Telematics and Informatics 48 (2020) 101345

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Sentiment analysis of extremism in social media from textual information

Muhammad Asif[a], Atiab Ishtiaq[a], Haseeb Ahmad[a,*], Hanan Aljuaid[b], Jalal Shah[c]

[a] Department of Computer Science, National Textile University, Faisalabad, Pakistan
[b] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia
[c] Department of Computer System Engineering and Sciences, Balochistan University of Engineering and Technology, Khuzdar, Pakistan

A R T I C L E I N F O

Keywords:
Extremism
Multilingual
Lexicons
Multinomial Naïve Bayes Linear Support Vector Classifier

A B S T R A C T

Uncertainty in political, religious, and social issues causes extremism among people that are depicted by their sentiments on social media. Although, English is the most common language used to share views on social media, however, other vicinity based languages are also used by locals. Thus, it is also required to incorporate the views in such languages along with widely used languages for revealing better insights from data. This research focuses on the sentimental analysis of social media multilingual textual data to discover the intensity of the sentiments of extremism. Our study classifies the incorporated textual views into any of four categories, including high extreme, low extreme, moderate, and neutral, based on their level of extremism. Initially, a multilingual lexicon with the intensity weights is created. This lexicon is validated from domain experts and it attains 88% accuracy for validation. Subsequently, Multinomial Naïve Bayes and Linear Support Vector Classifier algorithms are employed for classification purposes. Overall, on the underlying multilingual dataset, Linear Support Vector Classifier outperforms with an accuracy of 82%.

1. Introduction

The social network has now emerged as an essential part of an individual's life. It has changed the way of living in the 21st century. Globalization has played an important role by the beginning of the last era of the twentieth century in linking diverse people around the world. Through online communication, people around the globe have started to understand the norms, culture and traditions of each other. As a result of this, similar-minded people have started working together to achieve a common goal (Yadav and Manwatkar, 2015).

Search engines and social networks are entirely different sources of data that can reveal imperative information (Lali et al., 2016). For making social relations with other people and sharing personal or real-life situations, social networking sites such as Facebook, LinkedIn, and Twitter are commonly used by people. Along with informative and refreshing contents, these social networking sites are also responsible for spreading hate speech (Barnidge et al., 2019). On these social networking sites, users share their views about different events, news, and products (Haider et al., 2018). Facebook platform framework was launched in May 2007 for developers to create an application. Now, it also provides the opportunity to exchange information while communicating with other people.

* Corresponding author.
E-mail addresses: asif@ntu.edu.pk (M. Asif), atiab.sheraz@gmail.com (A. Ishtiaq), haseeb_ad@hotmail.com (H. Ahmad), haaljuaid@pnu.edu.sa (H. Aljuaid), engrjalalshah@yahoo.com (J. Shah).

https://doi.org/10.1016/j.tele.2020.101345
Received 3 October 2019; Received in revised form 3 January 2020; Accepted 10 January 2020
Available online 15 January 2020
0736-5853/

Currently, Facebook is listed among the most widely used social networks since, out of 35 million active users of social media in Pakistan, 32 million are hold by Facebook (30 Pakistan's Digital and Social Media Marketing Stats and Facts, 2019). This interesting enhancement has opened many horizons as well as challenges for scholastic and research communities as well (Akter and Aziz, 2017). Currently, social networks are being widely used for various purposes. Even Al-Rawi pointed out the sale of drug "fentanyl" on social media.

People use these social sites to express their thoughts individually or by joining different groups and pages. People perform various social activities suited to their cognitive views. Different studies are performed to analyze the causes of addiction and the impact of Facebook on people (Foroughi et al., 2019). As from past years, due to the impoverished neighborhoods and absence of education, the extremist thoughts are thriving in the media, political circles, elite circles, and educational establishments of Pakistan. Extremism is a sociological wonder that has laid the foundations in social, financial, and political disparities; moreover, its religious roots rose above every single ordinary clarification and cures. The more religious general public has the more harmful impacts of extremism (Jawaid, 2018).

According to Davies, extremism is provoked "when you do not allow for a different point of view; when you hold your own views as being quite exclusive and when you don't allow for the possibility of difference" (Davies, 2020). The author further added to this definition that " when you want to impose your view on others while using violence if necessary." Extremism is also defined as " activities (beliefs, attitudes, feelings, actions, strategies) of a character far removed from the ordinary" (Coleman and Bartoli, 2003). There is a type of extremism that is "violent extremism" that is defined as the beliefs and actions of people who support or use ideologically-motivated violence for further radical ideological, religious, or political aims (DHS, 2019). Violent extremism is rapidly increased in recent years, and the main reason for this increment is the internet and social media (Alava et al., 2017).

People post and comment about their views, opinions, and response to different events and products on Facebook regularly. Due to this, Facebook has become a valuable source of sentiments. Sentiment analysis is an efficient and effective way of finding the people view, opinion, and the response regarding any product, incident, and an event (Can et al., 2018). Sentiment analysis also helps to computationally find and cluster the views showed in a piece of text (Prabowo and Thelwall, 2009). For example, a news agency can retrieve a timely response on news of any incidents that happened by evaluating people's responses and opinions on that news in the form of comments, tagging, and posting. These views also include foul, overtly sexiest and racist language and threats termed as extreme speech (Johnson, 2018). Different social groups and pages are created on Facebook to exploit such extremism, while the others are formed to oppose such activities openly. We can also mine people's views from such posts and comments. For instance, from the news "terrorists attack on school" We can analyze the extreme sentiment by performing sentiment analysis (Shahzad et al., 2017).

On Facebook, in addition to English and Urdu, people of Pakistan mostly write the Urdu language in English format that is called Roman Urdu to show their sentiments. These views also include extreme speech. By performing multilingual based sentiment analysis we can identify the views of people regarding extremism and other related thoughts to take action.

Lexicon and machine learning based approaches are used to perform sentiment analysis. In the formen, sentiment lexicons are used to express sentiments, e.g., "attacked" shows extreme sentiment, and "happy" shows moderate sentiment. Lexicons polarity can be domain-dependent. This approach generally uses a dictionary of lexicons to determine the sentiment orientation (Ding et al., 2008). This approach is very useful in analyzing the text of the documents and sentences. For instance, "I love Pakistan" shows a moderate sentiment. In the later approach, we train a sentiment classifier with the help of labeled data to predict the sentiment (Medhat et al., 2014). This is the frequently used approach for sentiment analysis. Large datasets are usually processed with machine learning classifiers. However, it is very hard and time-consuming to do manual labeling of text for machine learning.

Text categorization can be done as subjective and objective approaches. Any opinion, review, and discussion is referred as the subjective approach. Any neutral text that is based on fact belongs to the objective category. The objective of sentiment analysis is to classify subjectivity related text toward different domain labels (Syed et al., 2010). The sentiment analysis can be performed on the comments and posts gathered from social networking sites to reveal the areas of interest of people.

The extremist behavior of people is increasing due to the different political, social, religious situations of Pakistan (Ollapally, 2008). Due to this, people favor violence and become the cause of violent extremism (Odekon, 2015). All extremist behaviors are not caused by violence, but if it is decided that anxiety, fear, and violence are acceptable to attain changes like social, political, or ideological goals, this is called violent extremism. This paper focuses on the extraction of violent extreme sentiment from the Facebook multilingual text of different users. This sentiment extraction report can help the government and organizations to monitor the people's views in different situations and help to create remedies to cope with the trend of extreme behavior among people.

For underlying research, a method is presented that classifies the multilingual text into any four classes of extremism, i.e., moderate, neutral, low extreme, and high extreme based on sentiments of text. For this purpose, at first, comments and posts of users are extracted from different news pages of Facebook for the topic of extremism. The extracted dataset is comprised of political, terrorism, and different social issues like rapes, target killing based contents. News pages are used for data extraction because every incident, happening, and people's views can be found on such news pages in the form of posts and comments. Data is extracted by using Data Miner Scrapper from different pages of news agencies, including ARY news, Ptv news, Dawn, The News, Samaa, Express, Dunya news, and Geo. This text data includes posts and comments in Urdu, English, and Roman-Urdu, depending upon the choice of language by the people.

This paper applies the both subjective and objective sentiment analysis approaches to the extracted Facebook data. Due to the unavailability of a multilingual extreme lexicon dictionary, the lexicon dictionary is created from scratch comprising of lexicons that show a different level of extreme and moderate sentiments. Each lexicon is assigned a weight between $+5$ and -5 to show the sentiment orientation of lexicons toward extremism as well as moderation, respectively. With the help of these lexicons, each post and comment gets a score based on the matching of sentiment lexicons. After setting the threshold, each post and comment is labeled in any four classes, i.e., moderate, neutral, low extreme, and high extreme. This dataset is then classified with the help of machine learning classifiers. Sentiment classifier is trained with the help of automatically labeled data to assign sentiment polarities to new Facebook data. It is believed that this method is desirable for practical application because of its automation since no manual effort is required for labeling large amount of data. It is also desirable since it performs multilingual sentiment analysis by using the most commonly used languages in Pakistan.

1.1. Related concepts

Data mining. A widely accepted definition of data mining defines it as “the discovery of interesting, unexpected, or valuable structures from large datasets” (Hand, 2007). Data mining has two diverse aspects. One aspect deals with the structure of large-scale data called ‘global’, and the purpose of this aspect is to represent the features’ distributions. Other aspect deals with small-scale structures known as ‘local,’ and the purpose of this aspect is to identify anomalies and to decide whether these are genuine or occurred by chance.

The preparation of data is an extensive phase of data mining. Abundant of work has been put forward, addressing data quality (Zhang et al., 2003). Task-relevant data should be appropriately distributed, it should not contain incorrect or missing values, and all features must be important while having maximum information gain. This requires paying special attention regarding the following scenarios, i.e., not to:

- disguise data hidden patterns that could be useful,
- lower the performance,
- give a lower quality result.

The world is connected to many types of links, such as emails, web pages, and online social networks, etc. Social network mining and mining of communities is an active research area of the current era (Yang and Wu, 2006). For social networks mining, several challenges include but are not limited to followings:

- Static structures of social networks
- Social networks dynamic behavior.

Sentiment analysis. Sentiment analysis is the branch of study that examines views, attitudes, feelings, assessments, evaluations, and sentimentalities of people regarding different entities such as products, facilities, societies, characters, matters, occasions, topics, and their attributes. There are also many names and somewhat different jobs, e.g., “sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, review mining,” etc. Though, they are now all under the shade of opinion mining and sentiment analysis. The word sentiment analysis normally used in industry, but analysis and opinion mining are commonly active in the educational field (Liu, 2012; Nasukawa and Yi, 2003).

Humans and machines can only be distinguished by sentiments or emotions. Several researchers are trying to construct techniques to create feelings in machines. In parallel, others are also working to automatically extract particular news, products or any other required aspect of life (Islam and Dey, 2016). At present, sentiment analysis through natural language processing is the most challenging job being widely researched by the scholastic community.

Social media usage is increased exponentially in the past few years, with the proliferation of functionalities that are available online (Mittal et al., 2016). Social networking sites are typically used to state the particular life opinions and experiences (Go et al., 2009). Social networking sites are categorized into E-commerce websites such as Amazon, Flipkart, and online social media websites, including Facebook, Twitter, and Sina Weibo. The other categories include academic websites such as Google Scholar, DBLP, and Research gate, on the other hand, professional websites, include LinkedIn, Stack Overflow, and Quora. Some among these social networks are recreational, while the others provide a platform for the users to be in touch with each other by distributing the information, experiences, and views in many forms like tweets, images, status, emoticons, forums, posts, ratings, comments, videos, and blogs.

The intensive use of social media is also influencing sentiments of people in favor of or against specific topics including government, extremism, education or financial policies, organization, etc. Therefore, to recognize sentiments behind the posts on social media forums, there is the ultimate necessity of a well-organized, effective, and efficient technique (Islam et al., 2016).

For sentimental analysis, subjective lexicon and machine learning based methods are used (Kaur and Gupta, 2013).

Lexicon based sentiment analysis: Opinion words are utilized in numerous sentiment classification tasks. Positive opinion words are utilized to express some desired states, while negative opinion words are utilized to express some undesired states. There are likewise opinion expressions and idioms, which together are called opinion lexicons (Khoo and Johnkhan, 2018). There are three main approaches to arrange or gather the opinion words list including manual, corpus and dictionary-based approaches. The manual methodology is extremely tedious, and it isn't utilized alone. It is normally joined with the other two automated methodologies as the last check to avoid the missteps that occur because of automated techniques.

Machine Learning based Sentiment Analysis: Machine learning provides the system with the ability to learn automatically; there is no need for programming the system (Pang et al., 2002). It focuses on the development of a program that accesses data and uses it for learning. The process of learning starts with some initial data or instructions, and it checks for patterns in data for making effective decisions in futures. Basically, it allows learning automatically without human interference (Tong and Koller, 2001). Supervised learning is a type of machine learning algorithms, which depends on labeled data. Supervised algorithms apply to what is learned from the labeled data (Kotsiantis, 2007). First, the training data is analyzed, the algorithms then make predictions based on learning. This output can be compared with the correct outputs to detect errors. There are many supervised learning algorithms, i.e., Naïve Bayes, Support Vector Machines, Logistic Regressions and Neural Networks, etc. (Caruana and Niculescu-Mizil, 2006). Sometimes it becomes difficult to create the labeled training data, so unsupervised is employed to overcome this problem (Gentleman and Carey, 2008; Ko and Seo, 2000). Examples of unsupervised learning algorithms are clustering, lexicon-based approaches, etc.

Scikit-learn: This bundle centers around bringing machine learning to non-masters utilizing a broadly useful high-level language. Scikit-learn (previously scikits.learn) is a free machine learning library for the Python programming language. It provides different classification, regression, and clustering algorithms including Support Vector Machines, Random Forest, Gradient Boosting, K-means and similar others. It is intended to interoperate with the Python numerical and logical libraries NumPy and SciPy.

Data Miner Scraper It is a tool to scrape data from HTML pages and subsequently to export it to an excel spreadsheet. This tool can extract any table or list from a webpage. It allows importing data in different formats including XLS, CSV, XLSX, or TSV. It permits to scrap 50,000 recipes from different web pages. By using one of the thousands of scraping recipes, it can exchange the majority of the trendy websites to CSV with a single click. It also provides the facility to scrap data manually, depending on need. Thus, the underlying tool is enabled to fulfill the following requirements:

- To extract social media profiles, emails, and user IDs from groups and pages.
- To extract email and addresses.
- To find contact info from professional social profiles.
- To analyze posts for likes, comments, connections, and contacts.

In short, the above concepts and techniques are helping us to perform extreme sentiment analysis. The data miner scraper tool is used to collect data from the Facebook. Lexicon based sentiment analysis is employed to create the label dataset. For data classification, two machine learning algorithms are used. Scikit python library is also employed to use inbuilt methods for data classification in python language.

2. Literature review

For keyword matching, Shashank H. Yadav and Pratik M. Manwatkar applied pattern matching algorithm for violent keyword detection from social networking comments and to prevent it from publishing on the social platform (Yadav and Manwatkar, 2015). This approach does not need any human interference, as was required by earlier works (Chen et al., 2012). More precisely, the technique is used to confine the violent words by detecting and then preventing these automatically. Thao T. Nguyen and Alla G. Kravets presented an analysis of Facebook comments while defining the module that depends on the structures of Facebook graph API (Nguyen and Kravets, 2016) The goal of this study was to recognize the interest area of the targeted user. This approach recognized some users' interests while comparing them to their friend's interests to better reflect the overlapping interests. Himanshi Agrawal and Rishabh Kaushal employed two text mining methods to measure the connection between posts and comments of two public pages Wikipedia and India-forum.com (Agrawal and Kaushal, 2016).

Antonio Teixeira and Raul M. S. Laureano presented sentiment analysis by employing different tools (Teixeira and Laureano, 2017). The authors utilized the data of Facebook fashion page. Along with sentiment analysis, the authors also explain to extract data from Facebook using open-source tools.

To recognize political sentiments, posts are also analyzed (Pang et al., 2002). In political postings, several keywords are found dominant. First, the unique words' dictionary is prepared that utilizes political or nonpolitical posts and comments. Subsequently, the system is trained using the Naïve Bayes algorithm. Each word is extracted from posts and then matched with the words of the dictionary for classification, to retrieve the sentiments stated in the posts or comments. Sounthar Manickavasagam and B.Yinayaga Sundaram examined the number of likes on users' cover photos on Facebook to identify their gender and influence (Manickavasagam, 2014). By using the clustering coefficient, triadic census, and degree analysis, the authors extracted the influencers.

Márton Miháltz, Tamás Váradi presented the approaches and outcomes of a project that gathers and examines the public comments written in reply to political posts on Facebook using natural language processing and social psychological approaches (Mihaltz and Varadi, 2016). The main objective of the study was to search for emotional approaches and social actions.

According to (Ahmed and Diesner, 2012), online social networking sites (SNS) are growing in Pakistan, India, and Bangladesh in terms of usage. These social sites are also impacting the individual, professional, and public life of millions of people. Authors discussed that how people response on SNS with respect to different time events. Results displayed that different religious matters and the social caste system have a great impact on growing SNS. Also, more people were involved in SNS due to its accessibility in local languages and through cell phones.

The work presented in (Masiha et al., 2018) explored the connection between the use of Facebook and the political participation of youth in Pakistan.. Data on youth political activities was gathered by using questionnaires. It was concluded in the study that substantial association was present between the use of Facebook and political participation. Denzil Correa, Ashish Surekarevealed that online extremism is expanded and has become a major and growing concern to the society, governments, and law enforcement agencies around the world. Their research shows that various platforms on the Internet are being used for hateful intent. Such platforms are being used to form hate groups, prejudiced societies, spread extremist agendas, provoke anger, or violence. Automatic detection of online radicalization is a technically challenging problem because of the vast amount of data.

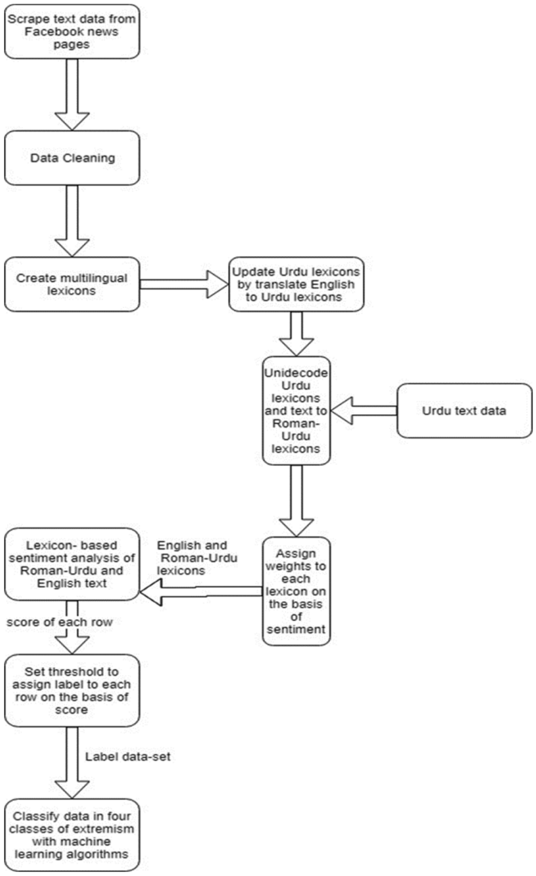

Fig. 1. Workflow of Methodology.

Different crawlers are used to crawl information from online social websites, Fredrik Erlandsson et al. presented the crawler named *SINCE* and compared it with the other available crawlers (Erlandsson et al., 2015). The presented crawler varies expressively from comparative crawlers in terms of efficiency and depth of crawling. Using the underlying crawler, one can parse all communication associated with each post. Joseph Mei and Richard Frank proposed an approach (Mei and Frank, 2015) to build a web-crawler to collect data from extremist websites. This crawler uses sentiment-based classification rules which let the crawler to make decisions on the material of the webpage. At first, contents from 2,500 web pages were collected for each of the four different sentiment-based classes: pro-extremist websites, anti-extremist websites, neutral news sites, and sites with no discussion of extremism. Then speech tagging was used to find the most frequent keywords in these pages. The result showed an 80% success rate on differentiating between the four classes and a 92% success rate at classifying extremist pages.

Albeit of such nuance works, the sentiment analysis on multilingual data in the context of extremism was still missing, which motivated us to perform such analysis in the context of Pakistan.

3. Methodology

The work presented in this research will reveal the extreme sentiments of users from the posts and comments they write on Facebook. Various public pages are identified containing extreme content written in diverse language including Urdu, English and Roman Urdu. A big portion of the population likes these pages and comments over these posts. These posts and comments are very helpful for the extraction of the sentiment of people. In this research, we propose an approach that is helpful for the analysis of extreme sentiment. The multilingual data that shows extremism is collected from Facebook's news pages by using the Data Miner Scrapper tool to perform sentiment analysis. The detailed methodology is depicted in Fig. 1.

To avoid consumption of time, and the labor of manual labeling of collected text for sentiment analysis, the lexicon-based technique is used. This technique is used on comments and posts to label them among four classes including moderate, neutral, low extreme, and high extreme. Domain related lexicons are created manually due to the unavailability of a multilingual extreme lexicon dictionary. Also, the dictionary is updated by adding more Roman-Urdu lexicons by translating English lexicons to Urdu and Unicode them into Roman-Urdu. Finally, an updated lexicon based dictionary is used for text labelingSubsequently, Multinomial Naïve Bayes and Linear Support Vector Classifier are employed to classify the posts and comments. By deploying this mechanism, the text involved in the propagation of extremism is identified through sentiment analysis.

3.1. Data collection

By using Data Miner Scraper, posts and comments are collected from Facebook news pages including ARY news, Ptv news, Dawn, The News, Samaa, Express, Dunya news, and Geo. Ptv news has16 889 079 fans and at 1st in all news agencies, ARY News has 16 640 629 fans, Express News has14 651 939 fansSamaa has 10 424 790 fansand Geo has 10 022 126 fans (Most popular Facebook pages in Pakistan | Socialbakers, 2018). These fans comment on these news pages on different news and show their sentiments. Data miner scrapper has inbuilt recipes to scrape data from different websites. Moreover, it also provides the facility of creating recipes. For Facebook, data miner scraper creates a recipe that scrapes posts and its comments along with URL address, date of posts, and comments writer. Subsequently, with the help of field experts, some initial lexicons of every category are created. These lexicons are then used to search for posts and comments related to our interest and for data collection purposes. Data is collected in the spreadsheets of the past eight to nine years. For all news pages, different spreadsheets are created for posts and comments.

3.2. Data cleaning

Errors and inconsistencies of data are detected and subsequently removed at this stage to improve the data quality. This mechanism is called data cleaning (Rahm and Do, 2000). Such inconsistencies could be a result of the integration of data from heterogeneous sources, or by entering false spellings during data entry. Information may be missed or may contain illogical data. In web-based data or information systems, data cleaning has become essential because diverse sources may hold superfluous data.

Data cleaning is done in three steps, including screening, diagnosis, and editing (Van Den Broeck et al., 2005).

- During the *screening* phase, lack or excess of data, outliers, strange patterns, and inconsistencies are identified.
- *The diagnosis* phase identifies errors, missing data, true extreme, and normal form.
- During *the editing* phase, correction and deletion are performed.

Three widely used methods for data cleaning include preprocessing, string comparison, and analytical linking (Winkler, 2003).

Preprocessing: At this stage, following actions are performed for data cleaning.

- Reply to comments are not parsed.
- The extraction of the text included in the URL is removed.
- URL in comments and posts is removed.
- Some posts and comments are blank (as they may contain images), therefore, not used.
- Removed emojis and other characters.
- All data is combined in one spreadsheet.
- Posts and comments text are converted into lowercase.

3.3. Creation of lexicons

Sentiment lexicons are generated by undergoing the following steps:

- Took around 1800 positive/negative English lexicon lists from GitHub source for initialization.
- Asssigned weights to the lexicons according to the extremism domain.
- We have updated the list by adding domain-specific (extreme) lexicons manually along with their weights.
- We have constructed the Urdu lexicons list related to extreme and moderate sentiment.
- We have updated the lexicons with Roman Urdu, English, and Urdu lexicons.

English sentiment lexicons: The lexicon-based approach relies upon sentiment words that express positive or on the other hand, negative sentiments. Words that encode an alluring state (e.g., \great“ and \good”) have a positive extremity, while words that encode a bothersome state have a negative extremity (e.g., \bad“ and \awful”). Experts have assembled sets of opinion words and expressions for modifiers, intensifiers, verbs, things separately (Khan et al., 2015).

We got some underlying opinion lexicon (positive/negative) from the GitHub source. These lexicons help for initialization. At that point, the lexicons are enhanced according to our sentiment classes, i.e., high extreme, low extreme, neutral, and moderate. Words that encode terrorism state (e.g., bomb blast, a terrorist attack) are categorized as high extreme sentiment, while words that show negativity along with less extremism (e.g., hurt, accident, fight) are counted in the low extreme class. The neutral class contains other domains related to sentiments like support, showbiz, etc. and the words that show alluring state along with kind of positivity (e.g., zindabad, peaceful) are included as moderate sentiments.

Hashtags of Facebook show sentiments of users. Hashtags are the way for including new settings and metadata to microblogs. Few sentiment labels introduce sentiment to the Facebook information, for instance, \#peace“, and \#attacks” hashtags show extreme and moderate sentiments. We include opinion hashtags into our opinion lexicons manually. There are additionally numerous words whose polarities rely upon the situation. For updation of lexicons, the dataset of posts and comments is tokenized based on the number of occurrences and the lexicons related to extremism and moderate added to the dictionary. Different English lexicons that represent a different level of extremism are shown in Table 1.

Table 1
English lexicons with weights.

Lexicons	Scores
Abuse	−2
Terror	−4
Fight	−3
Awesome	2
Peace	3

Urdu sentiment lexicons: In lexicon-based sentiment analysis, we require a sentiment encoded lexicons developed from a tremendous measure of content. The most appropriate and simple source for content gathering is online assets, for example, web journals, internet-based life destinations, online news or electronic diaries, and so forth (Pang and Lee, 2009). Although, Urdu is an exceptionally rich language, but its accessible assets on the web are reserved. A broad and far-reaching reviews are produced in English. In spite of the fact that the main approaches used to deal with English content (lexicon based and machine learning) can also be utilized for Urdu text, however, changes and adjustments are necessary because of huge orthographic, morphological, and syntactic contrasts between the two dialects as portrayed in going before area (Annett and Kondrak, 2008).

Urdu is also a widely used language on Facebook. Urdu lexicon dictionaries are not versatile than English because of a lack of sentiment analysis on Urdu. For extreme sentiment analysis on Urdu, Urdu lexicons are created from scratch. Two approaches are used for the creation of Urdu lexicons. First, the translation of already created English lexicons to Urdu with the help of google translator. Secondly, creation of Urdu lexicons manually from the gathered text.

Roman-Urdu sentiment lexicons: Roman-Urdu is also widely used for writing posts and commenting in Pakistan. In Roman Urdu, Urdu words are written with English chracters. It is mostly used because it is an easy way to type Urdu on electronic media. For the creation of Roman Urdu lexicons, first, Roman-Urdu lexicons are manually extracted from a text corpus. In Table 2, different Roman-Urdu lexicons are shown that are collected manually from the text along with their weights. These weights are assigned according to their extreme sentiments.

Unicode Urdu lexicons are also used to create Roman-Urdu lexicons. Unidecode is a Python module that is used to convert Urdu text into Roman-Urdu. Table 3 shows unidecode module that is used to convert Urdu text into Roman-Urdu text.

In this approach, each Urdu word is mapped with similar English sounds alphabet. Table 4 shows the Unicode data generated by using the procedure depicted in Table 3.

Assigning weights: Semantic orientation refers to a proportion of subjectivity and feeling in the content. It normally catches an evaluative factor (positive or negative) and intensity or quality (the degree to which the word, expression, sentence, or record being referred to is positive or negative) (Taboada et al., 2011).

In the proposed system, weight are assigned to the lexicons for depicting their strength. Scores are assigned manually according to the extreme sentiment of lexicons. The lexicons are weighted between −5 to + 5. This lexicon based system classifies text among four classes of extremism, i.e., highextreme, low extreme, neutral, and moderate. The lexicons that show sentiment related to high extreme are given the highest weight, i.e., “terrorists” is assigned with −5 score, “attack” is assigned with −4 score. The lexicons depicting less extreme are assigned less score , for instance, −2 is assigned to “injured” lexicon and similarly positive scores are assigned to moderate and neutral class lexicons. Table 5 depicts the English and Roman-Urdu lexicons, along with their scores.

Dataset and lexicons format: Multilingual lexicons are including Urdu, Roman-Urdu, and English words. Data is stored in the CSV file. Since, CSV format does not support urdu words, therefore, Urdu words arechanged into Roman-Urdu. The presented approachprovides a facility to Unicode the Urdu data into the Roman-Urdu format to eliminate this problem. By using the python Unidecode module, Urdu text is changed to Roman-Urdu format as depicted in Table 6.

Now, data-set and lexicons are composed of Roman-Urdu and English. Roman-Urdu data is the collection of both Urdu encoded to Roman Urdu and pure Roman Urdu data.

3.4. Feature selection

Tf-idf terms weighting In the text corpus, some words that are meaningless in the sense of context, for instance ‘the’, ‘a’, ‘is’ are are found with high frequency. Similarly, in Roman Urdu, some words like ‘ka’, ‘ki’, ‘hay’, etc. are useless for sentiment analysis. If we give weight on the basis of the count, important words with lesser count will be overlooked for sentiment analysis. Thus, Tf-idf is employed for assigning weight to important words for appropriate classification of the contents (Medina and Ramon, 2015).

Table 2
Roman-Urdu lexicons with Weights.

Lexicons	Score
shndr	2
shukar	2
qatal	−3
hamla	−3

Table 3
Code use to convert Urdu lexicons to Roman-Urdu Lexicons.

```
Step1: Import unidecode
Step2: Unidecode.unidecode("Urdu text")
```

Table 4
Urdu lexicons Unidecode to Roman-Urdu Lexicons.

Lexicons	Unicode to Roman-Urdu
بچاو	Bchw
زخمی	Zkhmy
کمزور	Khmzwr
حملہ	Hmlh
مرنا	Mrna

Table 5
Assign weights to Lexicons.

Lexicons	Score
Terrorist	−5
Attack	−4
Guilty	−3
Qatal	−3
Hacked	−1
Honor	2

Table 6
Assign Weights to Lexicons.

Urdu text	Roman-Urdu text
دکھا دی نا اپنی اوقات شریف فیملی نے———شریف فیملی اب ہر ہربہ استعمال کرے گی عاجزانہ درخواست ہے پاک آرمی سے کہ ان شرپسندوں کے خلاف ایکشن لے لے ورنہ یہ گرے ہوئے لوگ مزید کسی بھی حد تک نقصان پہنچائیں گے	dkhh dy n pny wqt shryf fymly ny……..shryf fymly b hr hrbh staml khry gy ajznh drkhwst hy pkh army sy khh n shrpsndwn khy khlf ykhshn ly ly wrnh yh gry hwyay lwg mzyd khsy bhy hd tkh nqsn phnchyayn gy
پاکستان کو اپنے فیصلے خد کو کرنے چاھیں-چین کے ساتھ-مل کر-کشمیر میں کتنا ظلم ھورہا ہے -کسی کچھ نہیں بولا -کشمیر کا ساتھ اللہ آپ کی مدد کرے گا-آمین	pkhstn khw pny fysly khd khw khrny chhyn. chyn khy sth.ml khr.khshmyr myn khtn zlm hwrh hy. khsy khchh nhyn bwl .khshmyr kh sth llh ap khy mdd khry g.amyn

Basically, Tf stands for term frequency, while Tf-idf refers to term frequency times inverse document frequency.

$$Tf - idf = tf\,(t,\,d) \times idf\,(t), \tag{1}$$

where, t id depicting the term and d is depicting the document.

Stopwords removal: Words such as 'and', 'the', 'him', 'wo', 'ya', 'aur', etc. are stopwords of English and Roman-Urdu language, these words are uninformative for representing the meaning of a text. These words are removed in classification to avoid the wrong prediction pattern (El-khair, 2006).

Common vectorizer: This process can be done with the help of 'Count Vectorizer' in python by using its library. Words that are not present in training text will be completely ignored in a future call to transform. It gives the facility to extract words in n-gram (Dey et al., 2018).

$$Bigram\ vectorizer = Count\_vectorizer\,(ngram_{range} = (1,\,2)) \tag{2}$$

Univariate selection: Measurable tests can be utilized to choose the feature having the most grounded associations with the output variable (Hira and Gillies, 2015). For this purpose, the scikit-learn library can be used that provides the SelectKBest class, which can further be utilized with a suite of various measurable tests to choose an explicit number of features. It works by choosing the best features dependending on univariate statistical tests. It may also be viewed as a preprocessing venture to an estimator.

Chi-squared test is employed to choose four of the best features from the extreme dataset. SelectKBest eliminates all but the features that have a maximum score. Set the value of k to select the number of high score features. SelectPercentile selects features according to the maximum score's percentile.

Principle component analysis: PCA utilizes linear algebra to change the dataset into a reduced form, thus, it may be viewed as a data reducing system. A property of PCA is that you can pick the number of dimensions or principal segments from the transformed outcome (Abdi and Williams, 2010).

For feature selection, the truncated singular value decomposition method is used because of it appropriate handling of the sparse matrix. This transformer basically performs linear dimensionality reduction. Unlike PCA, this estimator does not focus the information before processing the singular value disintegration, which makes it compatible with "scipy. sparse" frameworks productively.

Features on the basis of importance: This method is used to decide the important features with the help of trained supervised classifier. For important feature selection, Linear Support Vector Classifier with L1-based feature selection is used. L1 based feature selection is pushed with the L1 standard that has sparse arrangements as considerable evaluated coefficients are zero. At the point when the objective is to lessen the dimensionality of the information to use with another classifier, this can be utilized to choose the non-zero coefficients (sklearn.svm.LinearSVC, 2019).

3.5. Sentiment analysis

Now, data of approximately 20,000 rows (476,050 words) and 4300 lexicons word count are prepared for analysis. First step is to label a dataset for classification with the help of the presented procedure. This approach assigns a weight to each post and comments according to the weights of lexicons present in these posts and comments. A multilingual lexicon dictionary compares with each row of the dataset. If row's words match with the lexicons then sum up that certain lexicons' weights to find the overall score of the row. Each comment and post now gain some numeric positive or negative value. With the help of threshold, each post and comment label is classified in any four levels of extremism, i.e., moderate, neutral, low extreme, and high extreme. Now, the dataset is labeled for the classification of extreme data with the help of machine learning algorithms.

Lexicon-based sentiment analysis: In this step, lexicon-based analysis is performed. Data is tokenized, each word is compared with the lexicons dictionary. If it is matched, then weights of the lexicons are added to the sentence for finding an overall score of the sentence. Lexicons dictionary, text data, and the negation word list are needed to find the overall score of the sentence. Following are the steps to find the overall score of each post and comment.

Now, word_list of lexicons and negation words are loaded with the help of the algorithm shown in Table 7. In Table 8 the row-wise text is taken from a spreadsheet. Tokenization of the row into words is performed and then compared it with word_list. Subsequently, a score of matching lexicons is added to get the overall score of row-wise text.

Each row has its score according to the lexicons match with the text that gets through the approach depicted in Table 8. Then with the help of threshold value, each row is assigned with a label. If result score is greater or equal to -2 than the text assigned low extreme or if result $< = -3$ than highly extreme. If the result score is greater than 0, then moderate category is assigned. If the result score is equal to 0, than the respective text rows assigned as moderate.

Fig. 2 shows the scores and the label of each row of text. After this, a labeled dataset is created. This approach helps to avoid manual labeling of data. This system helps to label huge datasets in no time. Table 9 shows the size of different classes of the extreme dataset after labeling.

3.6. Classification

The labeled dataset is created in the previous step. Now with the help of two machine learning algorithms including Linear Support Vector Classifier and Multinomial Naïve Bayes the data is classified.

Naïve Bayes algorithm and its variant: One of the most popular machine learning algorithms for many years is Naive Bayes. Its simplicity makes it an attractive framework to be used for different tasks. It has obtained comparable performances in the tasks, though it has idealistic independence assumption based learning (Myaeng et al., 2006). That is why for the investigation of Naïve Bayes, many interesting works are put forward. Particularly in (Domingos and Pazzani, 1997), it is demonstrated that surprisingly the performance of Naïve Bayes is quite well for classification tasks.

In traditional machine learning, it is assumed that the distribution of training and test data must be identical. The transfer-learning algorithm is proposed for text classification that depends on the EM-based Naive Bayes classifier (Pan et al., 2008).

For text sentiment classification, there is a need for a filter that could build a model by learning and subsequently could predict the probability of a class level either as moderate, neutral, low extreme and high extreme. For this purpose, a method based on Bayes rule called Multinomial Naïve Bayes is incorporated. During the testing phase, each word extracted from the test dataset is compared with the keywords list. Total probability of extremism is originated by multiplying each matched words with the conditional probability of extremism. Likewise, the total probability of moderation is originated by multiplying each matched words with the conditional probability of moderation. Hence, the appropriate class of the text is assigned by comparing these probabilities. In actual terms, there is a need to find maximum aposteriori probability for the data with the help of Bayes theorem. Bayes rule is applied to a document *d* and an appropriate class c.

Table 7
Initialize and load the lexicons list.

Step1: word_list ← {}
Step2: negations ← set()
Step3: Function load_negations(self, filename)
Step4: Function load_wordlist(self, filename)
Step5: word_list ← {row['word']: float(row['score']) for row in reader}

Table 8
Assign weights to each row.

```
Step6: FUNCTION analyze(self, sentence)
Step7: token ← sentence_clean.split()
Step8: scores ← defaultdict(float)
Step9: words ← defaultdict(list)
Step10: for i, token in enumerate(tokens):
Step11: IF token in word_list AND not is_prefixed_by_negation:
Step12: core ← word_list[token]
Step13: scores[score_type] ← scores[score_type] + score
Step14: result ← scores['positive'] + scores['negative']
```

new 3 d map of milky way will r	0	Neutral
popular singer nazia iqbal accu:	2	Moderate
bayern munich clash with real r	2	Moderate
pm abbasi expresses anger ove	-4	highly extreme
parents of terminally ill uk toddl	-2	low extreme
mickey arthur hopes ball tampe	-1	low extreme
world class salah threatens ron	-2	low extreme
russia willing to play role in pak	10	Moderate
jamaat e islami to part ways wi	0	Neutral
rawalpindi corps commander sa	-4	highly extreme
at gaza s largest hospital, the v	-4	highly extreme
trump, apple ceo cook to talk tr	0	Neutral
full schedule of icc cricket worl	-2	low extreme
gautam gambhir steps down as	0	Neutral
jit finds rao anwar responsible f	-4	highly extreme
icc apologises as india rape gu	-2	low extreme
road rage in karachi claims one	-15	highly extreme
naeemul haq says iqbal s state	-5	highly extreme
respect the vote slogan is mear	4	Moderate

Fig. 2. Class labeling on the basis of the overall score of each row.

Table 9
Dataset Size of different extreme's classes.

Moderate dataset	5279
Neutral dataset	4315
Highly Extreme dataset	6912
Low Extreme dataset	2991
Total dataset	19,497

- $P(x1, x2 \ldots xn|cj)$ can only be predictable if an available training example is large.
- $P(cj)$ can be predictable from the training examples of class frequency.

For the classification of text, Multinomial Naïve Bayes Classifiers is applied.

$$c_{NB} = argmax_{c\varepsilon C} P(c_j) \prod_i P(x_i|c_j) \tag{3}$$

In python, scikit-learn provides the Multinomial Naïve Bayes library. In the proposed method, this library is used for text classification. The call of the library of Multinomial Naïve Bayes from Sklearn module is made as follows:

sklearn.naive_bayes.MultinomialNB(alpha = 1.0, fit_prior = True, class_prior = None) (4)

With the help of this library, data is classified in any of the four classes of extremism.

Support Vector Machines (SVM) and its variant: Support vector machines is a supervised machine learning algorithm. It is mostly used in a classification problem, but it can be used for regression as well. In this algorithm, each data is plotted as a point in n-dimensional space (Joachims, 1999). Subsequently, the hyperplane is found that segregates the classes. But if it is not linearly segregate, then SVM gives the facility of kernel trick. It takes low dimensional input space and changes it to higher, thus, it changes non-separable to separable.

Another method used for the classification of data in SVM is Linear Support Vector Classifier. It segregates our data into four classes by using linear hyperplane. In this proposed method linear kernel is used, which is suitable for text classification.

In Python, scikit-learn is the widely used library for the implementation of machine learning algorithms. It follows the same structure; first import library, then object is created, subsequently, model is fitted for making predictions. This following call imports the Linear Support Vector Classifier library from scikit-learn module in python.

Table 10
The accuracy of MNB.

Alpha	n-gram	Accuracy
0.01	(1,2)	0.66
0.01	(1,1)	0.639
1e-5	(1,2)	0.63

$$\text{from sklearn.svm import LinearSVC} \tag{5}$$

In Linear Support Vector Classifier parameter kernel is set to linear and implemented in liblinear term rather than libsvm. As libSVM workes for both linear and non-linear, its training timing complexity is around $O(n^2)$ to $O(n^3)$ (Fan et al., 2005). But liblinear works with linear kernel and also regularizes. Its training time complexity is O(n). It has the flexibility in choosing penalties and loss functions. It supports both dense and sparse inputs, and also handles multiclasses. So, it is best suited for our dataset. The tolerance of stopping criteria is set to *1e-3*, while penalization and primal optimization are set to false.

$$\text{LinearSVC(penalty} = \text{penalty, dual} = \text{False, tol} = 1\text{e} - 3) \tag{6}$$

4. Results and discussions

4.1. Multinomial Naïve Bayes

Table 10 shows that by changing different parameters, i.e. n-gram, alpha, tf-idf, and stop-words, the best achieved result by Multinomial Naïve Bayes on our dataset is 66%.

Confusion matrix: It is a technique to summarize the performance of classification. Alone accuracy misleads if number of observations are imbalanced in each class. It gives an idea of our model for getting the right and differentiating it from the error. It clearly shows that the correct classification of the low extreme class is less due to which its precision and recall perform poorly. In Table 11, it can be observed that high extreme class's true positive classification is higher than in other classes.

Precision and recall: Precision is also called positive predictive value. It is the fraction between the relevant instance among the retrieved instances. The recall is the sensitivity and it is the fraction between retrieved relevant instances over total relevant instances. In classification, precision is true positive (tp) divided by a total number of labeled(tp + fp) belonging to that class. Recall in classification total true positive (tp) divided by instances that actually belong to the class (tp + fn).

$$Precision = \frac{tp}{tp + fp} \tag{7}$$

$$Recall = \frac{tp}{tp + fn} \tag{8}$$

Table 12 shows the precision and recall of different classes. It can be observed that precision and recall of low extreme class is not good. Overall precision and recall are 0.65 and 0.66, respectively.

Graphical visualization of precision and recall can be observed with the help of Fig. 3. In which low extreme class line is lower than all other classes.

In Fig. 4 graph shows the number of correctly and incorrectly classified classes of test data. The test dataset contains 3900 posts and comments. Training data consist of 15,597 posts and comments in which moderate contain 5279, Neutral contain 4315, High extreme contains 6912 and low extreme contain 2991 posts and comments.

Receiver Operating Characteristic (ROC): It is used to evaluate the output quality of the classifier. ROC curve is drawn with the true positive rate and false positive rate. True positive rate is plotted on the y-axis while false positive rate is plotted on X-axis. The 'ideal point' is the top left corner of the plot, where the false positive rate is zero, and the true positive rate is one. This is not realistic, but it means a large area under the curve is better.

ROC is usually used for binary classification to study the output quality of the classifier. To find ROC for multi-label classification, it is compulsory to binarize the output. One curve is drawn per label but considering each indicator as a binary prediction. The ROC curve for Multinomial Naïve Bayes can be seen in Fig. 5. This curve shows a sudden change in behavior.

Table 11
Confusion matrix of MNB.

	Moderate	Neutral	High Extreme	Low Extreme
Moderate	799	98	108	60
Neutral	201	405	199	73
High Extreme	50	26	1211	89
Low Extreme	131	75	199	176

Table 12
Precision and Recall.

Classes	Precision	Recall	F1
Moderate	0.68	0.75	0.71
Neutral	0.67	0.46	0.55
Low extreme	0.44	0.30	0.36
Highly extreme	0.71	0.88	0.78
Avg/total	0.65	0.66	0.65

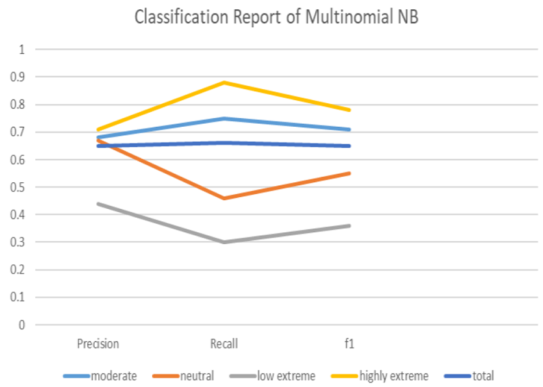

Fig. 3. Classification Report of Multinomial Naïve Bayes.

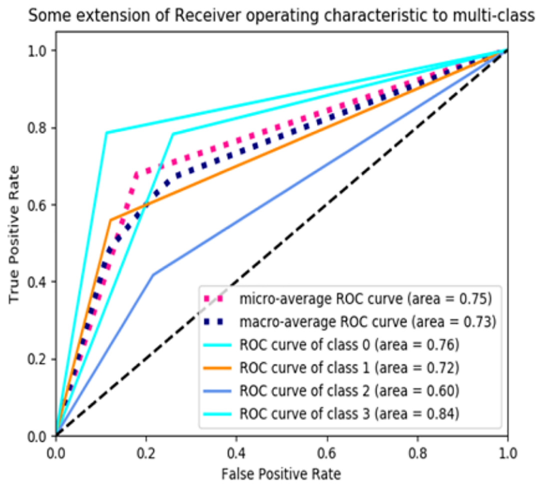

Fig. 4. Correctly vs. Incorrectly Classification of text with the help of MNB.

4.2. Linear Support Vector Classifier

It can be observed from Table 13 that Linear Support Vector Classifier gives much higher accuracy than Multinomial Naive Bayes. It gives approximately 82% accuracy for multilingual extreme text classification. It is the most suitable classifier for the underlying situation. It can be observed that by adjusting different parameters of Linear Support Vector Classifier. More precisely, it gives the best result on *uni-gram* with penalty *L1* and tolerance $1e-3$.

Confusion matrix With the help of Table 14, it can be observed that Linear Support Vector Classifier performs better than Multinomial Naïve Bayes.

Precision and recall: The precision and recall of the Linear Support Vector Classifier are depicted in Table 15. This classifier improves the precision of low extreme class but not same in the case of recall. Other classes precision and recall improve very well due to this SVM modification. In support vector machine, the Linear Support Vector Classifier modification performs well for text classification.

Graphical representation: Visually, it can be observed that precision and recall of moderate, neutral, and highly extreme is near to one that is the ideal situation. But, the recall of low extreme is low. The second graph in Fig. 6 shows the number of different classes

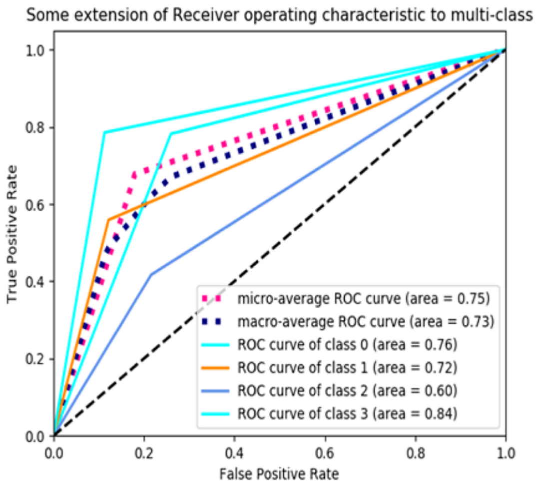

Fig. 5. ROC curve of Multinomial Naïve Bayes.

Table 13
The accuracy of Linear SVC.

n-Gram	Penalty	Tolerance	Stop words	Accuracy
1,1	L1	1e-3	none	82.1
1,1	L1	1e-3	true	80.1
1,1	L2	1e-3	none	80
1,2	L1	1e-3	none	81
1,2	L1	1e-5	true	80.6

Table 14
Confusion matrix of Linear SVC.

	Moderate	Neutral	Low Extreme	Highly Extreme
Moderate	975	69	5	16
Neutral	74	756	21	27
Low Extreme	71	145	173	192
Highly Extreme	11	27	57	1281

Table 15
Precision and recall of linear SVC.

Classes	Precision	Recall	F1
Moderate	0.86	0.92	0.89
Neutral	0.76	0.86	0.81
Low extreme	0.66	0.33	0.44
Highly extreme	0.87	0.93	0.90
Avg/total	0.81	0.82	0.81

of test data for classification. Incorrect classification by Multinomial Naïve Bayes is very low.

In test data, how many sentences from each class classified correctly and incorrectly can be visualized in Fig. 7.

Receiver Operating Characteristic (ROC): It also shows the classifier performance. Its left above corner is the ideal point. Due to the Linear Support Vector Classifier, the classes curve is more bent toward the ideal point in Fig. 8. That shows the good performance of the presented approach.

4.3. Comparison between Multinomial Naïve Bayes and Linear Support Vector Classifier

On the basis of both classifier's results, we can infer that Linear Support Vector Classifier shows better performance than Multinomial Naïve Bayes. Linear Support Vector Classifier is a suitable classifier for an underlying classification.

The precision and recall for both classifiers are visualized in Fig. 9.

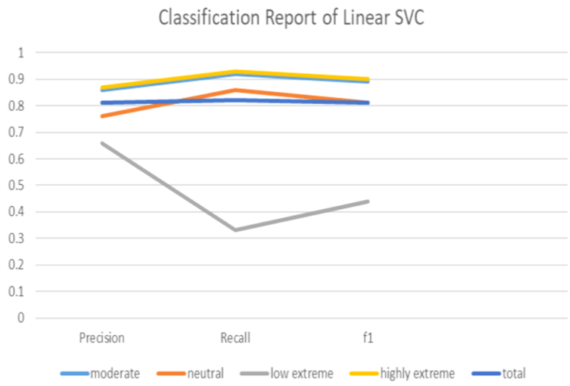

Fig. 6. Classification report of Linear Support Vector Classifier.

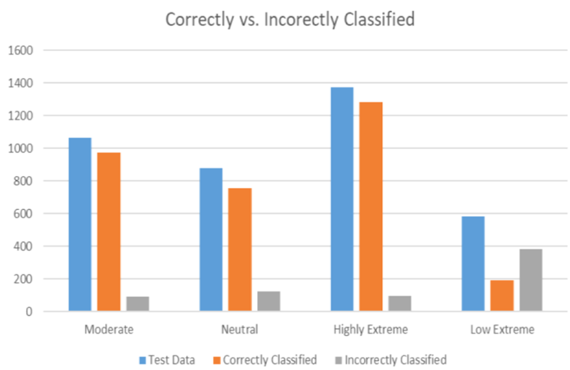

Fig. 7. Correctly vs. Incorrectly classified data.

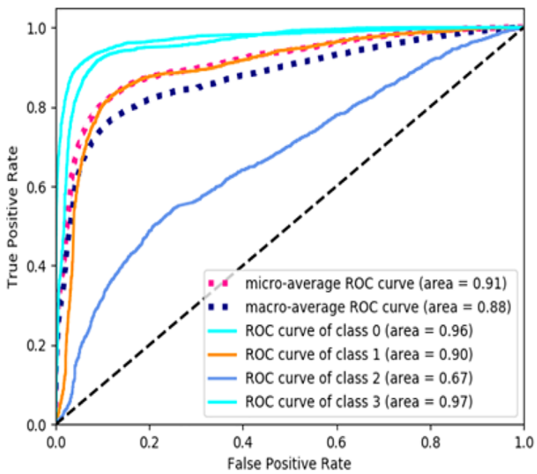

Fig. 8. ROC curve of Linear SVC.

4.4. The accuracy of Linear Support Vector Classifier by using different feature selection and extraction methods

In the univariant selection method, while using 'select k best' the accuracy of the classifier is 80%. In this method, selected value of k is 4000. In 'select percentile' when percentile is set 100%, then the accuracy of Linear *Support Vector Classifier* is 82.23% that is more than the comparative method.

For the principal component analysis, due to sparse matrix, truncated Singular Value Decomposition gives an accuracy of 73.8%, when 800 components are selected. This method gives a little improvement in accuracy. When the number of components is reduced to 500, it gives an accuracy of 72%.

By important feature selection method, on L1 selection it gives an accuracy of 82.33%. This is the best accuracy among the

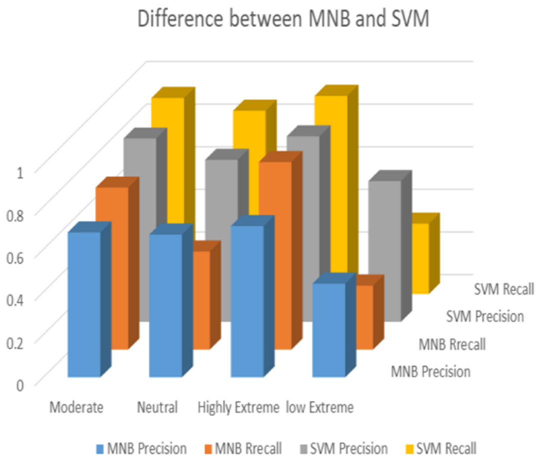

Fig. 9. Difference between Multinomial Naïve Bayes and Linear Support Vector Classifier.

incorporated feature selection methods. For feature extraction methods, when tf-idf is employed, the accuracy of the classifier is reduced to 72.8%. For the count vectorize method while using the unigram, it gives a good accuracy of 82.1%, which is better than n-gram.

Table 16 depicts the accuracies when all these feature selections and extraction methods are set to true one by one with the count vectorize method.

4.5. Comparison with different works

In Table 17, the systematic literature comparison with different work on sentiment analysis is presented. The authors performed sentiment analysis by using different datasets, languages, and techniques. In Syed et al. (2010), the authors perform sentiment analysis by using only the lexicons based technique on movies and products reviews. These reviews are in the Urdu language and the method is achieving average accuracy of 75%. In Khan et al. (2017), analysis is also performed on the Urdu language that is using social media data. Lexicon and machine learning based techniques are used by giving an accuracy of 67%. In Sharf et al. (2018), sentiment analysis is performed on social media data involving just the Roman-Urdu language. The classification achieves an accuracy of 80%. All remaining works in Table 17 are focused on the English language while using different techniques.

The proposed work is novel because it deals with different multi-languages (i.e., Urdu, English, and Roman-Urdu) simultaneously. These are the commonly used languages in Pakistan. Moreover, the analysis in the context of extremism is also absent in previous works. The uncertainty in political, religious, and social issues is basically causing extremism among people who leave their sentiments on social media. For this purpose, data is collected from different social media news channels and the multilingual lexicons dictionary is created in the context of extremism. Data labeling is performed by using a multilingual lexicon dictionary. The labeled dataset is classified by using Multinomial Naïve Bayes and Linear Support Vector Classifiers. By applying different feature extraction methods and tuning the parameters of classifiers, the results are analysed in detail.

4.6. Comparison between supervised and unsupervised classifiers

The classification is also performed by employing various classifiers, however, Linear Support Vector Classifier still performed the best as depicted in Fig. 10. SGD classifier which is an extension of SVM also performed well but little less than Linear Support vector Classifier. Further, it can be observed from Fig. 10 that KNN did not perform well. Moreover, KNN takes 0.039 s as training time and 2.633 s as testing time while giving only 26.9% accuracy. It is also found that Random Forest classifier takes high training and testing time comparative to other algorithms. Thus, the best choice for underlying task is Linear Support Vector Classifier.

Table 16
Accuracy on different features methods.

Feature selection method	Accuracy
Univariate Selection	Select k best = 80% Select percentile = 82.23%
Principle Component Analysis	Truncated SVD = 73.8%
Features importance	L1 = 82.33
Tf-idf	72.87%
Count vectorize	Unigram = 82.1%

Table 17
Comparison with different works.

Paper	Language	Sentiment analysis	Classes	Technique	Accuracy
[20]	Urdu	Lexicon based	Movie, product	Senti-unit	72%, 78%
[58]	English	Lexicon based	Multiple	SO-call	80%
[70]	Urdu	Lexicon +ML	Twitter positive/negative	Naive Bayes + LR + lexicons	Approx. 67%
[72]	English	Lexicon based	Stress and relaxation	tensiStrength	Approx. 80%
[73]	English	ML + bagging	Book, movie, shopping	MCS-based prediction systems.	Approx. 82%
[71]	Roman Urdu	Pos-tagging +ML	Social media Data	Neural Network	Approx. 80%
Propose work	Urdu, English, Roman Urdu	Lexicons based + ML	Social media extremism data	Multilingual lexicons + MNB + Linear SVC	Approx. 82%

4.7. How to deal with repeating lexicons

In some situations when the same sentence repeats in the text i.e., “I hate this text” hate has −3 lexicon score. So the whole sentence attains sentiment level −3. The second sentence is, “I hate this text. I hate this text.” would then have a sentiment level of −6, because the word “hate” occurs twice. But, it is just repeating the sentence with the same meaning. In our scenario, we tackle this problem as described follow.

First, we set the limit of repeating words in a sentence. We set this limit up to 3, if the same lexicon repeats, then we only consider the first three lexicons. After 3rd ocuurene their score is set to 0. Secondly, we minimize the score after 1st occurrence of lexicons. At second occurrence of same lexicon, 1 will be subtracted from the score. On the third occurrence, 2 is subtracted from the score. As in above example score of hate becomes −2 at second occurrence and total sentiment level of sentence would be −5. If hate lexicon is found three times in the text, then sentiment level would be −6, not −9. Even if it presents four, five or so on times then sentiment level will be −6 not −12. After 3rd occurrence its score is set to 0.

We introduce a new variable Counts[token] in Table 18 that handles the occurrence of same lexicon and makes decrement in score on the basis of number of occurrence.

Table 19 shows a new dataset after applying the above technique. It can be seen that some texts change their class because of the same lexicon limit.

Fig. 11 shows the accuracy of different algorithm on this dataset. Linear Support Vector Classifier shows the accuracy of 81%, and Multinomial Naïve Bayes shows 66% accuracy on this data.

5. Validation

5.1. Survey-based validation

The action of inspection or verifying the validity or accuracy of something is called validation. This part describes the validation process of the assigned labels to the dataset. Our lexicon based labeling system labels the text on the basis of lexicons matching and then gives an overall score to the text. On the basis of the score, each sentence is catogorize in any of the four classes. For validation purposes, a dataset of 25 similar random posts and comments is created and it is then labled by 109 different people. These people belong to different fields; some were students, teachers that belong to different domains, i.e., Computer Science, Literature, and Phycology. The survey was provided in the form of multiple choice questions. Each question had 4 options same as our labeled classes, i.e., high extreme, low extreme, neutral, and moderate. A total of 109 random people filled this survey while assigning each

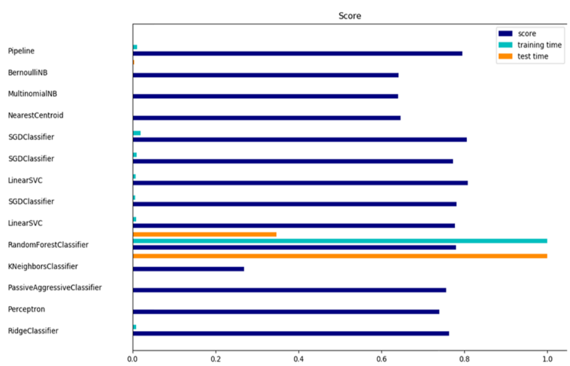

Fig. 10. Accuracy of supervised and unsupervised classifiers.

Table 18
Assign weights to each row.

Step6: FUNCTION analyze(self, sentence)
Step7: token ← sentence_clean.split()
Step8: scores ← defaultdict(float)
Step9: words ← defaultdict(list)
Step10: for i, token in enumerate(tokens):
Step11:if token in counts and counts[token] < 3:counts[token] + = 1
Step12: IF token in word_list AND not is_prefixed_by_negation:
Step13: score ← word_list[token]
Step14:if score < 0: score = score + (counts[token]-1)else: score = score-(counts[token]-1)
Step13: scores[score_type] ← scores[score_type] + score
Step14: result ← scores['positive'] + scores['negative']

Table 19
Dataset Size of different extreme's classes after setting the limit.

Moderate dataset	5446
Neutral dataset	4088
Highly Extreme dataset	7814
Low Extreme dataset	2149
Total dataset	19,497

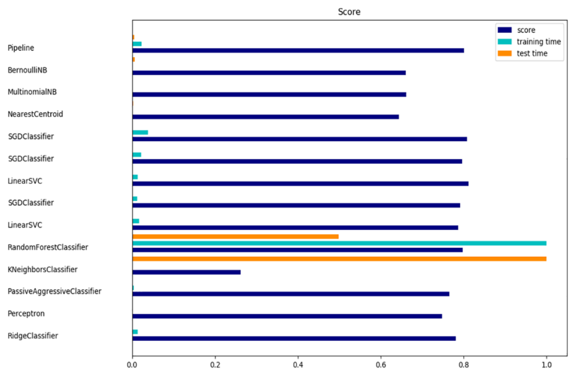

Fig. 11. Accuracy of supervised and unsupervised classifiers on new dataset.

text a class on the basis of their understanding. The dataset that is used for this form is shown in Fig. 12. The survey labeling result is also compared with our system's assigned labeling.

Among 25 posts and comments, 22 posts and comments got the same labels as assigned by the proposed system. Only three comments got the change labeling from the people. First, the comment is, "aese nahin bolty bhai may be sab ne inki service ko ya inki quality ko like kia ho aese kisi ko bura bhala nahin kehty", our system labels it as "moderate" but some people labeled it as "neutral". The second comment is, "they deserve to die like dogs what they are doing in yemen they will definitely land in hell" that is changed from "low extreme" to "high extreme" class. The third comment label is changed from "moderate" to "low extreme". On the basis of people reviews, our lexicons based labeling gots an accuracy of 88%.

During validation, we faced some challenges, as it is not easy for any person to assign lable to entire data for validation purposes. Due to which the size of the survey is reduced to 25. Also, people find it difficult to understand Roman-Urdu words. Because every person uses different spellings to write in Roman-Urdu, there is no spelling correction technique for Roman-Urdu. So, we personally assisted the people in understanding the text. Before conducting the survey, we explained the people the meaning of every label, so that they could label the data with full understanding.

5.2. K-fold cross-validation

Finding out the parameters of a prediction function and testing it on the same data is a methodological mistake: a model that would merely repeat the labels of the samples that it has just seen would have a perfect score but would fail to predict anything useful on yet-unseen data. This is called overfitting. To avert it, it is usual practice when performing a machine learning experiment to carry out part of the un-seen data as a test set X_test, y_test. Cross-validation performs division of data into corresponding subsets, executes the analysis on one subset (called the *training set*), and validate the analysis on the other subset (called the *validation set*). To minimize inconsistency, in many approaches several series of cross-validation are done using different divisions, and the validation results

DATA SET	LEXICON BASED	PEOPLE VIEW
zynab ky qatel dhondny walo ko rao anwar jysy dhashtgard police nazer nahi aty pehly inki illag to karwaly pir hum manlyngy ky ins	highly extreme	highly extreme
i don t know that why pakistan are in the favor of terrorist taliban in afghanistan even they are killing children destroying schools h	highly extreme	highly extreme
pakistan to have a complete tour of south africa in december pakistan to have a complete tour of south africa in decemberpakistan	neutral	neutral
asifa bhutto addresses party workers in lyari using hologram technologyasifa bhutto addresses party workers in lyari using hologran	neutral	neutral
but unfortunately u r a coward u also did not open thecorruption files of khursheed shah u indicated on the floor	low extreme	low extreme
it is not good for pakistan becoz any foreign army can come on our soil killed extremist persons we r sovereign country andshould r	highly extreme	highly extreme
chief justice zindabad	moderate	moderate
arey yaar bhatta nahi mila hoga yeahi to aik lanat karachi mein hai	low extreme	low extreme
waiting for the day when imran khan represents pakistan as a pm that day we can hold our heads high a man of poise personality ar	moderate	moderate
every one terrorist is only islamic person	highly extreme	highly extreme
pm shahid khaqan along with leaders from 52 countries join the opening of commonwealth heads of government meeting london :	moderate	moderate
40 militants killed 43 hideouts destroyed in operation khyber 3 ispr army troops apprehend six facilitators of terror in hangu | pakis	highly extreme	highly extreme
the committee headed by mr abbasi always against the police officers	low extreme	low extreme
aese nahin bolty bhai may be sab ne inki service ko ya inki quality ko like kia ho aese kisi ko bura bhala nahin kehty kk	moderate	neutral
sitting in a mixed gathering by choice scantly dressed by choice singing dancing by choice and she s offended at harassment if you c	low extreme	low extreme
they deserve to die like dogs what they are doing in yemen they will definitely land in hell	low extreme	highly extreme
islam don t permit to celebrate valentine day	moderate	low extreme
misha shafi is member of israeli communist group never trust her ali zafar remained mr asia he won t go for such cheap act	low extreme	low extreme
she was killed for money p honour killing even her brother admitted why is the hype over honour killing then i wonder when woul	highly extreme	highly extreme
usama khalid bachey 2013 aur 2018 mein bohat farq hai imature thinking these days all the recruitments done throw nts	neutral	neutral
most of them were candidates of pmln they contested independently as directed by ecp	neutral	neutral
most of it comes frm tax money gov ads	neutral	neutral
still this female has no shame	low extreme	low extreme
allah bless them	moderate	moderate
attack on mosque kills 50, hurt 100 attack on mosque kills 50, hurt 100peshawar a bomb attack on a mosque in darra adam khel on fr	highly extreme	highly extreme

Fig. 12. Shows the Survey dataset along with Lexicon-based labeling and people labeling.

are joined (e.g. averaged) over the series to give an evaluation of the model's predictive performance. K-fold is a non-exhaustive cross-validation method. In K-fold cross-validation, the novel example is arbitrarily divided into K equal-sized subsample. A single subsample is reserved as the validation data for testing the model, and the other K-1 subsamples are utilized as training data. The process of cross-validation is repeated K times, with each of the K subsamples used precisely once as the validation data.

For our model, 10-fold cross-validation is also used. Linear Support Vector Classifier resulted the best accuracy of 82% on the underlying dataset. This model is cross-validated by using a 10-fold method that gives 10 estimations including 0.78392217, 0.73002049, 0.75179487, 0.81384615, 0.84153846, 0.85120575, 0.84556183, 0.83119548, 0.75731144, 0.77207392. Average of these 10 estimation is 80%.

6. Conclusion

Sentiment analysis is a useful method to detect the sentiments of people on different events. Social media is a significant source of data, and researchers perform sentiment analysis on social media text to check trends on different events. Due to the different circumstances, youth is facing the challenge of extremism. Social media is a platform where users easily express their sentiments. In the Pakistan region, Roman-Urdu is the famous social media language that is used to express the sentiments. Along with this, users are also typing the text in Urdu, and English. So, to detect the reasons and cause of extremism in Pakistan, there was a need to perform sentiment analysis on social media multilingual text that could help the organizations to take action. Currently, Roman-Urdu is the new social media language that lacks in research, therefore, sentiment analysis becomes very challenging. Since people usually use Urdu, English, and Roman-Urdu. So, to perform sentiment analysis on multilingual data becomes very challenging and also it becomes difficult to analyze the extreme sentiments from multilingual data. To deal with these challenges, in this research, we proposed methods that effectively find the extreme sentiment from multilingual data. For this purpose, first we created a new multilingual lexicon dictionary manually that is comprised of extreme lexicons of different levels (i.e., moderate, neutral, low extreme, and high extreme) and also validated it with the help of a domain experts. Subsequently, the collected multilingual data is labeled into different classes of extremism with the help of a multilingual extreme dictionary. Third, classification is performed by using different supervised and unsupervised algorithms and concluded that supervise algorithms perform better than unsupervised algorithms. In supervised algorithms, Linear Support Vector Classifier resulted in the highest accuracy of 82%. In unsupervised algorithms, KNN classified with an accuracy of 26%. It is also concluded that news pages are a big source of extreme content. By labeling dataset into different classes of extremism, it was seen that highly extreme dataset is bigger than any other available dataset of similar type. In this work, we also dealt with the problem of the same sentence repetition by setting the allowable occurrence of same lexicon. Overall, the proposed work can be used by stakeholders for taking preemptive actions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

All the co-authors have agreed for the inclusion of their names.

References

30 Pakistan's Digital and Social Media Marketing Stats and Facts. [Online]. Available: https://www.soravjain.com/30-amazing-digital-and-social-media-marketing-stats-and-facts-of-pakistan-2018 (accessed 08.12.2019).
Abdi, H., Williams, L.J., 2010. Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics.
Agrawal, H., Kaushal, R., 2016. Analysis of text mining techniques over public pages of facebook. In: Proc. – 6th Int Adv. Comput. Conf. IACC 2016, pp. 9–14.
Ahmed, S., Diesner, J., 2012. Information network analysis to understand the evolution of online social networking sites in the context of India, Pakistan, and Bangladesh. ACM Web Science Conference Workshop.
Akter, S., Aziz, M.T., 2017. Sentiment analysis on facebook group using lexicon based approach. In: 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2016, pp. 8–11.
Alava, S., Frau-Meigs, D., Hassan, G., 2017. Youth and Violent Extremism on Social Media: Mapping the Research. pp. 1–167.
Annett, M., Kondrak, G., 2008. A comparison of sentiment analysis techniques: polarizing movie blogs. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 25–35 LNAI.
Barnidge, M., Kim, B., Sherrill, L.A., Luknar, Ž., Zhang, J., 2019. Perceived exposure to and avoidance of hate speech in various communication settings. Telemat. Inf. 44, 101263.
Can, E.F., Ezen-Can, A., Can, F., 2018. Multilingual sentiment analysis: an RNN-based framework for limited data. arXiv Prepr. arXiv1806.04511.
Caruana, R., Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. In: ACM International Conference Proceeding Series, pp. 161–168.
Catal, C., Nangir, M., 2017. A sentiment classification model based on multiple classifiers. Appl. Soft Comput. J. 50, 135–141.
Chen, Y., Zhou, Y., Zhu, S., Xu, H., 2012. Detecting offensive language in social media to protect adolescent online safety. In: Proc. - 2012 ASE/IEEE Int. Conf. Privacy, Secur. Risk Trust 2012 ASE/IEEE Int. Conf. Soc. Comput. Soc. 2012, pp. 71–80.
Coleman, P.T., Bartoli, A., 2003. Addressing extremism: defining extremism. In: Position Pap. New York Columbia Univ. Int. Cent. Coop. Confl. Resolut. 2, pp. 1–13.
Davies, L., 2020. Education against extremism. In: Education Against Extremism, pp. 1–21.
Dey, A., Jenamani, M., Thakkar, J.J., 2018. Sentiment wEight of N-grams in dataset (SEND): a feature-set for cross-domain sentiment classification. 2017 9th International Conference on Advances in Pattern Recognition, ICAPR 2017.
DHS, 2019. Targeted Violence and Terrorism Prevention. [Online]. Available: https://www.dhs.gov/tvtp (accessed 12.12.2019).
Ding, X., Liu, B., Yu, P.S., 2008. A holistic lexicon-based approach to opinion mining. p. 231.
Domingos, P., Pazzani, M., 1997. On the optimality of the simple bayesian classifier under zero-one loss. Mach. Learn. 29, 103–130.
El-khair, I.A., 2006. Effects of stop words elimination for arabic information retrieval : a comparative study. Inf. Sci. (Ny) 4 (3), 119–133.
Erlandsson, F., Nia, R., Boldt, M., Johnson, H., Wu, S.F., 2015. Crawling online social networks. In: 2015 Second Eur. Netw. Intell. Conf. pp. 9–16.
Fan, R., Chen, P., Lin, C., 2005. Working set selection using second order information for training support vector machines. J. Mach. Learn. Res. 6, 1889–1918.
Foroughi, B., Iranmanesh, M., Nikbin, D., Hyun, S.S., 2019. Are depression and social anxiety the missing link between Facebook addiction and life satisfaction? The interactive effect of needs and self-regulation. Telemat. Inf. 43, 101247.
Gentleman, R., Carey, V.J., 2008. Unsupervised machine leargn. In: Bioconductor Case Studies, pp. 137–157.
Go, A., Bhayani, R., Huang, L., 2009. Twitter sentiment classification using distant supervision. Processing 150 (12), 1–6.
Haider, S., Tanvir Afzal, M., Asif, M., Maurer, H., Ahmad, A., Abuarqoub, A., 2018. Impact analysis of adverbs for sentiment classification on Twitter product reviews. Concurrency Comput.
Hand, D.J., 2007. Principles of data mining. Drug Saf. 30 (7), 621–622.
Hira, Z.M., Gillies, D.F., 2015. A review of feature selection and feature extraction methods applied on microarray data. Adv. Bioinf.
Islam, T., Bappy, A.R., Rahman, T., Uddin, M.S., 2016. Filtering political sentiment in social media from textual information. In: 2016 5th Int. Conf. Informatics, Electron. Vision, ICIEV 2016, pp. 663–666.
Islam, S., Dey, J.J., 2016. Supervised approach of sentimentality extraction. In: 19th Int. Conf. Comput. Inf. Technol. pp. 383–387.
Jawaid, A., 2018. Rise in extremism.
Joachims, T., 1999. Svmlight: Support vector machine. SVM-Light Support Vector Machine, University of Dortmund 19 (4).
Johnson, B.G., 2018. Tolerating and managing extreme speech on social media. Internet Res. 28 (5), 1275–1291.
Kaur, A., Gupta, V., 2013. A survey on sentiment analysis and opinion mining techniques. J. Emerg. Technol. Web Intell. 5 (4), 367–371.
Khan, A.Z.H., Atique, M., Thakare, V.M., 2015. Combining lexicon-based and learning-based methods for twitter sentiment analysis. Spec. Issue IJECSCSE (National Conf. "Advanced Technol. Comput. Networking" – ATCON). p. 9477, 2105.
Khan, M.Y., Emaduddin, S.M., Junejo, K.N., 2017. Harnessing english sentiment lexicons for polarity detection in urdu tweets: a baseline approach. In: Proceedings – IEEE 11th International Conference on Semantic Computing, ICSC 2017, pp. 242–249.
Khoo, C.S.G., Johnkhan, S.B., 2018. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. J. Inf. Sci. 44 (4), 491–511.
Ko, Y., Seo, J., 2000. Automatic text categorization by unsupervised learning. In: Proceedings of the 18th conference on Computational linguistics, pp. 453–459.
Kotsiantis, S.B., 2007. Supervised machine learning: a review of classification techniques. Informatica (Ljubljana) 31 (3), 249–268.
Lali, M.I.U., Mustafa, R.U., Saleem, K., Nawaz, M.S., Zia, T., Shahzad, B., 2016. Finding healthcare issues with search engine queries and social network data. Int. J. Semant. Web Inf. Syst. 13 (1), 48–62.
Liu, B., 2012. Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. 5 (1), 1–184.
Manickavasagam, S., 2014. Exploring gender based influencers using Social Network Analysis. In: Sixth Int. Conf. Adv. Comput. pp. 224–228.
Masiha, S., Habiba, U., Abbas, Z., Saud, M., Ariadi, S., 2018. Exploring the link between the use of facebook and political participation among youth in Pakistan. J. Polit. Sci. Public Aff. 06 (01).
Medhat, W., Hassan, A., Korashy, H., 2014. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J.
Medina, C.P., Ramon, M.R.R., 2015. Using TF-IDF to Determine Word Relevance in Document Queries Juan.
Mei, J., Frank, R., 2015. Sentiment crawling: extremist content collection through a sentiment analysis guided web-crawler. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1024–1027.
Mihaltz, M., Varadi, T., 2016. TrendMiner: large-scale analysis of political attitudes in public facebook messages. 6th IEEE Conference on Cognitive Infocommunications, CogInfoCom 2015 – Proceedings.
Mittal, S., Goel, A., Jain, R., 2016. Sentiment analysis of E-commerce and social networking sites. In: Proc. 10th INDIACom; 2016 3rd Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2016, pp. 2300–2305.
Most popular Facebook pages in Pakistan | Socialbakers, 2018. [Online]. Available: https://www.socialbakers.com/statistics/facebook/pages/total/pakistan (accessed 12.12.2019).
Myaeng, S.H., Han, K.S., Rim, H.C., 2006. Some effective techniques for naive bayes text classification. IEEE Trans. Knowl. Data Eng. 18 (11), 1457–1466.
Nasukawa, T., Yi, J., 2003. Sentiment analysis: capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003, pp. 70–77.

Nguyen, T.T., Kravets, A.G., 2016. Analysis of the social network facebook comments. In: IISA 2016 – 7th Int. Conf. Information, Intell. Syst. Appl. pp. 1–5.
Odekon, M., 2015. Futures without violence. The SAGE Encyclopedia of World Poverty.
Ollapally, D.M., 2008. The politics of extremism in South Asia.
Pan, S.J., et al., 2008. Transferring naive bayes classifiers for text classification. Adv. Neural Inf. Process. Syst. 22 (PART 2), 29–34.
Pang, B., Lee, L., 2009. Opinion mining and sentiment analysis. Comput. Linguist. 35 (2), 311–312.
Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, pp. 79–86.
Prabowo, R., Thelwall, M., 2009. Sentiment analysis: a combined approach. J. Informetr.
Rahm, E., Do, H.H., 2000. Data cleaning: problems and current approaches. Bull. Tech. Comm. Data Eng. 23 (4), 3–13.
Shahzad, B., Lali, I., Nawaz, M.S., Aslam, W., Mustafa, R., Mashkoor, A., 2017. Discovery and classification of user interests on social media. Inf. Discov. Deliv. 45 (3), 130–138.
Sharf, Z., Saif, D., Rahman, U., 2018. Performing natural language processing on roman urdu datasets. IJCSNS Int. J. Comput. Sci. Netw. Secur. 18 (1), 141–148.
sklearn.svm.LinearSVC—scikit-learn 0.22 documentation. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html (accessed 08.12. 2019).
Syed, A.Z., Aslam, M., Martinez-Enriquez, A.M., 2010. Lexicon based sentiment analysis of Urdu text using SentiUnits. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 32–43 LNAI, no. PART 1.
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M., 2011. Lexicon-basedmethods for sentiment analysis. Comput. Linguist. 37 (2), 267–307.
Teixeira, A., Laureano, R.M.S., 2017. Data extraction and preparation to perform a sentiment analysis using open source tools: the example of a Facebook fashion brand page. Iberian Conference on Information Systems and Technologies, CISTI.
Thelwall, M., 2017. TensiStrength: stress and relaxation magnitude detection for social media texts. Inf. Process. Manag. 53 (1), 106–121.
Tong, S., Koller, D., 2001. Support vector machine active learning with applications to text classification. J. Mach. Learn. Res. 2, 45–66.
Van Den Broeck, J., Cunningham, S.A., Eeckels, R., Herbst, K., 2005. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS Med. 2 (10), 0966–0970.
Winkler, W.E., 2003. Data cleaning methods. ACM SIGKDD Work. Data Cleaning, Rec. Linkage, Object Consol.
Yadav, S.H., Manwatkar, P.M., 2015. An approach for offensive text detection and prevention in Social Networks. ICIIECS 2015–2015 IEEE International Conference on Innovations in Information, Embedded and Communication Systems.
Yang, Q., Wu, X., 2006. 10 challenging problems in data mining research. Int. J. Inf. Technol. Decis. Mak. 05 (04), 597–604.
Zhang, S., Zhang, C., Yang, Q., 2003. Data preparation for data mining. Appl. Artif. Intell. 17 (5–6), 375–381.