computers

MDPI

Review

A Review of Urdu Sentiment Analysis with Multilingual Perspective: A Case of Urdu and Roman Urdu Language

Ihsan Ullah Khan [1], Aurangzeb Khan [1,2,*], Wahab Khan [1], Mazliham Mohd Su'ud [2], Muhammad Mansoor Alam [3], Fazli Subhan [2,4] and Muhammad Zubair Asghar [5]

1 Department of Computer Science, University of Science and Technology, Bannu 28100, Pakistan; ihsanullah1974@gmail.com (I.U.K.); wahabshri@gmail.com (W.K.)
2 Faculty of Computer and Information, Multimedia University, Kuala Lumpur 50050, Malaysia; mazliham@mmu.edu.my (M.M.S.); fsubhan@numl.edu.pk (F.S.)
3 Riphah International University, Rawalpindi 74400, Pakistan; m.mansoor@ripah.edu.pk
4 Faculty of Engineering and Computer Sciences, National University of Modern Languages-NUML, Islamabad 44000, Pakistan
5 Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan 29050, Pakistan; zubair@gu.edu.pk
* Correspondence: Aurangzeb.saadatkhan@mmu.edu.my

Abstract: Research efforts in the field of sentiment analysis have exponentially increased in the last few years due to its applicability in areas such as online product purchasing, marketing, and reputation management. Social media and online shopping sites have become a rich source of user-generated data. Manufacturing, sales, and marketing organizations are progressively turning their eyes to this source to get worldwide feedback on their activities and products. Millions of sentences in Urdu and Roman Urdu are posted daily on social sites, such as Facebook, Instagram, Snapchat, and Twitter. Disregarding people's opinions in Urdu and Roman Urdu and considering only resource-rich English language leads to the vital loss of this vast amount of data. Our research focused on collecting research papers related to Urdu and Roman Urdu language and analyzing them in terms of preprocessing, feature extraction, and classification techniques. This paper contains a comprehensive study of research conducted on Roman Urdu and Urdu text for a product review. This study is divided into categories, such as collection of relevant corpora, data preprocessing, feature extraction, classification platforms and approaches, limitations, and future work. The comparison was made based on evaluating different research factors, such as corpus, lexicon, and opinions. Each reviewed paper was evaluated according to some provided benchmarks and categorized accordingly. Based on results obtained and the comparisons made, we suggested some helpful steps in a future study.

Keywords: preprocessing; feature extraction; classification

Citation: 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* **2022**, *11*, 3. https://doi.org/10.3390/computers11010003

Academic Editor: Paolo Bellavista

Received: 13 October 2021
Accepted: 10 November 2021
Published: 27 December 2021

1. Introduction

The Indo-Pak subcontinent is one of the most significant markets for all types of products. Urdu- and Hindi-speaking people react to any online product or event using Roman Urdu, Roman Hindi, pure Urdu, or pure Hindi text. South Asia or the Indian subcontinent is one of the largest markets for all such types of organizations. In this highly populated area and homeland of approximately 1.95 billion people (https://www.worldometers.info/world-population/southern-asia-population/, accessed on 14 August 2021), companies are attracted to selling their products and understanding people's experiences, feelings, and emotions about their businesses. It is the best strategy to express personal experiences and feelings on the Internet by using the local language. This subcontinent is also very rich in languages, and more than 451 languages are spoken in the Indi-Pak subcontinent. Out of the languages taught in this subcontinent, Hindi and Urdu are the dominant languages. More than 588 million people speak Urdu and Hindi,

which is larger than the number of people who speak English (The world's languages, in 7 maps and charts—The Washington Post). The Internet is a massive repository of data that contain reviews, blogs, and sentiments about products, events, videos, and news. Through this data, companies can determine their outcomes and can make better decisions about improving their products. In this paper, we provide a detailed study related to investigating Roman Urdu and Urdu sentiments, collecting data sets, preprocessing techniques, classifying methods, and comparing results of various researchers.

1.1. Sentiment Analysis for Product Review

Sentiments provide an easy way to access people's feelings about any product or any policy. There is no restriction on people giving their views about any event or any product. According to Encyclopedia Britannica (Urdu language | History, Script, and Words | Britannica), Urdu is a national language of Pakistan, and more than 100 million speak Urdu and use it with Roman text to express their views about products, events, services, issues, and topics. Sentiments given by people can be divided into three levels, which are document level, sentence level, and aspect level. We give this example of a sentence written in Roman Urdu to express someone's feelings about a mobile phone.

"*mere pass Samsung M31 hai, bakwaas mobile hai*"

In the above sentence, the word "*bakwass*" in Roman Urdu is used to mean futile or useless in the English language; therefore, this sentence's polarity is considered negative or "*Canon zabardast camera hi*". Here the word "*Zabardast*" is used for the English word "great "or "awesome"; therefore, this sentence is considered positive.

Similarly, some sentences contain a combination of English and Roman Urdu parts, such as:

"*Go for Dell. Because mae hp lae k pachta rha hu*" or "*Awesome laptop, mera HP*"

Special preprocessing techniques were required in which the English portion was taken separately and the Roman Urdu portion was taken separately with their subjectivity.

Finally, the polarity of sentences was obtained by combining both English and Urdu parts.

Similarly, people can express their views in pure Urdu text. Here we give an example of an Urdu sentence about a product: "مرا HP ہي ٹاپ لپ بہترین اک"

In the above sentence, "بہترین", which is equivalent to the English word "best", shows the positive nature of the sentence. Similarly, another Urdu sentence that shows negative subjectivity is as follows: "ہي موبائل بیکار اک گلکسی سنگ سام".

Here, the Urdu word "بیکار", which means "futile" in English, hence converting the whole sentence into negative form. Although the ratio of writing Urdu text is less compared to Roman Urdu text, the percentage grows day by day by introducing Urdu norms and better functions.

1.2. Research Inspiration

This study was conducted based on the following realities:

i. The customers entered multilingual sentences (in Roman Urdu, pure Urdu, or a combination of both Roman Urdu and English) to express their opinions about any products.
ii. Countries, such as India and Pakistan, are a vast market for online business, and this business spreads day by day, hence motivating us to perform a study on people's opinions using languages other than English about some online products.
iii. Advancement in multilingual sentiments compelled us to perform a comprehensive survey on Roman Urdu and Urdu sentences relating to the product reviews. This goal was achieved by searching the relevant studies, then identifying, summarizing, and evaluating.

1.3. Our Commitments and Contributions

Our contributions and commitments for this paper are as follows:

a. To discuss the importance of sentiment and multilingual sentiment analysis.
b. To discuss various linguistic strategies used in multilingual sentiment analysis, especially in Urdu and Roman Urdu text.
c. To evaluate presently available techniques related to multilingual sentiment analysis.
d. To identify problems that occur in multilingual sentiments.
e. To compare classification techniques used for multilingual sentiments.
f. To suggest future work.

1.4. Relation to Previous Work

Deduction of multilingual sentiments is a challenging task. Minimal work has been conducted on multilingual emotions. A more significant part of emotion analysis work has been completed in the English language only. Very limited work has been performed in Urdu and Roman Urdu because of their resource-poor quality [1]. Languages other than English are often addressed as resource-poor languages due to the unavailability of proper resources [2]. In Figure 1, we present the mechanism or the process that we adopted to consider related articles for the writeup of this survey study. In addition, we also present an overview of a few research papers out of the selected papers that present their work in Roman Urdu and Urdu. [3] in their research presented a state-of-the-art review on multilingual opinion mining. The research work included steps, data preprocessing, features extraction, and using classifiers. A number of classification techniques were applied on the English language, as well as other languages. Eleven models were implemented for testing corpora. The results obtained were not satisfactory as reported by the corresponding authors. Researchers found the primary problem was the lack of lexical recourses for any language other than English. [4] Muhammad Bilal et al. [4] presented research containing 300 Roman Urdu sentiments, out of which 150 were positive and 150 were negative. Out of three classifiers, applied Naïve Bayes outperformed decision tree and KNN by using a WEKA platform. Ghulam et al. [5] in their research, used long short-term memory (LSTM), a model of a deep learning neural network, and hence showed very high accuracy on sequential data. Research showed that deep learning provides a capability through which long-range information is captured and solves gradient attenuation problems. Comparative results showed higher accuracy of the adopted LSTM model than naïve Bayes, RS, and SVM over 3 K data. Rafique et al. [6] presented research in which popular sentiment analysis techniques, namely naïve Bayes, support vector machine (SVM), and logistic regression with stochastic gradient descent (LRSGD) were applied over Roman Urdu sentiments. Results given in the research article showed that overall performance of the linear regression (LR) model with SGD and SVM was much better than naïve Bayes. There was a trivial difference between the performance of both LRSGD and SVM, but SVM gave the best results on unigram, bigram, and TF-IDF features sets and achieved 87.22% accuracy. Akhtar et al. [7] presented a novel hybrid technique used for Hindi data sets. This technique is applicable for all resource-poor languages. Combining a support vector method of machine learning and a CNN (convolutional neural network) of deep learning, this novel approach showed very high performance over Hindi data sets, as well as on English language data sets. Chhajro et al. [8] presented research on multitext classification of Urdu and Roman Urdu text using machine learning and NLP preprocessing techniques. Data were collected through different online sites using the Beautiful Soup web-scraping tool and preprocessed using various preprocessing techniques to clean the data and remove noise. Five different machine learning classifiers (naïve Bayes, linear regression, SVM, LSVM, and random forest) were applied on collected data sets. Results showed that LSVM outperformed the other machine learning algorithms with an accuracy of 96%.

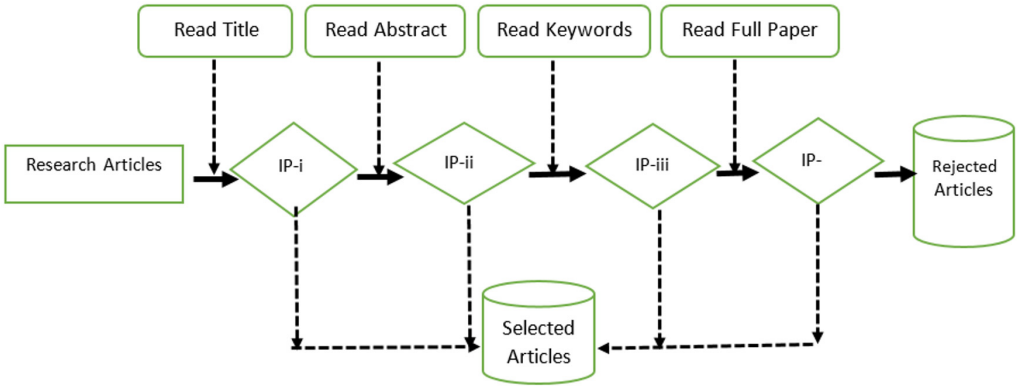

Figure 1. Flowchart of searching and filtering of research paper.

2. Methodology

This survey paper was conducted using the following methodology:

A survey for this paper was performed by collecting related articles from various online resources. Research question criteria were fixed for selection and rejection of these articles. Research articles selected for this survey were filtered purely on given research questions.

2.1. Research Questions

The survey was conducted using the following research questions:

Research Question 1: What is text preprocessing, and why is it important in sentiment analysis? How is multilingual text (Urdu and Roman Urdu) processed using the various techniques mentioned by the researchers in the enlisted articles?

Research Question 2: What is feature extraction? What feature extraction methods are used for the Roman Urdu and Urdu corpus in the selected research papers?

Research Question 3: What is sentiment classification, and how are various classification techniques used to evaluate multilingual (Urdu and Roman Urdu) sentiments in the selected articles?

2.2. Technique and Criteria for Acceptance and Rejection

To obtain the most relevant research articles, different research queries were given using keywords, such as "multilingual sentiment analysis", "multilingual opinion mining", "Roman Urdu sentiment analysis", "Urdu sentiment analysis", "multilingual opinion mining on product review," "Roman Urdu preprocessing", "Urdu preprocessing", "sentiment classification on various products", "Urdu part of speech tagging", and "Roman Urdu classification method."

The selection and rejection or inclusion/exclusion principles of articles followed the criteria mentioned in [9–11], which also refer to Figure 1.

IP1: Inclusion Principle: Select an article that is entirely related to a given title or contains one or a few words related to a given title.

IP2: Inclusion Principle: Select an article that contains an abstract related to some classification technique for multilingual sentiments (Urdu and Roman Urdu) for a product review.

IP3: Inclusion Principle: Select an article that shows some new techniques in multilingual sentiment analysis.

IP4: Inclusion Principle: Select an article that shows the collection of corpora of Urdu and Roman Urdu sentiment with a preprocessing technique.

The exclusion principle (EP) for selected articles is as follows:

EP1: Exclusion Principle (EP): Discard the article that is against the criteria described in IP1 to IP4.

2.3. Study Quality Evaluation

The procedure mentioned in article [9] was adopted for selected articles to maintain quality. Every research article was assessed by the research questions given as already

mentioned in Section 2.1. An Excel spreadsheet was created, and each quality assessment question was added with a predefined rating: value "1" was given to those research papers in which answers were entirely explained, value "0.5" was given to those questions in which answers were partially explained, and value "0" was given to those questioned in which answers were not explained. There were a total of four research questions that were used for quality assessment.

The results indicated in Table 1 showed assessment questions on six selected research articles discussed in Section 1.4 of this paper. Out of a total quality score of 4, research article 6 received a 4.0 with the normalized score of 1, and article 1 received a score of 3.5 with a normalized score of 3.5, while research articles 2 and 3 received a score of 3 with a normalized score of 0.75, and research articles 4 and 5 received a score of 2.5 with a normalized score of 0.63. The threshold for the quality score was 0.5. Any articles below this score were rejected as they did not fulfill the quality criteria.

Table 1. Assessment questions on six selected research articles discussed in Section 1.4 of this paper.

Quality Assessment Criteria	Research Questions	Sample Studies						Remarks
		S1 Chhajro et al. [8]	S2 Dashtipour et al. [3]	S3 Bilal et al. [4]	S4 Ghulam et al. [5]	S5 Rafique et al. [6]	S6 Akhtar et al. [7]	
QA1	The article provided details about preprocessing technique(s) used for multilingual (Urdu and Roman Urdu) sentiments.	1	1	1	0	0.5	1	S4 explained no preprocessing techniques in their research. S5 partially defined it.
QA2	The article provided details of feature extraction techniques used for multilingual sentiment analysis.	1	1	0.5	0.5	0.5	1	S3, S4, and S5 partially discussed feature extraction techniques in their research articles.
QA3	The article provided a classification technique for multilingual sentiments especially for Roman Urdu and Urdu using some classifiers.	1	1	1	1	1	1	All researchers used one or more classification techniques along with proper plate form.
QA4	The article gave enough relevant data to evaluate multilingual sentiments.	0.5	0	0.5	0.5	0.5	1	S2 did not have its own data set; similarly, S1, S3, S4, and S5 consisted of limited data sets. Only S6 contained enough data.
Summation (out of 4) Normalized score (0–1)		3.5 0.88	3 0.75	3 0.75	2.5 0.63	2.5 0.63	4.0 1.0	

2.4. Survey Execution

One hundred twenty-three (123) research articles were retrieved following the criteria mentioned in Section 2.3 from different research articles databases, such as IEEE Xplore, ScienceDirect, SpringerLink, and Wiley. Seventy-four (74) articles were selected in the first phase after applying the insertion criteria and in the second phase using rejection criteria; finally, 34 articles were selected for the survey in this paper.

2.5. Survey Classification

Survey classification presented a detailed summary of selected research articles conducted on Urdu and Roman Urdu sentiments. The idea behind this survey was to mention all the related tasks that help fill the research gap and find ways through which Urdu and Roma Urdu sentiment analysis is performed in a much more accurate way. This survey was done to explore the ideas of various authors about preprocessing, feature extraction, and application of various classifiers with their advantages and disadvantages in Urdu and Roman Urdu text. The survey was conducted and categorized according to the research questions that follow.

3. Research Question 1

What is text preprocessing for multilingual sentiment analysis (Roman Urdu and Urdu sentiments), and what type of techniques are used for text preprocessing by the researchers in the enlisted articles?

What is text preprocessing, and why is it important in sentiment analysis? How is multilingual text (Urdu and Roman Urdu) processed using the various techniques mentioned by the researchers in the enlisted articles?

Text preprocessing in sentiment analysis refers to preparing input data to be provided for the next stage of evaluation and checking. Text preprocessing is a challenging task, and for languages other than English, it becomes harder due to unavailability of required resources, because each language contains its own word segmentations, speech tagging, and grammatical hurdles. In this part, our focus was to check Roman Urdu and Urdu preprocessing techniques discussed in selected articles. In general, preprocessing for multilingual sentiments are divided into three steps, which are listed below and graphically shown in Figure 2.

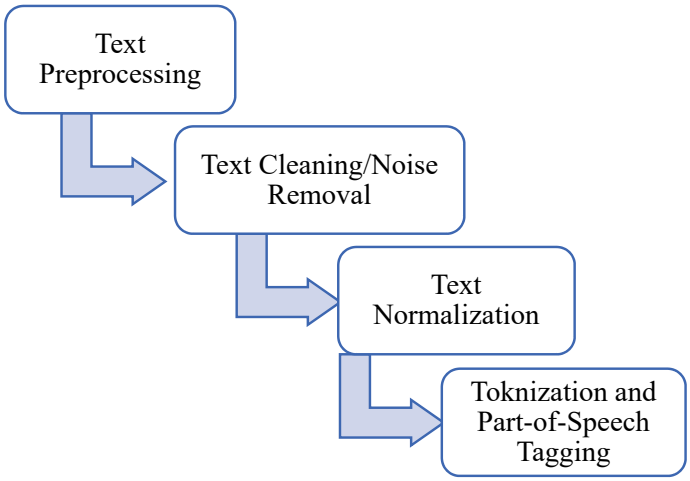

Figure 2. Flow of preprocessing techniques.

a. Text Cleaning/Noise Removal;
b. Text Normalization;
c. Tokenization and Part-of-Speech Tagging.

3.1. Text Cleaning/Noise Removal

Typically, text collected from the Internet contains a lot of noise in term of HTML tags, scripts, punctuation, and advertisements. Eliminating all these helped to reduce noise in the text, which ultimately enhances to some extent the performance and accuracy of the classification models used for text. Preprocessing is a very crucial step in multilingual sentiment analysis. Text cleaning process is depicted in Figure 3.

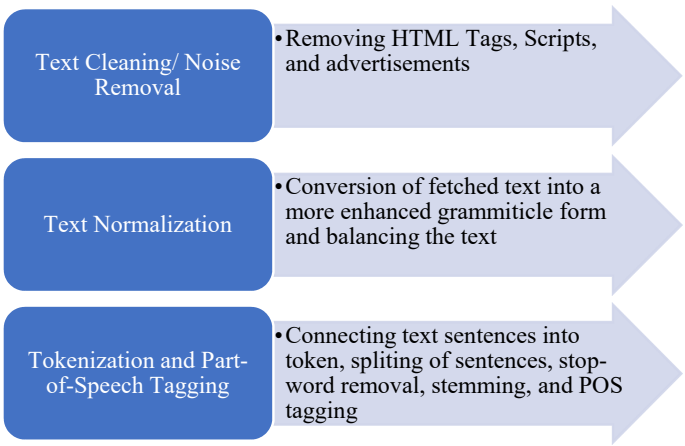

Figure 3. Steps involved in preprocessing techniques.

Dashtipour et al. [3] describe the process of text cleaning and noise removal for multilingual sentiment analysis. The first part of the paper described the initial process of text collection and preprocessing of multilingual sentiment analysis and preprocessing techniques applied on various languages other than English. According to Bilal et al. [4], preprocessing is obtained by applying various text cleaning steps on extracted data. Web crawlers are used for the extraction of contents, and extracted content is stored into some sequence and manually given the subjectivity that rejects all non-relevant data. Alam et al. [12] in their research performed preprocessing in two steps. The first step consisted of removing noise from the data and converting long sentences into short ones of less than 30 words. In the next step, one-to-one correspondence was made between Roman Urdu text and pure Urdu text. If we take this example of one-to-one correspondence, either input or output sentences must contain the same text language: thus, in Roman Urdu "*Yeh kuta hai*" was transliterated to pure Urdu as یہ کتا ہي. However, all sentences do not have the same length. Urdu versions of text consist of lengthier sentences than in RU. In the following example, "*Yhe Faisalabd hy*" is translated as یہ فِیصل آباد ہي. Comparing these sentences shows the RU sentence consists of three words, whereas the Urdu sentence consists of four words. The Urdu language has its own specific grammatical structure; therefore, there is space between آ با د and فِیصل, which shows that Urdu contains multiple words as compared to Roman Urdu, which could potentially mislead the model. The preprocessing step in the research paper of [6] focused on collecting data, labeling data as positive, negative, or neutral, and eliminating words that are not necessary, such as stop words (words that occur frequently but not as important for analysis), punctuation marks, and numerical characters. Khan et al. [13] focused their studies on customers' automobile reviews using Roman Urdu text. The data preprocessing portion of this research paper included data extraction, removal of stop words, conversion of all uppercase words into lowercase, development of a corpus of 2000 sentiments, and creation of an ARFF (attribute-relation file format) file for further processing. Although research conducted by Bose et al. [14] focused on food product reviews using English text, a detailed discussion was provided on preprocessing techniques. The main steps involved in preprocessing were removing all URLs (e.g., www.abc.com), hashtags (e.g., #topic), screen names (e.g., @username), symbols, punctuation, numbers, duplicate sentences, and stop words; changing text into lowercase; and replacing words with their stems or roots. All these steps helped in the removal of noise from the extracted text and helped in cleaning the data. Khan et al. [15] in their research discussed a space problem in Urdu sentiments in the preprocessing stage. When Urdu text is converted into English or Roman Urdu, it leads to some problems. The Urdu word "انکا" consists of two words, but the algorithm considered it a single word. Similarly, the word "د ا نشمند" ("danishmand", "intelligent") is basically one word, but after tokenization, it was taken as collection of two words, i.e., "د ا نش" and "مند", which created another problem for the algorithm. To overcome these problems, different preprocessing strategies, such as noise removal, detection, word tokenization, and sentence boundaries are used.

3.2. Normalization

Textual data from social sites and other user-generated content is used for analysis and decision making. Since users are free to express their opinions without using basic grammar and lexical rules, most of the time, this textual data consist of informal language, which differs from everyday language use. Such texts require conversion into a more advanced grammatical form, which is furthered by a NLP analysis tool. As per Wikipedia, "text normalization is the particular type of process in which text is transferred into a single canonical form that it might not have had before". The process of normalization is performed before applying a model to make the text consistent before processing.

Text normalization requires a complete awareness about the type of text, what kind of procedure is adopted for text processing, and the afterword process. The authors of [16,17] performed a study comparing normalization methods of social media text in different

languages, such as Chinese, Arabic, Japanese, Polish, Bangla, Dutch, and Roman Urdu to obtain the best results. A model was proposed along with an algorithm to normalize Roman Urdu text. As per their study, algorithms for normalization of Roman Urdu text were based on phonetic algorithms. They suggested using a machine learning technique to produce better results in the future. Posadas-Durán et al. [18] used multilingual Twitter text for sentiment analysis.Alam and ul Hussain [12] presented research that included a normalization process using tokenization and each word's frequency, which was built separately for both the RU and Urdu. There is lot of variation in Roman Urdu, as there are many words for a single Urdu word. For example, for the word, یہ, the top five Roman Urdu variants are *yeah, yeh, yeah, ye, and yah*. Urdu-speaking persons use Roman Urdu for expressing their feelings about any product or event. Urdu language has a lack of a standard lexicon, and there are many spellings of one word that are used in Roman Urdu, e.g., the word خوشی kushi (happiness) can also be written as kushai, khooshi, khoshi, and khshi. Specifically, it creates two main problems: the first one is that one word contains different spellings, and the second one is that one word can be used for two different meanings, e.g., "bahar" can be used for both "outside" and "spring". Khan and Malik [13] expressed normalization in their research as follows: Before going to the classification phase, all the string attributes were transformed into a set of attributes, depending on the word tokenizer using the StringToWordVector. The training data can determine attributes. Sentiments were divided into "good", "bad", "positive", or "negative." A classifier must be trained by a set of rules from the training corpus before the testing process.

3.3. Analysis of Natural Language

Analysis of natural language consists of the following terms.

3.3.1. Tokenization

It is a process in which long text is converted into words and symbols, placing all words and symbols in double quotes [19,20]. Word tokenization is an initial step for higher-order natural language processing tasks, such as part of speech, named entity recognition, parsing, etc., and independent NLP tasks. The ULP researchers use various techniques for different Urdu word tokenization issues. They have achieved remarkable results and contributed to the ULP research community. The common techniques used for Urdu word tokenization are dictionary/lexicon, linguistic knowledge-based, and statistical/machine learning.

3.3.2. Sentence Splitting

It is the process that determines sentence boundaries. It is another important preprocessing for various higher-order language analysis processes, such as tokenization, named entity recognition, part of speech tagging, parsing, and information retravel. It is a very challenging task in Urdu as Urdu uses various marks for sentence boundaries.

3.3.3. Stop Word Removal

Words that commonly occur but have no significant meaning in any given language are called stop words. Removal of these words normally improves the overall performance of the sentiment analysis models [20]. In Urdu, stop words are referred as conjunction words or haroof jaar (خروف جار). Stop words have no role in text classification and are usually considered meaningless; these words are simply functional words for any language. Since these words are meaningless, all stop words are eliminated from the corpus to reduce the size. All languages have a list of predefined stop words, so they are eliminated using the predefined list. The main advantage of the elimination of stop words is that in IR, only relevant documents are returned.

3.3.4. Stemming

This process converts words into their root form; for example, the word "working" is changed to its root form "work" [21]. Stemming is considered a core data preprocessing

text analysis process. The objective of stemming is to shrink the token into its parent word or root word. Stemming is usually performed when dealing with textual data prior to IR, DM, and NLP. Stemming consists of reducing a given word to its stem, base, or root, e.g., the stem of دردمند("dardmand", "sorrowful") is درد("dard", "pain").

4. Research Question 2

What feature extraction methods are used for the Roman Urdu and Urdu corpus in the selected research papers?

Sentences are a rich source for feature extraction, and the process through which these features are extracted is called feature extraction. In sentiment analysis, the term text feature extraction is used for the creation of lists from extracted data and the conversion of them into feature sets that are further used by the classifier. Mehmood et al. [22] in their research on Roman Urdu discussed the various aspects of Urdu sentiments, as well as variations in writing Roman Urdu words, as there is no basic structure for Roman Urdu words; e.g., the word "bakwaas" can be written in many ways, as "bakwas", "bakwass", "bkwass", "bkwass", etc. have different levels of features extracted on different word levels using uni-grams, bi-grams, uni- and bi-grams, and uni-, bi-, and tri-grams. In the next step, character-level features were extracted with and without word boundaries. Each subcategory used bi-gram, tri-gram, four-gram, five-gram, and six-gram features. In the third level of feature extraction, a union was made between word level and character level. Various machine learning and deep learning algorithms were applied in which "voting" outperformed over 13 other classifiers with a maximum accuracy of 80.5% with the uni-bi-tri feature. Mehmood et al. [23] in other research proposed a novel technique for assigning weights to different Roman Urdu terms called discriminative feature spamming technique. The main idea behind their work was to assign weight during the feature selection procedure. Roman Urdu is like other languages in that it is a resource-poor language and contains no specific rules for writing words and sentences. This weighting technique was compared with other weighting techniques, such as binary weighting, row term frequency, and TF-IDF. Eleven thousand reviews were collected through different online shopping sites. Word level, character level, union level (a combination of word and character levels), and stylistic features were applied on the collected data. DFST was applied with term utility criteria (TUC) using various machine learning algorithms, and the highest accuracy was achieved through the "voting" algorithm. Manzoor et al. [24] proposed a deep learning neural network model for Roman Urdu sentences. A preprocessing mechanism was applied on 10,000 Roman Urdu sentences and assigned the subjectivity as positive or negative. A process of normalization further refined the sentences and more than 3000 sentences were selected for checking out of the 10,000. A bidirectional LSTM in a deep neural network was applied with self-attention to handle complex sentences and variation of words in Roman Urdu. Results showed that self-attention bidirectional LSTM had an accuracy of 68.4%, a precession of 68.4%, and a recall of 68.5% on the preprocessed data set, where an accuracy of 69.3%, a precession of 69.3%, and a recall 69.4% were obtained from normalized data sets. Mehmood et al. [25] in their research worked on an analysis of Roman Urdu sentiments. The research process started with the collection of Roman Urdu sentiments, and manual provision of subjectivity was applied to these sentences. A feature extraction process was performed in the second phase by using uni-gram, bi-gram, and a combination of both uni-gram and bi-gram. Various machine learning algorithms were applied on selected review. Each machine learning classifier was applied with each feature selection, and results showed that naïve Bayes outperformed over other classifiers with a feature selection of uni-gram-bi-gram. Iqbal et al. [26] in their research presented a lexicon-based 5031 tweets (out of which 2673 were positive, 1923 were negative, and the remaining were taken as neutral) of Roman Urdu related to the Pakistan general election in 2018. Preprocessing techniques were applied to remove noise from the data, translate English words and tweets into Roman Urdu, and normalize and tokenize the tweets. Positive, negative, and neutral sentiments were analyzed separately using a lexicon-based model. As per authors, given

results obtained 98% accuracy from the positive class, 94% accuracy from the negative class, and 96% accuracy from the neutral class. For processing multilingual text either in Urdu, Roman Urdu, or mixing any other language, the two major lexical resources are corpus and lexicon. First, we define the Roman Urdu corpus, then the Urdu corpus.

4.1. Roman Urdu Corpus

Bilal et al. [4] collected 300 sentiments, out of which 150 were positive and 150 were negative, using easy web extractor software. Alam and ul Hussain [12] collected up to 5 million Roman Urdu sentences and up to .1 million Urdu sentences by crawling. Roman Urdu sentences were converted into Urdu. The complete Roman Urdu to Urdu parallel corpus collection consisted of 0.113 million lines. Rafique et al. [6] in their study prepared a Roman Urdu collection. Their collection consisted of 806 comments, out of which four were positive and 406 were negative. They manually assigned the subjectivity to all the collected sentiments. Sharf and Rahman [16] in their research made an analysis of data collected from different websites, such as Twitter, Reddit, and Urdu poetry. Their data set consisted of 10 input files from these sources. Approximately 280,000 sentiments in Roman Urdu were collected as corpus from different social and newspaper sites. Sharjeel et al. [27] in their research used an Urdu news corpus that was further distributed into source documents and derived documents. One thousand two hundred documents were used for the purpose of evaluation, including 227 words from source documents and 254 words from derived documents. Rafique et al. [6] expressed in their research that preprocessing that includes removing noise and cleaning of text is not enough to achieve better results. The text after preprocessing needs further processing using extraction features that may improve the results. In combination with the other researchers' already given features, they developed a list of different features necessary for the normalization of data. These features are expressed as follows:

a. N-Gram: An n-gram is a contiguous sequence of words from a given text. When the value of n is 1, it refers to uni-grams; when the value of n is 2, it refers to bi-grams. For example, in the sentence, "I live in Pakistan", the uni-grams are "I", "live", "in", and "Pakistan", and the bi-grams are "I live", "live in", and "in Pakistan". TF-IDF: A statistical measure to determine the importance of words in a document.
b. OneR Attribute: It uses simple association rules to find out only one main attribute involved in the principal prediction component: axes that give the data the maximum variation called principal component.
c. Gain Ratio Attribute: It is the ratio of information gain to intrinsic information. The purpose of using this ratio is to reduce bias toward multivalued attributes. Khan and Malik [13] in their study collected sentiments comprised of 2000 automobile reviews in Roman Urdu having equal polarity of positive and negative reviews. One thousand six hundred reviews were used for training the machine, and the remaining 400 were used for testing the accuracy of the models trained via different classifiers.

4.2. Urdu Corpus

Akhter et al. [28]'s area of research is "Automatic Detection of Offensive Language for Urdu and Roman Urdu", in which a Roman Urdu data set consisted of 0.147 million people's comments collected from multiple videos from YouTube. This data set is available in a comma-separated file (CSV) format. There is a lack of a standard data set of Urdu that can be used for offensive language detection; therefore, they developed a data set of 2171 comments and designed a data set of Urdu language from YouTube videos. Hashim et al. [29] in their research proposed a word embedding neural network approach that was used to represent RU sentences in a more effective way as compared to all previous approaches. The main contribution was to develop an approach that used large RU data sets as they used in their research. Normalization of text was performed by collecting words of the same style and assigning them a single word. A process of normalization was performed using linguistic rules given by Sharf and Mansoor [17] that created 100 new

rules based on word phonetics. To illustrate this point, all words including "kesi", "kesy", "kesyy", "kesiy", and "kesii" were transformed into "kese" by considering the phonetics of word-ending characters (e.g., I, y).

4.3. Sentiment Lexicon

According to Dzakiyullah et al. [30], for unlabeled data, a lexicon-based approach is the best one. Due to the poor resource nature of Urdu language, only a few lexican are available for this language. Naz et al. [31] in their studies developed an application of POS tagging by assigning tags to the words based on surrounding words and character affixes using handcrafted rules. Iqbal et al. [26] in their study described the linguistic resources used for sentiment analysis. They used a lexicon of 3900 words containing adjectives, verbs, adverbs, and nouns for improved sentiment analysis. For example, in the sentence, "***Mujhe ye phone buhat acha aur sasta lagta hai***" ("I think this phone is very good and inexpensive"), to analyze the sentence, the authors took the intensifier "buhat" ("bht", "very") into consideration along with the adjectives "acha" ("aa", "good") and "sasta" ("ssta", "inexpensive").

Figure 4 is the consolidated graphic representation of the major steps involved in a sentiment analysis tk.

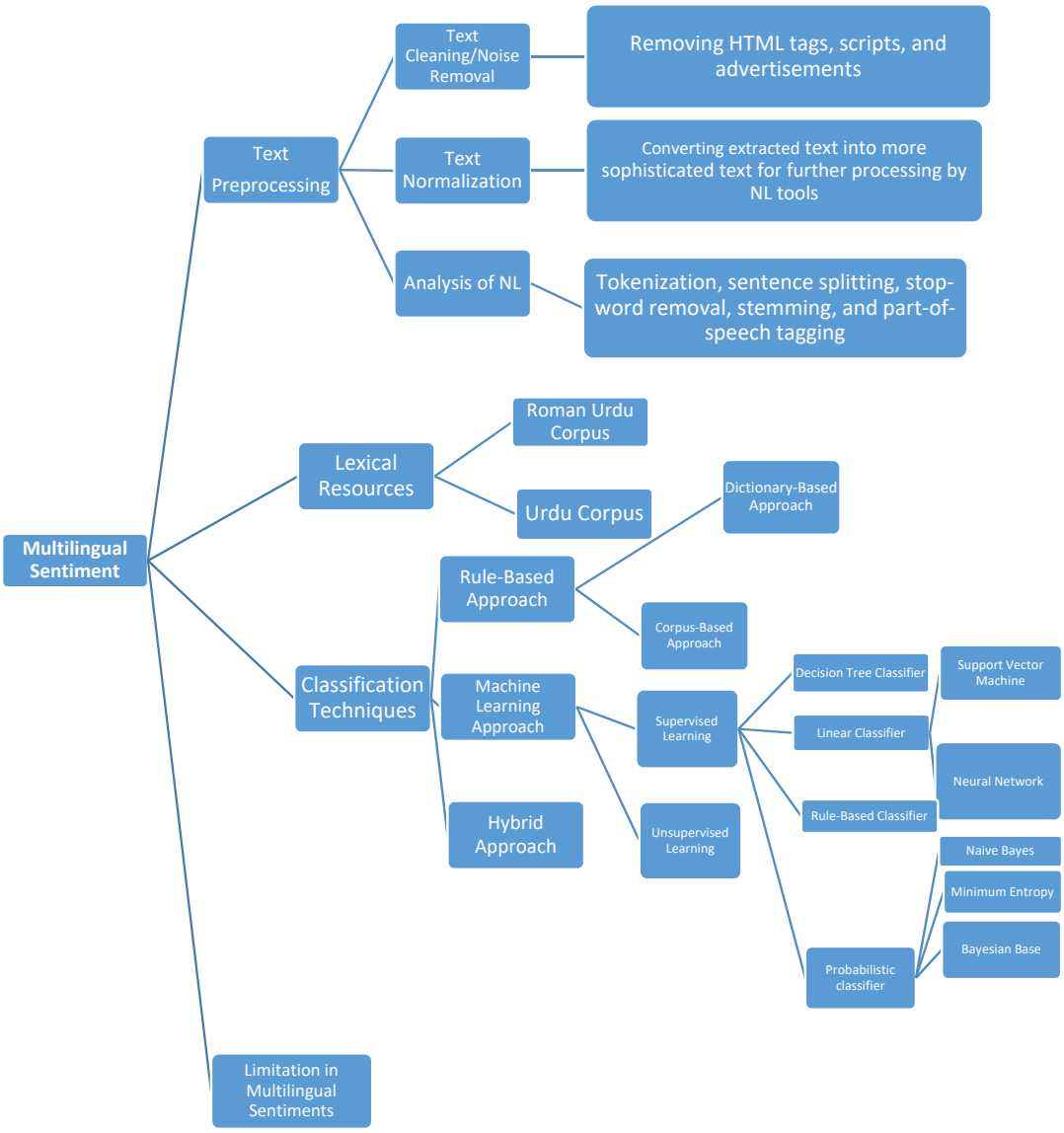

Figure 4. Graphic representation of the sentiment analysis process.

4.4. Features for Sentiment Analysis

It is observed that an optimal feature set has a greater impact on performance than the learning model chosen. Many advanced sentiment analysis systems employ a range of linguistic features, spanning from semantic information about words to lexical and syntactic structure information. Below some common features are listed:

(a) Word presence and frequency: Individual words or syllable n-grams, along with respective frequency counts, are analyzed in this sort of feature. Sometimes it employs term frequency ratings to highlight the relative importance of features or offers the terms binary weighting. The most commonly n-gram features include uni-gram, bi-gram, and tri gram.
(b) Part of speech tags: These features are also known as language-dependent features. One of the approaches for creating a more particular feature in a document is part of speech (POS). The existence of a grammatical structure, such as adjective or negative, may be identified employing the POS-based feature in a document. The fundamental marker of feeling or opinion in a writing is the adjective and negation.
(c) Words and phrases of opinion: These are terms that are widely used to convey views, such as "excellent" or "awful", "like" or "dislike". On the other hand, some sentences offer opinions simply employing opinion words.
(d) Negative terms: The emergence of derogatory words may cause a shift in viewpoint attitude, such as "not good" is comparable to "awful".

4.5. Feature Extraction Methods

Generally, feature extraction methods fall into two broad categories:

- Lexicon-based approaches;
- Statistical approaches.

The lexicon-based approaches are dependent on human efforts. All the annotations are performed by the domain experts manually. The main advantages of this approach are the degree of accuracy and the availability of clean data. However, manual annotation is a tedious job and requires a lot of effort, as well as time. The statistical approach is a widely adopted approach and is considered state of the art. All annotation is achieved automatically with the help of learning models. The main advantages are its speed and fast development; however, low accuracy rate is its limitation. The main approaches that are used include bag of words (BOWs), point wise mutual information, chi-square methods, latent semantic indexing, hidden Markov model (HMM), LDA, etc.

5. Research Question 3

What is sentiment classification, and what type of classification techniques are used to evaluate multilingual sentiments?

What is sentiment classification, and how are various classification techniques used to evaluate multilingual (Urdu and Roman Urdu) sentiments in the selected articles?

Sentiment classification is the process of gathering people's opinions and emotions from different online programs on the Internet and assigning them values as positive, negative, or neutral. Through natural language processing (NLP), the subjective data are interpreted, which helps to understand customer feelings about any product, service, or brand. NLP's primary goal is to obtain a solution to the problem using computer techniques and linguistics that convert human given text into a format understandable by the computer.

Sentiment classification is divided into three approaches:

i. Rule-based/lexicon based approach;
ii. Dictionarybased approach;
iii. Corpus-based approach;
iv. Machine learning approach;

v. Hybrid system.

5.1. Rule-Based/Lexicon-Based Approach

This technique contains a series of manual rules for each tag. The rule-based method depends on the lexicon of each language, which has positive and negative words. Any text's polarity is determined by calculating the number of positive or negative words in a text. If a phrase contains more positive than negative words, the phrase is called positive. However, this system has some limitations, including adding new words and finding polarity of complex type sentences.

The rule-based or lexicon-based approach is further divided into two types, i.e., dictionary-based approach and corpus-based approach.

5.2. Dictionary-Based Approach

This approach is used to find the polarity in the sentence level or document level manually or by using software, such as WordNet. Start counting the words using signaling sentiment words, such as negations, then find the frequency of these words. In a multilingual context, these words are then translated into English for assignment of values against each word. Since each word is collected manually according to its polarity, it is considered a simple approach. Even though this technique is considered less accurate, the quality of the algorithm depends on the performance of work done for the collection of words for the specific language.

5.3. Corpus-Based Approach

It is a data-driven approach, which not only accesses sentiments by using labels, but also takes advantage of context used in machine language algorithms. A corpus-based approach uses seed words related to opinion, which can be used for finding other opinionated words when using a vast corpus [32]. This approach is divided into two subtypes: statistical-based approach and semantic-based approach. The first is used by collecting the polarity of words by number of occurrences, which is also called the frequency of words. A semantic-based approach uses similar sentiment values to semantically close terms [21,33].

5.4. Machine Learning Approach

This system uses machine learning algorithms and artificial intelligence to predict sentiments. A trained data set is used for such type of predictions. By knowing the sentence's polarity, the machine learning system converts text data into vectors and locates a predefined pattern associated with each vector that is negative, positive, or neutral. The system becomes intelligent through given data and starts making their predictions for classification. The accuracy of the system improves by providing more accurate data sets. Machine learning is further divided into two types:

- Supervised learning;
- Unsupervised learning.

Supervised Learning

This method uses labelled data, which is also called trained data sets, for analysis of sentiments. There are different techniques involved in supervised learning, which are as follows:

- Decision tree classification;
- Linear classification;
- Neural network-based classification;
- Probabilistic classification.

Decision Tree Classification

The decision tree uses a data-mining technique called divide and conquer. A DT consists of a structural graph that includes root nodes, branches, and leaf nodes. Every internal node represents a test on an attribute, every branch represents the outcome of a test, and a leaf node contains a class label. The node that exists on the topmost position is called the root node [34]. The decision tree is very attractive in representing data, which depends on some attributes; moreover, a DTC (decision tree classifier) does not require any domain knowledge. Understanding the different steps in the decision tree are quite simple. DTC performs well on trained data sets [35]. An example of DTC with some trained data is provided in Figure 5.

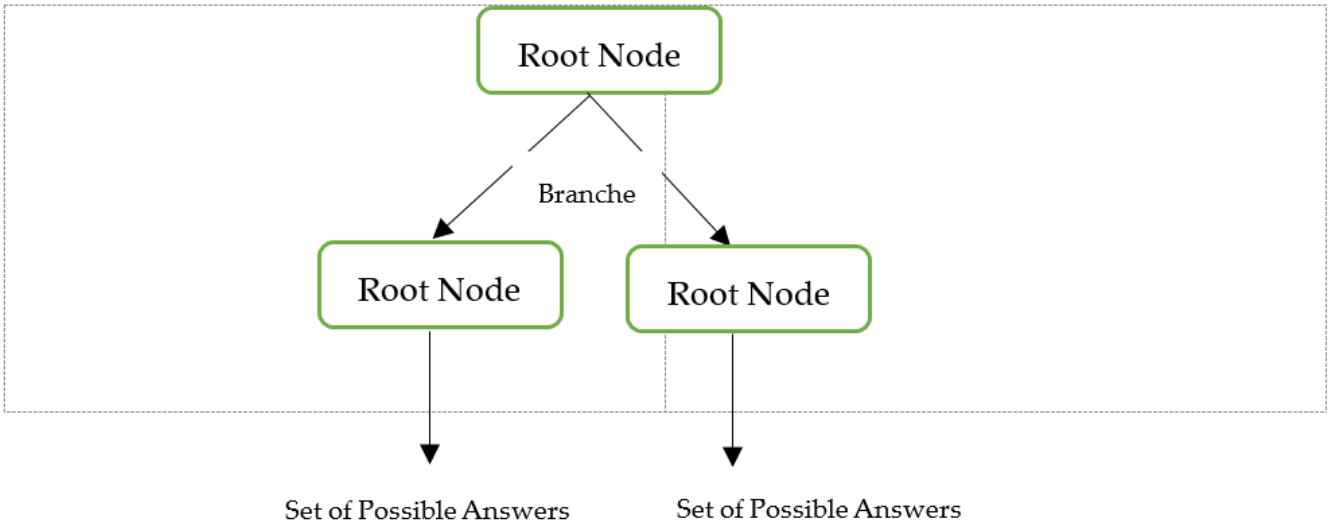

Figure 5. Decision tree generation example.

Linear Classification

Linear classification is one of the most important techniques used in machine learning and data mining. Compared with nonlinear classification techniques for some data in a rich dimensional space, this technique shows very high performance, including training and testing speed. A linear classifier can be characterized by a score, linear on weighted features, giving a prediction of outcome:

$$y = f(w, x) = f\ \left(\sum w_i, x_i\right) \quad (1)$$

Two approaches are used in the linear classification, which are:

a. Support vector machine;
b. Neural network.

Probabilistic Classification: Support Vector Machine

It is a type of machine learning algorithm used for handling mathematical and engineering problems, including object recognition, speaker identification, handwriting digit recognition, face detection in images, and target detection.

Take S of point x_i, such that:

$$x_i \in R_n \ where\ i = 1, 2, 3 \ldots\ldots n \quad (2)$$

Each point can take two values, and thus is given a label y_i, such that:

$$y_i a?? \{-1, 1\}$$

A support vector machine performs classification by creating an n-dimensional hyperplane that optimally divides the data into two categories. Consider the objects in the illustration on the left (see Figure 6). We can see that the objects belong to two different classes. The separating line (two-dimensional hyperplane) on the picture on the right is a decision plane that divides the objects into two subsets, such that in each subset, all elements are similar.

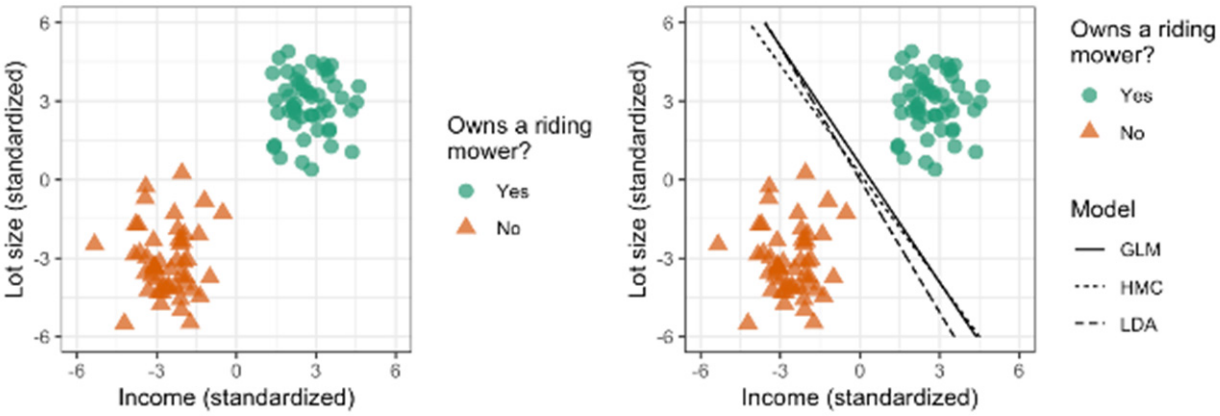

Figure 6. Support vector machine class separation.

SVM is a linear classifier, and all linear classifiers are based on the linear discriminant function of Equation (3) below:

$$f(x) = \mathrm{a}w, x\mathrm{a} + b \tag{3}$$

SVM is used as a linear divisor for different classes. SVM divides the data across a decision boundary determined by only a small subset of data. The data objects must have features x_1, x_2 . . . .. x_n and class label y_i. Some advantages of SVM are:

1. Support vector method performs relatively well because there is a clear gap of separation between classes.
2. It is more effective when used in high dimension space. Its performance becomes more effective in cases where several dimensions are more significant than the number of samples.
3. Memory is managed more efficiently in this technique.

Disadvantages of SVM are as follows:

1. SVM does not perform well when the data set is large.
2. Noisy data can also affect the performance of SVM.

Neural Network

In literature, neural networks and artificial neural networks (ANN) are two words that are used interchangeably. Artificial neural networks are advanced information-processing models that are pepped up after inspiration from the natural human nervous system. They process information like the human brain does. Currently, neural networks (NN) have shown record-setting performance in diverse areas including NLP. The neural network uses neurons, representing the vector over line X_i, which are word frequencies in the ith documents. The linear function of a neural network is p_i = A. X_i where A means the set of weights associated with each neuron. An experiment performed to check the performance of both SVM and NN on a 7-year data set found that NN outperforms SVM [36]. The research was performed to check the thermal infrared (2.5 to 14.0 µm) hyperspectral emissivity spectra for the classification of 13 different plant species. Currently, the most common traditional and deep learning based NN models used for sentiment analysis tasks are listed below:

- Autoencoder;

- Perceptron;
- Multilayer perceptron (MLP);
- Feed forward perceptron;
- Restricted Boltzmann machine;
- Convolutional neural network;
- Recurrent neural network;
- Long short-term memory (LSTM);
- Gated recurrent neural network.

5.5. Hybrid Approach

A hybrid system is obtained by the combination of both rule-based and machine learning systems. This system provides a way through which any specified model learns to detect sentiments from a series of tagged examples, then compares the results with a lexicon to improve accuracy.

6. Classification Techniques Used by Various Authors

Medhat et al. [37] in their research paper discussed the various classification techniques used by different scholars for the years 2010 to 2013. They focused their research on multilingual sentiments, including Chinese, Spanish, Dutch, Italian, and Japanese. It provided a broad overview of various sentiment algorithms and applications. Interest in languages other than English is growing, as there is still a lack of resources. According to the authors, WordNet is the most common technique used as a lexicon source for languages other than English. Dorle et al. [38] examined various English and Chinese language-classification techniques in their research paper. They discussed the various issues during the multilingual process, including the polarity shift problem, data sparsity, and binary classification. Routray et al. [39] discussed linguistic and statistical approaches, as well as machine learning approaches. Formulas for obtaining accuracy, recall, precision, F-measure, and finding relative errors were discussed in their paper. Hasan et al. [40] discussed the deep learning approach called the LSTM (Long Short-Term Memory) model for Roman Urdu sentiment analysis. The authors used naïve Bayes, SVM, and deep learning (LSTM) for calculating precision, recall, F1 score, and accuracy. The result showed that deep learning methods (LSTM) outperformed other classifiers. The accuracy obtained from the deep learning model was 95%, where precision was 97%, recall was 92%, and 94% was the F1 score. The total data set contained 300 sentiments of Roman Urdu. Alam and ul Hussain [12] in their research paper applied the deep learning LSTM classifier for finding the sentencing behavior in Roman Urdu. They collected a large amount of data, out of which 0.113 million lines of data of Roma Urdu were converted into Urdu text form. The creation of a one-to-one mapping dictionary between Roman Urdu to Urdu was another part of their research discussed in this research paper. Rafique et al. [6] used supervised machine, naïve Bayes, LRSGD (logistic regression with stochastic gradient descent), and SVM classification techniques. For this section, we first reviewed some of the most significant surveys and reviews proposed in the AutoML literature to better contextualize our research effort in relation to them. We then illustrated some remarkable AutoML applications in specific domains, highlighting commonalities and differences from the work presented here. Overview of selected studies on multilingual sentiment analysis is provided in Table 2.

Table 2. Overview of selected studies on multilingual sentiment analysis (Urdu and Roman Urdu).

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
1	Syed et al. [41]	Text preprocessing subjectivity classification Sentiment classification on Urdu text	Space insertion and space deletion Document-level classification Lexicon-based subjectivity analysis	450 user reviews of the movies	Average of 69% F-measure on all corpus	Extension of annotated lexicon Noun phrases consideration No mechanism for handling implicit negation
2	Bilal et al. [4]	Text preprocessing Subjectivity classification Sentiment classification Comparison of three different classifiers on Roman Urdu	Sentiment classification of Roman Urdu opinions Compression of naïve Bayesian, decision tree, and KNN classification techniques	300 sentiments, out of which 150 were positive and 150 were negative	97.33% accuracy shown by naïve Bayesian over other classifiers	Classifier checked on limited data sets Require checking on more processed and normalized large data sets
3	Ghulam et al. [5]	Application of long short-term memory model on selected Roman Urdu text	Deep learning technique LSTM for checking Roman Urdu sentences for polarity	300 sentences of Roman Urdu	95.18% accuracy with 97% precision Recall 92% and F1 score 94%, as compared to machine learning techniques, such as NB, RF, and SVM	Preprocessing techniques not mentioned Significantly less information regarding collection of data Methodology was not transparent in terms of data applied
4	Alam and ul Hussain [12]	Develop a deep learning classifier for Roman Urdu sentences using LSTM technique	Complete data were randomly shuffled and divided into a train (75%) and test set (25%) For seq2seq architecture, three-layer LSTMs were chosen, and maximum sequence length was set to 15	0.113 million Roman Urdu sentences	50.68 and 48.6% on evaluation metric BLEU	In future, more data sets to be collected for checking their accuracy in Roman Urdu

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
5	Rafique et al. [6]	Develop a classifier for Roman Urdu sentiments using different features using machine learning algorithms	Collection of Roman Urdu sentiments from various online sources Preprocessing Comparison of different classifiers over Roman Urdu text Construction of eight features, such as uni-gram, TF, IDF, bi-gram, etc. Application of three multilingual classifiers over eight features	806 different Roman Urdu sentiments were collected from other websites, out of which 406 were negative and 400 were positive	SVM gives the best results on the features uni-gram + bi-gram + TFID Accuracy = 87.22%	Data set was limited No proper arrangement of data set Use of deep learning for Roman Urdu included in their future plan
6	Sharf and Rahman [16]	Development of lexical normalization techniques for Roman Urdu Comparison of different lexical normalization techniques used for different languages	Developed an algorithm for Romanizing Urdu Lexicon-based approach Rule-based approach Machine learning and phonetic algorithms	Approximately 300,000 tweets of Roman Urdu collected from various websites	Lexical normalization results depended on the number of the tweet's success rate of the phonetic algorithm Biography shows 91% success rate	The initial stage work focused on the normalization of the Roman Urdu text Data collected only from some specific categories The central part of the data consisted of a paragraph Moving towards deep learning is a future plan
7	Sharf and Mansoor [17]	Roman Urdu discourse development Parsing and tagging Roman Urdu data sets	Calculation of precision, recall, and F1 scores over several sentences with discourse Calculation of success rate of false negative and false positive	15,000 sentences of Roman Urdu from Twitter, Reddit, Urdu poetry, and social worker biographies	The high value of the F1 score was obtained from the majority of data sets with discourse	Roman Urdu data collection is a challenging task The majority of sentences collected related to biographies but not general In the future, they plan to use neural network for the Roman Urdu data set to get more accuracy

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
8	Naqvi et al. [42]	Roman Urdu data collection Comparing nine popular classification methods over Roman Urdu sentiments Development of unique algorithms with the combination of the above classifiers	Calculation of precision, recall, and F score and accuracy of multipoint NB, SGD classifiers, and linear SVC for supervised learning, and k-nearest neighbor for unsupervised learning	22,000 sentiments of Roman Urdu collected, out of which 4500 were labeled positive, 4900 negative, and 13,000 sets as neutral	The best result was obtained from SVM with F1 score accuracy of up to 81% Worst result computed from a decision tree with F1 value of 38%	The combination of nine classifiers made ambiguity in the system More accurate results may be obtained using deep learning techniques with Roman Urdu
9	Sharjeel et al. [27]	Roman Urdu corpus-based study Development of corpus of Urdu news text reuse Linguistic analysis of corpus	Corpus of Urdu News text reuse (counter) using three annotators: wholly derived, partially derived, and non-derived	1200 documents collected from various news sites	Naïve Bayes on classification reported F1 results on COUNTER corpus for the binary and ternary classification using word grams to overlap others	Planning to use character n-grams Corpus will be evaluated on another state-of-the-art semantics for the Urdu language
10	Khan and Malik [13]	Collect customer reviews about automobiles in Roman Urdu Preprocessing Assign subjectivity	WEKA classifier on data sets Multinomial naïve Bayes performed best among other classifiers in terms of precision, recall, F-measure, and accuracy	2000 automobile reviews in Roman Urdu, out of which 1000 were positive and 1000 were negative	89.75% accuracy on multinomial naïve Bayes	Enhance the study on a specific area of the Internet, for example, the engine of an automobile Multicase sentiment classification required to count neutral sentiment
11	Dzakiyullah et al. [30]	Explore the unique features in Urdu text Collection of data Preprocessing Polarity identification	Compared the various algorithms based on already given values	Not mentioned	Only provided the review of previous work done on Urdu sentiments	Introductory level paper related to pure Urdu text Missing classifier Data set missing

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
12	Bose et al. [14]	Preprocessing Classification of unstructured data Customer product reviews of the six most popular products on Amazon	AFINN, ANEW, EMDEX, Lab Mat, and SentiWordNet with lexical classifier	568,454 number of sentiments related to fine food from Amazon web server collected from October 1999 to October 2012	Not mentioned	Data showed some graphs, but their style was confusing There was no proper table that showed the results
13	Naz et al. [31]	Part-of-speech tagging Urdu part-of-speech tagging Different approaches for part-of-speech tagging	Brill's transformation-based learning (TBL)	4323 Urdu sentences having 123,775 tokens from CRULP	Brill transformation approach outperforms uni-gram and bi-gram with a backoff method with 90% training fraction and keeping corpus size of 32,133 number of tokens 100% precision, 95% recall, and 94.435% F-measure obtained	Results depended on corpus fraction and also on the size of the corpus
14	Suri et al. [43]	Comparison of various approaches for multilingual sentiment analysis for Twitter database Apache Hadoop MapReduce	Comparison of various classifiers over Hadoop MapReduce	Twitter data set, 1000 Facebook posts, Amazon customer reviews on products	Results showed that Twitter is beneficial in the prediction of items, product sales, and quality of services, etc.	Each classifier showed some limitations concerning the data set available Preprocessing played a vital role in data accuracy
15	Raza et al. [44]	Urdu parsing Top-down and bottom-up parsing techniques Preprocessing Normalization	Phrase structure Parsing on Urdu text	2850 sentences	74.48%	Lack of using their own data sets Survey type paper No clear suggestion for improving the process of parsing
16	Abbas [45]	Urdu parser development Obtain the morphological rich, context-free grammar	URDU.KON-TB tree-bank, tree-bank parser	1400 annotated sentences	87% F-score, which outperformed exiting parsing work of Urdu on tree-banking approach	Very impressive working regarding generating Urdu parser using URDU.KON-TB tree-bank
17	Shah [46]	Collection of tweets Preprocessing Feature extraction Product review extraction	Support vector machine with KNN	2477 words and phrases from Twitter using search AI	Accuracy rate 93.33%	Other classifiers were required to check the data

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
18	Daud et al. [47]	Roman Urdu opinion mining system (ruomis)	Ruomis system consists of crawling of site, translating Roman Urdu reviews into English language, identifying the opinion polarity, and giving rating in graphical form	1620 comments Roman Urdu on three different mobile phones were collected and labeled as positive, negative, or neutral	Precision 27.1%, recall 100%, and F1 score 42.7%	Although results were not very satisfactory, their efforts for translating Roman Urdu text into English format were appreciable More work required in preprocessing and authors are interested in creating semantic dictionary in future
19	Mehmood et al. [48]	A precisely extreme multichannel hybrid approach for Roman Urdu sentiment analysis	Hybrid approach by combining machine learning (SVM, LR, and NB) along with deep learning approaches (CNN and RNN)	Data set consisted of 3241 mobile-related Roman Urdu sentences collected through website WhatMobile.com	84% accuracy, 81.68% precision, 82.84% recall 82.21% F1	In future, their focus is to apply generated neural word embedding for other natural language processing tasks, such as cyberbullying detection and machine translation
20	Hasan et al. [40]	Opinion within opinion: Segmentation approach for Urdu sentiment analysis	SEGMODEL (segmentation model)	Two data sets: D1 = 443 product reviews of cars and cosmetic products D2 = 401 product reviews of an electronic device	75% accuracy obtained from an average of both D1 and D2	
21	Akhter et al. [28]	To explore the offensive language in Urdu and Roman Urdu script Data set preparation Feature selection in classification models	Uni-gram, bi-gram, and tri-gram Combination of Urdu and Roman Urdu sentences Machine learning models of Bayes, tree, rule-based, KNN	2171 Urdu sentences and 10,000 Roman Urdu sentences		Comparison of 17 models from seven machine learning techniques to process and detect offensive language

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
22	Khattak et al. [49]	Finding all the related work for resource-poor Urdu language in text processing, corpus, and sentiment lexicon Comparison and evaluation of different preprocessing techniques used for Urdu language	Qualitative and quantitative evaluation approach to compare the various articles in Urdu language	Researcher data sets used in various research papers	LSTM technique of deep learning outperformed the other approaches to get the best accuracy for Urdu sentiment analysis	Urdu idioms and proverbs need to be addressed correctly A comprehensive Urdu dictionary is required for obtaining the best results
23	Iqbal et al. [26]	Data collection of Roman Urdu tweets Data preprocessing Translation and labeling of Urdu tweets Normalization and tokenization of tweets	Lexicon-based approach on sentence-level classification	A total of 2673 positive, 1923 negative, and 426 neutral tweets were collected	In the approach presented here, the accuracy of 98% for positive class, 94% for negative class, and 96% for a neutral class was achieved with a data set containing 4177 tweets	Authors aim to build their own lexicon, making it more generic and applying the proposed approach to a larger data set for further validation
24	Shakeel Ahmad [50]	Develop the Internet of Things-based classification model used for Urdu sentiment analysis Presented Urdu tweet crawling framework used for crawling Urdu tweets from Twitter in real time	Acquiring Urdu tweets in real-time using Twython Cleaning Urdu tweets using tweet preprocessor API Sentiment classification of Urdu text using IBM Watson's IoT-based natural language understanding toolkit	10126 number of Urdu tweets using Twython (https://twython.readthedocs.io/en/lates/, accessed on 15 August 2021)	Proposed Urdu tone analysis system IoT-based showed up to 92%, which outperformed all other techniques, such as SVM, KNN, RF, and naïve Bayesian classifiers	In future, emojis and slang terms will be included, which may enhance the system's overall performance
25	Mehmood et al. [22]	Developing a large Roman Urdu corpus Identifying (a) word-level features, (b) character-level features, and (c) feature union (word + character level). Applying classification algorithms	Proposed a confidence-based voting technique for Roman Urdu sentiment analysis called Word Voting or wVoting, a machine learning algorithm	11,000 reviews in Roman Urdu collected from six different domains	wVoting outperformed all other techniques, such as LR, NB, and ANN, and wVoting had an accuracy rate of 82.16%	Roman Urdu sentiment analysis is a challenging task because of the variety of writing styles Research should be continued to obtain more fruitful results using more normalized words

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
26	Mehmood et al. [25]	Collection of Roma Urdu corpus Preprocessing Feature selection Appling novel type technique for analysis of Roman Urdu Comparison of various results with various features	wVoting, which used discriminative feature spamming (DSFT) for Roman Urdu sentiment	11,000 reviews in Roman Urdu collected from six different domains	wVoting with discriminative feature spamming (DSFT) outperformed all other techniques, such as LR, NB, and ANN, and wVoting had accuracy rate of 83.95%	
27	Mehmood et al. [23]	Data Preprocessing Uni-gram and bi-gram feature selection application of five algorithms for feature reduction Computing of accuracy by applying the five algorithms	Naïve Bayes, logistic regression, support vector machine (SVM), k-nearest neighbors (KNN), decision tree (DT)	Data set consisted of 779 reviews, out of which 412 were positive and 367 were negative	Results showed high accuracy of naïve Bayesian classifier with average of 67.58% In second linear regression, showed average accuracy of 66.17%	
28	Manzoor et al. [24]	Preprocessing Normalization Creation of lexical variation of Roman Urdu table for normalization of Roman Urdu words	Self-attention biLSTM deep learning technique	Collected 20,000 sentences, out of which 10,000 Roman Urdu sentences (negative and positive reviews) were processed, and more than 3000 sentences were normalized	Accuracy on preprocessed data was obtained = 68.4% Accuracy on normalized data was obtained = 69.3%	They tried to enhance the efficiency of SA biLSTM and bring it to use for language inference and generation tasks and vocabulary
29	Hashim and Khan [29]	Sentence-level sentiment analysis using Urdu nouns Sentence boundary identification Urdu corpus creation and data cleaning Part-of-speech tagging	Sentence-level lexicon-based classifications	1000 Urdu opinions were collected from various online sources related to politics, current affairs, sport, health, and entertainment	86.8% accuracy was counted from the given Urdu text using sentence-level, lexicon-based classifier	This research focused only on subjective nouns of news articles In the future, their focus is to analyze products and movie reviews

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
30	Akhtar et al. [7]	Develop a hybrid approach to use for resource-poor languages, such as Roman Urdu, Hindi, and Urdu	Hybrid that combined both the features of machine learning and deep learning (SVM + CNN)	More than 13,000 views collected from various online sites and Twitter Hindi	Results showed accuracy up to 74% in some cases when used with proposed approach	It worked only for aspect-based sentiment classification In future, they will work on aspect term extraction
31	Chhajro et al. [8]	Classification of Roman Urdu and Urdu news text using natural language processing and machine learning classification model	KNN, linear SVM logistic regression, naïve Bayes classifier, and random forest classifier	Collection of 10,500 new texts in Urdu and Roman Urdu language	Linear SVM outperformed other classifiers with accuracy of 96.1%	In future, they will extend their work to other categories of Roman Urdu and Urdu news data
32	Khan et al. [51]	Comparison made on machine learning and deep learning approaches for part-of-speech tagging in Urdu using language-independent features set	CRF (conditional random field) model, SVM (support vector machine), DRNN (deep recurrent neural network), HMM (hidden Markov model) DRNN model also included LSTM-RNN and LSTM-RNN with CRF output	CLE (Centre of Language Engineering) data set and Bushra Jawaid data set	In CLE data set CRF model gave more accurate result than other models and the accuracy of this model was 83.52. In Bushra Jawaid Data set LSTM-RNN has more accurate result and its accuracy was 88.7	Their planning is to propose such a sophisticated deep learning based Urdu POS tagging system which will base on semi-supervised learning
33	Latiffi et al. [52]	Enhance the ontology-based approach Combination of ontology based on formal concept analysis (FCA)	Combination of ontology based on formal concept analysis (FCA), a process of obtaining a formal ontology or a concept hierarchy from a group of objects with their properties and k-nearest neighbor (KNN) to classify the reviews	Technique applied on English language data sets; however, it can also be used for other languages	Authors believe that this technique is the most suitable for all type of sentiment analysis	They expect to refine and investigate for implied sentiment classification

Table 2. *Cont.*

Study No	Study	Objective(s)	Techniques Utilized	Data Set/Source	Results	Future Work and Limitations
34	Mukhtar et al. [53]	Developed a lexicon-based model for analysis of Urdu sentiments using adjectives, nouns, and negations, as well as verbs, intensifiers, and context-dependent words	Made new Urdu lexicon and applied this approach to find accuracy, F-measure, precision, and recall	A total of 6025 sentences were collected from 151 various Urdu blogs	89.03% accuracy with 86% precision, 90% recall, and 88% F-measure obtained from the model	They will extend the lexicon by adding more words and also considering adverbs

In this model, of more than 806 sentiments of Roman Urdu, 400 were positive and 406 were negative, and they used WEKA for their classifiers. According to the results obtained, support vector machine outperformed when using uni-gram + bi-gram + TF-IDF as a feature set. The accuracy obtained was 82.22%. Sharf and Mansoor [17] presented research in which nine popular classification methods were compared and checked using five classification algorithms. The six algorithms included support vector machine (SVM), random forest classification (RFC), decision tree, regression, perceptron, and k-nearest neighbor. Data algorithms checked 9400 Roman Urdu sentiments, out of which 4500 were positive and 4900 were negative. The algorithm showed an accuracy level of up to 74%. Sharjeel et al. [27] in their research introduced COUNTER (corpus of Urdu news text reuse), which has become a standard benchmark for Urdu reuse text. Twelve thousand (12,000) Urdu documents were used as the corpus. A naïve Bayes classifier was used in the WEKA environment with three annotators at the document level with three rewrite classes: wholly derived, partially derived, and non-derived. Results showed that the GST (greedy string tiling) method with a minimum match length of 1 (mMl) was most effective in text reuse detection in a given corpus. Khan and Malik [13] applied supervised machine learning classifiers on 1000 positive and 1000 negative Roman Urdu reviews related to automobiles using the WEKA environment. A multinomial naïve Bayes classifier showed the best results in accuracy, precision, recall, and accuracy. Syed et al. [41] presented research on Urdu language by creating a sentiment-annotated lexicon to include information about the subjectivity of a word/phrase and its orthographic, phonological, syntactic, and morphological aspects. The classification accuracy for SentiUnits was 75%. Naz et al. [31] presented Brill's transformation-based learning (TBL) approach for resolving Urdu language problems. Uni-gram and bi-gram models were used to tag the data initially. The corpus size was 123,775 tokens with an accuracy of 84%. The method automatically deduced rules from a training corpus with accuracy comparable to other statistical techniques. Soni et al. [54] in their research used an unsupervised lexicon-based method on the SentiWord platform. The SentiWord model contained two methods: the first one was SWN (AAC) or SentiWordNet (adverb + adjective combination) and the second one was SWN (AAAVC) or SentiWordNet (adverb + adjective and adverb + verb combination). Abbas [45] in his research presented an Urdu language parser called URDU.KON-TB tree-bank. Dynamic programming algorithms, which are called early algorithms, were extended to accomplish Urdu parsing needs. Through this extension, many problems relating to Urdu parsing were solved. Many papers have been published in the last few years on Urdu, Roman Urdu, and other local languages of Indo-Pak to explore the sentiments in these languages to analyze people's opinions to inform decision making.

7. Conclusions

Due to the wide spread of the Internet all over the world and people's huge number of responses about various online products and events, it becomes an obligatory need of organizations to consider online sentiments given in any language and process these sentiments for decision making and improving the quality and standard of the products. This paper's primary goal was to concentrate on various sophisticated techniques used for text mining, preprocessing data, feature extraction, lexical resources, and classification techniques used for Urdu and Roman Urdu. This survey investigated the Roman Urdu and Urdu sentiment problems and discussed preprocessing, feature extraction, lexical resources, and sentiment classifiers in detail. We compared the various research work done in Roman Urdu and Urdu in the fields of preprocessing, feature extraction, lexicon, parsing, and classification. Multiple classifiers were discussed with their accuracy, precision, recall, and F–measure on various types of data sets. The data set collected was discussed with the classifier and classifier environment. We discussed the limitations of all previous work done with suggestions. Approximately all the research discussed above on Roman Urdu and Urdu used either a lexicon-based approach or machine learning to find the sentence's polarity. However, due to the lack of a proper corpus and fixed dictionary, it is necessary

to use the hybrid technique, including the features of both lexicon and supervised and unsupervised approaches of machine learning. The majority of the previous work of checking the Roman Urdu and Urdu sentences was performed in a lexicon-based approach. It is recommended that further work investigates a system using a combination of machine and deep learning techniques to obtain more accurate results. Similarly, Roman Urdu users' emotions relating to product reviews require comparative research by applying new machine learning and deep learning techniques.

8. Future Work

We aim to develop a unique classifier for detecting sentiments in product reviews in Roman Urdu and pure Urdu, then extend the work to local languages, such as Sarakai and Punjabi. There is little difference between these languages' used words.

9. Human and Animal Rights

This study did not involve any experimental research on humans or animals; hence, an ethics committee's approval was not applicable in this regard. Data collected from online forums are publicly available data, and no personally identifiable information of the forum users was collected or used for the study.

Author Contributions: Conceptualization I.U.K. and A.K.; Methodology, W.K. and I.U.K. formal analysis, I.U.K., A.K. and M.M.A.; investigation, M.M.S.; writing—original draft preparation, I.U.K. writing—review and editing, W.K., F.S. and M.Z.A.; visualization, F.S.; supervision, A.K.; project administration. M.M.A.; funding acquisition, N/A. All authors have read and agreed to the published version of the manuscript.

Funding: No funding was used for this research work.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Balahur, A.; Turchi, M. Multilingual sentiment analysis using machine translation. In Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, Jeju, Korea, 12 July 2012; pp. 52–60.
2. Denecke, K. Using sentiwordnet for multilingual sentiment analysis. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering Workshop, Cancun, Mexico, 7–12 April 2008; pp. 507–512.
3. Dashtipour, K.; Poria, S.; Hussain, A.; Cambria, E.; Hawalah, A.Y.; Gelbukh, A.; Zhou, Q. Multilingual sentiment analysis: State of the art and independent comparison of techniques. *Cogn. Comput.* **2016**, *8*, 757–771. [CrossRef] [PubMed]
4. Bilal, M.; Israr, H.; Shahid, M.; Khan, A. Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *J. King Saud Univ. Comput. Inf. Sci.* **2016**, *28*, 330–344. [CrossRef]
5. Ghulam, H.; Zeng, F.; Li, W.; Xiao, Y. Deep learning-based sentiment analysis for roman urdu text. *Procedia Comput. Sci.* **2019**, *147*, 131–135. [CrossRef]
6. Rafique, A.; Malik, M.K.; Nawaz, Z.; Bukhari, F.; Jalbani, A.H. Sentiment analysis for roman urdu. *Mehran Univ. Res. J. Eng. Technol.* **2019**, *38*, 463–470. [CrossRef]
7. Akhtar, M.S.; Kumar, A.; Ekbal, A.; Bhattacharyya, P. A hybrid deep learning architecture for sentiment analysis. In Proceedings of the COLING 2016 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 482–493.
8. Chhajro, M.; Khuhro, M.; Kumar, K.; Wagan, A.; Umrani, A.; Laghari, A. Multi-text classification of Urdu/Roman using machine learning and natural language preprocessing techniques. *Indian J. Sci. Technol.* **2020**, *13*, 1890–1900. [CrossRef]
9. Nazir, S.; Nawaz, M.; Adnan, A.; Shahzad, S.; Asadi, S. Big data features, applications, and analytics in cardiology—A systematic literature review. *IEEE Access.* **2019**, *7*, 143742–143771. [CrossRef]
10. Nazir, S.; Shahzad, S.; Mukhtar, N. Software birthmark design and estimation: A systematic literature review. *Arab. J. Sci. Eng.* **2019**, *44*, 3905–3927. [CrossRef]
11. Keele, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Version 2.3, EBSE Technical Report, Keele University and Durham University Joint Report; EBSE: Keele, UK, 2007; pp. 1–57.

12. Alam, M.; Hussain, S. Sequence to sequence networks for Roman-Urdu to Urdu transliteration. In Proceedings of the 2017 International Multi-Topic Conference (INMIC), Lahore, Pakistan, 24–26 November 2017; pp. 1–7.
13. Khan, M.; Malik, K. Sentiment classification of customer's reviews about automobiles in roman urdu. In Proceedings of the Future of Information and Communication Conference, Cham, Switzerland, 5–6 April 2018; Springer: Berlin/Heidelberg, Germany; pp. 630–640.
14. Bose, R.; Aithal, P.; Roy, S. Sentiment Analysis on the Basis of Tweeter Comments of Application of Drugs by Customary Language Toolkit and TextBlob Opinions of Distinct Countries. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *8*, 3684–3696.
15. Khan, K.; Khan, W.; Rehman, A.; Khan, A.; Khan, A. Urdu sentiment analysis. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 646–651. [CrossRef]
16. Sharf, Z.; Rahman, S.U. Lexical normalization of roman Urdu text. *Int. J. Comput. Sci.* **2017**, *17*, 213–221.
17. Sharf, Z.; Mansoor, H.A. Opinion mining in roman urdu using baseline classifiers. *Int. J. Comput. Sci.* **2018**, *18*, 156–164.
18. Posadas-Durán, J.-P.; Markov, I.; Gómez-Adorno, H.; Sidorov, G.; Batyrshin, I.; Gelbukh, A.; Pichardo-Lagunas, O. Syntactic n-grams as features for the author profiling task. In Proceedings of the CEUR Workshop, 2015 Working Notes Papers of the CLEF, Toulouse, France, 8–11 September 2015.
19. Chikersal, P.; Poria, S.; Cambria, E. SeNTU: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 647–651.
20. Rajagopal, D.; Cambria, E.; Olsher, D.; Kwok, K. A graph-based approach to commonsense concept extraction and semantic similarity detection. In Proceedings of the 22nd International Conference on World Wide Web Companion, Rio de Janeiro, Brazil, 13–17 May 2013; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2013; pp. 565–570.
21. Ravi, K.; Ravi, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowl. Based Syst.* **2015**, *89*, 14–46. [CrossRef]
22. Mehmood, K.; Essam, D.; Shafi, K.; Malik, M.K. Sentiment analysis for a resource poor language—Roman Urdu. *ACM Trans. Asian Low-Resour. Lang. Inf. Process (TALLIP)* **2019**, *19*, 54. [CrossRef]
23. Mehmood, K.; Essam, D.; Shafi, K.; Malik, M.K. Discriminative feature spamming technique for roman urdu sentiment analysis. *IEEE Access.* **2019**, *7*, 47991–48002. [CrossRef]
24. Manzoor, M.A.; Mamoon, S.; Tao, S.K.; Zakir, A.; Adil, M.; Lu, J. Lexical Variation and Sentiment Analysis of Roman Urdu Sentences with Deep Neural Networks. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 719–726. [CrossRef]
25. Mehmood, K.; Essam, D.; Shafi, K. Sentiment analysis system for Roman Urdu. In *Science and Information Conference*; Springer: Cham, Switzerland, 2018; pp. 29–42.
26. Iqbal, F.; Ayoub, A.; Manzoor, J.; Basit, R.H. Bilingual Sentiment Analysis of Tweets Using Lexicon. In Proceedings of the 7th International Conference on Language and Technology, UET, Lahore, Pakistan, 15–16 February 2020; pp. 71–78.
27. Sharjeel, M.; Nawab, R.M.A.; Rayson, P. COUNTER: Corpus of Urdu news text reuse. *Lang. Resour. Eval.* **2017**, *51*, 777–803. [CrossRef]
28. Akhter, M.P.; Jiangbin, Z.; Naqvi, I.R.; Abdelmajeed, M.; Sadiq, M.T. Automatic detection of offensive language for urdu and roman urdu. *IEEE Access.* **2020**, *8*, 91213–91226. [CrossRef]
29. Hashim, F.; Khan, M. Sentence level sentiment analysis using urdu nouns. In Proceedings of the 6th International Conference Conference on Language &Technology 2016; UET, Lahore, Paksitan, 17–18 November 2016; pp. 101–108.
30. Dzakiyullah, N.R.; Hussin, B.; Saleh, C.; Handani, A.M. Comparison neural network and support vector machine for production quantity prediction. *Adv. Sci. Lett.* **2014**, *20*, 2129–2133. [CrossRef]
31. Naz, F.; Anwar, W.; Bajwa, U.I.; Munir, E.U. Urdu part of speech tagging using transformation based error driven learning. *World Appl. Sci. J.* **2012**, *16*, 437–448.
32. Qiu, G.; Liu, B.; Bu, J.; Chen, C. Opinion word expansion and target extraction through double propagation. *Comput. Linguist.* **2011**, *37*, 9–27. [CrossRef]
33. Altınel, B.; Ganiz, M.C. Semantic text classification: A survey of past and recent advances. *Inf. Process. Manag.* **2018**, *54*, 1129–1153. [CrossRef]
34. Sharma, H.; Kumar, S. A survey on decision tree algorithms of classification in data mining. *Int. J. Sci. Res.* **2016**, *5*, 2094–2097.
35. Yang, H.; Fong, S. Optimized very fast decision tree with balanced classification accuracy and compact tree size. In Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications, Vienna, Austria, 29–31 August 2014; pp. 57–64.
36. El-Masri, M.; Altrabsheh, N.; Mansour, H. Successes and challenges of Arabic sentiment analysis research: A literature review. *Soc. Netw. Anal. Min.* **2017**, *7*, 1–22. [CrossRef]
37. Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [CrossRef]
38. Dorle, S.; Pise, N.N. Sentiment Analysis Methods and Approach: Survey. *Int. J. Innov. Comput. Sci. Eng.* **2017**, *4*, 7–11.
39. Routray, P.; Swain, C.K.; Mishra, S.P. A survey on sentiment analysis. *Int. J. Comput. Appl.* **2013**, *76*, 1–8. [CrossRef]
40. Hasan, M.; Ullah, S.; Khan, M.J.; Khurshid, K. Comparative Analysis of SVM, ANN and CNN For Classifying Vegetation Species Using Hyperspectral Thermal Infrared Data. *Remote. Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 1861–1868. [CrossRef]

41. Syed, A.Z.; Aslam, M.; Martinez-Enriquez, A.M. Lexicon based sentiment analysis of Urdu text using SentiUnits. In *Mexican International Conference on Artificial Intelligence, Pachuca, Mexico, 8–13 November 2010*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 32–43.
42. Naqvi, U.; Majid, A.; Abbas, S.A. UTSA: Urdu Text Sentiment Analysis Using Deep Learning Methods. *IEEE Access* **2021**, *9*, 114085–114094. [CrossRef]
43. Suri, N.; Verma, T. Multilingual Sentimental Analysis on Twitter Dataset: A Review. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *10*, 2789–2799.
44. Raza, A.A.; Habib, A.; Ashraf, J.; Javed, M. A review on Urdu language parsing. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 93–97.
45. Abbas, Q. Morphologically rich Urdu grammar parsing using Earley algorithm. *Nat. Lang. Eng.* **2016**, *22*, 775–810. [CrossRef]
46. Shah, M.S. A research paper on product review based on geographic location using SVM approach in twitter. *Int. Educ. Res. J.* **2017**, *3*, 690–692.
47. Daud, M.; Khan, R.; Daud, A. Roman Urdu opinion mining system (RUOMiS). *arXiv* **2015**, arXiv:1501.01386. [CrossRef]
48. Mehmood, F.; Ghani, M.U.; Ibrahim, M.A.; Shahzadi, R.; Mahmood, W.; Asim, M.N. A precisely xtreme-multi channel hybrid approach for roman urdu sentiment analysis. *IEEE Access* **2020**, *8*, 192740–192759. [CrossRef]
49. Khattak, A.; Asghar, M.Z.; Saeed, A.; Hameed, I.A.; Hassan, S.A.; Ahmad, S. A survey on sentiment analysis in Urdu: A resource-poor language. *Egypt. Inform. J.* **2021**, *22*, 53–74. [CrossRef]
50. Shakeel Ahmad, Y.D.A.-O. Applying IoT for Sentiment Classification and Tone Analysis of Urdu Tweets. *Int. J. Comput. Sci. Netw. Secur.* **2019**, *19*, 166–173.
51. Khan, W.; Daud, A.; Khan, K.; Nasir, J.A.; Basheri, M.; Aljohani, N.; Alotaibi, F.S. Part of speech tagging in urdu: Comparison of machine and deep learning approaches. *IEEE Access* **2019**, *7*, 38918–38936. [CrossRef]
52. Latiffi, M.I.A.; Yaakub, M.R. Sentiment analysis: An enhancement of ontological-based using hybrid machine learning techniques. *Asian J. Inf. Technol.* **2018**, *7*, 61–69. [CrossRef]
53. Mukhtar, N.; Khan, M.A. Effective lexicon-based approach for Urdu sentiment analysis. *Artif. Intell. Rev.* **2020**, *53*, 2521–2548. [CrossRef]
54. Soni, V.; Patel, M.R. Unsupervised opinion mining from text reviews using SentiWordNet. *Int. J. Comput. Trends Technol.* **2014**, *11*, 234–238. [CrossRef]