

Semantic Analysis of Urdu English Tweets Empowered by Machine Learning

Nadia Tabassum¹, Tahir Alyas², Muhammad Hamid^{3,*}, Muhammad Saleem⁴, Saadia Malik⁵,
Zain Ali² and Umer Farooq²

¹Department of Computer Science, Virtual University of Pakistan, Lahore, 54000, Pakistan

²Department of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan

³Department of Statistics and Computer science, University of veterinary and animal sciences, Lahore, 54000, Pakistan

⁴Department of Industrial Engineering, Faculty of Engineering, Rabigh, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

⁵Department of Information Systems, Faculty of Computing and Information Technology - Rabigh, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

*Corresponding Author: Muhammad Hamid. Email: muhammad.hamid@uvas.edu.pk

Received: 28 March 2021; Accepted: 29 April 2021

Abstract: Development in the field of opinion mining and sentiment analysis has been rapid and aims to explore views or texts on various social media sites through machine-learning techniques with the sentiment, subjectivity analysis and calculations of polarity. Sentiment analysis is a natural language processing strategy used to decide if the information is positive, negative, or neutral and it is frequently performed on literature information to help organizations screen brand, item sentiment in client input, and comprehend client needs. In this paper, two strategies for sentiment analysis is proposed for word embedding and a bag of words on Urdu and English tweets. Word embedding is a notable arrangement of procedures that can remember words linguistics dependent on the spread theory which expresses that word is utilized and happens within the same settings tend to indicate comparable implications. Bag of words is an approach used in natural language processing to retrieve information and features from written documents. For the bag of words, machine learning techniques like naive bayes, decision tree, k-nearest neighbor, and support vector machine is used to enhance the accuracy. For word embedding the neural network technique is proposed by the combination of recurrent neural network (RNN) with long-short term memory (LSTM) for sentimental analysis of tweets. Datasets of Urdu and English tweets are used for negative and positive classification tweets with machine learning techniques. The contribution of this paper involves the implementation of a hybrid approach that focused on a sentiment analyzer to overcome social network challenges and also provided the comparative analysis of different machine learning algorithms. The results indicate improvement while using the combination of RNN with the help of LSTM showed accuracy 87% on the Urdu dataset and 92% on the English dataset.

Keywords: Short term memory; natural language processing; tweets; support vector machine; word embedding



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Word embedding is a technique in natural language processing for text where words have similar content and similar meaning related to content are merged to vectors of numbers. They are normally created from an enormous text *corpus*. The inserting of a word catches two of its syntactic and semantic angles. Tweets are short, loud, and have novel lexical and semantic highlights that are unique concerning different sorts of text. Hence, it is important to have word embeddings adapted explicitly from tweets. They help to learn calculations accomplish better execution in NLP related undertakings by gathering comparable words together and have been utilized in heaps of NLP applications, for example, election classification, sentiment classification and tweet classification [1].

Bag-of words and bag-of-n-grams is a technique used in NLP for information extraction and finding the distance between words. For example words “Speed”, “run” and “food” are nearly far away from each other but speed closer to run and food semantically. In light of word embeddings, speed and run are near to each other. In this field, the word embedding portrays a model that is processed using a neural network and created from a huge body without management. The scholarly vectors unequivocally encode numerous semantics examples and a significant number of these examples can be expressed as direct interpretations. For instance, the outcome of a vector estimation “Islamabad” near to “Pakistan” and “Canberra” is nearer to “Australia” [2].

Tweets are boisterous, little, and have various characteristics from different kinds of content. Due to the benefits of applying word embeddings in NLP, and the individuality of tweet text there is a need to have word embeddings gained explicitly from tweets. Various tweets cover different points, and junk data is common on tweeter. Furthermore few tweets are semasiologically incomprehensible or unreasonable contain the only irreverence and pivot on regular gossips or commercials. We consider both junk and such tweets with no generous content as junk. Vector models based on these sorts of tweets will carry commotion to certain applications. To fabricate vector models without utilizing junk tweets, we utilize a junk channel to eliminate junk tweets.

In any case, few applications may require embeddings created from all tweets together with daily tweets and junk tweets. To fabricate precise installing models, a multi-word term is used to include expressions. An installing model for just words is a lot more modest than that for the two words and expressions, and it will be more productive for certain applications that just need embeddings for words. In this examination, we utilize an information-driven technique to deal with distinguishing expressions. To oblige different applications we produced four installing sets utilizing just tweet data, which are the four blends with junk tweets and expressions. Some applications require both text data and also expressions so that’s why we require word embedding that educates on both tweeter text data and random expression data. We construct four embedding informational indexes that were found out from the preparation information that consolidate tweeter and random data. These four embedding indexes are from tweeter data sets and random expression data [3]. In recent years, social-networking sites such as Facebook, Twitter, and Instagram have attracted a large number of users. Many people use social media to share their thoughts, views, and opinions about things, places, and people. Machine-learning, Lexicon-based, and hybrid methods of sentiment analysis are the most popular. Another classification has been suggested, which includes statistical, knowledge-based, and hybrid methods. There is a place for computationally interpreting opinions and emotions to conduct demanding research in broad areas. As a result, a gradual practice of extracting information from data available on social media has developed.

We presented our structure, which explains the process of collecting, analyzing, and categorizing Twitter opinions. We looked at tweets in the form of hashtags that users used to share their thoughts on current political events. After that, we Pre-processed the dataset by storing the retrieved tweets in a folder, translating Urdu tweets, and storing them in the database.

Twitter allows users to write tweets in a number of languages, including Urdu, English and other languages. Urdu tweets are considered more common in Pakistan than English tweets. It is a reality that people prefer to express their opinions and feelings in their native language. Mostly, sentiment analyzers are created for other languages and have handling of their related script form, morphological, and grammatical variations, English is ineffective for Urdu tweets for local natives. As a result of grammatical distinctions, Urdu should be learned as a separate language. The advantages of automated opinion detection in a native language (in this case, Urdu tweets) will help you do a lot of semantic analysis. As shown in Fig. 1, the semantic model can be divided into tweets like reviews dataset, training dataset and classification model for positive and negative class.

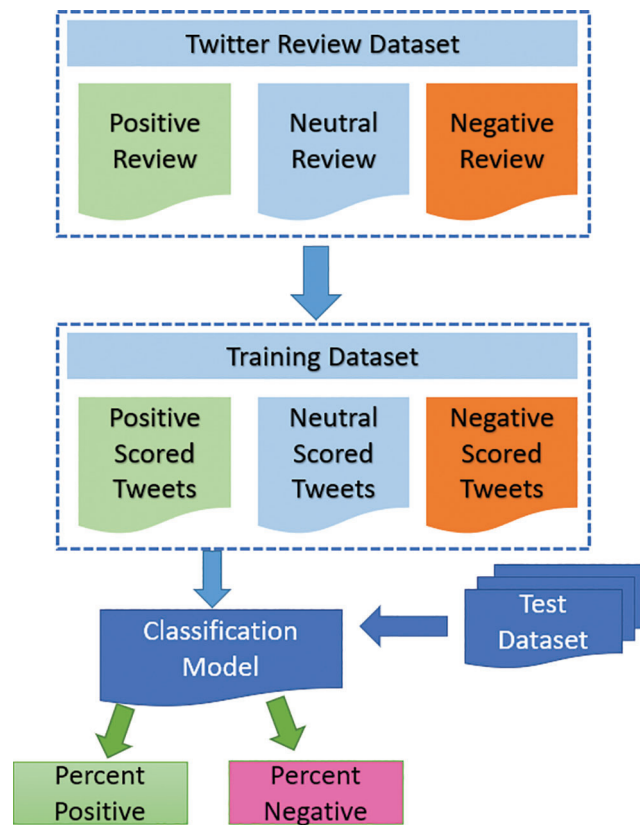


Figure 1: Tweet classification model

2 Literature Review

López et al. [4] investigates the utilization of word embedding in an unused situation to form a Vector Space Model (VSM) for web domain names. Pal et al. [5] portrays a framework that abuses information representation strategies and languages to clarify significant commerce exercises, developments of items inside the assembling organization to flexibly esteem added organizations. All the more especially he analyzes the depiction rationale formalism, which is used for information portrayal in a computable piece of Rationale related issues. Barros et al. [6] points to address such issues by supporting inactive investigation strategies for application program interfaces(APIs) that make strides in their measured quality. His methodology gets an item arranged perspective where the idea of the article is exemplified by the possibility of business substance.

Diac et al. [7] uses rules close by JSON to show the accompanying degree of linguistics, essentially dependent on the twofold association between boundaries of organizations. Organizations identify with an open cosmology to depict their convenience. Parallel relations can be shown between information or yield boundaries inside the advantages definition. The metaphysics incorporates a few relations and induction rules that offer assistance to infer unused relations between parameters of administrations. Abid et al. [8] introduced the IDECSE cutting edge approach for composite administrations designing. By considering linguistics web organizations, IDECSE addresses the test of totally automating the characterization, disclosure, and arrangement though decreasing progression time and incurred significant damage.

Barati et al. [9] focuses on a particular linguistic web information quality issue that shows inaccurate task among occasions and classes in metaphysics. Also, the paper tells the best way to find new classes which are not characterized in the metaphysics and how to put them in the various leveled structure of the cosmology. Calache et al. [10] presents visual documentation to speak the principal components of web services description language determination pivot on the semantic comment utilizing linguistic annotations for web services description language. He additionally portrays a graphical community linguistic explanation uphold device. Dantas et al. [11] presents restful linguistics web administration disclosure design dependent on linguistics interfaces with comments that linguistically depicts restful web administrations. This engineering empowers programming specialists to consequently find and settle on administration decisions to execute a decided undertaking. Maryam Amiri [12] shows a technique for the web administration determination among usefulness comparable web administrations dependent on web-service Policy linguistic coordination.

Lima et al. [13] sets up a computerized choice back framework competent for collecting, coordinating, and sharing tuberculosis wellbeing information within the Brazilian Bound together open wellbeing framework. It observes tainted patients and represents the solidified data of standard tuberculosis and this technique is safe for experts. Kumara et al. [14] proposes a comparability technique that learns space setting by AI to deliver models of setting for terms recovered from the web. To dissect outwardly the impact of area setting on the grouping results, the bunching approach applies a circular related space calculation. Eghan et al. [15] introduces a displaying strategy that exploits the linguistic web and its deduction administrations to catch and set up discernibility joins between information removed from various assets.

Natarajan et al. [16] initiates another web administration where the clients are given the alternative to choose the administration that fulfills the nature of administration specification from the arrangement of found administrations. To upgrade web administrations he offers a bunching based semantic assistance choice model and client special model. Shang et al. [17] proposes a semantic layered coordinating calculation. To confirm the attainability and viability of the various leveled coordinating calculation dependent on linguistics, a model framework named guard stage milestone was planned.

Ruminski et al. [18] presents a design for large-scale dispersed Augmented Reality administrations and an exploratory assessment of the foremost part of linguistics increased reality middleware. The design depends on a customer worker plan, which reinforces linguistics demonstrating and sculpturing significantly augmented reality introductions for a huge number of clients. Petrovic et al. [19] present the linguistics design-driven technique to the arrangement and adaptively of holder based applications in haze Computing. Demonstrating apparatuses, linguistics framework, direct improvement illustrate, diversion climate and establishment organization code generator utilizing the semantic remarks are shown. Marco [20] proposes a far-reaching approach that gives an unaided strategy to comment on autonomous tables, perhaps without header columns or other outside data. The methodology depends on the meaning of a setting made from the components inside the table to segregate among coordinating substances found in Graphs and make high-caliber comments.

Hammal et al. [21] proposes a conventional technique for their regularity checking by first deciphering synthesized administration into imparting mechanism utilizing an iterative cycle driven by control structures of these dialects. Lampropoulos et al. [22] present a work that how augmented reality base work can be improved by coordination of deep learning, linguistic web, and informative graphs. By the combination of these features can give creating contemporary, easy to understand, and client-focused clever applications. Nataliaia [23] introduces an information-based web administration model that incorporates the innovations of the semantic web to consider the information, their handling, and surmising.

3 Proposed Methodology

Our system consists of five steps which are Dataset Loading, Pre-Processing, Feature Extraction, Classification, and Evaluation. All steps working and details are shown in our proposed model which we can see in Fig. 2.

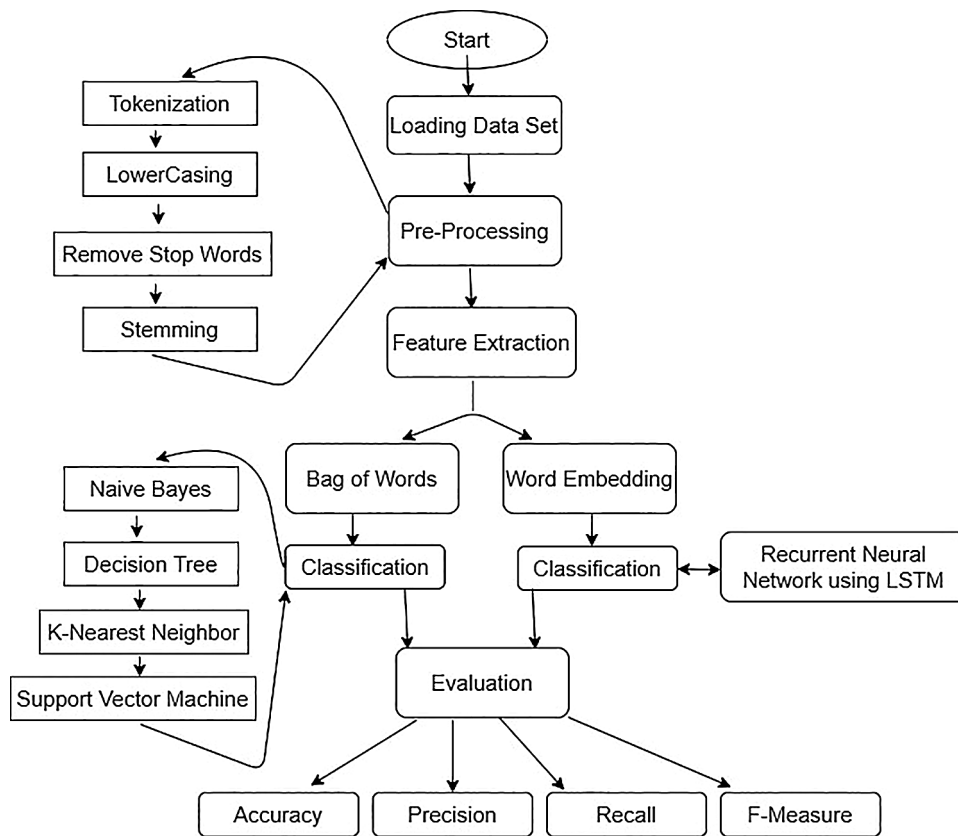


Figure 2: Proposed hybrid model for tweets

3.1 Dataset Loading

We used Urdu and English twitter dataset and the dataset consist of two columns which are tweets and labels including the positive and negative label. Documents from the data set were divided into 80-20, 80% of the data from each document used for training, and the remaining 20% used for testing.

3.2 Preprocessing

The Twitter analysis is purely based on the author's writing style. It is also observed in the writing survey that it is not feasible to clean the data set by removing special characters nor to correct the grammatical error and their preference of word suffixes and prefixes, later capitalization all provide important information about the author. So, removing or correcting such things will reduce the number of features related to the particular author.

3.2.1 Tokenization

It divides sentences or a paragraph into little units like words or characters by ignoring white spaces. And store each token with their frequency of occurrence in the dataset with the help of the Natural Language Toolkit (NLTK).

3.2.2 Remove Stop Words

In the English language “a, are, an, is, the, this” are the stop words. Which have a high recurrence in the content of language but stop words have negligible lexical information. So, we prefer to expel these words from the dataset before doing any further processing. But on account of the Urdu language, we don't have such a stop words list. So, we include limitations that the selected word should not appear in every document. We discount all words occurring in documents by more than 70%.

3.2.3 Lowercasing

All upper case text is converted into the lower case before doing any further processing. As in the Urdu language, there is only one case so no need to apply lowercasing.

3.2.4 Stemming

The process of separating the root word from the given words is called a stem word. Rule-based stemmer was utilized with the assistance of NLP tools to stem dataset words.

3.3 Feature Extraction

It is a process of extracting numerical information from the documents and that information is called a feature. For training, a model only best-fit features are selected for better results.

3.3.1 Bag of Words

Bag of words is a classic model and implemented for machine learning where the text is considered as a set of words having a repetition of body phenomena. On the other hand, the characteristic of a document is represented as the repetition of text that shows to make a lexicon, and this lexicon comprises character n-grams, word n-grams, and other different features that are taken out from the text. If we utilize all words of lexicon then it can expand the body length which is hard for enumeration. For feature selection, we used a term-document recurrence strategy.

3.3.2 Word Embedding

Word embedding is a technique in natural language processing for text and executed for the neural network where words have similar content and similar meaning related to content are merged to vectors of numbers.

3.4 Machine Learning Classification

For the classification, we applied two methodologies first one is machine learning and the second one is a neural network. And also we portray the algorithms which we use in both methodologies. There are four algorithms which are Naïve Bayes, Support Vector Machine, Decision Tree, and KNN (K-Nearest Neighbor) using in the machine learning approach. All of these are discussed below.

3.4.1 Naive Bayes

Naïve Bayes is a managed machine learning model utilized to figure the back likelihood of a category. The goal is to foresee future articles once a bunch of articles is given to each category.

3.4.2 Support Vector Machine

Support vector machine (SVM) is a supervised machine learning model. SVM is utilized to arrange direct and non-direct information. It utilizes non-direct planning to change over the preparation information into a higher measurement.

3.4.3 Decision Tree

One of the main widely use AI algorithms is a decision tree. The popularity behind the decision tree is that it can regulate any kind of data. The working of this algorithm is that it isolates the data into more modest components and gains the design for classification and then results display in the form of a tree that is very useful to understand the phenomena.

3.4.4 K-Nearest Neighbor

K-nearest neighbor (KNN) is a non-parametric machine learning algorithm that doles out an unclassified typical point to the grouping of the nearest set of recently characterized points. Text documents are shown as vectors where frequency phenomena are represented by attributes in each document. Vector used to discover the similarity between documents. We apply the KNN set of rules to our data that will discover ways to classify new documents based on their distance to our recognized documents.

3.5 Deep Neural Network Classification

We use a recurrent neural network in a deep neural network approach which is described below. RNN is a strong neural network and it is not the same as other networks. The main factor which differentiates it from the others is that it has a restricted memory where the input also relies upon the past input as well. This working makes RNN a unique network for successive information where requests take granted and it is also affected for sentiment analysis where words change their logical definition. RNN also shows the best performance with the help of a hidden layer that recalls previous information for the grouping of data. Long Short-Term Memory (LSTM) is the upgraded version of RNN in which the performance of memory is essentially broadened. In our proposed work we find the results of RNN with the help of LSTM. Accordingly, it is appropriate to have the capacity of gaining information from significant encounters that are exceptionally long-lasting slacks in the middle. In LSTM units utilized as assembling bars for RNN are indicate as LSTM network. LSTM empowers RNN memory for the input for a long time. Due to this working LSTM acquire data and perform fundamental task like perusing, composing, fetching and erasing in storage.

Fig. 3 shows an LSTM network that has one input that takes work embedding information data than three hidden layers that have relu function work like an initiation capacity and one output layer that has sigmoid function work like actuation work show characterized data into positive and negative classes. We found that with the help of three hidden layers that are the most important portion show the most elevated accuracy.

3.6 Evaluation

A confusion matrix is utilized to assess the model through accuracy, precision, recall, and f-measure. Accuracy is defined in Eq. (1), precision in Eq. (2), recall in Eq. (3), and f-measure in Eq. (4).

$$Accuracy = \frac{\text{Number of Correct Classified Tweets}}{\text{Total Number of Tweets}} \times 100 \quad (1)$$

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

$$F - \text{Measure} = \frac{(2 * \text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad (4)$$

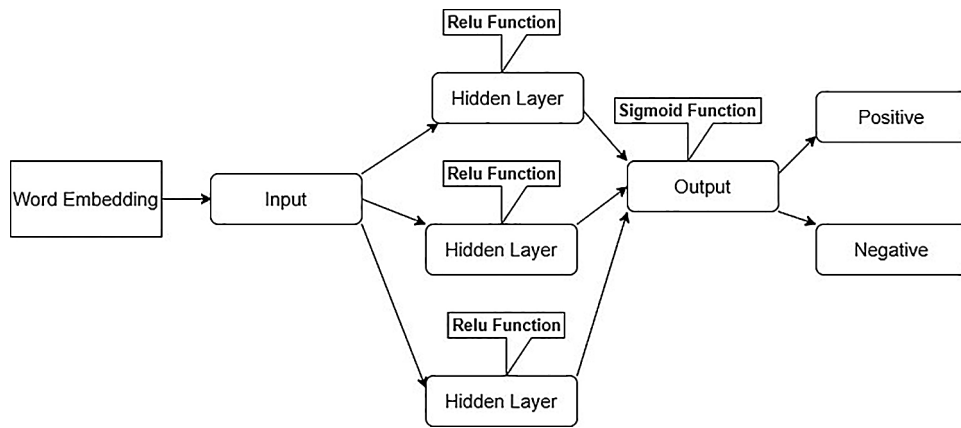


Figure 3: Proposed LSTM network for words

4 Results and Discussion

4.1 Datasets

Data set of Urdu and English tweets consist of two columns which are tweets and labels including the positive and negative label. Documents from the data set were divided into 80-20, 80% of the data from each document used for training, and the remaining 20% used for testing. The tweet datasets is a collection of 8 authors of which 4 for Urdu and 4 for English tweets. All of them were selected randomly from most recent 180 tweets, against each selected users from March 2020.

4.2 Experimental Setup

All experiments performed in Intel core i5@2.50Ghz working on windows 10 (64 bit) with 8 GB memory to test the performance and found the accuracy. Python 3.7 is used for machine learning approaches and Keras used for RNN-LSTM. We used confusion metrics such as recall, precision, and F1 measure through accuracy to illustrate aspects of a self-choice based on classifiers on [Tabs. 1](#) and [2](#) datasets.

4.3 English Dataset Results

[Tab. 1](#) shows the correlation between various calculations utilizing a few techniques for the English dataset.

[Fig. 4](#) shows algorithms accuracy of the machine learning and neural network in the form of percentages for the English dataset. Four machine learning algorithms that are naive bayes, decision tree, k-nearest neighbor, and support vector machine and one neural network that is RNN with the help of LSTM. Our results for the English dataset show that RNN-LSTM accomplished the largest accuracy that is 92% and

Naïve Bayes has the least accuracy that is 62%. The decision tree and support vector machine showed equal and comparable accuracy that is 86% and 87%.

Table 1: Accuracy for the english dataset

Methods	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	62%	78%	59%	62%
Decision Tree	86%	86%	86%	86%
K-Nearest Neighbor	84%	84%	84%	84%
Support Vector Machine	87%	84%	82%	78%
RNN-LSTM	92%	82%	81%	81%

Table 2: Accuracy for the urdu dataset

Methods	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	68%	72%	70%	70%
Decision Tree	81%	79%	77%	77%
K-Nearest Neighbor	80%	70%	68%	67%
Support Vector Machine	82%	79%	77%	77%
RNN-LSTM	87%	82%	80%	80%

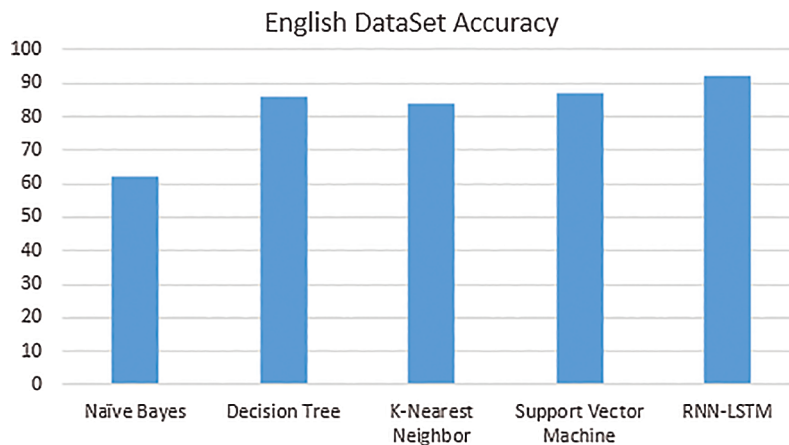


Figure 4: English dataset performance on different algorithms

4.4 Urdu Dataset Results

Tab. 2 showed the correlation between various calculations utilizing a few techniques for the Urdu dataset.

In Fig. 5 showed algorithms accuracy of the machine learning and neural network in the form of percentages for the Urdu dataset. Four machine learning algorithms that are naive bayes, decision tree, k-nearest neighbor, and support vector machine and one neural network that is RNN with the help of LSTM. Our results for the Urdu dataset show that RNN-LSTM accomplished the largest accuracy that is

87% and Naïve Bayes has the least accuracy that is 68%. The decision tree and support vector machine showed equal and comparable accuracy that is 81% and 82%.

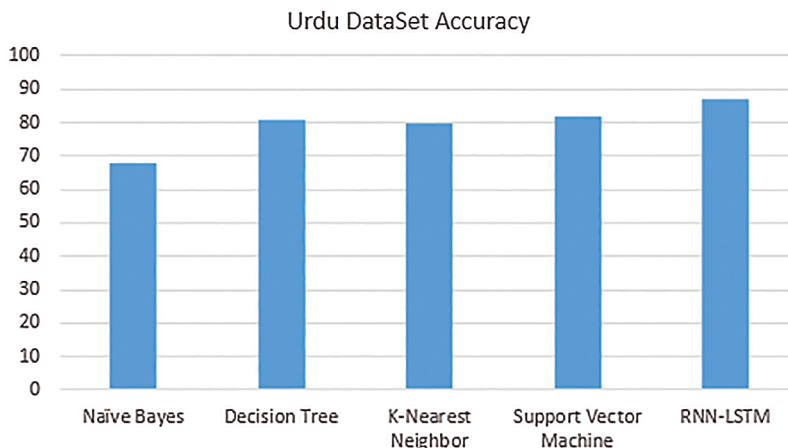


Figure 5: Urdu dataset performance on different algorithms

5 Comparative Study

In [Tab. 3](#) showed the accuracy performance for both the English and Urdu dataset using machine learning algorithms and recurrent neural networks with the help of long short term memory.

Table 3: Accuracy comparison for the dataset

Methods	English DataSet Accuracy	Urdu DataSet Accuracy
Naïve Bayes	62%	68%
Decision Tree	86%	81%
K-Nearest Neighbor	84%	80%
Support Vector Machine	87%	82%
RNN-LSTM	92%	87%

We attained an overall 92% accuracy on the English dataset and 87% accuracy on the Urdu dataset. In [Tab. 3](#), we can see clearly that accuracy for the English dataset is better than the Urdu dataset. But naïve bayes gives better accuracy 68% for the Urdu dataset and the English dataset is 62%. This could be a good reason that accuracy differs because of different datasets, in terms of size and content. But in this study, we achieved very good accuracy at a very small dataset. Hence, we could state that the accuracy could be increased if we increase the size of our dataset.

6 Conclusion

This paper focuses on the adoption of different sentiment analyzers with machine-learning algorithms to assess the solution to learning about urdu & English tweets with the best accuracy rate. In a sentiment analysis based on a lexicon, semantic orientation is based on words, phrases or sentences measured in a text. The main motive of this paper is the analysis of twitter sentiment by applying RNN-LSTM and different machine learning algorithms that are naïve bayes, decision tree, k-nearest neighbor, and support

vector machine. Our finding demonstrated that in all of the algorithms, RNN-LSTM produces better results and shows the highest accuracy 87% for the Urdu twitter dataset and 92% for the English twitter dataset.

Acknowledgement: Thanks to our teachers and families who provided moral support.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding this study.

References

- [1] X. Yang, C. Macdonald and I. Ounis, "Using word embeddings in twitter election classification," *Information Retrieval Journal*, vol. 21, no. 2, pp. 183–207, 2018.
- [2] S. Xiong, H. Lv, W. Zhao and D. Ji, "Towards twitter sentiment classification by multi-level sentiment-enriched word embeddings," *Neurocomputing*, vol. 275, pp. 2459–2466, 2018.
- [3] R. Rashdi and S. Keefe, "Deep learning and word embeddings for tweet classification for crisis response," *ACM*, vol. 10, pp. 1–6, 2019.
- [4] W. López, J. Merlino and P. Bocca, "Learning semantic information from Internet domain names using word embeddings," *Engineering Application for Artificial Intelligence*, vol. 94, pp. 823–832, 2020.
- [5] K. Pal and A. Yasar, "Semantic approach to data integration for an internet of things supporting apparel supply chain management," *Procedia Computer Science Elsevier*, vol. 175, pp. 197–204, 2020.
- [6] A. Barros, C. Ouyang and F. Wei, "Static analysis for improved modularity of procedural web application programming interfaces," *IEEE Access*, vol. 8, pp. 128182–128199, 2020.
- [7] P. Diac, L. Țucăr and A. Netedu, "Relational model for parameter description in automatic semantic web service composition," *Procedia Computer Science Elsevier*, vol. 159, pp. 640–649, 2019.
- [8] A. Abid, M. Rouached and N. Messai, "Semantic web service composition using semantic similarity measures and formal concept analysis," *Multimedie Tools and Applications*, vol. 79, no. 9, pp. 6569–6597, 2020.
- [9] M. Barati, Q. Bai and Q. Liu, "Automated class correction and enrichment in the semantic web," *Journal of Web Semantics*, vol. 59, no. 3, pp. 533–544, 2019.
- [10] M. Calache and C. Farias, "Graphical and collaborative annotation support for semantic web services," in *Proc. IEEE Int. Conf. of Software Architecture Companion ICSA-C 2020 Germany*, pp. 210–217, 2020.
- [11] J. Dantas and P. Farias, "An architecture for restful web service discovery using semantic interfaces," *Journal of Web Semantics*, vol. 16, no. 1, pp. 1–24, 2020.
- [12] W. Matching, "A procedure for web service selection using ws-policy semantic matching," *Journal of Advanced Computer Engineering Technology*, vol. 6, no. 2, pp. 70–78, 2020.
- [13] V. Lima, F. Pellison, F. Bernardi, R. Rijo and D. Alves, "Proposal of an integrated decision support system for tuberculosis based on semantic web," *Elsevier*, vol. 164, pp. 552–558, 2019.
- [14] B. Kumara, I. Paik and Y. Yaguchi, "Context-aware web service clustering and visualization," *International Journal of Web Services*, vol. 17, no. 4, pp. 32–54, 2020.
- [15] E. Eghan, P. Moslehi, J. Rilling and B. Adams, "The missing link – a semantic web based approach for integrating screencasts with security advisories," *Information Software Technology*, vol. 117, pp. 234–240, 2020.
- [16] B. Natarajan, M. Obaidat, B. Sadoun, R. Manoharan, S. Ramachandran *et al.*, "New clustering-based semantic service selection and user preferential model," *IEEE System Journal*, vol. 23, pp. 1–9, 2020.
- [17] K. Shang, "Semantic-based service discovery in a grid environment," *Journal of Intelligent Fuzzy System*, vol. 39, no. 4, pp. 5263–5272, 2020.
- [18] D. Rumiński and K. Walczak, "Large-scale distributed semantic augmented reality services – a performance evaluation," *Graphical Models*, vol. 107, pp. 1–15, 2020.
- [19] N. Petrovic and M. Tomic, "SMADA-Fog: Semantic model driven approach to deployment and adaptivity in fog computing," *Simulation Modeling Practice and Theory*, vol. 101, no. 3, pp. 33–54, 2020.

- [20] M. Cremaschi, F. Paoli, A. Rula and B. Spahiu, "A fully automated approach to a complete semantic table interpretation," *Future Generation Computer Systems*, vol. 112, no. 1–2, pp. 478–500, 2020.
- [21] Y. Hammal, K. S. Mansour, A. Abdelli and L. Mokdad, "Formal techniques for consistency checking of orchestrations of semantic web services," *Journal of Computer Science*, vol. 44, pp. 165–175, 2020.
- [22] G. Lampropoulos, E. Keramopoulos and K. Diamantaras, "Enhancing the functionality of augmented reality using deep learning, semantic web and knowledge graphs: A review," *Informatics*, vol. 4, no. 1, pp. 32–42, 2020.
- [23] N. Kulykovska, S. Skrupsky and T. Diachuk, "A model of semantic web service in a distributed computer system," *Journal of Healthcare Engineering*, vol. 2608, pp. 338–351, 2020.