

## Urdu Sentiment Analysis Using Deep Attention-Based Technique

Naeem Ahmed<sup>1</sup>, Rashid Amin<sup>1,2\*</sup>, Huma Ayub<sup>3</sup>, Muhammad Munwar Iqbal<sup>1</sup>, Muhammad Saeed<sup>1</sup>, Mudassar Hussain<sup>4</sup>

<sup>1</sup>Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan

<sup>2</sup>Department of Computer Science, University of Chakwal, Chakwal, Pakistan

<sup>3</sup>Department of Software Engineering, University of Engineering and Technology, Taxila, Pakistan

<sup>4</sup>Department of Computer Science, University of Wah, Wah Cantt, Pakistan

\*Corresponding Author: Rashid Amin. Email: rashid.amin@uoc.edu.pk

### Abstract:

Sentiment analysis (SA) is a process that aims to classify text into positive, negative or neutral categories. It has recently gained the research community's attention because of the abundance of opinion data on the internet. Deep learning techniques are widely used for language processing but are seen as black boxes, and their effectiveness comes in interpretability. The major goal of this article is to create an Urdu SA model that can comprehend review semantics without the need of language resources. We design an attention-based neural network for the review level Urdu SA. For better results, we used a transfer learning approach that uses pre-trained embedding's. The Visualization of attention weights is also measured that uncovers the black box of the models and confirms their intuition, which aids in the interpretation of the model's learned representations. The proposed model is tested and evaluated in terms of accuracy and F1 score. The proposed model archives 91% accuracy and 88% F1 score, respectively.

**Keywords:** Deep learning; Urdu language; Natural language processing; Sentiment analysis

### 1. Introduction

Sentiment analysis (SA) tries to summarize attitudes about a particular subject, service, or entity by manipulating textual data of people or group of people. Sentiment classification categories text inputs into one of two categories (positive or negative) Yang, L, et al., (2020). Where SA is useful is the trend analysis, social net analysis, public opinion categorization, policy evaluation, and decision-making direction. Currently, various machine learning (ML) models i.e., support vector machine (SVM), Random forest Naïve bayes (NB), Decision tree etc. are being used for sentiment analysis of various languages. learn hierarchical data Deep learning techniques are also being used for SA of different languages with huge amounts of data Soleymani, M, et al., (2017). Neural network models can representations by incorporating several processing layers. Emotion detection (ED) and analysis Asghar, M, et al., (2017) is used to determine the writer's implicit emotions from text. Various tasks like speech, facial expressions, posture, gestures, and text have all been used to identify emotion. Different approaches for emotion identification have been studied, such as the keyword methodology, which focuses on keywords in the text and includes a parser and an emotion lexicon. In a hybrid technique, classification models are trained utilizing extensive linguistic knowledge from lexicons, thesauri, and keyword and learning approaches Almani, N.M., et al., (2020).

Sentiment detection is important in affective computing Cambria, E, et al., (2017) (AC), which is focused with developing computer systems that can recognize and respond to human affective states Khotima, D.A.K, et al., (2018). Affect sensitive systems are used in a variety of fields, including education, gaming, psychological health, and customer service.

Urdu is a widely spoken language globally, with over 100 million speakers. Many people have recently

started using Urdu in their tweets, reviews, and comments. As a result, sentiment analysis for the Urdu language has grown important Daud, A., et al., (2017). Several approaches and methods are available for text mining and sentiment analysis in English. However, sentiment analysis in other languages such as Urdu, Hindi, Arabic, and others has rarely been studied. The Urdu language's complex morphological structure and distinct script from English are two primary reasons for its lack of tools Mukhtar, N, et al., (2020).

Natural language processing (NLP) neural networks are often implemented for various tasks to improve accuracy. Each sentence is considered a collection of tokens, i.e., characters or words Khalid, K, et al., (2017). The ability of neural networks to extract multi-level representations of input text allows them to control local and long-range dependencies. In various field like voice recognition, computer vision, NLP, and other fields, neural networks abstract information from raw input. One-dimensional Convolutional Neural Networks are used in these networks to model time delayed sequential data.

Transfer learning is also used for solving various problems and allows applications for NLP, bioinformatics and computer vision Tan, C, et al., (2018). The main aim of transfer learning (TL) models is to build effective, reliable and accurate models in the target domain by using the data of source domin. Because of this technique we can minimize the training time and enhamce the accuracy of model easily. This technique is often used when we want high-performing models but have less data or when training data is expensive.

SA models are facing various challenges. These challenges include sarcasm detection, negations, compound phrases, repetition of words etc. Like some other languages, the Urdu language is widely used by individuals for data sharing on internet. It is clear from the literature study that techniques used in SA for other languages cannot be used to deal with Urdu language. Urdu sentiment analysis is becoming popular since few years because of its increasing rate on internet. The main objective of this study is to analyze and investigate the attention based deep learning technique for Urdu sentiment classification.

This study proposes a deep attention approach for generating Urdu review representations that do not rely on external resources like handmade features or sentiment lexicon. The study's first phase data is scraped from various blogging and social media websites. These websites include daily Pakistan, hamariweb.com, BBC Urdu, and many other Urdu blogs. Some part of the standard IMDB is translated into Urdu using google trans library. This data is preprocessed to tokenize and normalize the Urdu sentences. The reviews length was fixed to 500 words for effective model development. In order to generate Urdu review representations without the need of outside resources like hand-crafted features or sentiment lexicon. The study suggests a deep attention technique by using a transfer learning approach to generate successful models. In addition, a review-level sentiment analysis algorithm is designed to look at the polarity of reviews by choosing the most informative phrases that support a particular sentiment in any review. The models' "black box" is also revealed via visualizing attention weights, which supports users' intuition and facilitates the understanding of the model's learnt representations. The key contributions of the study is listed below:

- For sentiment analysis of Urdu, we suggest an attention-based deep learning model.
- Creating large Urdu data sets using various blogs data for Urdu SA.
- The use a deep attention approach for generating Urdu review representations that does not rely on external resources.
- To use the transfer learning technique for building effevtive Urdu SA model by selecting the most informative phrases that reflect a certain sentiment in a review.

Visualizing attention weights for understanding the most informative terms for uncovering the BlackBox of DL models. This study uses the transfer learning approach to develop effective models. To begin, the

acquired data set is preprocessed to normalize and tokenize it. Secondly, a review-level sentiment analysis algorithm is designed to examine the polarity of reviews by selecting the most informative phrases that contribute to a certain sentiment in any review. The experiments were carried out using a transfer learning method that included pre-trained embeddings. Visualizing attention weights also uncovers the black box of the models and confirms their intuition, which aids in the interpretation of the model's learned representations. The paper is further divided into following sections. Section 2 discusses the related work for the proposed problem. The problem statement and formulation are discussed in section 3 while section 4 discusses the insights about the data set used in the study. Section 5 presents the proposed methodology and the results are explained in section 6, while section 7 concludes the article.

## 2. Related Work

Formerly, machine-learning approaches were used to apply several algorithms to the issue of classification tasks of sentiments. Bibi, R, et al., (2019). proposed a supervised machine learning technique for Urdu SA. They used a decision tree algorithm for the Urdu SA. Their approach is composed of two main phases. In the first phase, data were preprocessed, and unnecessary data was removed. Stop words, hashtags, and other unnecessary words were removed in this step. In the next phase, feature vectors were generated. For this purpose, positive and negative comments and POS tags were identified. Finally, the decision tree algorithm is applied to the data for sentiment classification. The Urdu tweets were used in this study as a data set. The proposed decision tree model got 90% accuracy.

Mukhtar, N, et al., (2018) presents research that focuses on sentiment analysis on the Urdu language. They gathered data from various blogs and preprocess it. After preprocessing, supervised machine learning techniques, i.e., Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), were used for sentiment classification. After comparing the performance of these models, they cannot get satisfactory results. Then feature extraction was analyzed, and 152 features were extracted. Then, after classifiers' training, they got a 67% accuracy level by using the KNN algorithm. So, in this study KNN outperform SVM and decision tree in sentiment classification of the Urdu language.

Syed, A.Z, et al., (2014) present a sentiment analysis article using sentiUnits. This article uses the identification of SentiUnits from the data by using shallow parsing. SentiUnits are expressions that have information about the sentiment expressed in the sentence. Lexicon of Urdu language is made in this article and then used for sentiment classification. This paper highlights the linguistic, i.e., grammar and morphology of the Urdu language and technical aspects of this problem. The evaluation of the system is done with various test text data and got satisfactory results. This article can be used as a baseline for sentiment analysis.

Mukund, S, et al., (2012) present research on sentiment by using Urdu blog data. This research used structural correspondence learning to transfer Urdu sentiment data from Urdu news wire to Urdu blog data. As newswire data is written in Latin script and has code-mixing and code-switching behaviour, the data from these two platforms are not trivial. So, for making pivots, two oracles were used in this study. First was the Transliteration oracle, and the one was called translation oracle. The transliteration oracle was for script variation, while the translation was used for code-switching and codemixing behavior. They introduce a new part-of-speech tagging process that helps them identify words based on POS categories, representing code-mixing behavior. They evaluate their model against a supervised learning model and compare results based on various performance measures. After evaluation, they got a 59.4% precision value and 62.4% recall value for the proposed technique.

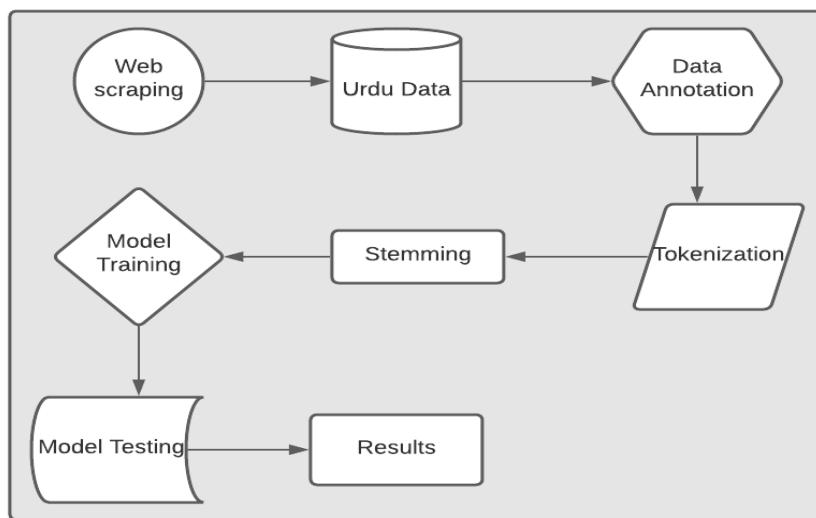
Mehmood et al. proposed an Urdu sentiment analysis system using RUSA data set Mehmood, K, et al.,

(2019). The data set contained 11,000 reviews of products. They presented three distinct techniques to achieve text normalization. RUSA dataset is used to establish the BL accuracies in the technique. The second and third research utilized six phonetic algorithms and TERUN to optimize the RUSA data set. The resultant data was used for the training of ML models. According to the empirical review, the results of TERUN were statistically significant and comparable to those obtained by phonetic algorithms. The TERUN word normalization technique was then generalized from a corpus-specific to a corpus-independent technique. The study concludes that text normalization enhances machine learning algorithms' accuracy rate. Another result was that a phonetic algorithm designed for one language will not generalize well to other languages unless it becomes properly updated to fulfill the phonological needs of its target languages.

Nasim, Z, et al., (2020) present an Urdu SA article combining various linguistic and lexical features. Their work focuses on building of Urdu SA system for Urdu tweets. A Markov chain model was used to design the approach in this paper. The help of Twitter API gathered the data set. The proposed model was trained on that data and model was able to predict people's attitudes based on their tweets. They also discussed about the challenges and limitations of Urdu SA systems. Their proposed model accurately predicted positive emotions because of less positive tweets in the data set.

Naqvi, U, et al., (2021) discussed many potential approaches available for sentiment analysis, but little work has been done on analyzing Urdu sentiments. This paper discusses the increasing rate of Urdu language on internet and need of Urdu SA systems. Their article outline and summarises the most recent SA updates and classification techniques used in the Urdu language. Various improvements were suggested in this article for Urdu SA. Figure 1 shows some proposed techniques for sentiment analysis of various languages. Masood et al. Masood, M, et al., (2022) used deep learning techniques to classify Urdu sentiments using a custom data set. The data set used in this study is 3995 Urdu sentences having 3 sentiment categories. They used LSTM model with 830 stem Urdu words after the preprocessing phase. After preprocessing the padding is performed for equalizing the length of vectors for training deep learning model. The results of this study shows that the proposed LSTM model achieved 86% and 89% accuracy and F1-score respectively.

A word level translation framework was proposed by Asghar, M, et al., (2019) to enhance the Urdu SA



*Figure 1:Techniques used for sentiment analysis*

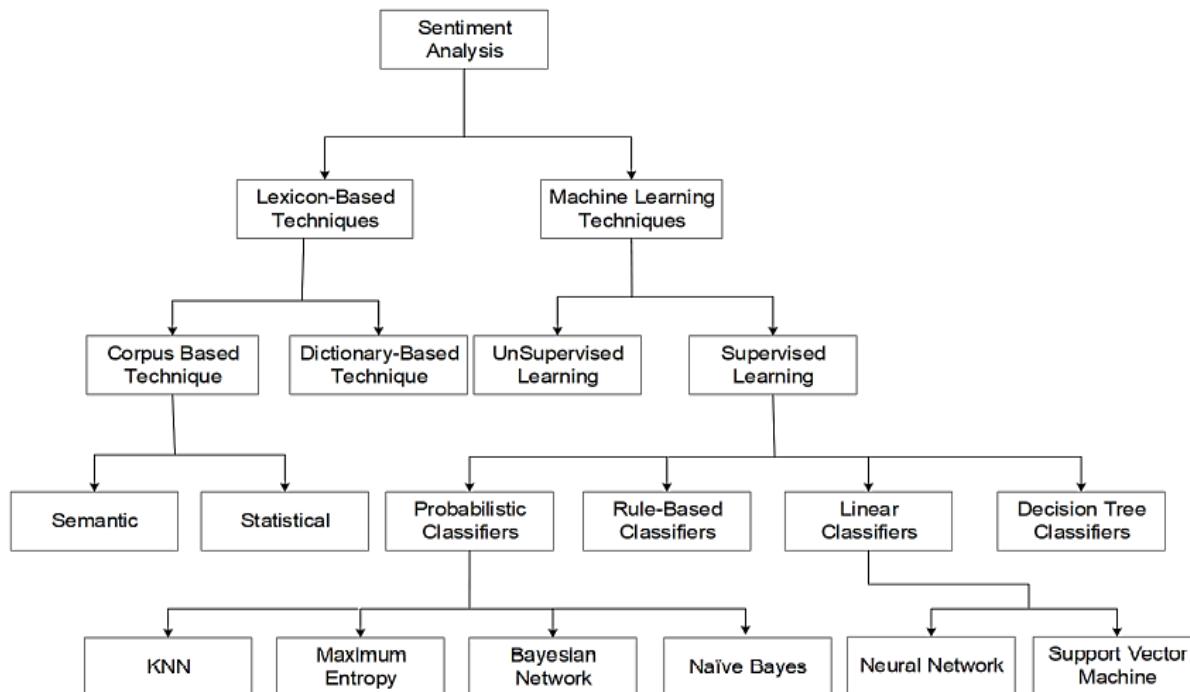


Figure 2: Process of sentiment analysis

lexicon. The framework was developed by combining different linguistic and lexicon resources, such as the English word list, SentiWord Net, the bilingual English-to-Urdu dictionary, Urdu grammar improvements, and a novel scoring mechanism. Their model consisted on three major modules, i.e., the collection of Words in English for an opinion, the translation of English words into Urdu, and sentiment scoring using SentiWordNet and manual scoring.

Hashim, F, et al., (2016) proposed a sentence-level Urdu SA method using nouns. They used Urdu news data for their lexicon-based method for Urdu SA. Urdu nouns are used for the detection of sentiments in sentences. Their proposed technique got 86.8 % accuracy and testing data set.

Naqvi, U, et al., (2021) Proposed deep learning techniques with different word vectors for Urdu sentiment analysis. They used various DL models, i.e., LSTM, CNN, BiLSTM and attention-based BiLSTM, to classify Urdu test sentiment analysis. In the sequential models, they used stacked layers and for CNN they used various filters with a single convolution layer. This study also analyzed the role of pretrained embeddings on an emotion classification problem. The attention-based Bi-LSTM model outperforms other models by achieving 77% accuracy and 72% F1-score in this study.

### 3. Problem Statement and Formulation

The proposed process of finding the polarity of piece of text can be explained as: Suppose, a review  $R_i = (W_1 + W_2 + W_3 + \dots + W_n)$ , which is made up of words "w" that are padded to a range of "n" using padding tokens. Without employing any constructed features, the suggested method uses a binary classification system to determine the category of any review.

The suggested problem is a deep attention-based neural model that can identify significant and insignificant terms in any review. The model takes some distributed vector representations as input for terms, model will give the entire review's distributed vector representation, which will be divided into positive or negative

classes using a linear classifier.

The initial stage in any NLP system should be text preparation and normalization Liu, H, et al., (2018). This phase removes noise from the data, including repeated and non-Urdu text such as URLs. The normalization procedure also transforms similar words into the same form. After that, all words are tokenized by the help of tokenizer. The step result is an integer sequence representing the input terms, with the complete data set encoded as integer values. In addition, all reviews are fixed to the same size by using the padding technique for shorter reviews.

#### 4. Dataset

There are not as many standard data sets for the Urdu language. A very few data sets are available for the concerned language, but they are very small. So the data is gathered from various blogging and social media websites. These websites include daily Pakistan, hamariweb.com, BBC Urdu, and many other Urdu blogs. Some part of the standard IMDB Kumar, H, et al., (2019) reviews data set is also translated into Urdu for building a larger data set for the training and evaluation of the proposed models. The data collection from these websites is done using the "Beautiful soup" Patel, J, et al., (2020) python library and stored in a .csv file. The final data set contains more than 25000 rows and two columns. The classification of the data set is shown in Figure 3.

Twenty percent of the data set is used for the testing phase, 10% is used for validation, while 70% is used for training the proposed models. The size of every comment is set to 500 words, as this was determined to be the average length of comments in data set. To decrease the length reviews shorter than the mean value were padded, and reviews bigger were shortened.



Figure 3: Distribution of dataset

#### 5. Proposed Scheme

The model's primary objective is to investigate how attention mechanisms are employed to identify the most meaningful terms contributing to the polarity of overall Urdu reviews. The proposed approach attempts to address the limits of global sentence representation by focusing on and emphasizing smaller data blocks, in this case review terms. The proposed process for Urdu sentiment analysis is illustrated in Figure 4.

The proposed model is consisting on various layers. In start, the model allows distributed representation by employing an embedding layer. This layer passes input distributed representation of words to a GRU-based layer Zhang, Z, et al., (2018) which creates reviews' hidden representations. Warping the entire sentence

into one single vector is not feasible. We are trying to represent excessive information in a restricted space and RNN layers cannot maintain the dependencies for more than a few time steps. As a result, to build a short-term memory, the attention layer is placed on top of the GRU layer, which retains the most important details from the sequence. The proposed model flow is illustrated in Figure 4.

The attention layer adds the GRU hidden representations for every word depending on the specified weights for the final vector representation. The proposed model's attention layer builds a distributed vector representation by using the essential terms of the data item. The weight is calculated at every hidden state at each time stamp is represented by " $h_{it}$ ". " $T$ " is the number of time steps in the input, while " $w_i$ " is the hidden representation. The attention layer's weights, " $i_w$ " and " $b_w$ ", are tuned during training to assign higher weights to a phrase's most essential words.

This phenomenon is illustrated in (1):

$$W_{it} = \text{Tanh}(I_w * h_{it} + b_w) \quad (1)$$

The updated hidden representation for the current word "Wit" is then passed to the Softmax method, which returns the normalized important weight  $\alpha_{it}$ .

$$\alpha_{it} = \frac{\exp(W_{it})}{\sum_{i=1}^j \exp(W_{jt})} \quad (2)$$

In the end, this representation is constructed as a weighted sum of the word annotations depending on the weights. It may be considered a high-level illustration of the informative words used in any review.

$$F_{it} = \sum_{i=1}^j (\alpha_{it} * W_{it}) \quad (3)$$

For the classification of reviews, the resultant vector representation "F" is transferred to a fully connected layer having a sigmoid activation function.

### 5.1 Model Details and Experiments

The transfer-learning technique is used at the embedding layer level in this study. The proposed model's embedding layer is weighted using a pre-trained model's final layer Naqvi, U, et al., (2021). As a result, the model inherits several common word patterns from the pre-trained model, and all it requires to discover additional relationships among words to categorize reviews. Various experiments were designed and assessed to see how effective pre-trained models were at solving the problem of Urdu SA. The baseline and the proposed model are trained and evaluated for the word embedding initialization with various test cases.

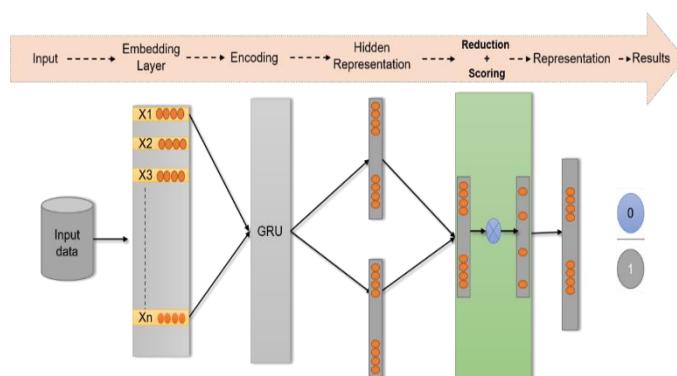


Figure 4: Proposed model for Urdu SA

The GRU layer and the final review representation on the GRU unit's dropout are set to 0.25, whereas between layers, it is set to 0.35. Table 1 shows the values of parameters for the proposed model.

*Table 1: Proposed model parameters*

Parameter	Values
Epochs	30
Batch Size	32
Loss Function	Cross_entropy
Optimization Function	Adam
Learning Rate	0.001
Output layer function	Sigmoid
Train test ratio	70%-30%

The first scenario is to use the weights of the retained model to apply transfer learning to the current model. No pre-trained models were employed in the second scenario. Glorot and Bengio Glorot, X, et al., (2011) uniform initialization were used to set the weights of the embedding layer. The experiments are carried out on a Dell PC with a 3.4GHz processor and 16GB of RAM. The suggested model is implemented using Python 3.8.8 by the help of the Keras package.

## 6. Results and Discussion

The proposed model output vector and attention weights vectors are presented for examining the black box of deep learning model's output. This was done to find the saliency characteristics contributing to the final classification. This is a missing feature in all sentiment classifiers for Urdu language. The approach utilized to assess reviews is divided into two categories. First, test the model's capability to identify the polarity of the entire review. The model's potential to determine the most informative terms in the given text is examined in the second stage.

Word context test was done on the test data to see whether the model could find the right class for the same term in different contexts or not. As shown in Table 2, for the reviews tested during this phase, the word "معلوماتی" might have a positive or negative sentiment.

After analyzing the proposed model, we identified that "فلم" is the most silent word in the review presented in Table 2. However, without considering the scores vectors that are given by the scoring function of the attention layer because we cannot ensure that the final representations are chosen by the most meaningful terms that determine the outcome of model. As shown in Table 2 one of the least informative terms in the review in the SAL scoring vector is "معلوماتی". The output vector of SAL for the proposed model also have the same important rank, but if we analyze the SAL score presented in Table 2, it can be seen that words that come before and after "اور" have almost the same scoring range. The existence of "اور" leads to the conclusion that the model may build a language's linguistic structure by predicting that the two terms have the same meaning.

The both models were trained for 30 epochs and tested for the evaluation purposes. The attention based proposed model got the accuracy of 91% while the baseline model achieves 86% accuracy. The models are

Table 2: SAL score

Original Label (Positive)	Predicted Label (Positive 0.78)
	The movie was interesting and informative.
SAL Output	<p style="text-align: center;">فلم دلچسپ اور معلوماتی تھی</p> <p>0.0141235 0.000087 0.00064 0.000043 0.000305</p>
SAL Score	<p style="text-align: center;">فلم دلچسپ اور معلوماتی تھی</p> <p>0.012235 0.0009243 0.012504 0.0007983 0.0007979</p>

also tested with F1-score values in which the proposed attention based model archives 88% F1-Score with pertained embedding while baseline model archives 85% F1-Score. The results of proposed attention based model is presented in Figure 5. As illustrated in Figure 5, the proposed model performance is low at the start. However, as epochs increase, the validation accuracy becomes higher than training; in the end, it got 90% validation and 88% training accuracy.

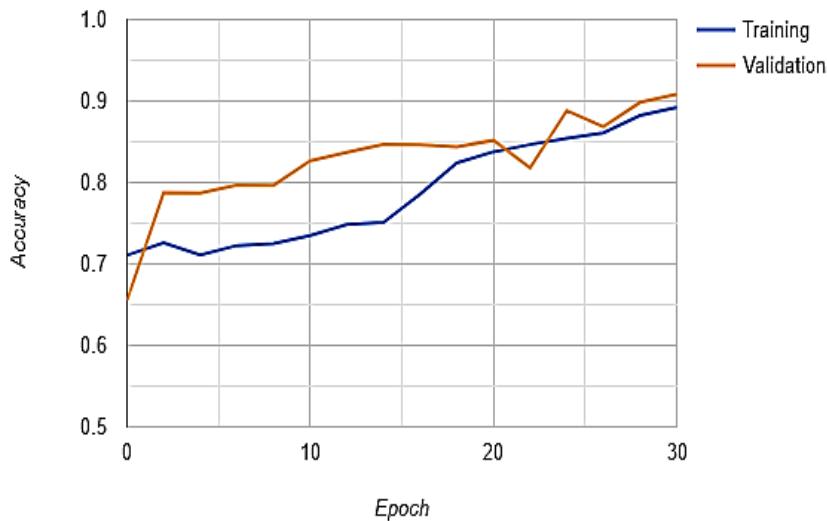


Figure 5: Proposed model accuracy

Compared to the suggested attention-based model, where the accuracy is about 91%, it achieves F-measure of 88% with pre-trained embedding. The comparison of both models is listed in Table 3.

Table 3: Comparison of baseline and proposed model

Model	Pre trained word Embedding		No pre trained word embedding	
	Accuracy	F1-Score	Accuracy	F1-Score
Baseline Model	86%	85%	84%	80%
Proposed Model	91%	88%	90%	87.3

A confusion matrix is a classifier performance assessment approach for machine and deep learning models. It is a table that shows how well a classifier performs on test data with known true values. Comparing actual

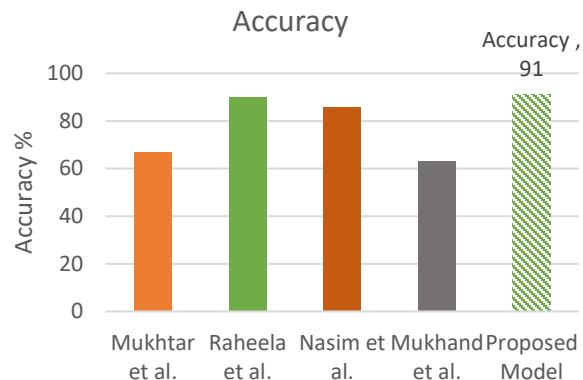


Figure 6: Accuracy of previous techniques

and anticipated classifications, the confusion matrix illustrates the quality of any classifiers. The proposed model is also evaluated using the confusion matrix metric. Figure 7 presents the confusion matrix for the proposed model. Various techniques have been used to classify Urdu sentiments in the last decade. Various authors used ML and DL techniques to classify various Urdu data sets. Table 4 presents the comparison of the proposed work with previous techniques.

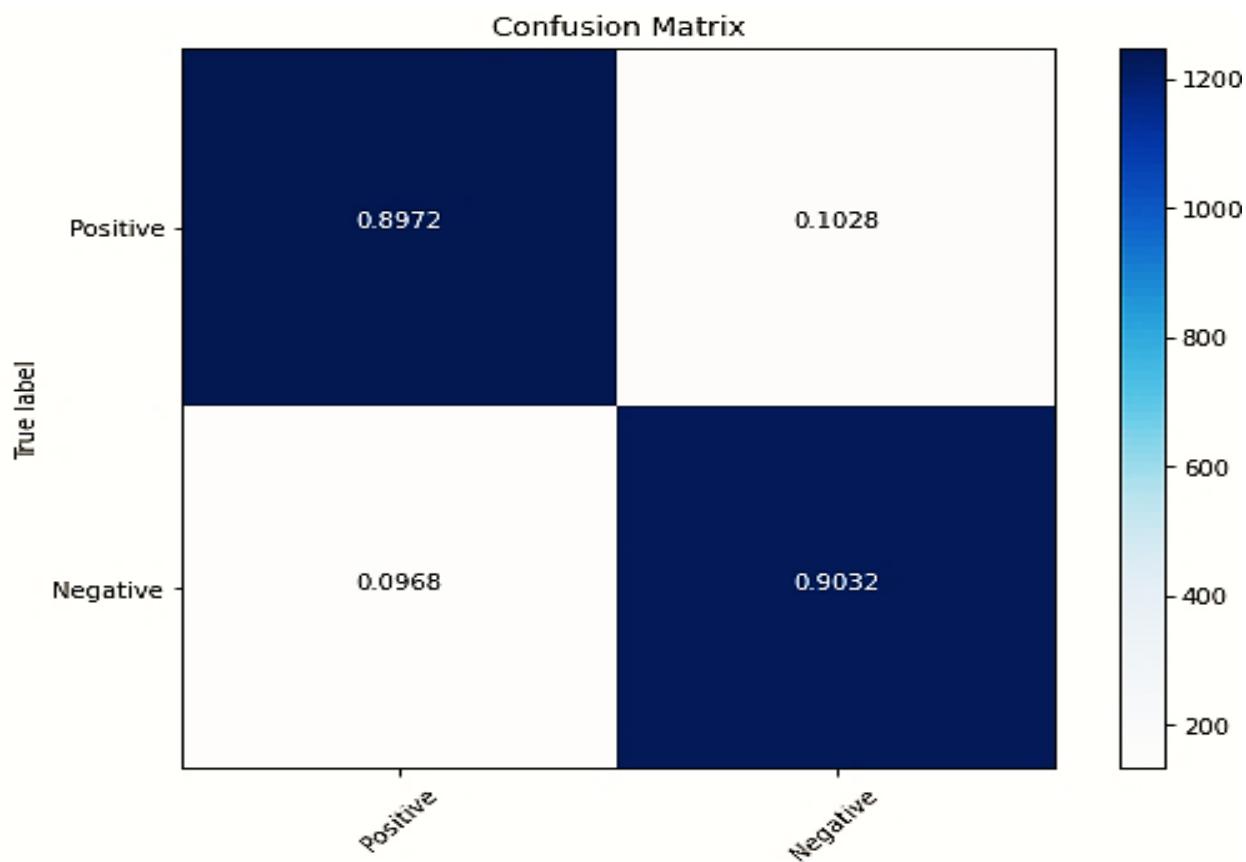


Figure 7: Confusion matrix for proposed model

*Table 4: Comparison with previous techniques*

<b>Author</b>	<b>Year</b>	<b>Technique</b>	<b>Accuracy</b>
Naqvi et al.	2021	LSTM, CNN, BiLSTM and attention based BiLSTM	77%
Hashim et al.	2016	Lexicon	86.8 %
Masood et al.	2022	LSTM	86.8 %
Safder et al	2021	SVM,CNN, LSTM, RCNN	84%
Mukhtar et al.	2016	SVM, KNN	67%
<b>Proposed Model</b>	<b>Proposed</b>	<b>Attention based Transfer Learning Model</b>	<b>91%</b>

## 7. Conclusion

Deep learning approaches have recently proven significant in various applications, including machine translation, image recognition, object detection, and natural language processing (NLP). The findings of this study's work are promising in Urdu SA. A deep learning model based on the attention mechanism is suggested in this work to categorize sentiment from Urdu text. First, preprocessing was done to normalize and tokenize the gathered data set. Secondly, a review-level sentiment classification algorithm is developed for analyzing the review's polarity by picking the most informative terms that indicate a specific sentiment in a review. The experiments were performed with the help of a transfer learning approach with pre-trained embeddings. Visualizing attention weights is also done to reveal the black box of the models and confirm their intuition, which helps interpret the model's acquired representations. The proposed system is evaluated based on accuracy and F1 score and compared with the baseline model. The findings show that the proposed deep learning model achieved 91% accuracy and 88% F1 score with pre-trained word embeddings. The model achieved 90% and 87.3% accuracy and F1-score, respectively, without pre-trained embeddings. In the future a large dataset by incorporating various Urdu dialects will get more reliable results. Moreover, the proposed technique can be trained on multilingual data for multilingual sentiment analysis.

## References

- Almani, N.M. and L.H. Tang. (2020). *Deep Attention-Based Review Level Sentiment Analysis for Arabic Reviews*, 6th Conference on Data Science and Machine Learning Applications (CDMA), IEEE.
- Asghar, M.Z., et al. (2017). Sentence-level emotion detection framework using rule-based classification. *Cognitive Computation*, 868-894.
- Bibi, R., et al. (2019). *Sentiment analysis for urdu news tweets using decision tree*, IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA), IEEE.
- Cambria, E., et al. (2017). *Affective computing and sentiment analysis*, in A practical guide to sentiment analysis, Springer, 1-10.
- Daud, A., W. Khan, and D. Che. (2017). Urdu language processing: a survey, *Artificial Intelligence Review*, 279-311.
- Glorot, X., A. Bordes, and Y. Bengio. (2011). *Deep sparse rectifier neural networks*, in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, JMLR Workshop and Conference Proceedings.
- Hashim, Faiza, and M. Khan. (2016). *Sentence level sentiment analysis using urdu nouns*, In Proceedings of the Conference on Language & Technology, vol. 2016.
- Khalid, K., et al. (2017). *Extension of semantic based urdu linguistic resources using natural language processing*, IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and

- Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE.
- Khotimah, D.A.K. and R. Sarno. (2018). *Sentiment detection of comment titles in booking. com using probabilistic latent semantic analysis*, 6th International Conference on Information and Communication Technology (ICoICT), IEEE.
- Kumar, H., B. Harish, and H. Darshan. (2019). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method, *International Journal of Interactive Multimedia & Artificial Intelligence*.
- Liu, H., Q. Yin, and W.Y. Wang. (2018). Towards explainable NLP: A generative explanation framework for text classification, *arXiv preprint arXiv:1811.00196*.
- M. Masood, F. Azam, M. Waseem Anwar and J. Ur Rahman. (2022). *Deep-learning based framework for sentiment analysis in Urdu language*, 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2), doi: 10.1109/ICoDT255437.2022.9787451, 1-7.
- M. Z. Asghar, A. Sattar, A. Khan, A. Ali, F. Masud Kundi, and S. Ahmad. (2019). Creating sentiment lexicon for sentiment analysis in urdu: The case of a resource-poor language, *Expert Systems*, e12397.
- Mehmood, K., et al. (2019). Sentiment analysis for a resource poor language—Roman Urdu, *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 1-15.
- Mukhtar, N. and M.A. Khan. (2018). Urdu sentiment analysis using supervised machine learning approach, *International Journal of Pattern Recognition and Artificial Intelligence*, 1851001.
- Mukhtar, N. and M.A. Khan. (2020). Effective lexicon-based approach for Urdu sentiment analysis, *Artificial Intelligence Review*, 2521-2548.
- Mukund, S. and R.K. Srihari. (2012). *Analyzing Urdu social media for sentiments using transfer learning with controlled translations*, in Proceedings of the Second Workshop on Language in Social Media, 1-8.
- Naqvi, U., A. Majid, and S.A. Abbas. (2021). UTSA: Urdu Text Sentiment Analysis Using Deep Learning Methods, *IEEE Access*, 114085-114094.
- Nasim, Z. and S. Ghani. (2020). Sentiment Analysis on Urdu Tweets Using Markov Chains, *SN Computer Science*, 1-13.
- Patel, J.M., Web. (2020). *Scraping in Python Using Beautiful Soup Library*, in Getting Structured Data from the Internet, Springer, 31-84.
- Soleymani, M., et al. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, 3-14.
- Syed, A.Z., M. Aslam, and A.M. (2014). Martinez-Enriquez, Associating targets with SentiUnits: a step forward in sentiment analysis of Urdu text, *Artificial Intelligence Review*, 535-561.
- Tan, C., et al. (2018). *A survey on deep transfer learning*, in International conference on artificial neural networks, Springer.
- U. Naqvi, A. Majid and S. A. Abbas. (2021). UTSA: Urdu Text Sentiment Analysis Using Deep Learning Methods, *IEEE Access*, doi: 10.1109/ACCESS.2021.3104308, 114085-114094.
- Yang, L., et al. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning, *IEEE Access*, 23522-23530.
- Zhang, Z., D. Robinson, and J. Tepper. (2018). *Detecting hate speech on twitter using a convolution-gru based deep neural network*, in European semantic web conference, Springer.