Tech Science Press

Multilingual Sentiment Mining System to Prognosticate Governance

Muhammad Shahid Bhatti[1,*], Saman Azhar[1], Abid Sohail[1], Mohammad Hijji[2],
Hamna Ayemen[1] and Areesha Ramzan[1]

[1]COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan
[2]Computer Science Department, University of Tabuk, Tabuk, Saudi Arabia
*Corresponding Author: Muhammad Shahid Bhatti. Email: msbhatti@cuilahore.edu.pk

Abstract: In the age of the internet, social media are connecting us all at the tip of our fingers. People are linkedthrough different social media. The social network, Twitter, allows people to tweet their thoughts on any particular event or a specific political body which provides us with a diverse range of political insights. This paper serves the purpose of text processing of a multilingual dataset including Urdu, English, and Roman Urdu. Explore machine learning solutions for sentiment analysis and train models, collect the data on government from Twitter, apply sentiment analysis, and provide a python library that classifies text sentiment. Training data contained tweets in three languages: English: 200k, Urdu: 200k and Roman Urdu: 11k. Five different classification models are applied to determine sentiments, and eventually, the use of ensemble technique to move forward with the acquired results is explored. The Logistic Regression model performed best with an accuracy of 75%, followed by the Linear Support Vector classifier and Stochastic Gradient Descent model, both having 74% accuracy. Lastly, Multinomial Naïve Bayes and Complement Naïve Bayes models both achieved 73% accuracy.

Keywords: Multilingual NLP; artificial intelligence; government; sentiment analysis; NLP; NLTK; ensemble technique; multilingual; twitter; data science

1 Introduction

Social media is playing a vital role in the estimation of the perception of public opinions. Social media content is widely available and easy to access to predict people's perspectives on any current issue [1]. Twitter users play a significant role in it as their sentiments are expressed in their tweets. Text processing is becoming a challenging machine learning task with the increasing number of social media users. The study is prompted to use this analysis to gauge communities' satisfaction with the government's policies and performance.

Furthermore, natural language processing and machine learning have been used to accomplish this task. Data was collected by fixing Urdu, English and Roman Urdu on geographical locations to complete this work. After collecting the data, the people living in this location about the

government's performance are assessed. Governments like to know the public's reaction to a specific policy. Still, by simply looking at their tweets, it is difficult to get the collective sentiment of the people, i.e., positive or negative. So, it was proposed [2] to develop a system, which would mine tweets using specified time and location. After that, the tweet's text is preprocessed for sentiment analysis to determine positive or negative responses to applied government policies. The primary significance of this result is to track the emotional response of people. Tweets are collected to find the happiness index by retrieving the dataset based on longitude and latitude [3]. In Microblogging, many people share their perspectives and suppositions on different everyday life issues concerning them straightforwardly or by implication through web-based media stages like Twitter, Reddit, Tumblr, Facebook [4,5]. In the paper [6], the authors researched that some areas have monopolistic electoral systems in developing countries like Pakistan. People are extorted to vote for the tribe leader having the most influential family background despite not providing necessities.

Similarly, seldom do people vote without understanding the manifesto of the party. Twitter has grown quite a hub for dispensing political affairs and raising voices against the unfulfilled obligations of the elected leadership. Political behaviour of parties and the public's response can be investigated and classified to predict elections and sentiments of people. In 2013, the general election result was forecasted, and the results were quite comparable to the actual results published by the Election Commission of Pakistan [7]. In the US, the tweets made by city governments use certain tones for sharing information with the public. For example, they adopt a neutral tone to push information to users to channel the information to the public.

On the other hand, if city governments seek support from residents, they adopt a more positive tone to encourage citizens for productive participation. In this way, they get more retweets, and their message gets regarded. In this manner, political bodies tend to adopt a specific tone for their target results. If they want to start a campaign and want public support, a positive temper is adapted and attached pictures with tweets, whereas their tone is neutral for information sharing. Through interaction on Twitter, citizens and government seem to share resources, the most important being information. The reaction of citizens is quite vital to the success or downfall of a campaign.

A review and relativistic research [8] of present techniques for sentiment analysis, including machine learning and dictionary-based methodologies, along with cross space and cross-lingual strategies and some assessment measurements. Exploration results show that AI techniques, for example, SVM and Naïve Bayes have the highest accuracy and can be viewed as the standard learning strategies. In contrast, dictionary-based techniques are compelling at times, which require not many exertions in human-marked reports. This survey mainly focused on sentiment analysis of the Twitter dataset, which helps analyse the information and filter them out either as positive, negative, or neutral. In the review articles [9–11], they zeroed in on the sentiment analysis of Twitter information. These articles show sentiment analysis types and techniques used to perform the Extraction of slang from tweets. To perform sentiment examination, one needs to perform different undertakings like subjectivity identification, sentiment analysis, perspective term extraction. These articles present the review of principle approaches used for sentiment ranking.

The objective of the research paper is to investigate preprocessing techniques and implement various machine learning paradigms. Well recognised for their performance on sentiment analysis of textual documents and provide a python package that predicts sentiment and opinion mining of data about governance. The paper's contributions are as follows:

- Urdu is rich and complicated in the structure that makes Urdu language processing even more challenging. In addition, the unavailability of the large annotated Urdu comments dataset was an essential obstruction for NLP in the Urdu language. Therefore, a large-sized multilingual dataset was developed that will be publicly available to overcome the Urdu language deficiencies. This dataset is annotated by using semiautomatic methods.
- In addition, hyperparameters tuning is implemented to use five different machine learning models efficiently.
- Furthermore, we conducted hyperparameter tuning and investigated their influence on the accuracy of the models.
- Finally, we provide a Python library for multilingual sentiment analysis focused on the three languages discussed earlier. The package can be accessed using the following URL: https://pypi.org/project/multilingualsentimentclassifier/.

The rest of this article is arranged as follows: the following section presents a thorough study on opinion mining of government data and machine learning solutions for sentiment analysis of multilingual datasets and a similar study on best solutions for sentiment analysis. Then the characteristics of Urdu, English and Roman Urdu languages are explained, followed by dataset description, data preprocessing, and steps followed for training models. Then the results are presented and explained along with the conclusion.

2 Literature Review

This section gives an insight into the work that has already been done on NLP of the multilingual dataset, sentiment analysis techniques and opinion mining on government. A user's past tweets were monitored, and an extraction technique was applied to get a clearer idea. Sentiment analysis on Twitter and a state-of-the-art unigram model as a baseline were proposed [12]. The report was an overall gain of over 4% for two classification tasks: a binary, positive *vs.* negative, and a 3-way positive versus negative versus neutral that presented a comprehensive set of experiments for both these tasks on manually elucidate data that was a random sample of stream of tweets. It also investigated two models, i.e., tree kernel and feature-based models, to show that both outperform the unigram baseline.

The messages posted by Italian users on Twitter were examined [13] to investigate idiosyncratic shocks of happiness. The data set comprises over 43 Million tweets posted daily in all the 110 Italian provinces. They constructed an index through inventive statistical techniques to measure happiness at the provincial level related to specific numbers of tweets and explore the determinants of satisfaction in the sample. In addition, tweets were used to predict customer sentiment [6], customer opinions using stream analysis which reads informative and exciting tweets based on their negativity or positivity, and pixel cell-based sentiment calendars, which visualise extensive data in a single view. Similarly, Twitter was used to investigate Spanish elections and determine if the conversations through tweets affect elections. A tool known as Tara tweet [14] was designed to get the related discussions of elections during 2011 and 2012. Before and after the elections held in the US and France in 2012, Twitter was used to compare the sentiments. The information was extracted through tweets [15,16] from Twitter accounts of several people for sentiment analysis. The data obtained from the US election to observe for the French elections compared to bring out the fact that social media nowadays has a significant influence on elections. The paper [17] gathered data through Twitter and did sentiment analysis of the Singapore presidential election with census correction. Reweighting techniques were used in addition to online data to predict the vote percentage a single candidate was received. In the paper [18], the authors proposed how

Twitter can predict the elections in three Asian countries of Malaysia, India, and Pakistan. They came up with how Sentiment information mined using machine learning models was appropriate for predicting election outcomes. The authors in [19] done research on five Latin American countries having mature e-participation capabilities. If the correct sentiment analysis technique is used, it can produce a very accurate forecast result. The supervised method by Hopkins and King was used [20] to predict the voting intention of people of the US and Italy in the 2012 elections. Data mining technique was used to collect data from Twitter, and preprocessing was applied to clean data. Finally, the best sentiment analysis technique was applied to forecast election results. The paper [21] introduced novel metrics for viewing public communication through Twitter. As twitter had become the primary path of discussing issues, politics, and other things, they collected data under #ausvotes hashtags to describe the election's coverage in this particular social media site. This work is mainly focused on the computer-aided analysis of election-related Twitter messages.

The main idea was to check how twitter covered the 2020 Australian federal election. Furthermore, researchers added how combining sentiment and volume information predicts vote effectively shares that are smaller and how real-world outcomes can be guessed for future work. Twitter can be used for analysing different opinions of the public on laws imposed locally in the urban region and helps to notice how an area moves towards a smart city for innovative governance. Ordinance-Tweet–SCC Mapping, Tweet Polarity Classification Algorithms [22] were applied. The tweets were further used for sentiment analysis of public reactions to ordinances.

An automated content analysis tool [23] for Malaysian legal firms and leaders to understand the public's sentiments was proposed. The technique used is Semantic Role Labelling (SRL) which generates a method to filter and classify the dataset to get more accurate results in sentiment detection approaches. As Natural Language Processing has been known to face problems when it comes to degraded text, this proposal was novel to use the SRL technique instead of the generally used NLP technique. The main idea was to help the Malaysian agencies for an improved outreach for future development. Big data analytics such as polls, surveys, and social media were used to build prediction models [24] for election forecasting. Different data sources were used to tap arrays of statistics to capture the winning capabilities of individual candidates. The Bayesian model was used to predict the seat share of parties in the National Assembly with 83% accuracy. This model is more accurate than any poll or survey was done before 2018 and shows that AI and machine learning models can significantly affect modern politics. LIWC text analysis software [25] was used to conduct a content analysis of more than 100,000 messages related to a specific political party. In addition, an analysis of tweets was done to find the political sentiments of the public. Microblogging message content was also used as an indicator of political opinion.

A Tweets Sentiment analysis model (TSAM) [26] can recognise the cultural interest and general individuals' suppositions concerning a get-together. Australian government political decision of 2010 occasion was taken to act as an illustration for investigation tests. The conclusion was principally intriguing of the particular political competitors, i.e., two essential priest applicants - Julia Gillard and Tony Abbot. Ongoing Irish General Election was used as contextual analysis [27] for examining the possibility of displaying political sentiment through the mining of online media. The approach consolidates sentiment analysis utilising administered learning and volume-based measures. Finally, an approach to the multilingual sentiment analysis dataset [28] and solution for multilingual sentiment analysis was proposed by executing calculations and getting the outcome. They contrast the exactness factor to locate the best solution for multilingual wistful examination.

Sentiment analysis of Urdu news tweets using supervised learning to classify tweets as positive, negative, and neutral was done using a Decision tree [29] as a classification algorithm. The

results of the proposed methodology were a great success in terms of accuracy and analysis. *Naive Bayes Algorithm* [30] was used to predict moods and sentiments of two different Indian schemes using Twitter. They performed polarity classification. In the paper [31], the authors used SVM, NB, and ME. They proposed a novel profound learning-based mixture for predicting polarity against text-only reviews. Pre-trained embedding helps in improving outcomes as they catch the logical importance of a word in a sentence. Lexicon-based approach [32] for sentiment label distribution brought up the fact that using character embedding through Deep CNN to enhance information based on top of the world through Twitter improves exactness in sentiment management. The semantic principles contribute to dealing with non-essential tweets to improve classification accuracy. Naive Bayes, SVM, CNN & RNN [33] and similar trials were conducted on various deep learning models. In past examinations, the authors chose three CNN structures and five RNN structures. They were thought about in the wording of AUROC utilising 13 datasets. Moreover, they studied the effects of two diverse information levels (word and character levels) on the grouping execution. Fig. 1 demonstrates the complete architecture of a system for sentiment analysis.
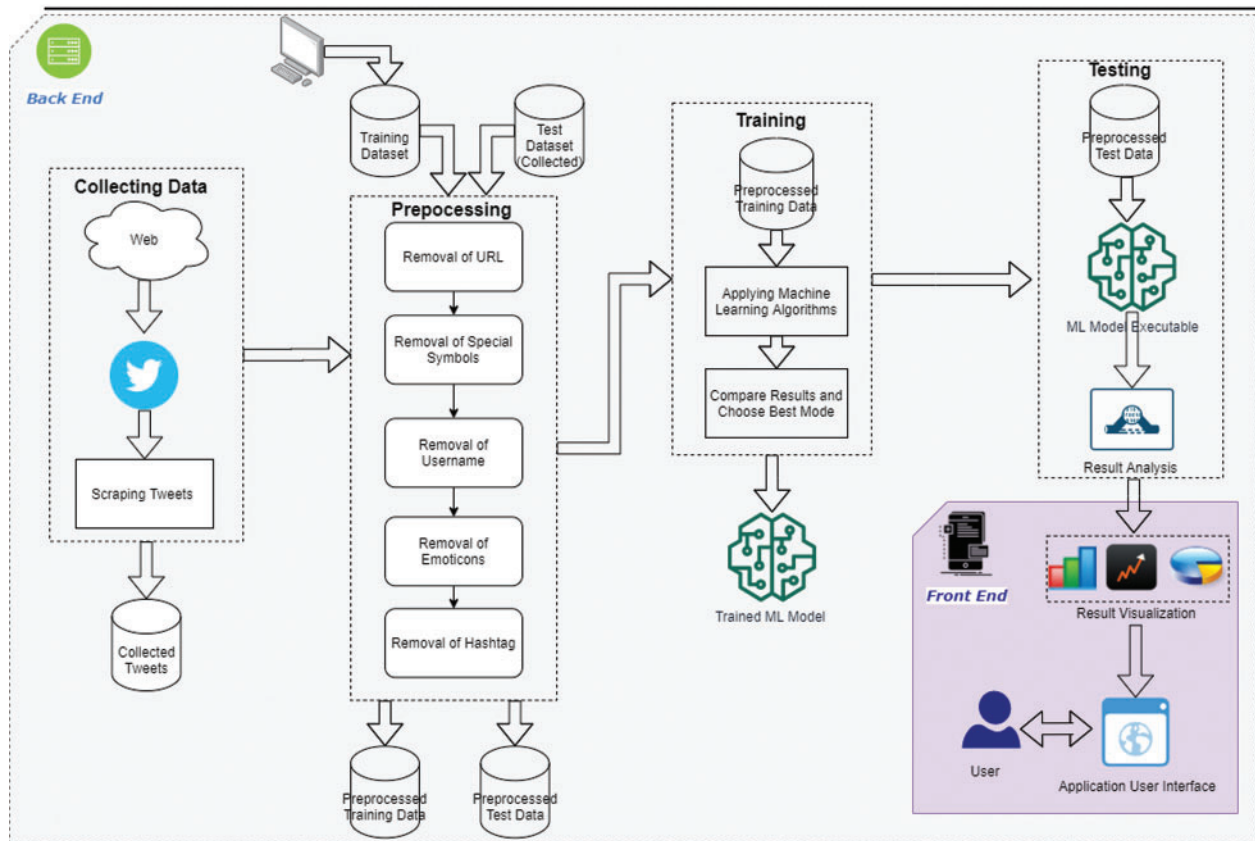

Figure 1: Complete flow of steps from data collection to sentiment prediction

3 Linguistics Features of Urdu, Roman Urdu and English

This paper tackles a multilingual dataset of three languages that are most widely used to communicate in Pakistan, i.e., Urdu, English and Roman Urdu. This section discusses the linguistic features of these three languages.

Roman Urdu is the Urdu language written using Latin scripts. It is an informal language used widely by internet users for communication. The linguistic features of Roman Urdu and English are pretty similar, and both have the same writing rules, such as the right to left and capitalisation of the first word of each sentence. However, English has 26 alphabets, each having upper and lower case forms. Roman Urdu is written using these alphabets, sometimes with unique characters, as shown in Tab. 1.

Table 1: Urdu Characters and their Counter Roman Urdu Characters

Letter	Name of letter	Transcription	Letter	Name of letter	Transcription
ا	alif	-	ی	ye	y, i
ب	be	B	ے	bari ye	ai or e
پ	pe	P	ض	zu'ād	ẓ
ت	te	T	ط	to'e	T
ٹ	ṭe	ṭ	ظ	zo'e	ẓ
ث	se	S	ع	'ain	'
ج	jīm	J	غ	ghain	Gh
چ	che	Ch	ف	fe	F
ح	baṛī he	H	ق	qāf	Q
خ	khe	Kh	ک	kāf	K
د	dāl	D	گ	gāf	G
ڈ	ḍāl	ḍ	ل	lām	L
ذ	zāl	Dh	م	mīm	M
ر	re	R	ن	nūn	N
ڑ	ṛe	ṛ	و	vā'o	v, o, or ū
ز	ze	Z	ہ	choṭī he	H
ژ	zhe	Zh	ھ	do chashmī he	H
س	sīn	S	ء	hamza	'
ش	shīn	Sh			
ص	su'ād	ṣ			

Urdu is a language containing 38 primary characters, as demonstrated in Tab. 1. Urdu is written in Nastaliq text style that is a complex and delicate style. An expression of Urdu is a mix of ligatures. A ligature is made out of a single or a few characters. For writing Urdu sentences, words are connected from right-to-left order. The composition of words, their ligatures, and characters are shown in Tab. 1.

Some critical features of the Urdu language are discussed below:

- **No capitalisation:** in contrast to the English language, there is no capitalisation or lowercase. Subsequently, it is troublesome to recognise proper nouns and the beginning of a sentence

in Urdu. For example: in علی کا تعلق پاکستان سے ہے (Ali lives in Pakistan), nouns and the start of the sentence cannot be identified.

- **Right to left:** The direction of Urdu writing is from right to left. For example: in سے ہے علی کا تعلق پاکستان the first word is considered the last word of the sentence.
- **Diacritics:** like Arabic, Urdu text may have not many diacritics. Zer (ِ) which means "under", zabar (َ) which means "over" also, pesh (ُ) which means "in front" are three normal diacritics. For example (Taer, Swim) and (Teer, Arrow) Similarly (Bal, Curl) and (Bill, Invoice).
- **No specific order:** In Urdu has a free words order without any restriction. The order of sentences in Urdu may be different, but the meaning would be the same. For example: اسامہ نے کپ دکان سے خریدا and اسامہ نے دکان سے کپ خریدا means exactly the same (Usama bought cup from the shop). Both sentences have separate order of words, but the purpose is precisely the same.
- **Context Sensitivity:** In Urdu, a character may have various shapes formed with different characters to frame a ligature or a word. For example, ب = ا + ب when ب is joined with ا it is با, whereas, ت = پ + ت when ت is joined with پ it is تپ.

4 Data and Methodology

This section describes the dataset used for training the models, the data collected for sentiment analysis on government, data cleaning and preprocessing techniques. It also describes the feature extraction methodology and the classifiers used for training. Fig. 1 describes the complete flow of steps followed from data collection to sentiment prediction.

4.1 Data Collection

This section describes the data collection process, and details of both training and testing dataset tweets about government are collected using different approaches.

4.1.1 Dataset

The training dataset was collected from an available online source. The Urdu and Roman Urdu datasets were relatively small, so Google Translator was used to translating the English dataset into Urdu. After that, a column of "Language" denoting the language of the text was added to each dataset, and all the data was merged into one dataset having columns "Text/Tweet", "Language", and "Sentiment". The Language column had 3 values, "en" for the English language, "ur" for Urdu and "in" for Roman Urdu and the Sentiment column had 2 values, positive and negative, which were later encoded to 0 and 1, respectively shown in Tab. 3. The data statistics of all 3 languages is given in Tab. 2. The dataset was not balanced; Roman Urdu had fewer instances than the other two, as shown in Fig. 2. It had 100,000 instances of English and Urdu languages but only 11,000 instances of Roman Urdu language. In Fig. 2, the blue bar represents positive instances, and the orange represents negative instances of each language.

Table 2: Data Description of all three languages

Dataset	Urdu	English	Roman Urdu
No of instances	200000	200000	11099
No of Unique Words	192429	164397	11056
Minimum Words	1	1	1
Maximum Words	57	40	55
Average No of Words in an instance	16	13	16

Table 3: Sample from training data

Tweet	Language	Sentiment
@Spitphyre well what to do…i have friends from all walks of life…i dont discriminate based on intelligence or evolutionary status	English	Positive
@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because I can't see you all over there.	English	Negative
سوچا تھا شہباز شریف وہاں سے کام آگے لے کر جائیں گے جہاں ہم نے چھوڑا چودھری پرویز الٰہی چودھری نے پنجاب تباہ کیا شہباز نے ترقیاتی کام کیے پنجاب کو سنوار دیا	Urdu	Positive
سہیل بخاری کا کہنا تھا کہ تعزیتی ریفرنس میں یاسین ملک کے قریبی دوستوں کی شرکت نہ کرنے پر انہیں بہت افسوس ہے	Urdu	Negative
Jab school teachers ne in ki awaaz suni to Mukesh aur Rafi ke songs ki farmaiysh karne lage aur hausla afzaai karne ke sath sath yeh dars bhi dete rehte ke singing na chorna	Roman Urdu	Positive
Aap ko phir buhat ghussa ayya k ye jo teacher party hai wo bachoon ko ghalat ratte marne ki training de rahe hain	Roman Urdu	Negative
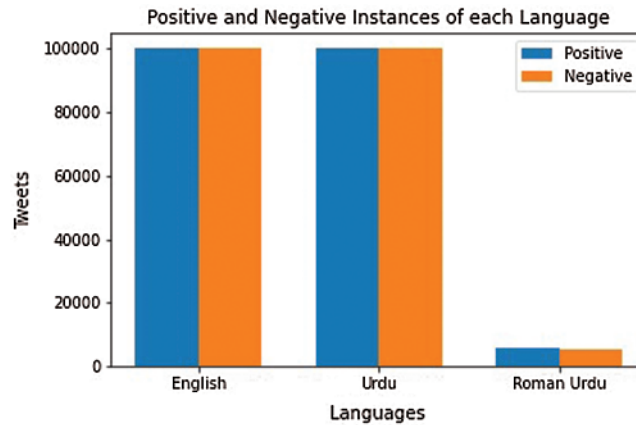
Figure 2: Frequency of positive and negative instances of each language

4.1.2 Governance Concerning Dataset

The data about the current government of Pakistan (2018–2023) was to be collected from Twitter, a very secure site. It does not easily allow to scrape its data. However, there are specific NLP tools like Tweepy, which is well known for fetching Twitter data but has certain limitations and constraints like access to only 3200 tweets in a timeline. The goal was to find a scraper that would scrape a large amount of data, so a scraped Twint, an open-source Python package, was used. The required data to be collected is quantitatively large sized. A big challenge was to manipulate this scraper to fetch tweets in the geographical location of Pakistan as back as 2018. A list of above 400 cities of Pakistan was given to the scraper to collect data within the geographical area of Pakistan. Some instances of the data collected are shown in Tab. 3. The dataset was collected on the three most influential political parties of the Pakistani government using their acronyms as keywords in the scraper:

- Pakistan Peoples Party (PPP)
- Pakistan Tehreek-e-Insaf (PTI)
- Pakistani Muslim League (PMLN)

The dataset of one million, three hundred and seven thousand tweets in Tab. 4 were collected due to scraping.

Table 4: No of Tweets Scraped

Dataset	Urdu	English	Roman Urdu
PTI	201852	409096	59143
PMLN	258102	201754	34494
PPP	14682	95541	33305

4.2 Data Cleaning and Preprocessing

The tweets were cleaned first to eliminate emoticons, hashtags, mentions, and other irrelevant information in sentiment prediction. That is followed by preprocessing, which involves steps like removing stop words and words with unusual symbols and normalizing tweets. The preprocessing of English and Roman Urdu was somehow the same as both of them share quite similar features. An open-source python library, i.e., Textblob, was used for processing. It performs different operations on textual data, such as noun phrase extraction, sentiment analysis, classification, and translation. The WordNetLemmatizer was used for grouping together the different inflected forms of a word so they could be analyzed as a single item.

On the other hand, cleaning Urdu text was done by removing punctuation, normalizing whitespace, removing accents, and replacing URLs and symbols. A good set of stop words for the classifiers to perform well was also required. While preprocessing, stopwords were removed. A list of Pakistani names in English, Roman Urdu, and Urdu see Tab. 6 was passed to the stop words list as a lot of the data scraped from Twitter had mentions of Pakistani names. A few sample stopwords are given in Tab. 5. After applying to preprocess, the data is split into a 30–70 ratio for the unbiased training and testing evaluation.

Table 5: Sample stopwords

Stopword	Language
the	English
me	English
and	English
we	English
آئے	Urdu
آج	Urdu
آخر	Urdu
أب	Urdu
آدھی	Urdu
ayi	Roman Urdu
hy	Roman Urdu
hai	Roman Urdu
main	Roman Urdu
ki	Roman Urdu
tha	Roman Urdu

Table 6: Example of familiar names

Names	Language
عائشہ	Urdu
زینب	Urdu
محمد	Urdu
داؤد	Urdu
Shan	Roman Urdu
Aatif	Roman Urdu
Ahmed	Roman Urdu
Ali	Roman Urdu

4.3 Feature Extraction and Classification Models

Preprocessing makes the tweets ready for Machine Learning Pipeline. The pipeline has three stages that convert each tweet in the dataset to token integer counts. Next, the integer counts are converted into weighted TF-IDF scores, and finally, the weighted scores are pass to the classifier for prediction. The three stages are word embedding, term weighting, and classification.

4.3.1 Feature Extraction

The Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer was used for word embedding and TF-IDF Transformer for term weighting. The TF-IDF weight is popular in text mining and information retrieval. This weight is a statistical metric for determining the importance of a word in a collection or corpus of documents. The importance of a word rises in proportion to the number of times it appears in the document, but the term's frequency counterbalances this in the corpus. TF-IDF can be successfully used for stop-word filtering in various subject fields. Term Frequency (TF) counts the number of times a term appears in a document see Eq. (1).

As a result, the term frequency is frequently divided by the length of the document words for normalisation:

$$TF(t) = \frac{Number\ of\ times\ term\ t\ appears\ in\ a\ document}{Total\ number\ of\ terms\ in\ the\ document} \tag{1}$$

Inverse Document Frequency (IDF) determines the significance of a term mathematically shown in Eq. (2). Therefore, when calculating TF, all the words are given equal weight. However, it is well known that some phrases, such as "is," "of," and "that," may appear frequently but have little meaning. As a result, the frequent terms must be scaled down while scaling up the rare ones by computing:

$$IDF(t) = log_2 \frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it} \tag{2}$$

4.3.2 Classification Models

After feature extraction, a classifier is passed to classify the tweets as positive or negative. According to our extensive research, the best classifiers for sentiment prediction of textual data are Multinomial Naïve Bayes (MNB), Complement Naïve Bayes (CNB), Logistic Regression (LR), Support Vector Machine-Linear Support vector classifier (LSVC), and Stochastic Gradient Descent (SGD).

To make predictions, Naive Bayes uses Bayes' theorem and probability theory shown in Eq. (3). For multinomial distributed data, MNB implements the Naive Bayes algorithm where vectors parametrise the distribution:

$$P(c|X) = P(x_1|c) * P(x_2|c) * \ldots * P(x_n|c) * P(c) \tag{3}$$

For each class labelled as y, where n is the number of features, and θ_{y_i} Is the probability of i appearing in a sample belonging to class y. For example, here is the mathematical representation is shown in Eq. (4):

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n} \tag{4}$$

where, $N_{yi} = \sum\limits_{x \in T} x_i$ is the number of times feature i appear in a sample of class y in the training set T, and $N_y = \sum\limits_{i=1}^{n} N_{yi}$. That is the total count of all features for class y.

The CNB algorithm is a variant of the regular MNB algorithm that deals with unbalanced data sets. The classification rule is shown in Eq. (5):

$$\hat{c} = agr\ \min_c \sum_i t_i w_{ci} \tag{5}$$

The Support Vector Machine (SVM) classifier is a discriminative classifier commonly used to identify anomalies and solve classification problems. Linear Support Vector Classifier (LSVC), an implementation of SVM, was used. This non-probabilistic linear classifier plots each data sample

as a coordinate point in multi-dimensional space and finds the best hyperplane to distinguish class boundaries effectively described mathematically in Eq. (6).

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1} \max\left(0, y_i\left(w^T \emptyset(x_i) + b\right)\right) \tag{6}$$

The supervised learning classification algorithm logistic regression is used to predict the probability of a target variable. Mathematically, $P(Y = 1)$ is predicted by a logistic regression model as a function of X see Eq. (7). It is one of the most basic and widely used machine learning algorithms. It can solve various classification problems, including spam detection, diabetes prediction, cancer detection. See the equation below:

$$p(y|X; W) = \sum_{i=1}^{i=n} (h_w(X))^y + (1 - (h_w(X))^{1-y}) \tag{7}$$

SGD estimator uses stochastic gradient descent learning to implement regularised linear models: the gradient of the loss is measured one sample at a time with the default learning rate $n^t = \frac{1}{\alpha(t_0+t)}$, and the model is modified along the way with a decreasing intensity schedule. In addition, the partial fit method in stochastic gradient descent (SGD) allows for mini-batch learning using the update rule. Mathematical representation is shown in Eq. (8).

$$w \leftarrow w - n\left[\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial L\left(w^T x_i + b, y_i\right)}{\partial w}\right] \tag{8}$$

All these classifiers are used to predict sentiments for tweets of each language. The proposed methodology is shown in Fig. 3.

Deep Learning Model: Five epochs on deep learning Long short-term memory (LSTM) were run. LSTM is a deep learning architecture that uses an artificial recurrent neural network (RNN). Since there may be lags of indefinite length between actual events in a time series, LSTM networks are well-suited to classifying, processing, and making predictions based on time series data. However, the accuracy after running was very low, i.e., 50.43%, so the deep learning solution for sentiment analysis was discarded. Below is the architecture of the deep learning LSTM model used in Fig. 4.

5 Results and Discussions

This section discusses the performance of our models and the approach used to predict sentiments in light of the results. The evaluation measures, accuracy, recall, precision, and F1 score are used as shown in Eq. (9).

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall},$$

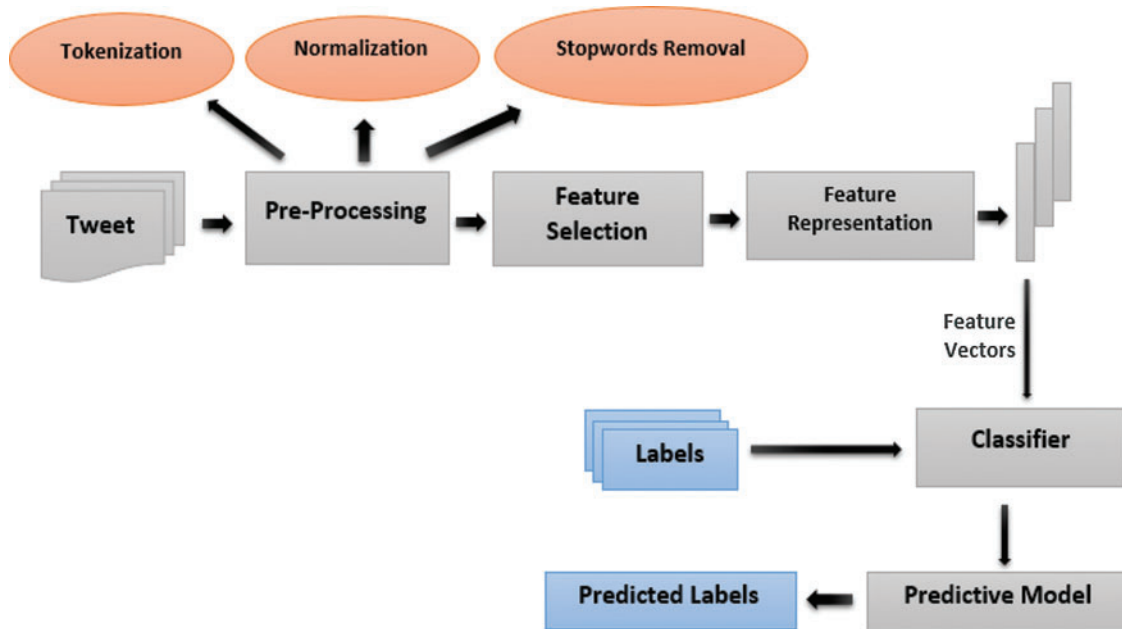$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

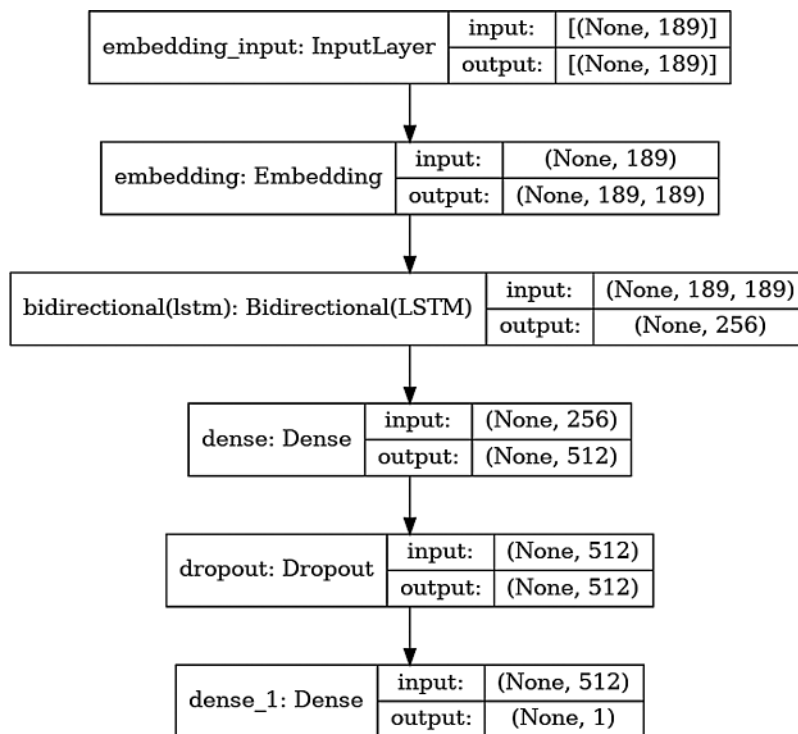**Figure 3:** Work Flow of proposed methodology for sentiment analysis

Figure 4: Architecture of Deep Learning Model

5.1 Comparative Analysis of Models

Average time and accuracy were calculated for ten iterations; results are shown below in Tab. 7. According to our research and experimental results, LR is easy to implement, understand,

and efficient to train compared to other classifiers. It provides good accuracy using saga as a solver for a simple dataset and also performs well when the dataset is linearly separable. As shown below in Tab. 7, the average accuracy score for LR was 75% which achieved the best out of 5 classifiers. Still, the average time for it was 28.28 s longer than the other classifier's running time. The accuracy score of MNB, CNB, LR, LSVC and SGD is shown in Tab. 7.

Table 7: Result analysis of different classifiers on ten iterations

Classifier	MNB	CNB	LR	LSVC	SGD
Average accuracy (approx.)	73%	73%	75%	74%	74%
Average time (s)	20.74	21.82	28.28	24.05	23.5

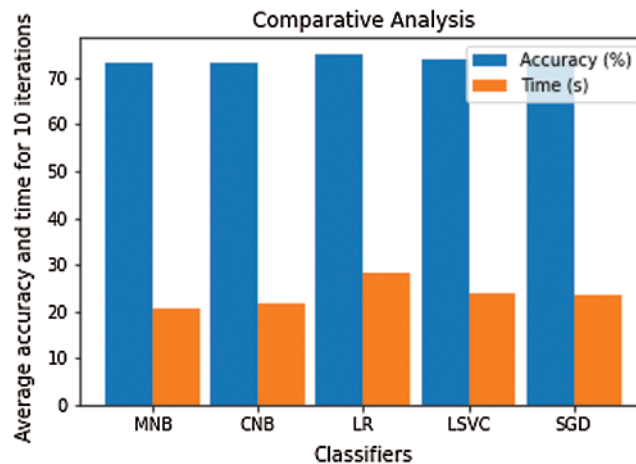Below in Fig. 5. is the graphical representation of the performance of different classifiers.

Figure 5: Performance of different ML classifiers

There were two different ways to go ahead with this:

- The first approach was to use a language detector. The drawback was that there is no language detector for Roman Urdu. So English and Urdu languages were to be detected using detectors. Furthermore, if a language is not detected as these two, it is automatically classified as Roman Urdu.
- The second approach was to label the language of each text before training so the models know the language of the text they were predicting. This approach was chosen for this study.

The average accuracies of the model trained using this technique is already discussed. Since the training data was unbalanced, to know if the models are even reliable, the LR model was tested on each language separately, and the results are shown in this Tab. 8. The model performed excellently for each language. The accuracy score for English is 77%, for Urdu is 84%, and for Roman, Urdu is 85%.

Table 8: Analysis of LR model on each language

Language	English	Urdu	Roman Urdu
Accuracy	0.77	0.84	0.85
F1-score	0.77	0.84	0.85
Recall	0.77	0.84	0.85

5.2 Ensemble Technique

Since all the classifier's accuracy was almost identical for all the classifiers, the ensemble technique was chosen. It is an elegant technique to produce improved Machine Learning results. The majority voting method for ensemble technique was used, in which every model makes a prediction (votes) for each test instance, and the final output prediction is the one that receives more than half of the votes. This method produces more accurate solutions than a single model would. 0 is a positive class, and 1 is the negative class in Tab. 9. All the five classifiers predicted the same in the first and last row, so the final sentiment for these rows is the same. Still, in the second row, one classifier, i.e., MNB, has classified the tweet as negative. The majority classified it as positive hence based on ensemble technique, the final sentiment is positive. Summary of results achieved using ensemble learning is applied in Tab. 10, which are improved and balanced.

Table 9: Prediction of the sentiment using the majority voting technique

Preprocessed tweet	Language	Prediction Model					Sentiment
		CNB	MNB	LR	SGD	SVC	
0 opinion officials tweet representatives commen…	En	0	0	0	0	0	0
1 Tweet account point agenda Reform Control Prov…	En	0	1	0	0	0	0
2 Punjab province Federal PTI suppertes hv Inter…	En	1	1	1	1	1	1

Table 10: Results achieved when ensemble learning is applied

	Precision	Recall	F1	Support
Negative	0.77	0.74	0.75	64064
Positive	0.73	0.76	0.74	59266
Accuracy			0.75	123330
Macro avg				123330
Weighted avg	0.75	0.75	0.75	123330

5.3 Multilingual Sentiment Classifier

As a result of this study, a python package was designed publicly available at sources like GitHub and Python Package Index (PyPI). The package is pip installable. It takes the text and language of the entered text as input and predicts the sentiment. It can predict sentiments of rows of a data frame as well. The workflow of the library is shown in Fig. 6.
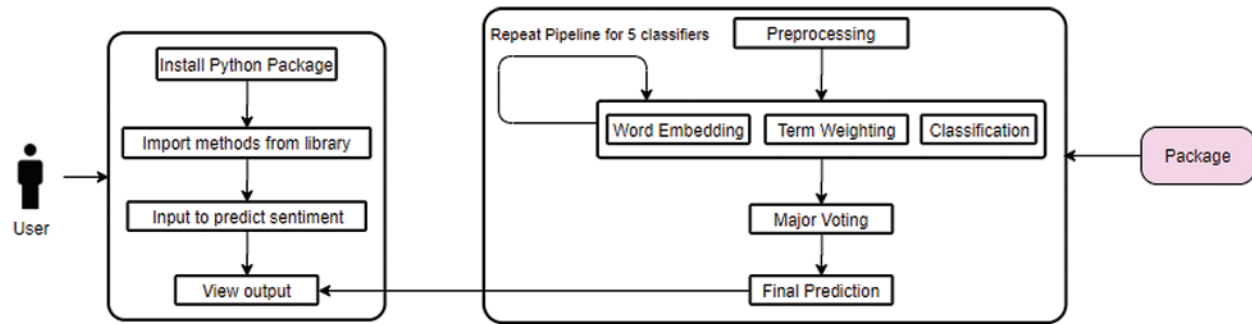

Figure 6: Workflow of Python Toolkit package

6 Conclusion

We conclude with some baseline accuracies discussed in Section 5 of this paper. The models were trained using an ML pipeline after text pre-pressing, which involved a set of steps like stopwords removal, special characters' removal, and normalisation. The pipeline also had several stages: word embedding, term weighting, and classification. Most of the algorithms had the same accuracies, so we opted for the ensemble technique and used majority voting to predict whether a sentiment is positive or negative. Again, LR topped with an accuracy of 75%. The possible reason is that LR uses probability theory to classify, and the algorithm identified the right features. Another reason can be that many people adopt sarcasm or irony in their tweets. Since the training dataset of Urdu was translated from English, Google translator may have misinterpreted the nature of the text while translating, which was then misclassified by our models, hence skewing the results. Also, the classification algorithms cannot always determine the sarcastic or ironic context of texts either, so we cannot say that LR performed best for a specific reason. That is the exact reason the majority voting approach was chosen for this dataset. More work can be done on the multilingual dataset of these three languages to design models that provide improved accuracy. The multilingual sentiment classifier intended for this research predicts input as negative or positive. The user has to input the language of the text manually they ask to predict. In the future, if a language detector for the Roman Urdu language is designed, the language detection for this API can be automated and handled in the backend.

Conflicts of Interest: The authors declare that they have no interest in reporting regarding the present study.

References

[1]   D. Quercia, J. Ellis, L. Capra and J. Crowcroft, "Tracking gross community happiness from tweets," in *Proc. of the ACM 2012 Conf. on Computer Supported Sooperative Work*, Washington, USA, pp. 965–968, 2012.

[2]   Y. W. Syaifudin and D. Puspitasari, "Twitter data mining for sentiment analysis on peoples feedback against government public policy," *MATTER: International Journal of Science and Technology,* vol. 3*, no.1, pp. 110–122, 2017.

[3]   K. Singh, S. K. Haris and B. Biswas, "Happiness index in social network," in *Proc. of Int. Conf. on Advanced Informatics for Computing Research*, Jalandhar, India, pp. 261–270, Springer, 2017.

[4]   Y. Chandra and A. Jana, "Sentiment analysis using machine learning and deep learning," in *Proc. 2020 7th Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, IEEE, pp. 1–4, 2020.

[5]   R. Sharma, N. L. Tan and F. Sadat, "Multimodal sentiment analysis using deep learning," in *Proc. of 2018 17th IEEE Int. Conf. on Machine Learning and Applications (ICMLA)*, Orlando, Florida, USA, IEEE, pp. 1475–1478, 2018.

[6]   M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Keim *et al.*, "Visual sentiment analysis on twitter data streams,". in *Proc. of 2011 IEEE Conf. on Visual Analytics Science and Technology (VAST)*, Providence, RI, USA, IEEE, pp. 277–278, 2011.

[7]   M. A. Razzaq, A. M. Qamar and H. S. M. Bilal, "Prediction and analysis of Pakistan election 2013 based on sentiment analysis," in *Proc. of 2014 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, Beijing, China, IEEE, pp. 700–703, 2014.

[8]   V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: A survey of techniques," in arXiv preprint arXiv:1601.06971, 2016.

[9]   R. Wagh and P. Punde, "Survey on sentiment analysis using twitter dataset," in *2018 Second Int. Conf. on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, IEEE, pp. 208–211, 2018.

[10]   H. Kaur and V. Mangat, "A survey of sentiment analysis techniques," in *2017 Int. Conf. on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, Tamil Nadu, India, IEEE, pp. 921–925, 2017.

[11]   Z. Li, Y. Fan, B. Jiang, T. Lei and W. Liu, "A survey on sentiment analysis and opinion mining for social multimedia," *Multimedia Tools and Applications,* vol. 78*, no. 6, pp. 6939–6967, 2019.

[12]   A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. J. Passonneau, "Sentiment analysis of twitter data," in *Proc. of the Workshop on Language in Social Media (LSM 2011)*, Portland, Oregon, pp. 30–38, 2011.

[13]   L. Curini, S. Iacus and L. Canova, "Measuring idiosyncratic happiness through the analysis of twitter: An application to the Italian case," *Social Indicators Research,* vol. 121*, no. 2, pp. 525–542, 2015.

[14]   J. M. Soler, F. Cuartero and M. Roblizo, "Twitter as a tool for predicting elections results," in *Proc. of 2012 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, Istanbul, Turkey, IEEE, pp. 1194–1200, 2012.

[15]   F. Nooralahzadeh, V. Arunachalam and C. G. Chiru, "2012 presidential elections on twitter–an analysis of how the us and French election were reflected in tweets," in *Proc. of 2013 19th Int. Conf. on Control Systems and Computer Science*, Bucharest, Romania, IEEE, pp. 240–246, 2013.

[16]   U. Yaqub, S. A. Chun, V. Atluri and J. Vaidya, "Sentiment based analysis of tweets during the us presidential elections," in *Pro. of the 18th Annual Int. Conf. on Digital Government Research*, Staten Island, NY, USA, pp. 1–10, 2017.

[17]   M. Choy, M. L. Cheong, M. N. Laik and K. P. Shung, "A sentiment analysis of Singapore presidential election 2011 using twitter data with census correction," in arXiv preprint arXiv:1108.5520, 2011.

[18]   K. Jaidka, S. Ahmed, M. Skoric and M. Hilbert, "Predicting elections from social media: A three-country, three-method comparative study," *Asian Journal of Communication,* vol. 29*, no. 3, pp. 252–273, 2019.

[19] R. B. Hubert, E. Estevez, A. Maguitman and T. Janowski, "Examining government-citizen interactions on twitter using visual and sentiment analysis," in *Proc. of the 19th Annual Int. Conf. on Digital Government Research: Governance in the Data Age*, Delft, Netherlands, pp. 1–10, 2018.

[20] A. Ceron, L. Curini and S. M. Iacus, "Using sentiment analysis to monitor electoral campaigns: Method mattersevidence from the United States and Italy," *Social Science Computer Review,* vol. 33*, no. 1*, pp. 3–20, 2015.

[21] A. Bruns and J. Burgess, "Ausvotes: How twitter covered the 2010 Australian federal election," *communication*," *Politics & Culture,* vol. 44*, no. 2*, pp. 37–56, 2011.

[22] M. Puri, A. Varde, X. Du and G. De Melo, "Smart governance through opinion mining of public reactions on ordinances," in *Proc. of 2018 IEEE 30th Int. Conf. on Tools with Artificial Intelligence (ICTAI)*, Volos, Greece, IEEE, pp. 838–845, 2018.

[23] S. S. Hasbullah, D. Maynard, R. Z. W. Chik, F. Mohd and M. Noor, "Automated content analysis: a sentiment analysis on Malaysian government social media," in *Proc. of the 10th Int. Conf. on Ubiquitous Information Management and Communication*, Danang, Viet Nam, pp. 1–6, 2016.

[24] M. Awais, S. Hassan and A. Ahmed, "Leveraging big data for politics: Predicting general election of Pakistan using a novel rigged model," *Journal of Ambient Intelligence and Humanized Computing,* vol. 12, pp. 1–9, 2019.

[25] A. Tumasjan, T. Sprenger, P. Sandner and I. Welpe, "Predicting elections with twitter: what 140 characters reveal about political sentiment," in *Proc. of the Int. AAAI Conf. on Web and Social Media*, Washington, D.C., vol. 4*, pp. 178–185, 2010.

[26] X. Zhou, X. Tao, J. Yong and Z. Yang, "Sentiment analysis on tweets for social events," in *Proc. of the 2013 IEEE 17th Int. Conf. on Computer Supported Cooperative Work in Design (CSCWD)*, Whistler, BC, Canada, IEEE, pp. 557–562, 2013.

[27] A. Bermingham and A. Smeaton, "On using twitter to monitor political sentiment and predict election results," in *Proc. of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011)*, Chiang Mai, Thailand, pp. 2–10, 2011.

[28] V. Goel, A. Kr Gupta and N. Kumar, "Sentiment analysis of multilingual twitter data using natural language processing," in *Proc. of 2018 8th Int. Conf. on Communication Systems and Network Technologies (CSNT)*, Nagpur, India, IEEE, pp. 208–212, 2018.

[29] R. Bibi, U. Qamar, M. Ansar and A. Shaheen, "Sentiment analysis for urdu news tweets using decision tree," in *Proc. of 2019 IEEE 17th Int. Conf. on Software Engineering Research, Management and Applications (SERA)*, Honolulu, HI, USA, IEEE, pp. 66–70, 2019.

[30] V. Srividhya and G. R. Meenakshi, "Comparison of sentiment analysis of government of India schemes using tweets," *International Journal of Computer Sciences and Engineering,* vol. 6*, no. 6*, pp. 998–1001, 2018.

[31] A. Aslam, U. Qamar, P. Saqib, R. Ayesha and A. Qadeer, "A novel framework for sentiment analysis using deep learning," in *Proc. of 22nd Int. Conf. on Advanced Communication Technology (ICACT)*, Phoenix Park, Korea (South), IEEE, pp. 525–529, 2020.

[32] H. Nguyen and M. Nguyen, "A deep neural architecture for sentence-level sentiment classification in twitter social networking," in *Proc. of Int. Conf. of the Pacific Association for Computational Linguistics*, Yangon, Myanmar, pp. 15–27, Springer, 2017.

[33] S. Seo, C. Kim, H. Kim, K. Mo and P. Kang, "Comparative study of deep learning based sentiment classification," *IEEE Access,* vol. 8*, pp. 6861–6875, 2020.