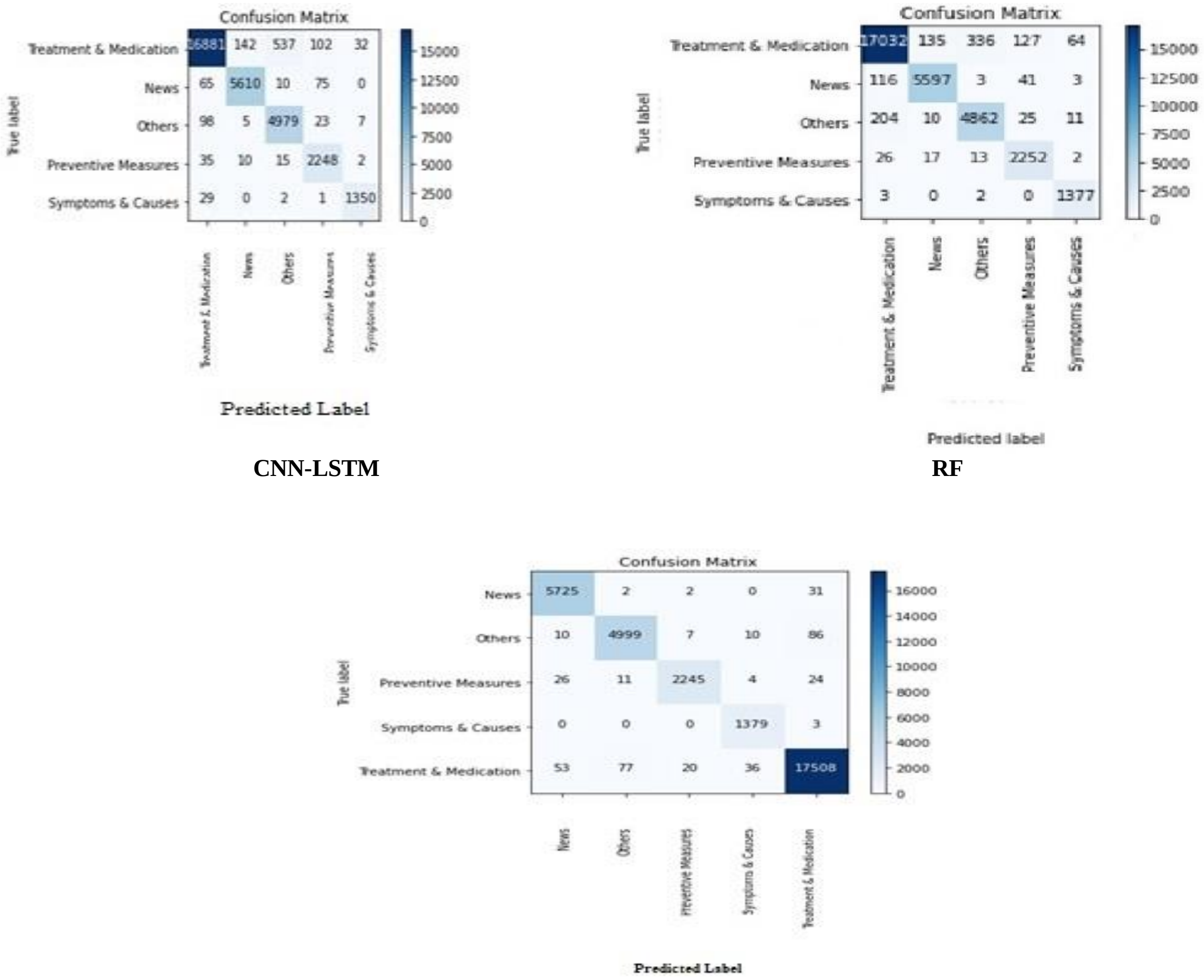
In Figure 4, describes After this multiclass classification, we applied three machine learning and two deep learning model to find accuracy, precision, recall, and F1 score. Drug review was used as a dataset then SVM and LR models provided an accuracy of 98% and **Figure 5** shows the confusion matrices for classification models with the highest accuracy for the analysis of all five labels.



CNN



LR



CNN-LSTM **RF**

SVM

Figure 5. Confusion matrices for classifiers with accuracy results.

4.2. Tweet used as a dataset

Table 3 describes when we applied K means labeling approach to the tweet review dataset we found 32046 reviews in five clusters 7% "treatment and medication", 32% "prevention", 5.30 % "symptoms and causes". 9.10% "News" and 47% categorized as "Other", **Figure 6** focuses about the number of instances of all five different labels by using Pie chart method **and Table 4** shows the Comparative performance analysis of applied three machine learning and two deep learning model to find accuracy, precision, recall, and F1 score. Tweets were used as a dataset then SVM and LR models provided an accuracy of 96% and 97% respectively

Table 3. Description of the Number of instances for each label.

NEWS	Treatment	Prevention	Symptoms	Others	Total
2943	2120	10298	1713	14972	32046
9.10%	7%	32.00%	5.30%	47%	-

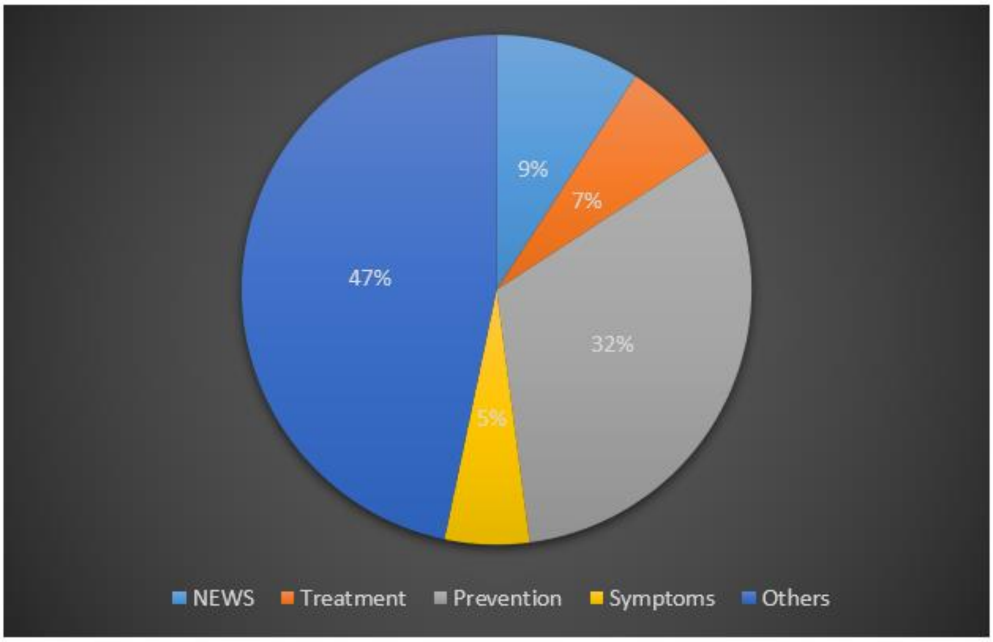


Figure 6. Pie chart representation of the Number of instances for each label.

Table 4. Comparative performance analysis of classification models in Accuracy, precision, recall, and F1 Score.

Models	Accuracy	Precision	Recall	F1 Score
RF	0.95	0.95	0.95	0.95
LR	0.97	0.97	0.97	0.97
SVM	0.96	0.96	0.96	0.96
CNN	0.94	0.93	0.94	0.94
CNN-LSTM	0.95	0.95	0.95	0.95

Figure 7 describes after this multiclass classification, we applied three machine learning and two deep learning model to find accuracy, precision, recall, and F1 score. Tweets were used as a dataset then SVM and LR models provided an accuracy of 96% and 97% respectively.

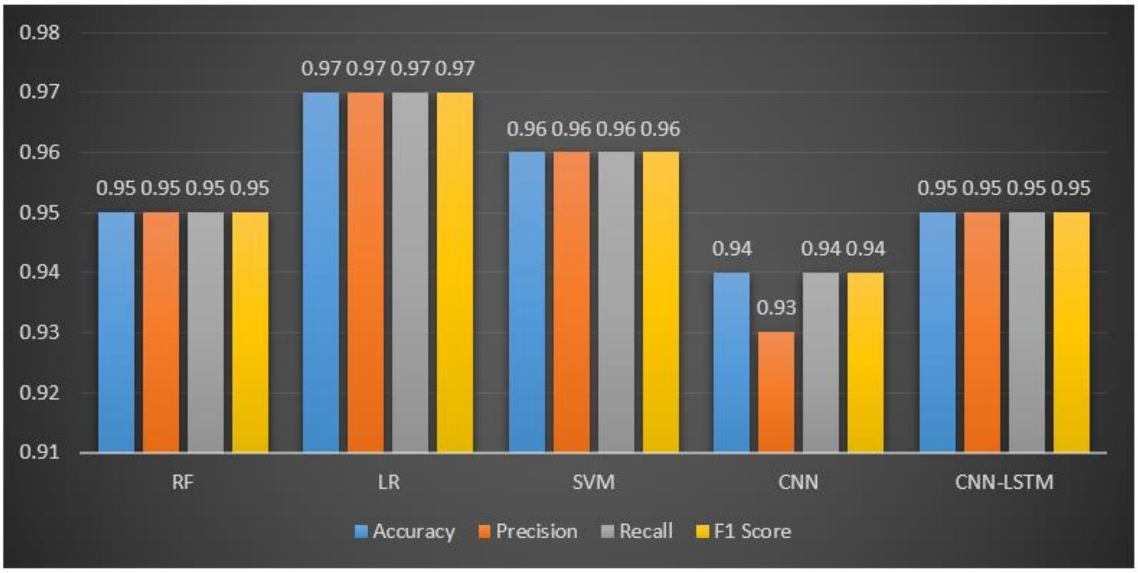
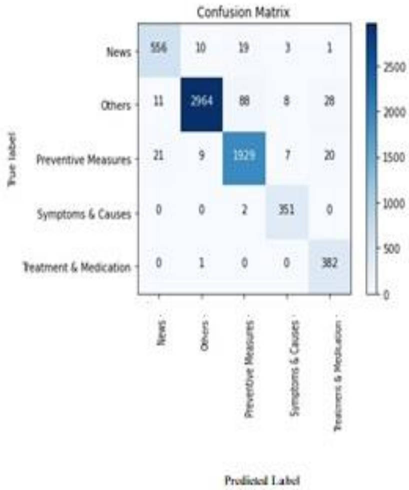


Figure 7. Accuracy, Precision, Recall, and F1 score comparison of classification models.

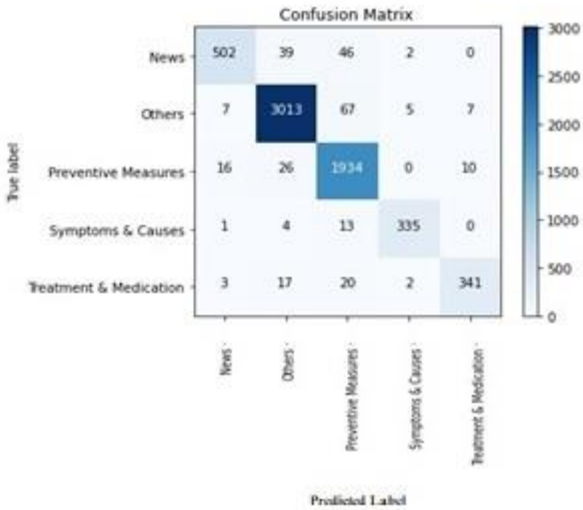
Figure 8 shows the confusion matrices for classification models with the highest accuracy for the analysis of all five labels.

CNN

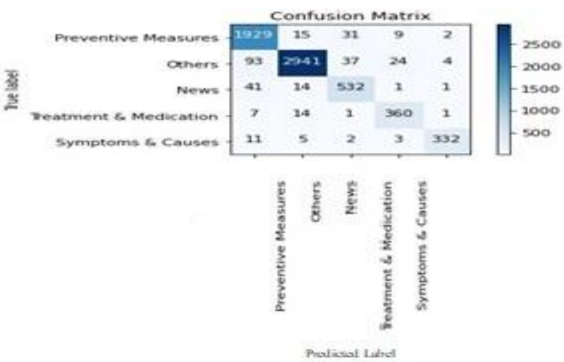
LR



SVM



RF



CNN-LSTM

Figure 8. Confusion matrices for classifiers with accuracy results.

In order to assess the hybrid approach's effectiveness for label categorization, its performance is compared to that of other cutting-edge methodologies. In order to compare different research that analyses the sentiment of medication reviews, this is done. The effectiveness of various sentiment analysis methodologies is shown in Table 5 and Results in Table 5 show that the present strategy performs better than previous works.

Table 5. Comparative investigation of existing approach with dataset.

References	DataSet	Classes	Feature	Method	Result			
					A	P	R	F
[37]	Drug Reviews (215063)	Multi	TF-IDF, Count Vectorizer	ANN, LSTM, GRU, SVM, LR, RF	0.93	0.95	0.95	0.9
[38]	Drug Reviews (50000)	Multi	Unigram and Bigram	Lexicon Combination and Information gain	0.91	0.76	0.52	0.62
[39]	Drug Reviews (5600)	Binary	SWN and Position Encoding	RF, SVM, NB, RBFN		0.65	0.58	0.62
[40]	Drug Reviews (26060)	Multi	Unigram, Bigram and Trigram	Corpus-based sentiment classification		0.89	0.79	0.83
Proposed	Drug Reviews (161297)	Multi	TF-IDF	RF, LR, SVM, CNN, CNN-LSTM	0.98	0.98	0.98	0.98
Proposed	Medical Reviews (32047)	Multi	TF-IDF	RF, LR, SVM, CNN, CNN-LSTM	0.97	0.97	0.97	0.97

5. Conclusion

We implemented a hybrid approach of unsupervised learning (K-means) and supervised learning (SVM, RF, LR, CNN, and CNN-LSTM). This makes it easier to classify tweets and medical texts into health-related categories that make sense for decision-making. Creators such as healthcare professionals, nurses, and even patients themselves from our collection of disease-related tweets and drug reviews, our approach identified approximately 44.3% and 66.8% of tweets and reviews, respectively, as preventive measures, etc. Categorized into categories, summarizes symptoms, causes, treatments, and drugs. The potential for curating useful decision-support information is attractive. We believed that social media had great potential to facilitate the sharing of health information and thus was an important source of targeted information. Based on our findings, health-related tweets are concentrated in PMs and TMs category, with social media users more interested in preventing disease outbreaks, treating symptoms, or finding ways to share them and suggest them.

Author contributions

Conceptualization, AK and SS; methodology, AK; software, AK; validation, AK and SS; formal analysis, SS; investigation, AK; resources, AK; data curation, SS; writing—original draft preparation, AK; writing—review and editing, SS; visualization, AK; supervision, SS; project administration, AK; funding acquisition, SS. All authors have read and agreed to the published version of the manuscript

Conflict of interest

The authors declare no conflict of interest

References

1. Khalifa NM, Elghany MMA, Elghany MMA. The Potential of Social Media in Emerging Supply Chain Management. International Journal of Service Science, Management, Engineering, and Technology. 2021, 12(4): 39-58. doi: 10.4018/ijssmet.2021070103
2. Wahi AK, Medury Y, Misra RK. Social Media. International Journal of Service Science, Management, Engineering, and Technology. 2014, 5(3): 1-15. doi: 10.4018/ijssmet.2014070101
3. O'Leary DE. Twitter Mining for Discovery, Prediction and Causality: Applications and Methodologies. Intelligent Systems in Accounting, Finance and Management. 2015, 22(3): 227-247. doi: 10.1002/isaf.1376
4. Gridach M, Haddad H, Mulki H. Empirical evaluation of word representations on Arabic sentiment analysis. In: Arabic Language Processing: From Theory to Practice: 6th International Conference, ICALP 2017. Fez, Morocco. 11-12 October 2017. Proceedings 6. Springer International Publishing. pp. 147-158.

5. Tariyal A, Goyal S, Tantububay N. Sentiment Analysis of Tweets Using Various Machine Learning Techniques. 2018 International Conference on Advanced Computation and Telecommunication (ICACAT). Published online December 2018. doi: 10.1109/icacat.2018.8933612
6. Kaur R, Kautish S. Multimodal Sentiment Analysis. Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines. Published online June 10, 2022: 1846-1870. doi: 10.4018/978-1-6684-6303-1.ch098
7. Bansal P, Kaur R. Twitter Sentiment Analysis using Machine Learning and Optimization Techniques. International Journal of Computer Applications. 2018, 179(19): 5-8. doi: 10.5120/ijca2018916321
8. Rathi M, Malik A, Varshney D, et al.. Sentiment Analysis of Tweets Using Machine Learning Approach. 2018 Eleventh International Conference on Contemporary Computing (IC3). Published online August 2018. doi: 10.1109/ic3.2018.8530517
9. Yasen M, Tedmori S. Movies Reviews Sentiment Analysis and Classification. 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). Published online April 2019. doi: 10.1109/jeeit.2019.8717422
10. Kumar A, Jaiswal A. Empirical study of twitter and tumblr for sentiment analysis using soft computing techniques. In: Proceedings of the world congress on engineering and computer science. 2017. Vol. 1, pp. 1-5.
11. Kamakshi P. Sentiment analysis on healthcare tweets. Indian J. Public Health Res. Develop., 2020. 11(6): 566-568.
12. Kaoud M. Investigation of Customer Knowledge Management. International Journal of Service Science, Management, Engineering, and Technology. 2017, 8(2): 12-22. doi: 10.4018/ijssmet.2017040102
13. Abuelenin S, Elmougy S, Naguib E. Twitter Sentiment Analysis for Arabic Tweets. Advances in Intelligent Systems and Computing. Published online August 31, 2017: 467-476. doi: 10.1007/978-3-319-64861-3_44
14. Al-Hadhrami S, Al-Fassam N, Benhidour H. Sentiment Analysis of English Tweets: A Comparative Study of Supervised and Unsupervised Approaches. 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). Published online May 2019. doi: 10.1109/cais.2019.8769550
15. Al-Khasawneh A. A Method for Classification Using Data Mining Technique for Diabetes. Psychology and Mental Health.: 738-761. doi: 10.4018/978-1-5225-0159-6.ch030
16. Alhanjouri M. Preprocessing techniques for Arabic documents clustering. International Journal of Engineering and Management Research (IJEMR), 2017. 7(2): 70-79.
17. Alsaeedi A, Zubair M. A Study on Sentiment Analysis Techniques of Twitter Data. International Journal of Advanced Computer Science and Applications. 2019, 10(2). doi: 10.14569/ijacsa.2019.0100248
18. Anitha Avula V, Asha A. Improving Prediction Accuracy Using Hybrid Machine Learning Algorithm on Medical Datasets.
19. Ariestya WW, Praptiningsih YE, Supriatin W. Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa. Jurnal Ilmiah FIFO, 2016. 8(1): 97-105. doi: 10.22441/fifo.v8i1.1304
20. Solanki VK, Cuong NHH, Zonghyu (Joan) Lu. Opinion Mining. Extracting Knowledge From Opinion Mining. Published online 2019: 66-82. doi: 10.4018/978-1-5225-6117-0.ch004
21. Gautam G, Yadav D. Sentiment analysis of twitter data using machine learning approaches and semantic analysis. 2014 Seventh International Conference on Contemporary Computing (IC3). Published online August 2014. doi: 10.1109/ic3.2014.6897213
22. Rajurkar, V. P. (2015). A Survey on Sentiment Analysis Techniques for Social Media Data. International Journal of Computer Applications, 120(10): 1–4. doi: 10.5120/21127-9044
23. Poongothai K, Vijayalakshmi R. Sentiment Analysis of Social Media Data using Machine Learning Techniques. International Journal of Advanced Science and Technology, 2021. 30(3): 3671–3681.
24. Priyanka P, Rekha T. Sentiment Analysis of Twitter Data using Machine Learning Techniques. International Journal of Computer Sciences and Engineering, 2020. 8(4): 589–595. doi: 10.26438/ijcse/v8i4.589595
25. Rajendran S, Kousalya M. Sentiment Analysis on Twitter Data Using Machine Learning Techniques. International Journal of Engineering Research in Computer Science and Engineering, 2017. 4(2): 63–67.
26. Sedhain, G. P., Bui, M., Nagarajan, V., & Raj, B. (2015). Hierarchical Attention Networks for Document Classification. Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1480–1489. doi:10.3115/N15-1162
27. Pershad Y, Hangge P, Albadawi H, et al.. Social Medicine: Twitter in Healthcare. Journal of Clinical Medicine. 2018, 7(6): 121. doi: 10.3390/jcm7060121
28. Statista. Most popular social networks as of January 2020, ranked by number of active users (2020). Available online: https://www.statista.com/statistics/272014/global-social-networks-ranked-bynumber- of-users/ Mention. Twitter engagement report 2018 (accessed on 18 November 2018).
29. Finfgeld-Connett D. Twitter and Health Science Research. Western Journal of Nursing Research. 2014, 37(10): 1269-1283. doi: 10.1177/0193945914565056
30. Gabarron E, Dorronzoro E, Rivera-Romero O, et al. Diabetes on Twitter: A Sentiment Analysis. Journal of Diabetes Science and Technology. 2018, 13(3): 439-444. doi: 10.1177/1932296818811679
31. Sedrak MS, Salgia MM, Decat Bergerot C, et al. Examining Public Communication About Kidney Cancer on Twitter. JCO Clinical Cancer Informatics. 2019, (3): 1-6. doi: 10.1200/cci.18.00088
32. Sinnenberg L, Buttenheim AM, Padrez K, et al. Twitter as a Tool for Health Research: A Systematic Review.

American Journal of Public Health. 2017, 107(1): e1-e8. doi: 10.2105/ajph.2016.303512
33. Chen K, Zhang Z, Long J, et al. Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Systems with Applications. 2016, 66: 245-260. doi: 10.1016/j.eswa.2016.09.009
34. Ali D, Missen MMS, Husnain M. Multiclass Event Classification from Text. He L, ed. Scientific Programming. 2021, 2021: 1-15. doi: 10.1155/2021/6660651
35. Divya, Y. A. (2018). Opinion Based Learning Model in Medical Sector. International Journal of Scientific Research in Computer Science. Engineering and Information Technology, 4(6), 607–609.
36. Emadi M, Rahgozar M. Twitter sentiment analysis using fuzzy integral classifier fusion. Journal of Information Science. 2019, 46(2): 226-242. doi: 10.1177/0165551519828627
37. Vijayaraghavan S, Basu D. Sentiment analysis in drug reviews using supervised machine learning algorithms. arXiv:23.11643.
38. Goeuriot L, Na JC, Min Kyaing WY, et al. Sentiment lexicons for health-related opinion mining. Proc. 2nd ACM SIGHIT Symp. Int. Health Informat. (IHI), 212, pp. 219-226.
39. Asghar MZ, Ahmad S, Qasim M, et al. SentiHealth: creating health-related sentiment lexicon using hybrid approach. SpringerPlus. 2016, 5(1). doi: 10.1186/s40064-016-2809-x
40. Liu S, Lee I. Extracting features with medical sentiment lexicon and position encoding for drug reviews. Health Information Science and Systems. 2019, 7(1). doi: 10.1007/s13755-019-0072-6
41. Go A, Bhayani R, Huang L. Twitter Sentiment Classification using Distant Supervision Processing. Journal on Processing, 2009. 1(12), 1–6.
42. Gohil S, Vuik S, Darzi A. Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health and Surveillance. 2018, 4(2): e43. doi: 10.2196/publichealth.5789