Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 06 Issue 02
2024

Research Article
https://doi.org/10.56979/602/2024

A Hybrid Machine Learning Model to Predict Sentiment Analysis on 𝕏

Fiza Malik[1], Muhammad Fuzail[1], Naeem Aslam[1], Ramla Sarwar[2], Kamran Abid[1], Muhammad Sajid Maqbool[2], and Anum Yousaf[1]

[1]Department of Computer science, NFC Institute of Engineering and Technology, Multan, Pakistan.
[2]Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan.
*Corresponding author: Muhammad Sajid Maqbool. Email: sajidmaqbool7638@gmail.com

Received: January 18, 2024 **Accepted:** February 20, 2024 **Published:** March 01, 2024

Abstract: Social media, particularly Twitter now 𝕏, have emerged as pivotal arenas for sentiment analysis due to their pervasive nature and significant impact on shaping opinions. Our research delves into Roman-Urdu sentiment analysis within the burgeoning realm of social media, addressing a significant gap in research. Leveraging machine learning techniques, it emphasizes the scarcity of sentiment analysis studies in this linguistic domain, specifically on platforms like Twitter. The methodology involves meticulous data collection from English and Roman-Urdu tweets, followed by comprehensive preprocessing to refine and enhance dataset quality in python. Feature extraction retrieves key characteristics like subjectivity and polarity, enabling a nuanced sentiment analysis. Our technique evaluates precision, don't forget, F1 rating, and accuracy metrics the use of a complete evaluation framework on 4 machine learning classifiers: Naïve Bayes (NB), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM) algorithms. Roman Urdu sentiment analysis has advanced way to the results, which show how nicely this method works to classify all three sentiments (Hate, Offensive, and Neither) in multilingual social media content.

Keywords: Machine Learning; Sentiment Analysis; Twitter; Roman-Urdu.

1. Introduction

The use of social media systems including Facebook, Twitter, YouTube, Tumblr, Reddit, and Quora to have interaction with loved ones and trade mind and experiences on a sure man or woman, item, service, or enterprise is growing in recognition [1]. End customers can now speedy look at a product's exclusive features or form evaluations about it with the aid of analyzing different customers' critiques which have been posted on-line way to the arrival of the net and the accessibility of consumer-generated information [2]. Sentiment analysis is presently one of the most researched subjects in text statistics mining and natural language processing. The region of gadget getting to know, or NLP, uses neural networks and computational language techniques to permit computers to recognise and interpret human language, thereby bridging the verbal exchange hole among humans and computers [3]. Sentiment evaluation (SA) is a methodological technique this is utilized to enhance performance metrics and offer facts to selection-making approaches via assessing human beings's emotions approximately items, services, and one-of-a-kind groups. On the opposite hand, the large use of social media, e-commerce, and assessment websites has resulted in a large quantity of multilingual and multicultural statistics, making thorough analysis in a unmarried language unfeasible. This hassle is effectively addressed by way of multilingual sentiment analysis (MSA), which makes use of resource-rich languages to guide pass-lingual sentiment analysis, codeswitching configurations, or the creation of language-neutral models that accommodate a number of linguistic contexts [4]. Research research propose that Twitter might be a beneficial supply of facts for assessing events that arise worldwide. With Twitter being the maximum popular social media website and information supply, human beings may additionally specific themselves on this platform and many other structures. Examining the content that is shared in this platform gives an possibility to extract crucial insights into the mind, emotions, and standard attitudes of the general public, which in turn offers a greater complicated knowledge of the views of the overall public [5].

The improvement of Internet offerings has led to the upward thrust of on-line structures as a main method of self-expression for users. This boom is especially significant while thinking about the context of on line discourse, where individuals freely talk and proportion evaluations on a wide range of subjects. Additionally, consumer-generated content is not unusual on a whole lot of e-commerce websites wherein a substantial number of people post critiques and feedback about a wide variety of services and products. Field specialists are also actively involved within the digital realm, frequently sharing knowledge about their fields thru apps like Twitter and different online systems. In the sector of training, sentiment and emotion evaluation will become critical for teachers and students alike. Notably, social media sites like Facebook and Twitter are vital channels for the trade of thoughts and viewpoints, and educational institutions are relying more and more on these websites to facilitate successful community participation and communique [6]. These days, Twitter serves as a brief-textual content microblog that allows many styles of content material—textual content, pics, films, and links—to be shared instantly. To put it every other manner, the platform sees more than 500 billion tweets posted annually, or approximately 6,000 tweets every 2nd. Users of the internet site are capable of engage with the facts thru the use of functions like "liking," "bookmarking," "replying," and "reposting." Moreover, users have the potential to make use of features like "quoting" and "replying" in reaction to the preliminary tweet, which provides to the platform's dynamic and interactive features [7]. One vicinity of text-mining studies that focuses on extracting critiques from written messages or correspondences is known as sentiment assessment (SA). Text ought to be classified into accurate, awful, and impartial sentiments as a part of the approach. To deal with the analytical problem, empiricists have advanced lots of approaches, together with lexical approach, hybrid strategies, and tool getting to know (ML) strategies. The lexicon-based totally completely technique uses a sentiment lexicon to determine the text's sentiment polarity. Conversely, hybrid era combines device studying and lexical procedures. Notably, the machine learning approach constructs its analytical framework by leveraging linguistic factors such as word frequency, term presence, and TF-IDF. This technique eschews reliance on prior information, such as sentiment lexicons, in its pursuit of sentiment polarity determination [8]. Sentiment analysis can be conducted through various modalities, including aspect-wise, sentence-wise, and document-wise approaches. Additionally, the classification process may unfold through either unsupervised or supervised methodologies [9]. Notably, on online platforms, individuals often contribute comments in languages divergent from English, Hindi, Urdu, Chinese, Arabic, Japanese, or other commonly studied languages. Despite the prevalence of sentiment analysis research predominantly conducted in English, non-native speakers frequently resort to Roman characters when expressing sentiments in their native languages. It is pertinent to acknowledge that the transliteration of Urdu words into Roman script lacks standardized linguistic norms, particularly in the widely spoken Subcontinental Urdu language [10].

Machine learning techniques have been applied on the Twitter platform for the purpose of categorizing and discerning instances of hate speech, encompassing thematic domains associated with immigration, refugees, sexism, racism, and misogyny. These classifications involve binary, multiclass, or a combination thereof. Predominantly, supervised learning methodologies, such as LR, SVM, and DT algorithms, have been extensively employed to predict the categorical classification of tweets. Additionally, advanced tree algorithms like Gradient Boosting and RF have found application in this context. Enhancements to these algorithms have been realized through the integration of techniques such as meta-learning, weighted and majority voting to augment their performance. Furthermore, optimization efforts encompassing parameter and hyper parameter fine-tuning have contributed to the refinement of these machine learning models [11].

Within the scope of this investigation, we proffer a methodology for discerning sentiment in tweets composed in both Roman-Urdu and English. The domain of sentiment analysis within the context of Twitter data remains an expansive realm with considerable avenues for further exploration. Our dataset is underpinned by English and Roman-Urdu dataset, collectively integrating 32,067 tweets from the aforementioned linguistic sources. The proposed methodology employs NLP techniques implemented in Python to adeptly extract optimized features from Roman-Urdu tweets, thereby generating a dataset conducive to interpretation by machine learning tools for subsequent model training. To make this approach viable, we apply the functionalities of the Google Colab tool, utilizing its capabilities for the construction of machine learning models across various classifiers. The study utilized four machine learning classifiers—NB, RF, DT, and SVM—to accurately classify users' emotional behavior into three categories: Neither (neutral), Hate (negative), and Offensive (positive). The subsequent sections of the paper are organized as follows: Section 2 reviews the literature focusing on machine learning approaches in diverse languages, particularly in sentiment analysis. Section 3 introduces the hybrid sentiment analysis model leveraging both NLP and ML techniques. In section 4 efficacy of this system is assessed. Lastly, Section 5 outlines the conclusions derived from the detection of sentiments.

2. Literature Review

Using NLP approaches in Python, this article examines public opinion and response to the Omicron form of the COVID-19 virus. Using 8,073 tweets from the third wave of the pandemic, the analysis discovered that the "Neutral" group had the highest forecast accuracy. Deep learning approaches such as Bidirectional Encoding Representation by Transformer might be used in future improvements [12]. Analyzing customer interests through Roman Urdu tweets and Google Map ratings in Pakistan's fashion business. It uses 15,000 tweets and 6,000 reviews to construct five clusters for each fashion firm. DistilBERT is used in the study for sentiment classification and sentiment analysis. The findings are experimentally validated using Cohen's Kappa, suggesting a modest agreement between human and machine-generated conclusions. This study provides an efficient method for determining user preferences in the fashion sector [13]. Multilingual sentiment analysis on Twitter, an area that has received less attention than English sentiment analysis. To convert non-English language tweets into English, the Multilingual Twitter Sentiment Analysis (MLTSA) technique employs NLP and SVM. When integrated with SVM, the algorithm demonstrated exceptional accuracy. The system employs Google Translator for translation and a diverse range of ML algorithms for data analysis. However, the study recognizes shortcomings, notably in dealing with code-switching and code-mixed phrases, which require additional investigation [14].

Twitter data from March to May 2020 to examine public attitude in nine major cities about the COVID-19 epidemic. It discovered that negative views peaked in mid-March, followed by a rise in positive feelings in early May. Positive attitudes and severe quarantine measures were shown to be associated with New York City and London. The findings emphasized the need of monitoring public mood for pandemic strategies and recommended that future studies take into account internet vernacular and complex emotions [15].

A unique approach that combines fasttext embedding, CNN, and LSTM models to maximize accuracy for early depression identification using textual data from social media. Chatbots might be used in future studies for virtual patient interaction and automated medical diagnostic systems [16].

With recent clustering methods that include K-means and DENCLUE approaches, Twitter is a great source for sentiment analysis. These algorithms categorize attitudes as good, negative, or neutral, providing a competitive advantage. The K-DENCLUE-IM method is especially efficient. Future studies will look into identifying sarcasm, irony, and emoticon use [17]. Using Twitter data, focus is to explore the importance of sentiment analysis in measuring public attitude during the COVID-19 epidemic. It examines sentiment analysis algorithms based on machine learning and discovers that ensemble models, such as BERT and RoBERTa, perform better. The research highlights the increased interest in using social media sentiment analysis to better understand public emotions and behaviors [18]. By analyzing sentiment in Instagram text, the study investigates emotional connections on social media sites. It surpasses prior models by developing a sentiment analysis model utilizing deep learning and NLP approaches. The model extracts features and evaluates sentiment, which helps with text sentiment analysis and analyzing social media platform sentiment trends [19].

To analyze face sentiment in social media photographs and videos, the paper introduces a hybrid ML approach that unifies CNN and SVM. The programme correctly assessed images of 25,400 faces from a major public event, attaining 89 percent accuracy for sentiment analysis. The efficacy of the model is related to its capacity to extract latent attribute information in areas such as the eyes and lips [20]. Focusing specifically at breastfeeding sentiment on Twitter, concentrating on nursing moms' experiences and contributing variables. It evaluates sentiment polarity using lexicon strategy and ML classifiers. Breastfeeding behaviors are influenced by health-related, social, psychological, and environmental variables, whereas positive elements include perceived advantages, maternal self-efficacy, education, and social support [21]. Exploring sentiment analysis in NLP and progressive advancements in deep learning systems. It underscores shift from classic RNN and CNN language models to transformer based language models. Transformer models, which are based on large amounts of text data, have transformed NLP by transferring information to downstream tasks. The study emphasizes continuing research to bridge the language resource gap between plentiful and limited languages, emphasizing the significance of developing multilingual systems capable of excelling at monolingual tasks [22].

Investigates sentiment analysis (SA) research by reviewing 112 papers and identifying major topics, methodology, and approaches. It demonstrates that LSTM and CNN algorithms are commonly utilized in SA, particularly in social media environments. Domain and language dependencies, opinion spam detection, and successful DL application are all challenges. Future directions include the development of interpretable models, the creation of additional datasets, and the investigation of dynamic sentiment analysis [23]. Cross-Lingual Sentiment Analysis (CLSA) in limited resource language, concentrating on the difficulties and solutions. It provides a comparative examination of pretrained language models, highlighting their benefits and drawbacks. The relevance of cross-lingual representation in minimizing language barriers in natural language processing is emphasized in the paper. However, it observes a reduction in the performance of pre-trained models in detailed settings and crosslingual tasks [24]. Cross-Lingual Sentiment Analysis (CLSA) research, highlighting its phases and Word Embedding techniques. It demonstrates that having a similar language family minimizes CLSA results, hence increasing model performance. Natural language processing has been transformed by PTMs such as BERT, GPT 2 and 3. CLSA's future path is heavily reliant on PTMs and ways to handle issues such as picking source languages and attaining consistent performance across varied language pairs [25].

The changing environment of Aspect-Based Sentiment Classification (ABSC) research, classifying models as knowledge-based, machine-learning, or hybrid. It explains these strategies in technical and understandable terms, showcases cutting-edge models, and investigates methodologies for input and output description. The paper also discusses present trends and future possibilities for ABSC research, providing useful insights on the field's growth and innovation [26]. Textual Emotion Analysis (TEA), with an emphasis on its roots in Deep Learning (DL) approaches. It defines emotions, categorization algorithms, and application areas, as well as identifying several DL technologies and learning methodologies for word and phrase representations. The work intends to address existing constraints and investigate future research areas in TEA, with the goal of contributing to the improvement of TEA methodology [27]. Using the Arabic ChatGPT tweets dataset, this study investigates ChatGPT's capabilities and limits. It utilizes the TextBlob Arabic Python module to classify tweets as favorable, bad, or neutral. DistilBERT, RoBERTa, and XLNet-based models are utilized, with a unique hybrid model displaying considerable performance benefits [28].

Aspect-Based SA to improve consumer review analysis on online platforms such as Twitter by proposing a dataset. It employs a variety of deep learning models, including transformer-based models such as AraBERT and GigaBERT, as well as LSTM and SVM. AraBERT and GigaBERT outperform other models in tasks such as aspect word identification, category classification, and polarity. The authors want to broaden the dataset and fix issues [29]. Multilingual language models to investigate sentiment analysis to measure the shift of sentiment detection skills from rich to constrained resources. It contrasts techniques based on Italian resources and knowledge transfer in English. The findings demonstrate the efficiency of information transmission mechanisms in multilingual situations, although they admit constraints such as limited sample sizes. Additional datasets and multi-class settings should be considered in future study [30]. A revolutionary word weighting mechanism presented that improves information retrieval and text classification tasks. DFST beats conventional systems in prediction accuracy, displaying considerable increases. Further research might be conducted to investigate its domain-specific effects [31].

The emphasis on the area of recognizing emotions and personalities from text, with an emphasis on sentiment and emotion analysis. It emphasizes the significance of collaboration between emotional and personality psychology, as well as cutting-edge approaches and future research avenues. It emphasizes the importance of using personality models with care [32]. The COVID-19 crisis has ignited discussions online, particularly concerning its effects on the elderly population. A novel approach named "TClustVID" analyses COVID-19 tweets using clustering-based categorization and topic extraction, finding important issues and categorizing them based on sentences [33]. On Twitter data, with an emphasis on Indian languages with distinctive code-mixing and multilingual properties, sentiment analysis is done. To increase sentiment classification accuracy, it employs composite kernel functions and Support Vector Machine (SVM). The study outperforms DL techniques in terms of performance and encourages future research for deeper learning models [34]. The study delves into the difficulties of interpreting sentiment variations on social media, specifically Twitter, and selects VADER as the most successful sentiment analysis classifier. Filtered-LDA, a novel framework for extracting interpretable emergent themes, and a sentiment reasoning panel are introduced [35]. Describing a unique architecture for recognizing and understanding Arabic words on social media that combines CNN and RNN and achieves outstanding accuracy. This model outperforms previous approaches, highlighting the significance of combining characterlevel CNN and subword-level RNN modules [36]. ML algorithms are applied in SA to interpret emotions in text data, categorizing information as positive, negative, or neutral. SVM, LR, and NB techniques enhance accuracy and recall. When compared to SVM and NB, Logistic Regression has the greatest accuracy [37]. Thorough research on the difficulties of language variety and limited resources in sentiment analysis on social media. It uses zero-shot transfer learning to assess the efficiency in cross-lingual context of word embeddings, attaining accuracy of 60 percent across two Hindi corpora [38]. Exploring sentiment analysis on online sites, focusing on diagnosing anxiety or depression states using machine and deep learning methodologies. It underscores text, emoticons, and emojis significance in sentiment analysis. Additional data sources such as biometric data, facial expressions, speech signals, and EEG signals should be considered for future study [39].

The COVID-19 rampant had a major influence on global health, showing societal attitudes during epidemics. This study looks at infectious illness literature from 2010 to 2020, with an emphasis on sentiment analysis and disease reduction. It underscores the significance of understanding public opinion, digital media, and sentiment evaluation, advocating for AI and ML fusion for readiness [40]. Highlighting the challenges faced in sentiment evaluation when identifying hate speech within Urdu tweets. To address challenges such as high dimensionality, sparsity, and class imbalance, researchers applied the Synthetic Minority Oversampling Technique, adoptive stop words filtering, and variable feature selection scheme. The paper underscores the necessity of resolving these issues and offers future research directions [41].

Table 1. Comparison of previous research work

Ref	Year	Language	Dataset	Method
(Bengesi et al., 2023)	2023	English	Monkeypox tweets.	ML Algorithms
(Fazal et al., 2023)	2023	English	Covid-19 Twitter Data.	ML Algorithms, NLP
(Mujahid et al., 2023)	2023	Arabic	Arabic tweet ChatGPT related.	BoW,TFIDF, LR,DT,RF,SVM, KNN
(Al-Jarrah et al., 2023)	2023	Arabic	4,880 Arabic tweets.	aspect-based sentiment analysis
(Chan & Yang, 2023)	2023	English	Instagram Dataset.	ML,NLP
(Susmitha et al., 2023)	2023	English	Twitter Dataset.	ML Algorithms
(Catelli et al., 2022)	2022	Italian	BERT.	Tripadvisor reviews
(Tahir & Asif Naeem, 2022)	2022	English	Google and twitter map reviews.	ML.LDA,LSA
(Jain et al., 2021)	2021	English	Airline quality, airline tweets.	ML ,NLP, DL models
(Rana et al., 2022)	2021	Roman-Urdu	RUSA 19, UCI Roman Urdu.	Rule-based Model
(Wijayanti & Arisal, 2021)	2021	Bahasa	Twitter Data.	CBOW and skip-gram
(Oyebode et al., 2021)	2021	English	19,551 tweets.	lexicon-based ML approach
(Ali et al., 2021)	2021	English	Twitter API.	ML Algorithms
(Wandabwa et al., 2021)	2021	English	Generic Tweets.	LDA
(Arun & Srinagesh, 2020)	2020	English	Twitter API.	ML Algorithms
(F. Mehmood et al., 2020)	2020	Roman-Urdu	DSL Roman Urdu.	ML,DL Models
(Oriola & Kotze, 2020)	2020	African	South-African Tweets.	ML Algorithms

(Rehioui & Idrissi, 2020)	2020	English	Tweets.	ML Algorithms
(K. Mehmood et al., 2019)	2019	Roman-Urdu	Websites and social media data.	ML Algorithms

Table 1, shows the comparison of existing approaches while Table 2 shows comparison with our proposed methodology.

Table 2. Comparison of proposed research with previous literature.

Ref.	Year	Technique	Dataset	Accuracy	Precision	Recall	F-1 score
(Jain et al., 2021)	2021	ML ,NLP, DL models	Airline quality, airline tweets.	90.2	87.8	87	87.6
(Tahir & Asif Naeem, 2022)	2022	ML. LDA, LSA techniques	Twitter & Google maps reviews.	80	N/A	N/A	73
(Mujahid et al., 2023)	2023	BoW, TFIDF, ML models	Arabic tweets. ChatGPT related	78	83	79	79
(Al-Jarrah et al., 2023)	2023	LR, KNN, Aspect-based sentiment analysis	4,880 Arabic tweets.	N/A	79	70	71
(Wijayanti & Arisal, 2021)	2021	CBOW & skip-gram	Twitter data.	80.9	80.5	98.9	88.8
(Shah et al., 2023)	2022	BERT	TripAdvisor reviews.	86	N/A	N/A	91
(Oriola & Kotze, 2020)	2020	ML algorithms	South African tweets.	92	69	58	61
(Oyebode et al., 2021)	2021	Lexicon-based ML approach	19,551 tweets.	N/A	74	73	73
(Susmitha et al., 2023)	2023	ML algorithms	Twitter data.	80	80	81	80
(Fazal et al., 2023)	2023	ML ,NLP technique	COVID-19 twitter data.	89.8	89	82	85
Proposed	2023	ML, NLP technique	Twitter Dataset	94	94	93	93

3. Materials and Methods

This chapter outlines the presented methodology for conducting sentiment analysis on Roman-Urdu tweets, exploring four Machine Learning (ML) algorithms for assessing emotional behavior classification. The study utilizes two distinct datasets containing English and Roman-Urdu tweets, employing ML algorithms such as NB, RF, DT, and SVM to categorize emotional states into Hate(negative), Offensive (positive), and Neither(neutral) categories. Figure 1 delineates the dataset creation and selection process.

3.1. Data Acquisition Process

Our selected two Twitter datasets—English and Roman-Urdu tweets—comprising attributes such as Text and Sentiment. The integrated dataset underwent the following processes of dataset downloading, preprocessing, feature extraction, and model evaluation.

3.2. Data Preprocessing

A pivotal stage involves fundamental tasks, including the removal of stop words, punctuation marks, special characters, URLs, and stemming. These methods aim to enhance the dataset quality by eliminating noise and redundant content from English tweets shown in Table 3. and Roman-Urdu tweets shown in Table 4.

3.2.1. Stop Word Removal

Python's NLTK Library is employed for stop words removal in English tweets, whereas a custom list is created for Roman-Urdu, encompassing non-technical phrases such as "ai," "ayi," "hy," "hai," "main," "ki," "koi," "tha," sy," "wo," "ko," "ky,"and others.

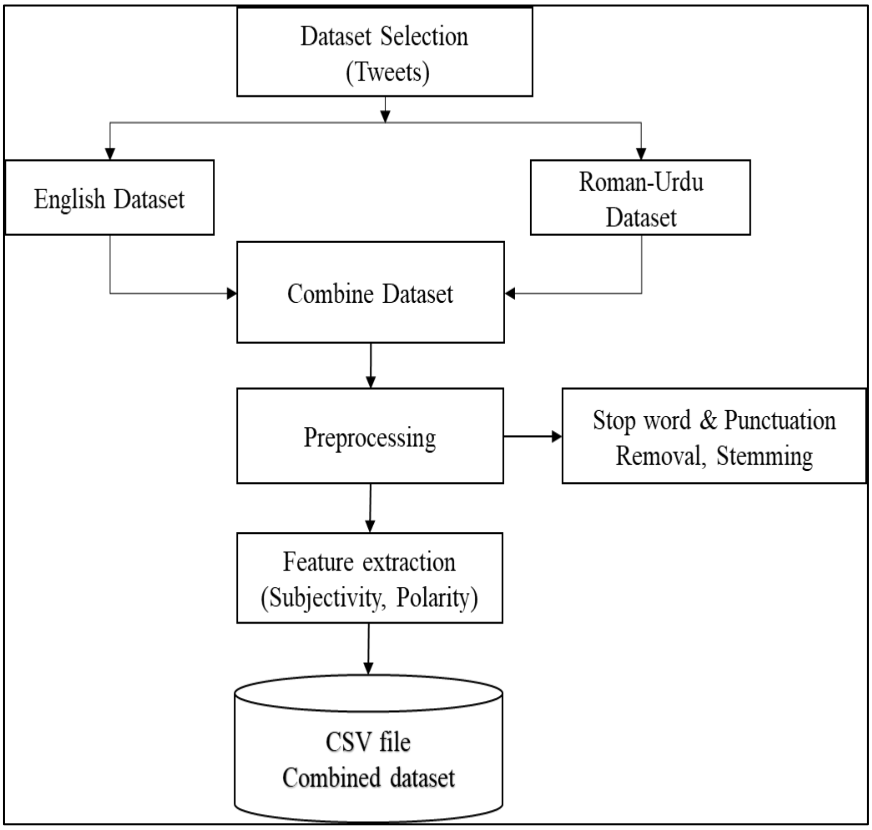

Figure 1. Dataset Selection and Creation Process

3.2.2. Punctuation Removal

Following forestall terms elimination, punctuation marks are eliminated the use of the NLTK package deal, an important step in text preprocessing that affects textual content-processing techniques reliant on word frequency.

3.2.3. Stemming

Stemming, a text normalization process, truncates terms to their root shape, reaping rewards text evaluation and system performance.

3.3. Feature Extraction

Characteristics that incorporates subjectivity and polarity are extracted from preprocessed tweets. Subjectivity and polarity values are forecasted by way of manner of those features, contributing to the best prediction of emotional conduct in tweets.

3.4. Assessment Framework

The assessment framework employed in this study integrates ML models (NB, SVM, RF, and DT) and metrics like F-1 score, recall, precision and accuracy. The framework aims to rigorously evaluate the effectiveness of ML models used in sentiment analysis for Roman Urdu and English tweets. It encompasses data cleaning, labeling, dataset splitting (75% training, 25% testing), and evaluation of metrics on all ML models based on accuracy. A comparative analysis of Accuracy along with other evaluation metrics is conducted on the basis of sentiments (Offensive, Hate, Neither). This comprehensive approach offers insights into the models' efficacy in identifying sentiments within multilingual social media content.

Table 3. English Tweets Preprocessing Steps.

English Tweets Preprocessing
Tweet Text:
"Just finished an amazing book called 'The Alchemist' by Paulo Coelho! Highly recommend it! 📚🌟 #booklovers"
Stop Words Removal:
"Just finished amazing book called 'The Alchemist' Paulo Coelho! Highly recommend 📚🌟 #booklovers"
Punctuation Removal:
"Just finished amazing book called The Alchemist Paulo Coelho Highly recommend 📚🌟 booklovers"
Special Character Removal:
"Just finished amazing book called The Alchemist Paulo Coelho Highly recommend booklovers"
Text after Stemming:
"Just finish amaz book call the alchemist paulo coelho high recommend booklov" Final Clean Text:
"just finish amaz book call alchemist paulo coelho high recommend booklov"

Table 4. Roman-Urdu Tweets Preprocessing Steps.

Roman-Urdu Tweets Preprocessing
Tweet Text:
"Khadi par 50 % ki sale hai wo bhi '20 November' tak hogi! Highly recommend it! 📚🌟 #blessedsale"
Stop Words Removal:
"Khadi par 50 % sale wo bhi '20 November' tak hogi! Highly recommend 📚🌟 #blessedsale"
Punctuation Removal:
"Khadi par 50 % sale wo bhi 20 November tak hogi Highly recommend 📚🌟 blessedsale"
Special Character Removal:
"Khadi par 50 % sale wo bhi 20 November tak hogi Highly recommend blessedsale"
Text after Stemming:
"Khadi par 50 % sale wo bhi 20 november tak ho high
Final Clean Text:
"khadi par 50 % sale wo bhi 20 november tak ho high recommend blesssale"

3.5. Used ML Models

Four ML models—NB, SVM, DT, and RF—are employed to assess tweet sentiments.

3.5.1. Evaluation Metrics

The four most commonly used evaluation metrics to assess the performance of ML models, each providing unique insights into the model's proficiency in classification tasks.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F-1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{4}$$

The results of the used models are calculated by the above-mentioned equations. The detail process is also shown in Figure 2.

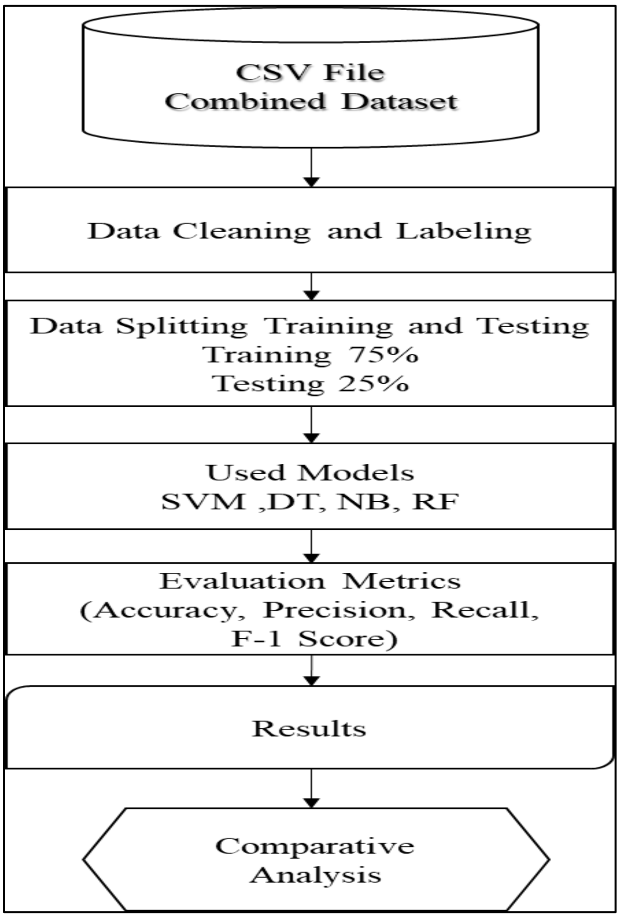

Figure 2. Assessment Framework.

4. Results

4.1. Models' Performance Assessment

All ML models used in this research to perform evaluation are NB, DT, RF, and SVM. The results of the evaluation are categorized as Accuracy, Precision, Recall, F-1 score.

4.1.1. Models' Accuracy

Accuracy of all four ML models used in this research are shown in Table 5. RF accuracy performance is 94 percent at Training while 93 at Testing dataset. Following with DT performing 93 percent at both training and testing dataset.

Table 5. Accuracy.

Models	Training Set	Testing Set
NB	80	80
DT	93	93
RF	94	93
SVM	74	73

NB is third at accuracy performance following with SVM the lowest performed model are shown graphically in Figure 3.

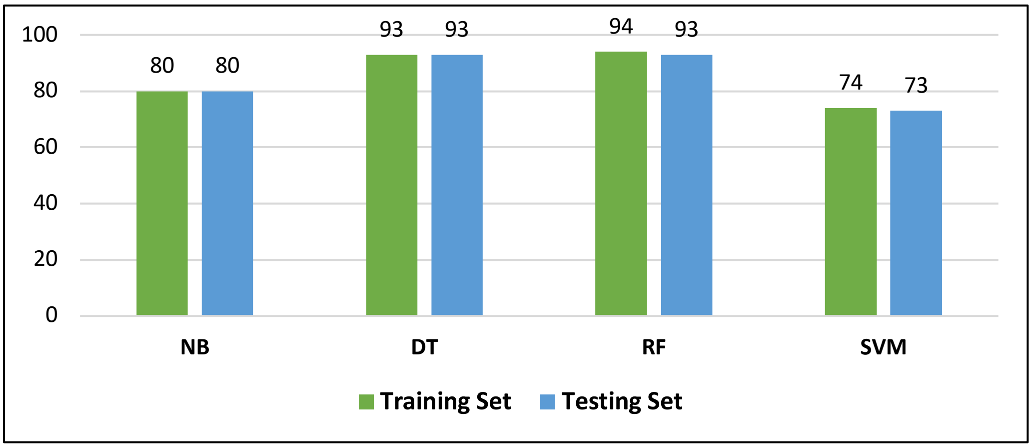

Figure 3. Accuracy of Models.

4.1.2. Models' Precision

Precision of all classes Hate (negative), Offensive (positive), and Neither (neutral) is shown in Table 6. DT is at maximum at Hate class with precision of 99 percent. NB, DT, and RF all three give 93 percent precision at Offensive class.

Table 6. Precision.

Models	Hate	Offensive	Neither
NB	57	93	0
DT	99	93	90
RF	96	93	90
SVM	50	87	19

NB and SVM give 57 and 50 precision at Hate class performing the poorest among all other classifier. NB and SVM performance along With outperforming DT and RF are shown in Figure 4.

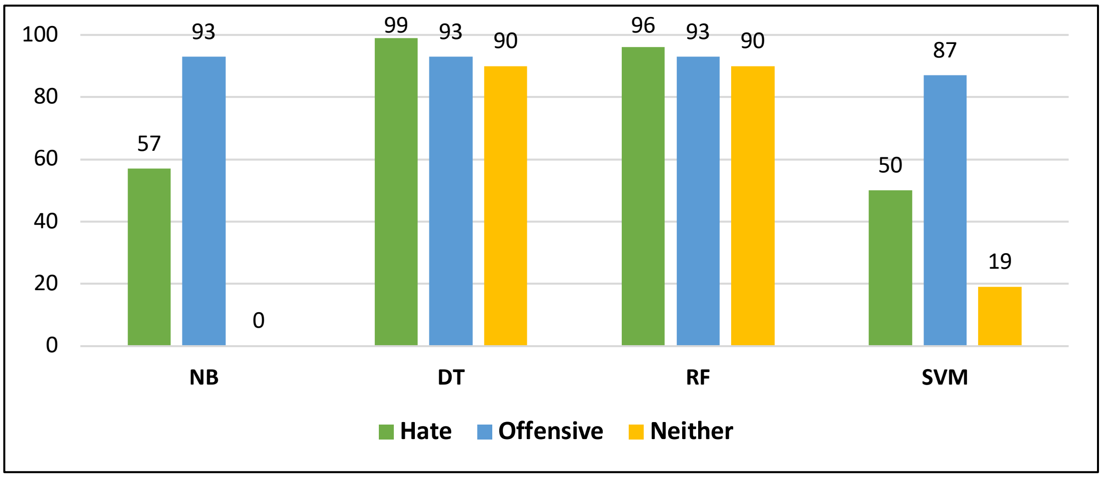

Figure 4. Precision of Models.

4.1.3. Models' Recall

Recall of all classes Hate (negative), Offensive (positive), and Neither (neutral) is shown in Table 7.NB is minimum at Neither class, following by SVM the second minimum at Neither class. DT and RF are maximum at Offensive class with recall of 99 percent. NB outperformed all other classifier at Hate class with the recall of 87 percent.

Table 7. Recall.

Models	Hate	Offensive	Neither
NB	83	96	0
DT	77	99	98
RF	77	99	94
SVM	54	96	08

Neither class giving zero to 08 percent recall of NB and SVM. Both NB and SVM classifier are shown graphically in Figure 5.

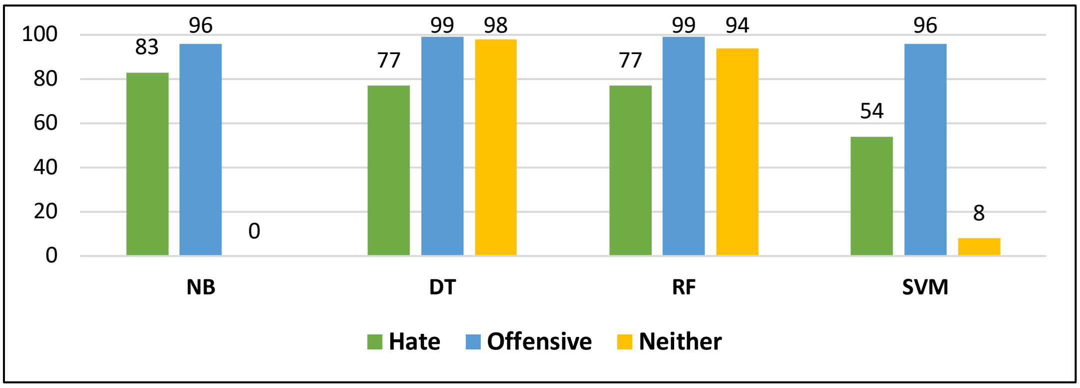

Figure 5. Recall of Models.

4.1.4. Models' F-1 score

F-1 score of all four ML models are shown in Table 8 below with sentiments including Hate (negative), Offensive (positive), and Neither (neutral). All classifiers showed highest F-1 score of Offensive class. While NB and SVM give the lowest scores of Neither class. DT and RF give the highest F-1 score for Hate class with DT outperforming RF.

Table 8. F-1 score.

Models	Hate	Offensive	Neither
NB	68	94	0
DT	87	96	93
RF	86	96	92
SVM	52	91	11

Overall performance of NB and SVM classifier is the lowest compared to DT and RF is shown graphically in Figure 6.

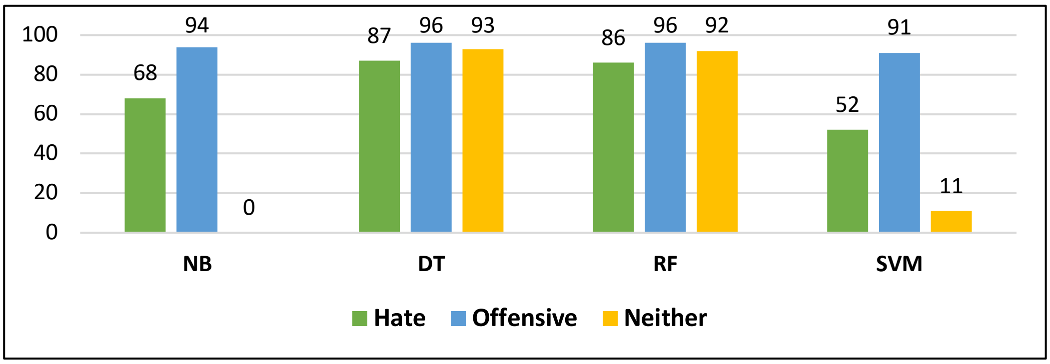

Figure 6. F-1 score of Models.

4.2. Sentiment Word Cloud

Sentiment word clouds are a visually appealing representation of the prevalent emotions found in the Roman Urdu and English Twitter collections. These clever and visually attractive word clouds provide a comprehensive representation of the most popular phrases associated with both positive and negative attitudes across a wide spectrum of social media content. By highlighting the dominant subjects and phrases within the dataset, the graphical representations aid in the intuitive grasp of the sentiment landscape within the analyzed social media discussion.

4.2.1. Offensive Sentiment Cloud

The word cloud depicting offensive close to positive feelings in Roman Urdu and English tweets is shown in Figure 7. It shows the most common terms linked to offensive sentiments.

Figure 7. Offensive Sentiment cloud.

4.2.2. Hate Sentiment Cloud

The hybrid dataset's most common terms associated with negative (Hate) attitudes are shown in Figure 8: negative sentiment word cloud.

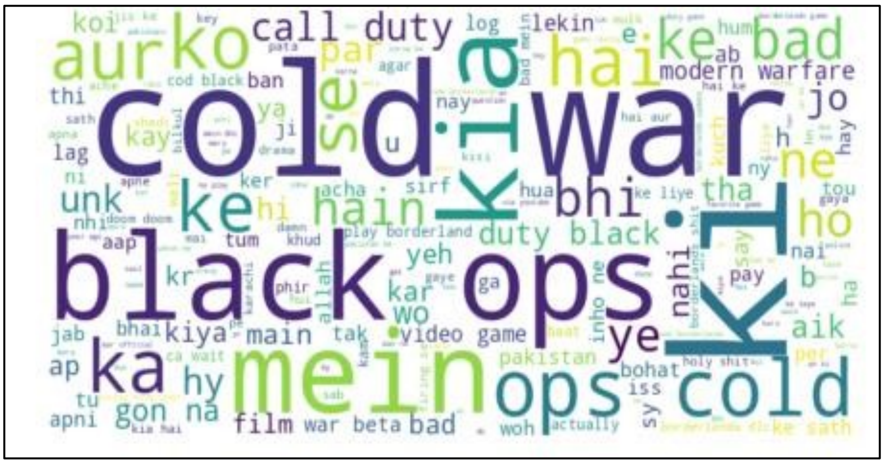

Figure 8. Hate Sentiment Cloud.

5. Conclusions

Using Machine Learning techniques, this work studies Roman Urdu sentiment analysis. Its goal is to identify roadblocks and give effective solutions. The study gathered and preprocessed English and Roman Urdu tweets. Decision Tree scored remarkably well in all evaluation metrics, according to the data. It outperformed Random Forest in both positive and negative sentiment categorization. Random Forest, on the contrary, was the most accurate model, with training and testing accuracy of 94% and 93%, respectively.

This study sets the basis for more effective sentiment analysis tools in Roman Urdu and English discussions. This study establishes the foundation for prospective investigation into sentiment analysis of Roman Urdu and English tweets. Prospective research efforts might concentrate on improving machine learning models to handle the many subtleties inherent in sarcasm, cultural context, and the vernacular idioms ubiquitous in social media chats. The fusion of neo deep learning models and sophisticated NLP methodologies has the potential to improve sentiment classification precision, particularly in addressing the complexities associated with ambiguous sentiments and contextually specific disparity.

Conflicts of Interest: No competing interests have been declared by the authors.

Funding: This research received no external funding.

Data Availability Statement: The dataset analyzed during this current study is available from the corresponding author on reasonable request. For data inquiries, email: sajidmaqbool7638@gmail.com.

References

1. Abdullah, Tariq, and Ahmed Ahmet. 2023. "Deep Learning in Sentiment Analysis: Recent Architectures." ACM Computing Surveys 55(8): 1–37.
2. Agüero-Torales, Marvin M., José I. Abreu Salas, and Antonio G. López-Herrera. 2021. "Deep Learning and Multilingual Sentiment Analysis on Social Media Data: An Overview." Applied Soft Computing 107: 107373. https://doi.org/10.1016/j.asoc.2021.107373.
3. Al-Jarrah, Ibrahim, Ahmad M. Mustafa, and Hassan Najadat. 2023. "Aspect-Based Sentiment Analysis for Arabic Food Delivery Reviews." ACM Transactions on Asian and Low-Resource Language Information Processing 22(7).
4. Alamoodi, A. H. et al. 2021. "Sentiment Analysis and Its Applications in Fighting COVID-19 and Infectious Diseases: A Systematic Review." Expert Systems with Applications 167(April 2020): 114155. https://doi.org/10.1016/j.eswa.2020.114155.
5. Alattar, Fuad, and Khaled Shaalan. 2021. "Using Artificial Intelligence to Understand What Causes Sentiment Changes on Social Media." IEEE Access 9: 61756–67.
6. Ali, Muhammad Z. et al. 2021. "Improving Hate Speech Detection of Urdu Tweets Using Sentiment Analysis." IEEE Access 9: 84296–305.
7. Arun, K., and A. Srinagesh. 2020. "Multi-Lingual Twitter Sentiment Analysis Using Machine Learning." International Journal of Electrical and Computer Engineering 10(6): 5992–6000.
8. Babu, Nirmal Varghese, and E. Grace Mary Kanaga. 2022. "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review." SN Computer Science 3(1): 1–20. https://doi.org/10.1007/s42979-021-00958-1.
9. El Barachi, May, Manar Alkhatib, and Sujith Samuel Mathew. 2021. "A Hybrid Machine Learning Approach for Sentiment Analysis of Partially Occluded Faces." 2021 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2021: 389–92.
10. Bengesi, Staphord, Timothy Oladunni, Ruth Olusegun, and Halima Audu. 2023. "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets." IEEE Access 11(January): 11811–26.
11. Beseiso, Majdi, and Haytham Elmousalami. 2020. "Subword Attentive Model for Arabic Sentiment Analysis: A Deep Learning Approach." ACM Transactions on Asian and Low-Resource Language Information Processing 19(2).
12. Braig, Niklas et al. 2023. "Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data." IEEE Access 11(November 2022): 14778–803.
13. Brauwers, Gianni, and Flavius Frasincar. 2022. "A Survey on Aspect-Based Sentiment Classification." ACM Computing Surveys 55(4).
14. Catelli, Rosario et al. 2022. "Cross Lingual Transfer Learning for Sentiment Analysis of Italian TripAdvisor Reviews." Expert Systems with Applications 209(August 2021): 118246. https://doi.org/10.1016/j.eswa.2022.118246.
15. Chan, Chia-Pang, and Jun-He Yang. 2023. "Instagram Text Sentiment Analysis Combining Machine Learning and NLP." ACM Transactions on Asian and Low-Resource Language Information Processing.
16. Fazal, Unaiza et al. 2023. "Sentiment Analysis of Omicron Tweets by Using Machine Learning Models." : 67–75.
17. Jain, Praphula Kumar, Vijayalakshmi Saravanan, and Rajendra Pamula. 2021. "A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents." ACM Transactions on Asian and Low-Resource Language Information Processing 20(5).
18. Kumar, Akshi, and Victor Hugo C. Albuquerque. 2021. "Sentiment Analysis Using XLM-R Transformer and Zero-Shot Transfer
19. Learning on Resource-Poor Indian Language." ACM Transactions on Asian and Low-Resource Language Information Processing 20(5).
20. Ligthart, Alexander, Cagatay Catal, and Bedir Tekinerdogan. 2021. 54 Artificial Intelligence Review Systematic Reviews in Sentiment Analysis: A Tertiary Study. Springer Netherlands. https://doi.org/10.1007/s10462-021-09973-3.
21. Maity, Shuverthi, and Kamal Sarkar. 2022. "Topic Sentiment Analysis for Twitter Data in Indian Languages Using Composite Kernel SVM and Deep Learning." ACM Transactions on Asian and Low-Resource Language Information Processing 21(5).
22. Mehmood, Faiza et al. 2020. "A Precisely Xtreme-Multi Channel Hybrid Approach for Roman Urdu Sentiment Analysis." IEEE Access 8: 192740–59.
23. Maqbool, M. S., Hanif, I., Iqbal, S., Basit, A., & Shabbir, A. (2023). Optimized Feature Extraction and Cross-Lingual Text Reuse Detection using Ensemble Machine Learning Models. Journal of Computing & Biomedical Informatics, 5(01), 26-40.
24. Kanwal, F., Abid, M. K., Maqbool, M. S., Aslam, N., & Fuzail, M. (2023). Optimized Classification of Cardiovascular Disease Using Machine Learning Paradigms. VFAST Transactions on Software Engineering, 11(2), 140-148.
25. Rafiqee, M. M., Qaiser, Z. H., Fuzail, M., Aslam, N., & Maqbool, M. S. (2023). Implementation of Efficient Deep Fake Detection Technique on Videos Dataset Using Deep Learning Method. Journal of Computing & Biomedical Informatics, 5(01), 345-357.
26. Abid, K., Aslam, N., Fuzail, M., Maqbool, M. S., & Sajid, K. (2023). An Efficient Deep Learning Approach for Prediction of Student Performance Using Neural Network. VFAST Transactions on Software Engineering, 11(4), 67-79.
27. Hasnain, M. A., Ali, S., Malik, H., Irfan, M., & Maqbool, M. S. (2023). Deep Learning-Based Classification of Dental Disease Using X-Rays. Journal of Computing & Biomedical Informatics, 5(01), 82-95.
28. Basit, A., Hanif, I., Maqbool, M. S., Qayyum, W., Hasnain, M. A., & Nazeer, R. (2023). Cross-Lingual Information Retrieval in a Hybrid Query Model for Optimality. Journal of Computing & Biomedical Informatics, 5(01), 130-141.

29. Mehmood, Khawar, Daryl Essam, Kamran Shafi, and Muhammad Kamran Malik. 2019. "Discriminative Feature Spamming Technique for Roman Urdu Sentiment Analysis." IEEE Access 7: 47991–2.
30. Mujahid, Muhammad et al. 2023. "Arabic ChatGPT Tweets Classification Using RoBERTa and BERT Ensemble Model." ACM Transactions on Asian and Low-Resource Language Information Processing 22(8).
31. Nandwani, Pansy, and Rupali Verma. 2021. "A Review on Sentiment Analysis and Emotion Detection from Text." Social Network Analysis and Mining 11(1): 1–19. https://doi.org/10.1007/s13278-021-00776-6.
32. Nanjing, Iceb Proceedings, Xudong Luo, Xudong Luo, and Sihan Yin. 2021. "A Survey of Cross-Lingual Sentiment Analysis Based on Pre- Trained Models A Survey of Cross-Lingual Sentiment Analysis Based on Pre-Trained Models."
33. Oriola, Oluwafemi, and Eduan Kotze. 2020. "Evaluating Machine Learning Techniques for Detecting Offensive and Hate Speech in South African Tweets." IEEE Access 8: 21496–509.
34. Oyebode, Oladapo, Richard Lomotey, and Rita Orji. 2021. "'I Tried to Breastfeed but...': Exploring Factors Influencing Breastfeeding Behaviours Based on Tweets Using Machine Learning and Thematic Analysis." IEEE Access 9: 61074–89.
35. Peng, Sancheng et al. 2022. "A Survey on Deep Learning for Textual Emotion Analysis in Social Networks." Digital Communications and Networks 8(5): 745–62. https://doi.org/10.1016/j.dcan.2021.10.003.
36. Qureshi, Muhammad Aasim et al. 2022. "Sentiment Analysis of Reviews in Natural Language: Roman Urdu as a Case Study." IEEE Access 10(March): 24945–54.
37. Rana, Toqir A. et al. 2022. "An Unsupervised Approach for Sentiment Analysis on Social Media Short Text Classification in Roman Urdu." ACM Transactions on Asian and Low-Resource Language Information Processing 21(2).
38. Rehioui, Hajar, and Abdellah Idrissi. 2020. "New Clustering Algorithms for Twitter Sentiment Analysis." IEEE Systems Journal 14(1): 530–37.
39. Safari, Faezeh, and Abdolah Chalechale. 2023. "Emotion and Personality Analysis and Detection Using Natural Language Processing, Advances, Challenges and Future Scope." Artificial Intelligence Review (0123456789). https://doi.org/10.1007/s10462023-10603-3.
40. Satu, Md Shahriare et al. 2021. "TClustVID: A Novel Machine Learning Classification Model to Investigate Topics and Sentiment in COVID-19 Tweets." Knowledge-Based Systems 226: 107126. https://doi.org/10.1016/j.knosys.2021.107126.
41. Shah, S. Muhammad Ahmed Hassan et al. 2023. "Arabic Sentiment Analysis and Sarcasm Detection Using Probabilistic Projections-Based Variational Switch Transformer." IEEE Access 11(June): 67865–81.
42. Susmitha, C. et al. 2023. "Sentimental Analysis on Twitter Data Using Supervised Algorithms." Proceedings - 7th International Conference on Computing Methodologies and Communication, ICCMC 2023: 921–25.
43. Tahir, Rida, and M. Asif Naeem. 2022. "A Machine Learning Based Approach to Identify User Interests from Social Data." 2022 24th International Multitopic Conference, INMIC 2022: 1–6.
44. Tejaswini, Vankayala, Korra Sathya Babu, and Bibhudatta Sahoo. 2022. "Depression Detection from Social Media Text Analysis Using Natural Language Processing Techniques and Hybrid Deep Learning Model." ACM Transactions on Asian and Low-Resource Language Information Processing.
45. Wandabwa, Herman M., M. Asif Naeem, Farhaan Mirza, and Russel Pears. 2021. "Multi-Interest Semantic Changes over Time in Short-Text Microblogs." Knowledge-Based Systems 228: 107249. https://doi.org/10.1016/j.knosys.2021.107249.
46. Wijayanti, Rini, and Andria Arisal. 2021. "Automatic Indonesian Sentiment Lexicon Curation with Sentiment Valence Tuning for Social Media Sentiment Analysis." ACM Transactions on Asian and Low-Resource Language Information Processing 20(1).
47. Xu, Yuemei, Han Cao, Wanze Du, and Wenqing Wang. 2022. "A Survey of Cross-Lingual Sentiment Analysis: Methodologies, Models and Evaluations." Data Science and Engineering 7(3): 279–99. https://doi.org/10.1007/s41019-022-00187-3.
48. Yao, Zhirui et al. 2021. "Comparing Tweet Sentiments in Megacities Using Machine Learning Techniques: In the Midst of COVID-19." Cities 116(July 2020): 103273. https://doi.org/10.1016/j.cities.2021.103273.