Global Scientific JOURNALS

SENTIMENT ANALYSIS OF URDU TEXT

University of Engineering and Technology, Peshawar.

KeyWords

Sentiment Analysis, Urdu Language Processing, Machine Learning, Deep Learning, Urdu Data, Classification.

ABSTRACT

The emergence of the internet allowed for the extensive and very easy information sharing on a various topics, like events, services, products and political viewpoints. The studies on sentiment analysis is increasing fastly,but many of them focus on English. The purpose of this research is to survey a state-of-the-art for lookup into the advancements and flaws in Urdu sentiment analysis, and to propose solutions. This research summarised the development of this field by classification of studies into three stages: text pre-processing, lexical resources, and sentiment categorization. Word segmentation, text cleaning, spell checking, and part-of-speech tagging are some of the pre-processing activities. An analysis of advanced lexical resources, such as corpuses and lexicons, was carried out and as well as the investigations into sentiment analysis components including negations , modifiers, and opinion words. Each of the reviewed studies' performance is reported. This paper presents the framework for future research on Urdu sentiment analysis, based on experimental results and ideas.

INTRODUCTION

The advancement in social media sites has enabled and promoted the sharing of information and opinions on subjects such as, facilities, guidelines, merchandises , and dilemmas [1]. The high-tech gadgets created to make decisions by individuals as well as organisations via sharing of information on social networks [2].

Resources for sentiment analysis (SA) are abundant in the English language. This comprises a wide range of natural language processing (NLP) tools, such as lexicons, parsers, part-of-speech taggers, and more [3]. Although the majority of today's SA systems are written in English [4], there are now a number of non-English SA appliances due to the increase in non-English web traffic. SA in a single language increases the chance that important information in literature written in other languages may be missed. It is necessary to design a flexible SA framework and functional SA instruments for the analysis of data in languages like Urdu.

Urdu, Pakistan's native language, is also widely spoken in India. Several factors make SA in Urdu challenging. The lack of recognised lexical resources in Urdu [5,6,7] is one of them. Urdu SA, as a result of this shortcoming, includes transferring information from an English language rich in resources to an Urdu language deficient in resources [8,9]. Rather than using a suitable text encoding technique, most Urdu websites use an illustrative layout. This situation creates challenges when attempting to structure a machine-readable corpus. The emotion lexicon is the most important part of creating a SA system in any language.

There exist many well established sentiment lexicons (like SentiWordNet) that makes English language a resource-full language . Where on the other side, URDU language lacking behind for having such sentiment lexicons. Problems related to word dissimilarities in morphology, segmentation, and vocabulary inconsistencies gets to other obstacles on the way to create a Urdu SA system with full functionality.

There have been very limited Studies focusing on Urdu SA. This shown the less interest of the language engineering and the limited resources available for Urdu language . The focus of the previous studies on various aspects of Urdu [14,10].These aspects were stemming, named entity recognition (NER), Urdu language morphology, concept searching, datasets and stop words identification. A

survey also performed by Singh [11] on Urdu sentiment analysis, the focus of which was on subjectivity analysis. However, this research have cover most details of Urdu text preprocessing, lexical resources and sentiment classification along with the tasks and techniques available for Urdu sentiment analysis in this survey.

Motivation

The following reasons served as the impetus for this survey. The official language of Pakistan is Urdu, which is also extensively spoken in the Indian subcontinent. A lot more information on the Urdu language is now available online. Different lexical resources are required for the SA in the resource-scarce Urdu language. This review aims to highlight the most recent work on sentiment lexicons, sentiment analysis techniques, corpora, and tools for text processing in the Urdu language. This research conducted a thorough survey by looking for, finding, summarising, and analysing pertinent papers as a result of the quick research breakthroughs achieved in Urdu SA.

The linked publications from various electronic repositories were searched for this survey. The quantity of obtained articles is filtered in the following phase by using inclusion and exclusion criteria. After thorough study, suitable works are chosen based on the research queries and the findings are provided.

Research Queries

The survey in this study is conducted by addressing the following research queries.

Query 1: What text pre-processing methods are employed in Urdu SA, and what methods are employed by researchers as stated in published articles?

Query 2: What distinct lexical resources are used for Urdu SA, and what methods are employed in their creation?

Query 3: What strategies have been effective for classifying the feelings in Urdu text? What are the best practises for classifying the sentiments in Urdu reviews?

Problem Solving Technique

To handle the above research queries, this research has used the method suggested by [12] to evaluate the quality of the chosen articles. Each chosen paper (article inclusion) was assessed using the following quality assessment (QA) queries:

QA1: One or more preprocessing methods utilised for Urdu SA are described in the paper.

QA2: One or more lexical resources and strategies needed for Urdu SA are described in the study.

QA3: The study explains sentiment analysis of Urdu text using cutting-edge methodology.

A rating of 1 ('query totally explained'), 0.5 ('query somewhat explained'), or 0 ('query not explained') is entered to the excel sheet for each of the aforementioned quality evaluation queries [13, 27].

Results from applying the aforementioned quality evaluation queries to the four studies are shown in Table 1. The "comments" column contains the explanation for each evaluation. The final normalised score and the total of the evaluations provide the quality assessment consequent normalised score for each study. The four studies S1, S2, S3, and S4 obtained the normalised scores of 0.83, 1.0, 0.66, and 0.5 correspondingly out of a total quality score of 3. The threshold was set at a quality score of 0. If at least one research partially addresses one of the quality evaluation queries, it is regarded appropriate to include it in the survey; otherwise, it is eliminated from the paper bank. A brief overview of a few preparation methods employed in Urdu SA is provided in Study S2. Regarding the lexical resources employed in Urdu SA, the research S1 doesn't make any new contributions. A brief overview of the sentiment categorization method employed in Urdu SA is provided in Study S2.

Table 1. A small collection of studies together with their quality ratings.

QA Criteria	Qeury	Study			
		1	2	3	4
		Afraz et al. [14]	Asghar et al. [5]	Mukhtar and Khan [8]	Sana et al. [15]
1	The article describes multiple and sigle preprcessing	1	0.5	1	1

	methods applied to Urdu SA.				
2	The article describes multiple or single lexical methods and materials necessary for Urdu SA.	1	1	0.5	0
3	The study explains sentiment categorization of Urdu Corpora using cutting-edge methodology.	1	0.5	1	1

LITERATURE REVIEW

Compared to other resource-rich languages like English, sentiment analysis in Urdu is still in its early phases of development. Additionally, little work has been done, which has a direct influence on the quantity of surveys and review papers that are now accessible. Anwar et al.'s survey on artificial Urdu language processing included a description of methods concentrating on Urdu corpus creation [3]. A variety of linguistic approaches were used, including named entity recognition, parsing, and part of speech tagging (POS). One of the first studies on Urdu language processing, which this work tries to rectify, lacked the appropriate methods for undertaking sentiment analysis in Urdu.

In their examination of numerous linguistic resources and preprocessing methods for the processing of the Urdu language, Daud et al. [10] covered best practises for a number of tasks, including sentence boundary recognition, tokenization, POS tagging, NER, and the creation of WordNet lexicons. Information retrieval, plagiarism detection, and categorization are just a few of the applications of Urdu language processing that are explored. The poll does not, however, concentrate on the sentiment analysis paradigm. Therefore, it is necessary to carry out a thorough survey with a sentiment analysis focus. Given that the emphasis is on sentiment analysis in the Urdu language and not merely simple-text processing, the survey that is conducted in this article is somewhat distinct.

In a recent review of Urdu sentiment analysis, Singh [11] placed a strong emphasis on subjectivity analysis and sentiment categorization. According to their research, there aren't many various Urdu linguistic tools available, such as named entity and POS taggers. They categorised the seventeen research on sentiment analysis in Urdu by approach and data sets. However, this research analyse 11 works along three dimensions in this study, including I text pre-processing, (ii) lexical resources, and (iii) sentiment analysis, which are further separated into several subcategories. This research then focus on sentiment analysis in Urdu. Additionally, this research have included information on the approach used, dataset, goal, limitations, and future objectives of the chosen studies. This research review current methods and give our own, as stated by the authors, original findings in this survey.

A study on multilingual sentiment analysis was carried out by Lo et al. [16] with a focus on languages with limited resources. For doing multilingual sentiment analysis, many methods and tools are examined and presented. Additionally, many difficulties are identified along with suggestions for future courses of action. Our planned survey, however, is distinct because we're concentrating on sentiment analysis in the Urdu language. By examining more than 14 works published in the field of Urdu sentiment analysis, Khan et al. [17] performed a survey on the subject. Machine learning, lexicon-based, and hybrid methods were used to characterize the approaches involved for Urdu SA. A thorough survey that can cover all facets of Urdu SA in terms of given queries and obtaining their replies is still required, nevertheless.

Word Sense Disambiguation (WSD) in Text Documents of Urdu language may be accomplished using three classifiers, including SVM, Decision tree, and Naive Bayesian, according to a method given by Abid et al. in [2]. On a dataset obtained from domestic and foreign news websites, they conducted trials and obtained an f-measure of 0.71. An adaptable window size for uncertain Urdu words, however, could improve the system's performance. For the purpose of classifying the sentiment of Urdu text, KNN, Lib SVM, and J48 were three supervised machine learning classifiers that Mukhtar et al. [53] assessed for performance. The findings collected demonstrate that KNN outperforms the others. With larger amounts of data, the system must be assessed using several statistical metrics including the Kappa statistic and Root Mean Squared Error. Ghulam et al. [20] proposed Based On Deep Neural Network name as Long-Short Time Memory model (LSTM) for sentiment classification in Roman Urdu while working on the language. The gradient attenuation problem can be solved by the model, and it can record data over extended time periods. Additionally, the suggested approach may capture both the semantics of the word order and contextual information. In comparison to current Machine Learning (ML) methodologies and lexicon-based techniques, experimental findings demonstrate the model's performance. Table 2 represents an overall overview of quality research arcticles already published.

The research has offered many Urdu sentiment analysis approaches, continuing the literature. The best performing approach must be chosen in order to create practical applications. It becomes challenging to directly compare the stated procedures, nevertheless, because of many considerations. First off, different datasets are used when testing such algorithms, which makes comparisons challenging. Additionally, the contributing writers provide the approaches in their works insufficiently in detail and at an abstract level, which may render them useless for upcoming researchers. The strategies presented in the published papers were applied to two datasets in this study while keeping in mind the aforementioned problems. Although we made every effort to perform this study in accordance with the experimental setup and procedure described in the papers, there were a few instances when we were forced to ignore specific aspects of the technique or make assumptions about what the authors had in mind.

In a well-known study by Ghulam et al. [20], it was revealed that they used the LSTM model for sentiment analysis in Roman Urdu, but they omitted to include specifics on the parameter settings and feature choices, which may not have been the same as the authors'. Additionally, they did not identify the tools needed to put their method into action; this study employed Python and the Anaconda framework. This research provided a qualitative comparison of the strategies in addition to the aforementioned straightforward and consistent execution. It has been noted that comparing approaches based on one or more widely used datasets may not be a fair comparison to techniques created for a particular area. For instance, Sana et al[15] .'s approach for classifying Urdu emotions in the electronics and sports area produced subpar results in our implementation since we used different datasets, such as books and drugs. On two cutting-edge datasets obtained from [5], which comprise happy and negative reviews on books and pharmaceuticals, this research carried out a quantitative evaluation of the various Urdu sentiment analysis methods. Using an Anaconda-based Jupyter notebook, the research put known methods into reality [21].

Table 2 displays the accuracy of each study's published results as well as the accuracy attained as a result of applying the approach to the two datasets used in our research. The findings show that the accuracy gained in our studies differs from that which has been reported. This discrepancy is mostly caused by the reported research' lack of implementation details, which made it difficult to replicate the precise implementation of the examined publications. Due to the use of various datasets, parameter settings, and tools, outcomes in our studies occasionally differ from those that have been previously reported.

Another research, whereas Ali and Ijaz [29] reported 93 percent accuracy, our studies produced 71 percent, in contrast to Bilal et al. [9] who reported accuracy of 86 percent. These discrepancies between reported and experimental accuracy were brought about by the researchers' use of various datasets. For instance, this research conducted trials on book and drug reviews while Rehman and Bajwa [23] employed news dataset; Mukhtar et al. [19] used seven distinct datasets while this research only used two. This study also made use of several lexical sources. Moreover, Sana et al. [15] employed an annotated corpus of Urdu tweets, whereas [22] used a constrained Urdu polarity lexicon, and our study used the Urdu emotion lexicon suggested by Asghar et al. [5]. A few of the aforementioned methods were created for Roman Urdu sentiment analysis. Continuing to recent literature , Ghulam et al. [20] used a neural network model for Roman Urdu sentiment analysis.

Table 2. *Literature review table of some quality research papers selected.*

Author	Methodology Used	Accuracy (%)
Asghar et al. [5]	Lexicon-based approach	82
Sana et al. [15]	Classical Machine Learning approachs	75
Mukhtar et al. [19]	Emerging Machine Learning approaches	87
Mukhtar and Khan [8]	Old Machine Learning approaches	84
Mukhtar et al. [26]	Rule based methods	67
Rehman and Bajwa [23]	Lexicon-based approaches	66
Ghlam et al. [20]	Sentiment Analysing using Deep Learning, LSTM model	95

Bilal et al. [9]	Classical Machine Learning methods supervised	79
Sharf and Rehman [36]	Natural language processing techniques for preprocessing	81
Ali and Ijaz [37]	Sentiment Analysis using supervised Machine Learning methods	93

DISCUSSION

Twenty papers on crucial aspects of Urdu sentiment analysis were taken into account in this systematic evaluation of the literature. This analysis discovered that each article in the study made use of lexicon- and corpus-based techniques. Additionally, it should be noted that most datasets and lexicons created for experimentation are done so manually. To enable experimentation in subsequent research, it is necessary to make such datasets and lexicons publicly accessible.

In reaction to RQ1, the research done in this study has discovered four key preprocessing activities that help with Urdu SA: words segmentation, text cleaning, POS tagging, and spell checking and correction. Even though pre-processing methods needed for Urdu SA have recently been researched, more research is necessary to improve these approaches so they can deal with social media-driven language concerns such emojis and slang phrases. For a particular corpus, text preprocessing methods for word segmentation, text cleaning, part of speech tagging, and spell checking and correction worked well. There are certain restrictions, though: i) Existing pre-processing techniques lack filtering of emoticons and slang terms, which can express an opinion; (ii) Lexicon-based Urdu spell checkers yield low (— in other words poor) accuracy because of the insufficient coverage of related words; and (iii) diacritics do not play a significant role in sentiment analysis, but some words can be distinguished only by using diacritics, which are very helpful in context-aware sentiment analysis or anaphora. Using supervised and unsupervised word sense disambiguation approaches, authors [23, 27] have demonstrated that it is possible to discriminate between two words in Urdu. Different online Urdu dictionaries can be connected with sentiment analysis algorithms to enhance the efficiency of spell checkers for the language.

In response to RQ2, this study has revealed crucial techniques for producing the lexical resources (lexicon, corpus) needed for Urdu SA. The methods used to create lexicons and corpora currently have the following drawbacks: There are several issues with the current Urdu lexicons, including (i) the limited number of entries in them that lack sentiment scores for opinion words, (ii) the dearth of publicly accessible sentiment annotated corpora for Urdu sentiment analysis [30, 31], and (iii) the critical role that domain-specific words play in the sentiment orientation of Urdu text. These, however, are disregarded [32, 35]. Therefore, it is necessary to extend the Lexicon by adding opinion word synonyms and POS tags to each item [6]. A lexicon should contain the domains in which such sentiment words are used since they might convey various meanings when spoken or written in different contexts. Since all of the existing datasets are in the news domain and point towards objective datasets, a computer-readable sentiment dataset of user attitudes is thus necessary. For terms that are not in the lexicon, a categorization system must exist. By utilising web crawling or bilingual translation techniques, it is possible to create an accurately annotated Urdu corpus from existing Urdu websites in a variety of sectors, including news, finance, politics, sports, and more.

In accordance to RQ3, this research identifies several sentiment classification strategies needed for Urdu SA, including semantic orientation using lexicon-based and corpus-based methodologies and subjectivity classification utilizing opinion lexicons. Various difficulties are noted as: I The effectiveness of Urdu sentiment analysis algorithms based on supervised learning and unsupervised approaches across many domains has not yet been established [24, 34]; I there is no suitable method for the identification and trying to find out the score of various types of renegotiations and enhancers, as modifiers and negations alter the sentiment score and sentiment categories [33]; (ii) lexicon-based opinion analysis is not entirely suitable to distinguish biassed sentences from objective sentences as the lexicons do not cover all the sentiment words [25]; (iv) Text can contain an emoji to convey a feeling. Slang terms are frequently used in reader-suggested reviews on social media sites, but they are not taken into account in Urdu sentiment analysis; (vi) the majority of the existing Urdu sentiment analysis systems work at the sentence level, not providing overall classification and scoring of the entire document [33]; and (vi) to the best of our knowledge, no work was performed on the sentiment analysis troubling exclamation mark analysis and creating chances [21, 27].

CONCLUSIONS

Applications for Urdu sentiment analysis are quickly catching the attention of individuals and organizations in order to get crucial input on goods, regulations, and events. Our goal in conducting this survey was to concentrate on cutting-edge methods for text pre-processing, lexical resources, and sentiment analysis of Urdu text.

This study examined an issue with sentiment classification in the Urdu language and explored three key modules, including pre-processing, vocabulary resources, and sentiment classification, which are necessary for creating sentiment-based applications. With a focus on the aforementioned components, this research examined several methods for creating SA applications that are based on Urdu. The following major aspects are recommended as suggestions for future work to acknowledge and include: (i) Correctly addressing Urdu idioms and proverbs is necessary to improve the accuracy of emotion categorization. , e.g. بہتی گنگا میں ہاتھ دھونا . The majority of evaluations submitted on social media websites are written in Roman Urdu text; (ii) to take advantage of the possibility; As a result, careful research employing supervised, unsupervised, and hybrid classification systems is necessary; (iii) word sense disambiguation is a neglected topic in Urdu text processing that demands more attention. (iv) Concept level sentiment analysis in Urdu is to be acknowledged as a problem; (v) Different machine learning and deep learning approaches need to be researched for Urdu sentiment analysis; and (vi) Emotion detection and classification in Urdu text is to be acknowledged as a tough task.

References

[1] Asghar, M.Z., Khan, A., Bibi, A., Kundi, F.M. and Ahmad, H., 2017. Sentence-level emotion detection framework using rule-based classification. *Cognitive Computation*, *9*(6), pp.868-894.

[2] Khana, W., Daudb, A., Nasira, J.A. and Amjada, T., 2016. Named entity dataset for Urdu named entity recognition task. *Organization*, *48*, p.282.

[3] Anwar, W., Wang, X. and Wang, X.L., 2006, August. A survey of automatic Urdu language processing. In *2006 International Conference on Machine Learning and Cybernetics* (pp. 4489-4494). IEEE.

[4] Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A.Y., Gelbukh, A. and Zhou, Q., 2016. Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cognitive computation*, *8*(4), pp.757-771.

[5] Asghar, M.Z., Sattar, A., Khan, A., Ali, A., Masud Kundi, F. and Ahmad, S., 2019. Creating sentiment lexicon for sentiment analysis in Urdu: The case of a resource-poor language. *Expert Systems*, *36*(3), p.e12397.

[6] Afraz, S.Z., Aslam, M. and Martinez-Enriquez, A.M., 2011. Adjectival phrases as the sentiment carriers in the Urdu text. *J Am Sci*, *7*(3), pp.644-52.

[7] Ijaz, M. and Hussain, S., 2007, August. Corpus based Urdu lexicon development. In *the Proceedings of Conference on Language Technology (CLT07), University of Peshawar, Pakistan* (Vol. 73).

[8] Mukhtar, N. and Khan, M.A., 2018. Urdu sentiment analysis using supervised machine learning approach. *International Journal of Pattern Recognition and Artificial Intelligence*, *32*(02), p.1851001.

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

[10] Daud, A., Khan, W. and Che, D., 2017. Urdu language processing: a survey. *Artificial Intelligence Review*, *47*(3), pp.279-311.

[11] Singh, V.K., 2015. A survey of sentiment analysis research in Urdu. *Ind J Sci Res Tech*, *3*(4), pp.63-5.

[12] Nazir, S., Nawaz, M., Adnan, A., Shahzad, S. and Asadi, S., 2019. Big data features, applications, and analytics in cardiology—a systematic literature review. *IEEE Access*, *7*, pp.143742-143771.

[13] Da Silva, F.Q., Santos, A.L., Soares, S., França, A.C.C., Monteiro, C.V. and Maciel, F.F., 2011. Six years of systematic literature reviews in software engineering: An updated tertiary study. *Information and Software Technology*, *53*(9), pp.899-913.

[14] Syed, A.Z., Aslam, M. and Martinez-Enriquez, A.M., 2014. Associating targets with SentiUnits: a step forward in sentiment analysis of Urdu text. *Artificial intelligence review*, *41*(4), pp.535-561.

[15] Sana, L., Nasir, K., Urooj, A., Ishaq, Z. and Hameed, I.A., 2018, November. Bers: Bussiness-related emotion recognition system in urdu language using machine learning. In *2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC)* (pp. 238-242). IEEE.

[16] Lo, S.L., Cambria, E., Chiong, R. and Cornforth, D., 2017. Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artificial Intelligence Review*, *48*(4), pp.499-527.

[17] Khan, W., Daud, A., Nasir, J.A., Amjad, T., Arafat, S., Aljohani, N. and Alotaibi, F.S., 2019. Urdu part of speech tagging using conditional random fields. *Language Resources and Evaluation*, *53*(3), pp.331-362.

[18] Abid, M., Habib, A., Ashraf, J. and Shahid, A., 2018. Urdu word sense disambiguation using machine learning approach. *Cluster Computing*, *21*(1), pp.515-522.

[19] Mukhtar, N., Khan, M.A. and Chiragh, N., 2017. Effective use of evaluation measures for the validation of best classifier in Urdu sentiment analysis. *Cognitive Computation*, *9*(4), pp.446-456.

[20] Ghulam, H., Zeng, F., Li, W. and Xiao, Y., 2019. Deep learning-based sentiment analysis for roman urdu text. *Procedia computer science*, *147*, pp.131-135.

[21] All things Anaconda, n.d Retrieved from https://www.anaconda.com/.

[22] Ali, A.R. and Ijaz, M., 2009, December. Urdu text classification. In *Proceedings of the 7th international conference on frontiers of information technology* (pp. 1-7).

[23] Rajput, Q., 2014. Ontology based semantic annotation of Urdu language web documents. *Procedia Computer Science*, *35*, pp.662-670.

[24] Kaplan, Ronald M., Jill Burstein, Mary Harper, and Gerald Penn. "Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics." In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 2010.

[25] Mukund, S. and Srihari, R.K., 2010, August. A vector space model for subjectivity classification in Urdu aided by co-training. In *Coling 2010: Posters* (pp. 860-868).

[26] Mukhtar N, Khan MA, Chiragh N, Nazir S, 2018. Identification and handling of intensifiers for enhancing accuracy of Urdu sentiment analysis. Expert Systems;35(6):e12317.

[27] Naseem, T. and Hussain, S., 2007. A novel approach for ranking spelling error corrections for Urdu. *Language Resources and Evaluation*, *41*(2), pp.117-128.

[28] Iqbal, S., Anwar, W., Bajwa, U.I. and Rehman, Z., 2013, October. Urdu spell checking: Reverse edit distance approach. In *Proceedings of the 4th workshop on south and southeast asian natural language processing* (pp. 58-65).

[29] Asghar, M.Z., Kundi, F.M., Ahmad, S., Khan, A. and Khan, F., 2018. T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme. *Expert Systems*, *35*(1), p.e12233.

[30] Durani, N. and Hussain, S., 2010. Urdu Word Segmentation, Human Language Technologies. In *The Annual Conference of the North American Chapter of the ACL, Los Angeles, California* (pp. 528-536).

[31] Afraz, S.Z., Aslam, M. and Martinez-Enriquez, A.M., 2011. Adjectival phrases as the sentiment carriers in the Urdu text. *J Am Sci*, *7*(3), pp.644-52.

[32] Muaz, A., Ali, A. and Hussain, S., 2009, August. Analysis and development of Urdu POS tagged corpus. In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)* (pp. 24-31).

[33] MY Khan., 2020. Urdu-Sentiment-Corpus, available at: https://github.com/MuhammadYaseenKhan/Urdu-Sentiment-Corpus/blob/master/urdusentiment- corpus-v1.tsv,last accessed 10-feb-2020

[34] Syed, A.Z., Aslam, M. and Martinez-Enriquez, A.M., 2011, November. Sentiment analysis of urdu language: handling phrase-level negation. In *Mexican International Conference on Artificial Intelligence* (pp. 382-393). Springer, Berlin, Heidelberg.

[35] Bilal, A., Rextin, A., Kakakhel, A. and Nasim, M., 2017, September. Roman-txt: forms and functions of roman urdu texting. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-9).

[36] Sharf Z, Rahman SU, 2018. Performing natural language processing on roman Urdu datasets. *International Journal of Computer Science Network Security*;18(1):141–8.

[37] Ali AR, Ijaz M, 2009. Urdu text classification. *In Proceedings of the 7th international conference on frontiers of information technology*,16 (p. 21). ACM.