Emotion Detection in Roman Urdu Text using Machine Learning

Adil Majeed*
adilmajeed350@gmail.com
AIM Lab, FAST-NUCES, Islamabad

Hasan Mujtaba
hasan.mujtaba@nu.edu.pk
AIM Lab, FAST-NUCES, Islamabad

Mirza Omer Beg
omer.beg@nu.edu.pk
AIM Lab, FAST-NUCES, Islamabad

ABSTRACT

Emotion detection is playing a very important role in our life. People express their emotions in different ways i.e face expression, gestures, speech, and text. This research focuses on detecting emotions from the Roman Urdu text. Previously, A lot of work has been done on different languages for emotion detection but there is limited work done in Roman Urdu. Therefore, there is a need to explore Roman Urdu as it is the most widely used language on social media platforms for communication. One major issue for the Roman Urdu is the absence of benchmark corpora for emotion detection from text because language assets are essential for different natural language processing (NLP) tasks. There are many useful applications of the emotional analysis of a text such as improving the quality of products, dialog systems, investment trends, mental health. In this research, to focus on the emotional polarity of the Roman Urdu sentence we develop a comprehensive corpus of 18k sentences that are gathered from different domains and annotate it with six different classes. We applied different baseline algorithms like KNN, Decision tree, SVM, and Random Forest on our corpus. After experimentation and evaluation, the results showed that the SVM model achieves a better F-measure score.

CCS CONCEPTS

• **Computer systems organization → Natural Language Processing**.

KEYWORDS

Datasets, Emotion detection, Text classification, Roman Urdu

ACM Reference Format:
Adil Majeed, Hasan Mujtaba, and Mirza Omer Beg. 2020. Emotion Detection in Roman Urdu Text using Machine Learning. In *35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW '20), September 21–25, 2020, Virtual Event, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3417113.3423375

1 INTRODUCTION

Language is a resource that can be used to communicate and convey information [25]. It is a way to express feelings as well. The methods of natural language processing have long been used to automatically detect the contents of data in the text [11]. In recent years, Artificial Intelligence-inspired work has focused on increasing efforts to develop emotion-based systems [20]. Emotions are important for several natural processes, modeled in artificial intelligence systems. Emotional research is important for the development of affective interfaces, which provide appropriate emotional responses, can make sense of emotional inputs, and promote online communication through animated affective agents [28][19]. Emotions are fundamental human characteristics and have been studied for several years by researchers in the areas of psychology, sociology, medicine, informatics, etc [10]. Emotions are feelings linked to a person's condition, mood, and relations. A man can portray his feelings with gestures, facial expressions, language, and text [9]. Now, for a couple of days, individuals interact on social media and demonstrate their feelings textually on various social sites [16]. Emotion performs a critical part in our day-to-day success, influencing many elements of our life, including social interaction, conduct, attitudes, and decision-making [8]. Figure 2 shows the abstract flow of the Emotion Detection system. The overall system has different stages which are preprocessing, feature extraction, and prediction model.

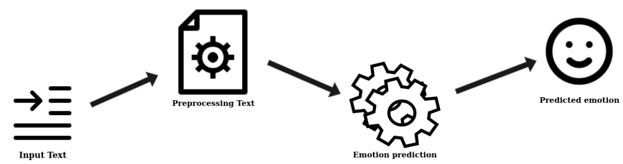

Figure 1: Problem diagram of emotion detection

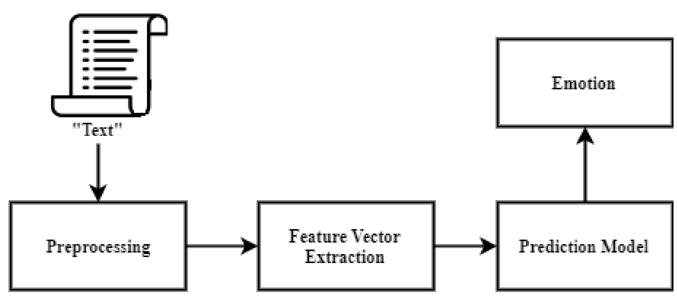

Figure 2: Emotion Classification Pipeline

Emotion Detection plays a crucial role in the growth of human-machine interfaces. Emotions are defined as intense feelings aimed at something or someone in reaction to inner or external occurrences of specific importance to the person [9]. And today, the internet has become the main tool through which individuals convey their emotions, feelings, and views [13]. Emotions are part of

ASEW '20, September 21–25, 2020, Virtual Event, Australia

ACM ISBN 978-1-4503-8128-4/20/09…$15.00
https://doi.org/10.1145/3417113.3423375

our daily lives which results in rapid actions and behaviors that maximize our survival and accomplishment opportunities [16]. When we communicate with other individuals, it's essential to offer those hints to assist them to know how we feel. Social communication is a significant component of our everyday life interactions and it is vital to be able to understand and respond to other people's feelings [30]. It enables us to react properly and create deeper relationships with our friends, family, and loved ones [23]. It also enables us to communicate efficiently with an angry client, or manage a hot-headed worker in a range of social circumstances. This data may be used by business analysts to monitor people's emotions and views about their goods [13]. The issue with most of the Sentiment Analysis being performed today is that the assessment only informs whether the response is positive or negative, but does not define the customers' precise emotions and their response intensity. Millions of individuals share, discuss, post, and comment on every case, news or activity around the globe using social media [8].
As computer usage and internet technology advances, people now use their computers, laptops, smart telephones, and tablets for web access, social networking, online businesses, electronic businesses, e-surveys, etc [10]. They now share their opinions, suggestions, comments, and feedback on something special, products, commodities, policy issues, and other viral news [18]. Most are openly shared and readily accessible from the internet. Of all these views, emotion detection came from its ability to be implemented in many different applications, which has innumerable applications for education, political prediction, brand marketing, and the interaction of human-robots [5][14]. Various applications can detect the emotions from text like improving the quality of the product, investment trend, mental health, relationship mapping, personality detection, etc.
We used Machine learning and NLP techniques for the Emotion Detection system. From Machine learning techniques we used a supervised method to annotate the corpus. After annotating the corpus our proposed model takes a labeled corpus to detect the emotion in the Roman Urdu text. The main contributions of this work are as follows.

- Data collection and improve data annotation
- Perform preprocessing to improve the quality of the corpus
- To make an efficient word embedding for Roman Urdu
- To apply different machine learning algorithm on our corpus.

2 IMPORTANCE AND CHALLENGES OF ROMAN URDU LANGUAGE

Roman Urdu text is the most commonly used language in Pakistan and India on social networking sites [15]. Recently, Roman Urdu has emerged into extra frequent language because people have a choice to discuss emotions and show their emotions in their own language [24]. In the past few years, technology has greatly reshaped the media and news industries. Social networking, forums, and mobile communication have become a common means of data exchange, and opinions [16]. Currently, data from comScore [1] showed that people in the U.S. spent 20% of their overall social media time online. Roman Urdu is the language of the Latin Urdu script. The latest survey conducted by Gallup Pakistan [15] shows that Urdu in Roman Script is the most commonly used language for SMS transmission among those who use a cellphone. Approximately 37% of people continue to use Urdu using Latin Script, while others use Urdu SMS and Arabic Script (15%). More than 7% receive text messages in English, while 29% say they don't send text messages to anyone. The other 2% gave no response. Below is an example in figure 3 of an Urdu sentence with its Roman written translation. Roman Urdu

Example

Urdu Text: آج میں بہت خوش ہوں
Roman Script: Aaj main bohat khush hon.
Translation: Today I am very happy.

Figure 3: Example in Urdu with its Roman Urdu and English Translation

also requires emotion identification. The figure 3 shows that it has a happy emotion. But there are some challenges of the Roman Urdu language to identifying those feelings in a text that can be a source of valuable data that can be used to explore how different individuals are responding according to the circumstances on social media. Sadly, the key challenge for doing research in the Roman Urdu language is the lack of standard tools for evaluation. The benchmark corpora plays an significant role for the creation of methods and techniques for many NLP tasks [21]. Emotional analyzers known as extra well-learned languages, similar to those in British, have not been made feasible for Urdu or Roman Urdu because of their changes in script, morphology, and language change [24].
The following are further complexity of Roman Urdu, which makes the development of an emotion detection system is a more challenging task.

- We need a benchmark corpus for the comparison, evaluation, and analysis of Emotion Detection systems for Roman Urdu. The main problem is there no standard corpus representing Roman Urdu is available.
- One word in Roman Urdu may point to two or more words in Urdu with different sounds. For example, "Sona" meaning "Sleeping" and "Sona" meaning "Gold".
- There are no standards in Latin script that represent the Urdu language. For example, "Aaj main bohat khush hon" and "Aaj main bhot khush hun represent "Today I am very happy".

To find out the emotion our model will look at the context and emotion words, then take a decision which emotion class does it belongs. Figure 4 shows an example of a sentence in English and the Roman Urdu language. In the English sentence, two emotion words such as succeeded and happy and also in Roman Urdu sentence have kamyaab and khush. These emotions words show that it has a happy emotion class. Our model does not make a decision only bases on emotion words it also looks at the context. In the same way, those sentences have no emotion then our model predicts an emotion class based on context.

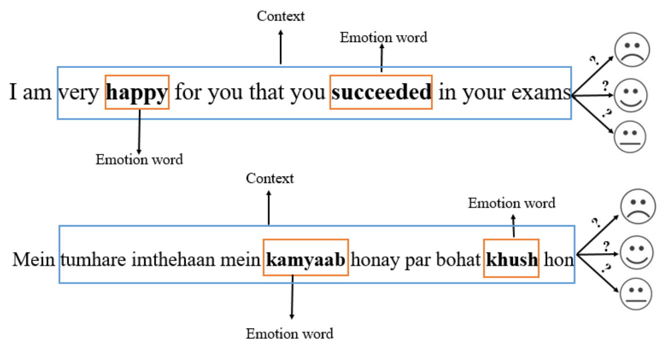

Figure 4: Example with Emotions Word and Context in sentence

3 RELATED WORK

Several people prefer to share their thoughts, feelings, and accomplishments through audio or video files in the present era of smartphones and social networking sites because of the inherent features of these components [16]. However, even though their popularity increases, many people still turn to text to communicate and interact with each other in daily life as well as on social networking platforms [14]. For people to express their feelings to other people, events, or things, the text is still the prime choice [3]. Due to the nature of the data, extracting emotion from text is even harder [12]. It is easier to detect emotions from the text if the words that describe a particular emotion are in the text explicit [29]. Most of the time, however, emotion is subtly expressed. In a single text piece, there are sometimes several emotions. Some texts have ambiguous emotions and words, some words have multiple meanings and many words mean the same emotion. Many texts are sarcastic, or slangs are used [27]. The on-line text has a common characteristic of several words, spelling errors, acronyms, and grammatically incorrect sentences [26]. Sometimes these limitations of textual information make it almost impossible to identify automatic emotions [12]. Work on the extraction of emotions from text is a very popular topic [7]. We have reviewed several papers that directly or indirectly concern the emotional detection of textual data from the languages of English, Arabic, Urdu, and Roman Urdu. The existing literature, however, contains very limited research on the Roman Urdu [31]. Researchers use different approaches to find emotions in textual data. The most common approach in this literature has been discussed. This section includes recent related work and focuses on emotion detection and sentimental analysis from the text.

Hasan *et al* [16] identifies emotions in a social network in several emotional classes by classifying text messages. They used two kinds of classifications to classify emotions, soft and hard classification. The Naive Bayes and logistic regression are probabilistic classifiers and can be used to categorize the emotions across the emotional classes by hard classification labels. The SVM returns the decision score it can use for soft labeling. For the second task, the author has been developing a two-stage framework which is known as EmotexStream to classify live tweet streams. The first phase distinguishes the tweets by the manner of a binary classification with explicit emotion from the tweets. The second phase uses Emotex to perform an explicitly emotionally fine-grained tweet emotional classification. Herzig *et al* [17] used a pre-trained word vector for emotion detection. They used five different datasets from different domains of emotion detection to increase the improvement. For this purpose, the researcher purpose ensemble approach consists of a linear model based on the BOW and non-linear model based on pre-trained word vectors. In the BOW approach, they used the SVM classifier with a linear model and represented every document as a BOW. In the Word Embedding approach, they used the SVM classifier with a non-linear model and represented a document to a fixed-size vector which can be combined from word embedding vectors. The results show that the ensemble approach outperforms from previous approaches for domain-specific datasets. Agrawal *et al* [4] has proposed an unsupervised, context-based approach to the detection of emotions from the text at the sentence level. The sentences were classified with semantic and syntactic dependency. There was no need for a noted dataset and no reliance on a lexicon. Furthermore, it is possible to improve the system's efficiency by suggesting the affected semantic relatedness steps, addressing polarity modifiers, negations, and slang.

Abdullah *et al* [2] proposed a model SEDAT, which predicts the emotion of sentence with intensity in the Arabic language. For the experiment, the model tested on the SamEval 18 Task 1 dataset. They use different word embedding for improving the accuracy and reducing the overfitting. The SEDAT composed of a combination of feedforward, CNN, and LSTM. For the evaluation of the proposed model, they compared the result with the SVM model and the results show that the proposed model outperforms.

Mehmood *et al* [21] focus on the sentimental analysis of Roman Urdu. For the sentimental analysis authors collect the datasets from different domains which consist of 11000 reviews. After cleaning the dataset, annotate the dataset using a multi-annotator methodology. The annotated data set has been used to create a sentiment analysis system using state-of-the-art algorithms. To show the effect on results, different ML algorithms were identified and used to enhance accuracy along with various kinds of word-level features, character-level features, and features union. Ghulam *et al* [14] focus on the sentimental analysis of Roman Urdu. To classify the sentiment from the Roman Urdu text, the authors develop Deep Neural Network long short-term memory (LSTM). For evaluation compared to the results with Machine learning, baseline methods like SVM, RF, and Naive Bayes are used. As the need for the era, there is a lot of work done on the Emotional Analysis in different languages i.e. English, Arabic, and Chinese[16] [2] [32]. According to the literature, we find out the research gap, given below.

- According to the best of our knowledge, the limited work is done in Roman Urdu.
- And which is focus on sentiment analysis in Roman Urdu, not the emotional analysis.

4 PROPOSED APPROACH

Our goal is to predict the emotions from the text. To achieve this goal we apply different machine learning algorithms i.e. KNN, Decision tree, random forest, and SVM to predict the emotion class according to text input. The proposed methodology is explained in this section.The source websites, blogs, and social media links containing Roman Urdu text were identified, the text was extracted using a semi-automatic methodology, and then the data was cleaned by removing undesirable information and stored in an Excel file. Since the extracted data had no emotion identifiers, the data were annotated in the next step using the multi-annotator technique mentioned. Secondly, The input text is first mapped to embedding (Word2Vec) for vector representations. These vector representations fed to our machine learning algorithms for classification.

4.1 Dataset

The benchmark corpora play an important role in the development of tools and techniques for several NLP tasks. For the comparison, evaluation, and analysis of Emotion Detection systems, we need a benchmark corpus. To execute Emotion Detection tasks many corpora are built for other languages i.e English, French, and other European languages. According to the best of my knowledge, there is no emotion-based labeled corpus publicly available for Roman Urdu. This work describes a novel benchmark for emotion-based corpus in the Roman Urdu language. Corpus is hand-labeled manually. Our corpus is made up of 18000 labeled sentences. The sentences are collected from different sources.

4.2 Corpus Collection

In this work to solve the problem, first of all, we make an emotion-based corpus for the Roman Urdu language [6]. For this purpose, identifying the data resources was the primary task to create a corpus. These are selected sites like hamariweb, youtube, facebook, whatmobile, pakistan.web, dramaonline, and twitter. We collected raw data using automatic and manual techniques like Selenium, Twint, etc. To maintain diversity in the corpus, we collect the data from different domains e.g. political, reviews, entertainment, sports, foods, and other miscellaneous websites. For the political and entertainment domain, we targeted twitter and facebook, for review data we targeted different product selling sites i.e. hamariweb, youtube, and masala.tv.

Table 1: Corpus overall statistics

Corpus Statistics	
Total Sentences	18,000
Total Words	2,74,708
Unique Words	39,063

4.3 Annotation

We want to develop an automated system such as Emotion Detection for the Roman Urdu text. For this purpose, we required an annotated corpus with emotions labels. There have been six fundamental emotions which are happy, sad, anger, fear, love and neutral. The annotated data sets are created by expert annotators as usual. The annotator team comprised four annotators. Annotation standards, principles, detailed instructions were given to annotators with multiple samples for possible difficult and common problems. We requested the annotators to follow these guidelines.

- Assign one emotion class, from the six emotion classes described above, to the given example.
- If any example doesn't falls in the above category of emotion, then annotate it as neutral class.
- If an example is a part of multiple emotions, then assign an emotion class which is closer to that example on the basis of context.

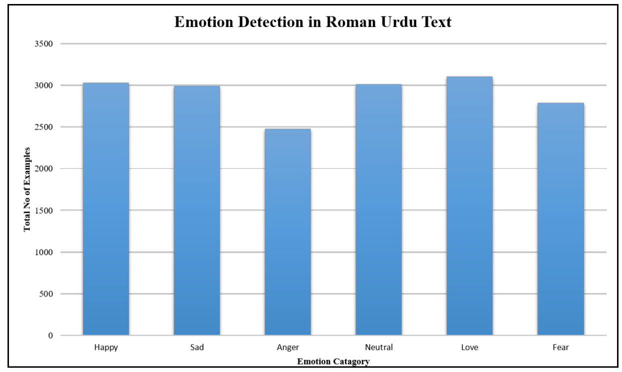

Figure 5: Frequency of corpus

Table 2: Annotated Sample of Corpus

Example	Emotion Class
Tumhara kya hal hai	Neutral
Main aj bohat khush hon k tum pass ho gaye	Happy
Mujhe yeh khabar sun kar bohat afsos howa	Sad
Mjy aaj tum per buhat gusa hai	Anger
Tum mujhe achay lagtay ho	Love
Mujhe Ali bhai se bohat dar lag raha hai	Fear

To maintain sentence diversity, we extracted data from multiple domains. During the annotation phase, we set a different count of examples from multiple domains belonging to different emotions. The total frequency of the labeled corpus is shown in figure 5.

4.4 Preprocessing

The corpus we collected from different online sources is raw, it has various issues, such as space problems, invalid characters, and inappropriate format, etc. So we first preprocess data to clean and convert it into the correct format so that a classification algorithm can be applied to the processed corpus. Some preprocessing steps like:

- Noise and Punctuation Removal
- URL Removal
- Removal of New Line
- Spelling Consistency
- Remove Extra Spaces
- Normalize Data

Table 3: Complete Corpus Results

Model	F1-Score	Precision	Recall	Accuracy
KNN	0.54	0.57	0.56	55.12
Decision Tree	0.43	0.44	0.44	44.29
Random Forest	0.55	0.64	0.56	56.53
SVM	0.69	0.7	0.7	69.54

Only preprocessing the text is not sufficient to achieve a better result for emotion detection from Roman Urdu. The preprocessed text is further processed to extract features that could improve the outcomes. To extract the features here we chose Word2vec, combined from various simple text features, and then pass these features to the ML algorithm to find out the emotions.

4.5 Word2Vec

Word2Vec [22] relies on the following concept that words with similar contexts have equivalent meaning. It begins with a randomly initialized set of word vectors and scans the corpus sequentially, keeping the window always focused on the context around each word. This algorithm typically has a fixed size to the window. Each window usually contains a target word in the center of the windows and surrounding words. Word2vec is a model for word embedding generation. We have been training our word2vec in Roman Urdu because until now no standard Word2Vec is available. Neural networks need a large amount of data for Word2Vec training and we have a huge amount of data to train it, it's around two million sentences. To train our Word2Vec we used the Gensim and FastText libraries. Specific parameters hypertuned to achieve productive results for the Word2Vec training like dimension size, learning rate, window size, etc.

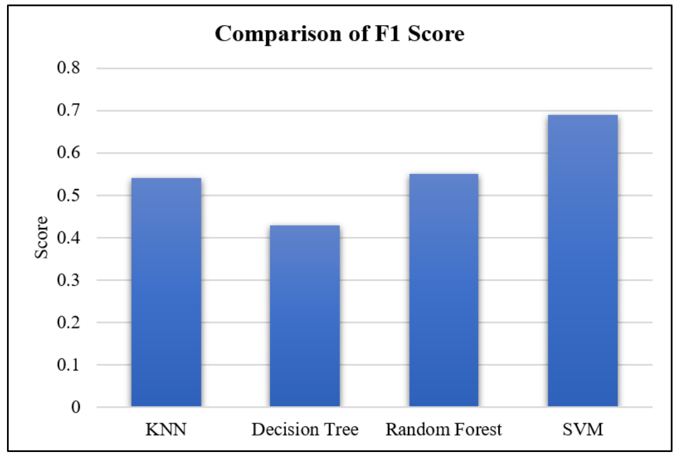

Figure 6: F1 Score

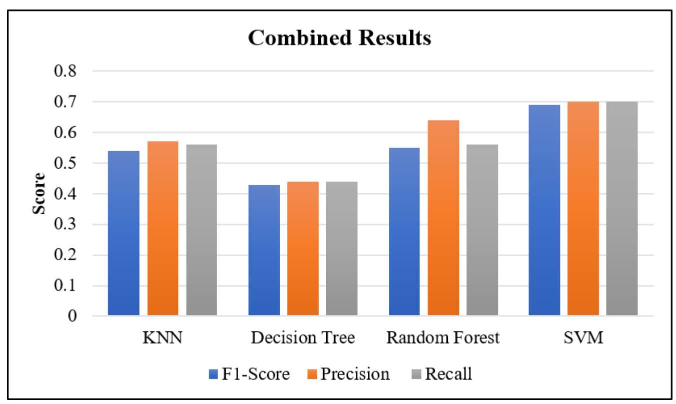

Figure 7: Overall Results

5 EXPERIMENT AND RESULTS

We have trained machine learning algorithms for emotion detection classification for Roman Urdu text on our corpus namely KNN, Random Forest, Decision Tree, and SVM. All algorithms were applied to a preprocessed corpus by selecting the Word2vec feature and obtaining results, we perform 20 experiments to find out the best algorithm in machine learning for this purpose and the performance of each classifier is measured. Results show that SVM model gives higher accuracy, precision, and recall as compared to Decision tree, KNN, and Random Forest. In the table, 3 the shows that all classifiers results are in term of F1 score, Recall, Precision and Accuracy and its shows that SVM model performs well than others.

6 CONCLUSION

There is a lot of work done for Emotion Detection in the English language and many other languages but the Roman Urdu language still needs to be explored. Roman Urdu is the most widely used language in the social media platforms for communication, suitable as a corporate language for NLP purposes. One major issue that is the absence of standard benchmark corpora for Roman Urdu Emotion Detection from the text, we made a comprehensive corpus that contains almost 18k annotated sentences. In this research, the main focus is to find the proper solution for Roman Urdu Emotion Detection. In this research, our novel contribution to the Roman Urdu Emotion Detection from the text is to made a benchmark corpus, word embeddings for Roman Urdu, and apply different state of the art machine learning algorithms for Emotion Detection. To evaluate the results we use multiple classifiers like baseline KNN, Decision tree, SVM, and Random Forest on our corpus. After the experiment and evaluation, the results show SVM model performs well as compared to other classifiers. We achieve an accuracy of 69.4% and an F-measure of 0.69 on Emotion Detection for Roman Urdu. In the future, we try to enhance the results by applying the start of art deep neural network like LSTM, CNN, and Attention models, more over we look into some others word embedding techniques like Bert and Elmo.

REFERENCES

[1] 2016. Nine trends in US media consumption: in charts. (2016).

[2] Malak Abdullah, Mirsad Hadzikadicy, and Samira Shaikhz. 2018. SEDAT: sentiment and emotion detection in Arabic text using CNN-LSTM deep learning. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 835–840.

[3] Basant Agarwal and Namita Mittal. 2014. Text classification using machine learning methods-a survey. In *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012*. Springer, 701–709.

[4] Ameeta Agrawal and Aijun An. 2012. Unsupervised emotion detection from text using semantic and syntactic relations. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE Computer Society, 346–353.

[5] Samar Al-Saqqa, Heba Abdel-Nabi, and Arafat Awajan. 2018. A Survey of Textual Emotion Detection. In *2018 8th International Conference on Computer Science and Information Technology (CSIT)*. IEEE, 136–142.

[6] Muhammad Umair Arshad, Muhammad Farrukh Bashir, Adil Majeed, Waseem Shahzad, and Mirza Omer Beg. 2019. Corpus for Emotion Detection on Roman Urdu. In *2019 22nd International Multitopic Conference (INMIC)*. IEEE, 1–6.

[7] Mubashar Nazar Awan and Mirza O Beg. 2020. TOP-Rank: A TopicalPostionRank for Extraction and Classification of Keyphrases in Text. *Computer Speech & Language* (2020), 101116.

[8] Laura L Carstensen, Monisha Pasupathi, Ulrich Mayr, and John R Nesselroade. 2000. Emotional experience in everyday life across the adult life span. *Journal of personality and social psychology* 79, 4 (2000), 644.

[9] Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM, 413–420.

[10] Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. 2019. SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text. In *Proceedings of the 13th International Workshop on Semantic Evaluation*. 39–48.

[11] Gobinda G Chowdhury. 2003. Natural language processing. *Annual review of information science and technology* 37, 1 (2003), 51–89.

[12] Kevin Crowston, Eileen E Allen, and Robert Heckman. 2012. Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology* 15, 6 (2012), 523–543.

[13] Bharat Gaind, Varun Syal, and Sneha Padgalwar. 2019. Emotion Detection and Analysis on Social Media. *arXiv preprint arXiv:1901.08458* (2019).

[14] Hussain Ghulam, Feng Zeng, Wenjia Li, and Yutong Xiao. 2019. Deep Learning-Based Sentiment Analysis for Roman Urdu Text. *Procedia computer science* 147 (2019), 131–135.

[15] Dr. Ijaz Shafi Gilani. 2009. Preferred Medium for Communicating with Others Via SMS: GILANI POLL/GALLUP PAKISTAN. (2009).

[16] Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu. 2019. Automatic emotion detection in text streams by analyzing Twitter data. *International Journal of Data Science and Analytics* 7, 1 (2019), 35–51.

[17] Jonathan Herzig, Michal Shmueli-Scheuer, and David Konopnicki. 2017. Emotion detection from text via ensemble classification using word embeddings. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. ACM, 269–272.

[18] Moin Khan and Kamran Malik. 2018. Sentiment Classification of Customer's Reviews About Automobiles in Roman Urdu. In *Future of Information and Communication Conference*. Springer, 630–640.

[19] Bo-hu Li, Bao-cun Hou, Wen-tao Yu, Xiao-bing Lu, and Chun-wei Yang. 2017. Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering* 18, 1 (2017), 86–96.

[20] Juan Martınez-Miranda and Arantza Aldea. 2005. Emotions in human and artificial intelligence. *Computers in Human Behavior* 21, 2 (2005), 323–341.

[21] Khawar Mehmood, Daryl Essam, Kamran Shafi, and Muhammad Kamran Malik. 2019. Sentiment Analysis for a Resource Poor Language—Roman Urdu. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)* 19, 1 (2019), 10.

[22] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[23] Saif Mohammad, Xiaodan Zhu, and Joel Martin. 2014. Semantic role labeling of emotions in tweets. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 32–41.

[24] Gule Zulf Nargis and Noreen Jamil. 2016. Generating an emotion ontology for Roman Urdu text.". (2016).

[25] David Nunan. 1999. *Second Language Teaching & Learning*. ERIC.

[26] Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies*. 380–390.

[27] Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. 1–10.

[28] Armin Seyeditabari, Narges Tabari, and Wlodek Zadrozny. 2018. Emotion Detection in Text: a Review. *arXiv preprint arXiv:1806.00674* (2018).

[29] Yang Shen, Shuchen Li, Ling Zheng, Xiaodong Ren, and Xiaolong Cheng. 2009. Emotion mining research on micro-blog. In *2009 1st IEEE symposium on web society*. IEEE, 71–75.

[30] Edward Z Tronick. 1989. Emotions and emotional communication in infants. *American psychologist* 44, 2 (1989), 112.

[31] Ahmed Uzair, Mirza O Beg, Hasan Mujtaba, and Hammad Majeed. 2019. Weec: Web energy efficient computing: A machine learning approach. *Sustainable Computing: Informatics and Systems* 22 (2019), 230–243.

[32] Liang Yang and Hongfei Lin. 2012. Construction and application of Chinese emotional corpus. In *Workshop on Chinese Lexical Semantics*. Springer, 122–133.