Societal Transformation: AI and Big Data Journal
Vol: 2(1): 25-37, 2024
DOI:10.20547/aibd.242102

A Novel Approach to Sentiment Analysis of Roman Urdu Data

Fatima Siddiqui[1]

Abstract: Sentiment analysis in Roman Urdu is increasingly crucial for enhancing consumer decision-making in diverse product domains. This paper addresses the challenge of extracting sentiment from product reviews written in Roman Urdu, leveraging Context-Free Grammar (CFG) to classify reviews into positive, negative, or neutral sentiments. Our study utilizes a dataset of online product reviews sourced from e-commerce platforms, focusing on the automation of sentiment classification. We propose a comprehensive methodology for sentiment extremity classification, demonstrating promising results through sentence-level analysis. This research contributes to advancing sentiment analysis in NLP, particularly in under-resourced languages like Roman Urdu, highlighting both methodological innovations and practical implications for e-commerce and beyond. Future work aims to further refine our approach and explore additional nuances in sentiment analysis for enhanced decision support.

Keywords— *Product review Analysis, Information Retrieval, Sentiment Analysis, Roman Urdu, CFGs*

1. Introduction

Sentiment analysis can be defined as the process of determining the attitudes or feelings of consumers, potential customers, or target audiences towards products or services. It is also called also known as opinion mining. The output can be classified as positive, negative, or neutral, and in some studies, it extends to other sentiments such as tension, depression, anger, vigor, fatigue, and confusion [2]. Sentiment analysis is a crucial tool for quickly connecting with consumer opinions. It is not a standalone topic but intersects with natural language processing (NLP), computational linguistics, and text analytics. However, NLP for the Urdu language is still in its infancy [8]. This research focuses on sentiment analysis of product reviews written in Roman Urdu. For a comprehensive taxonomy of sentiment analysis, interested readers are referred to [2].

Social media is a vast reservoir of data, with millions of people interacting daily and sharing their opinions, suggestions, critiques, appraisals, and inquiries about products and services

[1] Department of Computer Sciences (MSCS)
Karachi Institute of Economics and Technology

[4, 5]. These data are invaluable for producers, as transforming this raw information into actionable insights can help them design and create products that meet consumer demands. Sentiment analysis plays a key role in this transformation, especially for product reviews.

Analyzing customer reviews is essential for business success, as highlighted by [1]. Sentiment analysis significantly impacts the efficiency of e-commerce businesses. This research performs sentiment analysis on product reviews available in Roman Urdu. Typically, users provide reviews in English or Roman Urdu [6]. Given that Urdu is the native language of a large population, many users prefer to provide feedback in Roman Urdu due to the inconvenience of switc\hing keyboards and unfamiliarity with the Urdu script. This trend has led to the emergence of Roman Urdu as a significant medium of communication [7].

We explore the convenience of analyzing customer feedback sentiments—whether positive, negative, or neutral—in both Urdu and Roman Urdu. This paper also addresses the challenges posed by non-standard grammar and spelling, especially in Roman Urdu feedback.

The rest of the paper is organized as follows. The next section will present the literature review, followed by a discussion on the methodology adopted for this research. We then present case studies to highlight the efficacy of the proposed approach and conclude with a discussion on future work.

Contributions to the Literature

This paper contributes to the literature in several ways:

Focus on Roman Urdu: While extensive research exists on sentiment analysis in English, work on Roman Urdu is limited. This paper addresses this gap by focusing on sentiment analysis of product reviews in Roman Urdu.

Context-Free Grammar (CFG) Approach: Unlike most studies that employ machine learning techniques, this research utilizes a CFG-based approach for sentiment analysis in Roman Urdu. This novel approach is significant as it explores the syntactic structure of the language.

Handling Non-Standard Grammar and Spelling: The paper addresses the challenges of non-standard grammar and spelling in user-generated content, specifically in Roman Urdu. This is crucial for improving the accuracy and reliability of sentiment analysis in this context.

2. Literature Review

There has been extensive literature available on sentiment analysis in English language. For instance, [1] has performed text-based sentiment analysis. Most of the work on sentiment analysis can be divided as conventional and machine learning approaches. We will discuss sentiment analysis approaches in general. The next few paragraphs also present work on sentiment analysis in Urdu.

[1] have performed sentiment analysis using context free grammar. A CFG is a four-tuple comprising non-terminals, terminals, starting symbol and productions. The CFG is used by the authors to validate the proper formation of sentences. Then, emotions are classified as positive, negative or neutral. In another work, [2] have shown that support vector machines provide good accuracy for the sentiment analysis task. [3] performed sentiment analysis of tweets using a psychometric instrument. It is found that various socio-political events have a great impact on the mood of the public. In [5], deep learning based approaches have been advocated and an long short term memory model (LSTM) have been proposed for sentiment analysis on Roman Urdu. According to authors, they have achieved an accuracy of 95% for sentiment analysis.

Most of the sentiment analysis are performed on textual data. However, in a good number of studies videos and images have been analyzed to classify the sentiments such as in [5].

In a number of studies, Roman Urdu text's sentiment has been analyzed. In [6], Roman Urdu text were analyzed for various e-commerce shopping stores. Reviews are represented as vector space models. Then SVM were used to classify them. A survey on sentiment analysis approaches for Roman Urdu has been provided in [7]. A systematic literature review on sentiment analysis for Urdu text in general has been performed in [8]. The authors have gathered data about datasets, approaches, features used for sentiment analysis in Urdu language. In [9], the authors have analyzed sentiments of Pakistan Super League (PSL) anthems using Rapid Miner tool. Various machine learning algorithms were used such as Naïve Bayes, Gradient Boost Tree, Support Vector Machine, KNearest Neighbors, and Artificial Neural Network. A hybrid approach has been presented in [10] for sentiment analysis. A CNN and LSTM based architecture for Roman Urdu has been presented in [11]. In [13], sentiment analysis of Indo-Pak songs were performed using machine learning over a curated dataset. Hotel reviews in Roman Urdu were analyzed and sentiment analysis is performed in [14]. Opinion mining over Roman Urdu has been performed in [15]. A comprehensive study on approaches for sentiment analysis for Urdu has been presented in [12].

Based on the extensive literature review, we concluded that word on sentiment analysis for Roman Urdu text is under investigation. There are researches based on machine learning, however, CFG based approaches not considered. This paper employs a CFG based approach to sentiment analysis.

3. Methodology

In order to carry out sentiment analysis of Roman Urdu, a systematic approach has been followed. The various steps that are followed are: data collection, preprocessing, feature selection and classification.

Introduction to Roman Urdu

Before discussing the proposed methodology, we first provide an introduction to the language under examination. Urdu is one of the widely used language of the world spoken in Pakistan, India, Bangladesh and various parts of the world. Roman Urdu is also a popular script which can be easily written and understood by Urdu speaking people. The approaches that have been used to perform sentiment analysis are primarily machine learning or lexicon based approaches [5].

Now, we will discuss the proposed methodology. Figure 1 shows the proposed methodology.

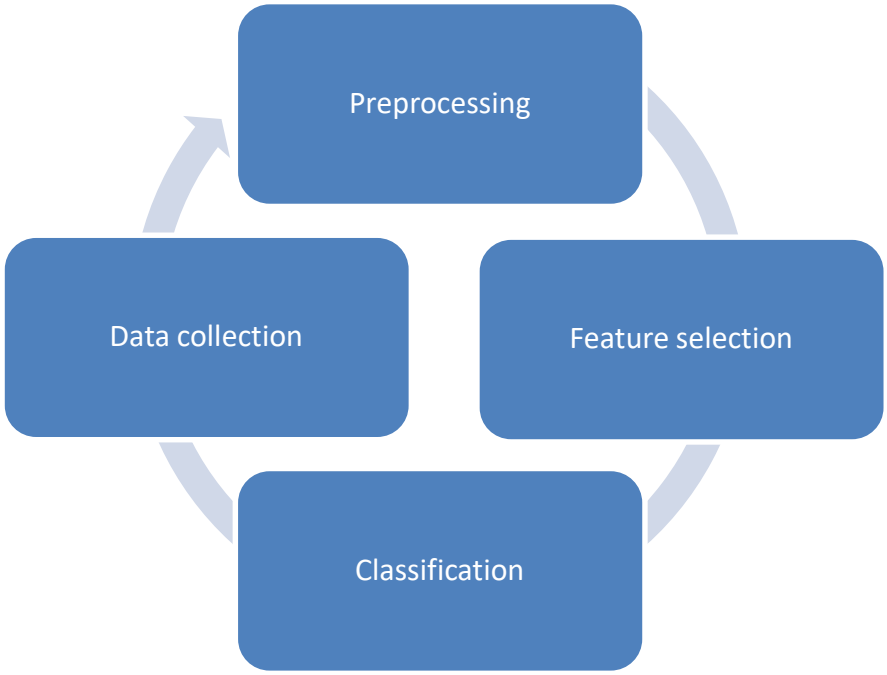

Figure 1: Proposed methodology

Data Collection

The first stage is to collect the customer feedback data more and more. So first we collect customer review and feedback from e-commerce websites on a given product. For this purpose, web scrapping has been performed. Then, we have to segregate the reviews and feedback in table based on the nature of feedback. We will filter the positive, negative and

neutral reviews separately.

Preprocessing

In this stage, we eliminated all the unwanted words and phrases which will possibly make the algorithm complicated. So, the outliers were removed. This is followed by normalization of the dataset. We removed special characters such as commas, full stop, color, preposition or any other special character which is unnecessary. This will make the string less sophisticated.

Feature Selection

The term feature selection refers to identify the true nature of the phrase. Some phrases can be written like this that the system can consider them negative regardless of being affirmative. For Example: “Mobile bura nhi hai”, In this example the system can pick the word “Bura” and consider this as a negative nature of sentiment but the phrase “Nhi hai” oppose the negativity. So the process of eliminating this confusion is called Feature Selection.

Classification

In order to perform classification, we used context free grammar (CFG). The CFG extracts text from the sources. The algorithm works at the sentence level and intended to isolate the tag citations. It begins with analyzing that textual citations are surrounded around years. It also finds each sentence for an author year token. Had it found such a token, the next process is then to extract whether it creates part of a citation or not. For the former case, it obtains the author names that accompany it.

Dataset details

Figure 2 shows the dataset details. The data was captured using web-scrapping and it comprises following fields:

- Product Id: The id of the product
- Customer Name: The name of the customer
- Date: The data in which review was recorded
- Rating: The rating for the product
- Spam/ Not: If the product is spam or not
- Reviews: The actual review about the product
- Sentiment: The actual sentiment about the product
- Features: The features of the product/ review

	Product_ID	Customer_Name	Customer_ID	Date	Rating	Spam(1) and Not Spam	Reviews	Sentiment	Features
0	[illegible]	itzmi7	6702	[illegible]	1	[illegible]	mil gya hai acha hai aur toota hov	Positive	['toota']
1	[illegible]	Ahidi	286	[illegible]	5	[illegible]	aram sy pocket mai aa jata hai aur	Positive	['aram', 'pocket', 'quality']
2	[illegible]	Muhammad	370	[illegible]	2	[illegible]	sb sy achi cheez apki customer ser	Positive	['customer', 'products', 'original']
3	[illegible]	Ghazanfar	1860	[illegible]	3	[illegible]	back cover acha hai sai cuts hai a	Positive	['back', 'cover', 'cuts', 'soft', 'protector', 'chota', 'size', 'khaas', 'for', 'width', 'touch', 'sensor', 'theek', 'combination']
4	[illegible]	Murtaza iqbal	3778	[illegible]	4	[illegible]	sho majmay thumb grip k khalil	Negative	['thumb', 'grip', 'cheez', 'stock', 'controller', 'wire', 'grips', 'pair', 'eternal', 'asan', 'bat', 'quality', 'asan']
5	[illegible]	Usman	5794	[illegible]	5	[illegible]	mai nyja lga bgyk ye madad karti	Positive	['madad', 'new', 'khel', 'pack']
6	[illegible]	Rehan	4446	[illegible]	4	[illegible]	mein neh shaurmein yeh kaasil ka	Positive	['neh', 'kaasil', 'bari', 'mansoobah', 'talaash', 'zaroorat', 'gift', 'set', 'match']
7	[illegible]	Ahab	485	[illegible]	5	[illegible]	iss ka pura package jis may naya sh	Negative	['package', 'naya', 'shipper', 'keemat', 'samsung']
8	[illegible]	Sikander	5264	[illegible]	1	[illegible]	bohat mehanga liya hai mein iss pr	Negative	['mehanga', 'khareedari', 'qeemat', 'kamiyanaj', 'qeemat', 'tamam', 'bekar', 'faraklif']
9	[illegible]	A compliment	34	[illegible]	5	[illegible]	yeh duniya mein har jagah hai koi k	Positive	['khandan', 'pehla', 'kaqeerqat']
10	[illegible]	Isah	6849	[illegible]	4	[illegible]	iss may iss ki cable nai h aur iss k	Negative	['cable', 'speaker']
11	[illegible]	Nasir Raza	3859	[illegible]	4	[illegible]	theek hai mein un head tamiratmi	Positive	['head', 'mijaz', 'mijaz', 'battery', 'waqt', 'qeemat', 'umdah', 'tarez']
12	[illegible]	Waseeb Rehman	5903	[illegible]	5	[illegible]	un head phone ke baarey mein mei	Positive	['head', 'tajurbah', 'pasand', 'order', 'guzashta', 'mijaz', 'madaq', 'saamne', 'awaz', 'kamer', 'mijaz', 'battery', 'tajreebat', 'ghanta']
13	[illegible]	Asfandyar	907	[illegible]	2	[illegible]	jaisay dekhnay ke baad khareeda g	Positive	['mandaar', 'head', 'awaz', 'mijaz', 'kamer', 'paisay', 'shaamil']
14	[illegible]	N	3804	[illegible]	2	[illegible]	iss mein koi za kikil shaamil nahi h	Positive	['shaamil', 'aks', 'khareed', 'marboot', 'head', 'taar', 'audio', 'durust']
15	[illegible]	Shoaib	5269	[illegible]	5	[illegible]	ya chez wkt p aa gai thi aur kaam b	Neutral	['wkt']
16	[illegible]	Talal Khan	5497	[illegible]	3	[illegible]	in k headphone zyda buri nai hai lal	Negative	['headphone', 'zyda', 'buri']
17	[illegible]	Qamar Ali	4221	[illegible]	5	[illegible]	maj koi acha phone nai janta mai	Positive	['parts', 'market', 'package', 'style', 'money']
18	[illegible]	Ammar gazi	768	[illegible]	1	[illegible]	maj ny order kya tha mi a in golder	Neutral	['order', 'mi', 'golder']
19	[illegible]	Zafar	6830	[illegible]	5	[illegible]	portrait mode bhot acha iss phone	Positive	['camera', 'zyda', 'update']
20	[illegible]	M Ahtisham	2758	[illegible]	5	[illegible]	bhot hi acha phone hai ja speed b	Positive	['speed', 'language', 'masla']
21	[illegible]	Hassan	285	[illegible]	5	[illegible]	achi timing hai sirf dino maj agar	Neutral	['timing']
22	[illegible]	Atif	1079	[illegible]	5	[illegible]	sirf thori se price zyda hai agar inte	Positive	['thori', 'zyda', 'international', 'market', 'muqabala', 'dekha']
23	[illegible]	Danish Abbas	1285	[illegible]	5	[illegible]	audio achi hai base k sath	Positive	['audio', 'base']
24	[illegible]	Saim	4924	[illegible]	5	[illegible]	shukriya achi quality k liya mujhy pa	Positive	['quality', 'audio']
25	[illegible]	Ahmed Raza	468	[illegible]	5	[illegible]	shukriya achi quality k liya aur iss ki	Positive	['quality', 'audio']
26	[illegible]	Ali Mansoor	686	[illegible]	5	[illegible]	iss ki quality achi hai aur baki sb b	Positive	['quality']
27	[illegible]	mohsin	6853	[illegible]	5	[illegible]	sb sy acha product achi price maj	Positive	['earphone']
28	[illegible]	Munir iqbal	3790	[illegible]	5	[illegible]	base achi hai baki sari cheez b achi	Positive	['base']
29	[illegible]	Aamir Naeem	128	[illegible]	5	[illegible]	audio ki quality achi hai aur baki sb	Positive	['audio', 'quality', 'cheez', 'balance']
30	[illegible]	Irfan Ahmad	2242	[illegible]	4	[illegible]	achi mobile hai lakin iss ki koi war	Positive	['warenty']
31	[illegible]	Rana	4381	[illegible]	2	[illegible]	maj ny iphone liya us py straches ha	Negative	['straches', 'box', 'band']
32	[illegible]	Sajid	4885	[illegible]	1	[illegible]	iss ki batteri ek issues hai aur cust	Positive	['batteri', 'issues', 'customer', 'response']
33	[illegible]	Zahid	6940	[illegible]	5	[illegible]	waqat a all behtareen phone ka ka	Positive	['waqata', 'behtareen', 'hawale', 'qeemat', 'misaal', 'kisan', 'bara', 'farakam', 'shakaal', 'cover']
34	[illegible]	[illegible]	7391	[illegible]	5	[illegible]	battery timing aur mobile bhot ach	Positive	['battery', 'timing']
35	[illegible]	Danish	1294	[illegible]	5	[illegible]	sb sy acha phone hai agar aap sy	Neutral	['compare']
36	[illegible]	Ali	589	[illegible]	1	[illegible]	mein nai phone bagair liya aur me	Negative	['dekha', 'haath']
37	[illegible]	Maryam	3847	[illegible]	5	[illegible]	sb sy acha product hai sirf din ma	Neutral	['din']
38	[illegible]	Abid	275	[illegible]	4	[illegible]	plz mujhy bta dik ya single ya dual	Neutral	['single', 'dual', 'sim']

Figure 2: Dataset details

Implementation details

The proposed approach has been implemented in python using the NLTK library. Figure 3 shows a snapshot of the implementation.

```
import nltk
import os
```

```
[ ] import nltk.corpus
```

```
[ ] from nltk.tokenize import word_tokenize
    from nltk.tokenize import RegexpTokenizer
```

```
[ ] from nltk.data import load
```

```
[ ] sent="The product is not good enough"
```

```
[ ] nltk.download('punkt')
    nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
True
```

Figure 3: Implementation details

4. Results

In order to understand the proposed methodology and its effectiveness, we will show few examples in this section.

Sentence 1: Ye Mobile bht acha hai

Parsing

The CFG will parse this sentence as shown in Figure 7.

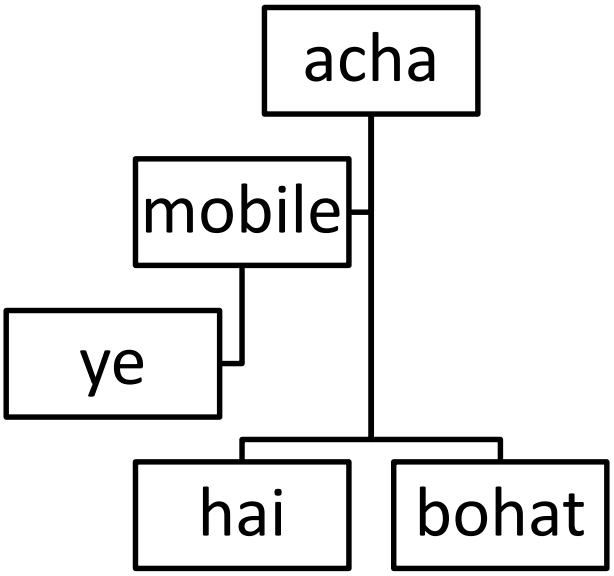

Figure 4: Parsing of the sentence

Assigning Polarity

We will calculate now the sentiment of the sentence. We can consider the most positive word as the root. Now according to our scale (0 - 1), 0 for the most negative number and 1 for the positive number. So now we give some values for the scale given above to our sentences as shown in Table 1.

Table 1: Assigning polarity to words

Ye	0.5
Mobile	0.5
Bht	0.4
Acha	0.8
Hai	0.5

Sentiment analysis

Now the highest positive or the most negative value will consider as governor and other are its subordinates
So “ACHA” is governor and other are its subordinates. According to our values, the sentiment is calculated as follows, as show in Figure 5.

```
if (Sgovernor>=0.55 and Sdependent>=0.4)
{
        if (Sgovernor<=Sdependent )
        {
        Stag = Sgovernor+Sdependent
        } else {
        Stag = (Sgovernor + Sgovernor*Sdependent) * 0.5
        }
}
if (Sgovernor>=0.55 and Sdependent<=0.4)
{
Stag=(Sgovernor − Sgovernor*Sdepdent)*0.5
}

if (Sgovernor<=0.55 and Sdependent>=0.4)
{
Stag=(Sgovernor − Sgovernor*Sdepdent)*0.5
}
if (Sgovernor<=0.55 and Sdependent<=0.4)
{
        if (Sgovernor >= Sdependent) {
                Stag = Sgovernor + Sdependent
        }
        else{
                Stag = [1-(Sgovernor − Sgovernor * Sdepdenent)] * 0.5
        }
}
```

Figure 5: Algorithm for the implementation

$$(\text{acha}_g\ 0.8)\ (\text{bht}_s\ 0.4) \quad = \frac{0.8+0.4}{2} \times 5 = 3 \text{ is the answer}$$

Table 3: Thresholds for sentiment analysis

Scores	Conclusion
$0<=S_{review}<2$	Sad and unsatisfied
$2<=S_{review}<3$	Indifferent, happy to use with sacrifices
$3<=S_{reivew}<=5$	Happy and satisfied

So according to Table 3, if the value is greater than 3 the sentence is negative.

Sentence 2: Game Acha hai Lekin Ads ziada hain.

Parsing

As the first step, we will first parse the sentence as shown in Figure 6.

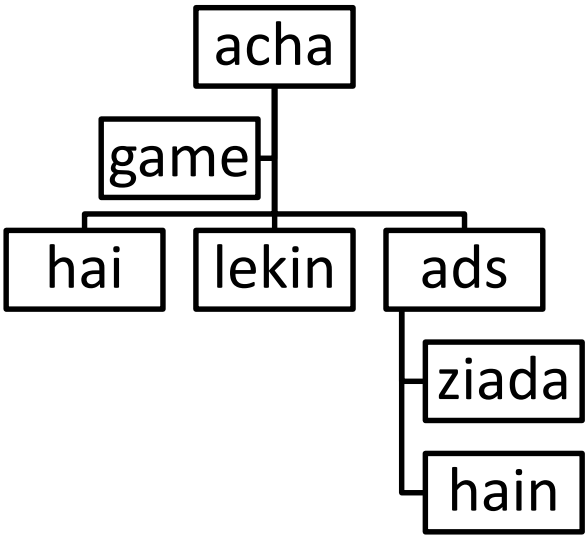

Figure 6: Parsing of the sentence

Assigning polarity to words

In the next step, polarity is assigned to each word as shown in Table 2.

Table 2: Polarity of words

Game	0.5
Acha	0.8
Hai	0.5
Lekin	0.6
Ads	0.3
Ziada	0.4
Hain	0.5

Sentiment analysis

According to the formula, we will pick the most positive word as the governor and the highest negative value will be also considered as governor,

Case: A

$(\text{acha}_g\ 0.8)\ (\text{game}_s\ 0.5) = \frac{0.8+0.5}{2} \times 5$

$= 3.25$

Case B

(Ads 0.3) (Ziada_s 0.4)

Now according to the formula if the value of governor

$$= (0.3 - 0.3\ x\ 0.5)\text{x}5$$

$$= \frac{0.75+3.25}{2}$$

$$= 2$$

So, the conclusion of the above sentence is “Happy to use with Sacrifice”.

5. Conclusion

This paper introduces a Context-Free Grammar (CFG) based approach to sentiment analysis for product reviews in Roman Urdu. The approach is demonstrated through two examples where the sentiments of several sentences were analyzed. The results indicate a good level of accuracy from the sentiment analysis algorithm. However, there are several limitations associated with the proposed approach, primarily stemming from the quality of Urdu Roman language or typographical errors which may lead to inaccuracies.

Moving forward, this model holds potential not only to assess product satisfaction locally but also to compare satisfaction scores across similar products in the global market. Such insights could potentially drive societal transformations in consumer behavior and product development strategies. Future enhancements to the model could include the ability to distinguish between positive and negative product features, providing valuable feedback to designers and marketers.

In summary, while this paper makes strides in applying CFG to sentiment analysis in Roman Urdu, there is room for improvement and further refinement of the approach to achieve greater accuracy and applicability in real-world scenarios.

Sentiment analysis has a great deal of significance specifically with the rise of big data and

explosive growth of user generated content on social media, mobile platforms etc. Massive amount of text in Roman Urdu are collected daily from social media that demands analysis. The use of the techniques proposed in this research can bring a number of societal transformations. It can be used for public opinion mining by the government, businesses can use the approach for understanding customer feedback, social stability can be ensured by law enforcement agencies.

References

[1] Nandi, B., Ghanti, M. and Paul, "Text based sentiment analysis", 2017 In International Conference on Inventive Computing and Informatics (ICICI) (pp. 9-13). IEEE.

[2] Shivaprasad T K, Jyothi Shetty, "Sentiment Analysis of Product Reviews", 2017 In International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE Xplore Compliant.

[3] Bollen, J., Mao, H. and Pepe, A., "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena". 2011 In Proceedings of the international AAAI conference on web and social media (Vol. 5, No. 1, pp. 450-453).

[4] Liu, B., "Sentiment Analysis and Subjectivity", 2010 In F. J. N. Indurkhya, Handbook of Natural Language Processing. Chicago,.

[5] Ghulam, H., Zeng, F., Li, W. and Xiao, Y., 20
19 In Deep learning-based sentiment analysis for roman urdu text. Procedia computer science, 147, pp.131-135. // Paper title is missing

[6] Noor, F., Bakhtyar, M. and Baber, J., "Sentiment analysis in e-commerce using svm on roman urdu text. In Emerging Technologies in Computing", 2019 In Second International Conference, iCETiC, London, UK, August 19–20, 2019, Proceedings 2 (pp. 213-222). Springer International Publishing.

[7] Qutab, I., Malik, K. I., & Arooj, H., "Sentiment Analysis for Roman Urdu Text over Social Media, a Comparative Study".In 2020 International Journal of Computer Science and Network, Volume 9, Issue 5

[8] Mashooq, M., Riaz, S., & Farooq, M, "Urdu Sentiment Analysis: Future Extraction, Taxonomy, and Challenges". 2022 In VFAST Transactions on Software Engineering .

[9] Aish, M.A., Mubeen, M., Iqbal, M.U. and Qaseem, A., "Sentiment Analysis of Roman

Urdu Reviews. UMT Artif". 2022 In Intell. Rev, 2(1), pp.00-00..
[10] Mehmood, F., Ghani, M.U., Ibrahim, M.A., Shahzadi, R., Mahmood, W. and Asim, M.N., "A precisely xtreme-multi channel hybrid approach for roman urdu sentiment analysis". 2020 In IEEE Access, 8, pp.192740-192759

[11] Khan, L., Amjad, A., Afaq, K. M., & Chang , H.-T. "Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media". 2022 In Applied Science MDPI.

[12] Khan, I. U., Khan, A., Khan, W., Su'ud , M. M., Alam, M. M., Subhan, F., & Asghar, M. Z. "A Review of Urdu Sentiment Analysis with Multilingual Perspective: A Case of Urdu and Roman Urdu Language. Computers MDPI". 2021 //Incomplete, Journal name missing

[13] Qureshi, M.A., Asif, M., Hassan, M.F., Abid, A., Kamal, A., Safdar, S. and Akbar, R.,. "Sentiment analysis of reviews in natural language: Roman Urdu as a case study". 2022 In IEEE Access, 10, pp.24945-24954.

[14] Arif, H., Munir, K., Danyal, A.S., Salman, A. and Fraz, M.M,. "Sentiment analysis of roman urdu/hindi using supervised methods". ., 2016 In Proc. ICICC, 8, pp.48-53.

[15] Bilal, M., Israr, H., Shahid, M. and Khan, A., "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques". 2016 In Journal of King Saud University-Computer and Information Sciences, 28(3), pp.330-344.

[16] https://github.com/Smat26/Roman-Urdu-Dataset/blob/master/Negative-and%20Positive-Words/pos.txt?cv=1