Intent Research Scientific Journal-(IRSJ)
ISSN (E): 2980-4612
Volume 5, Issue 3, March - 2026
Website: intentresearch.org/index.php/irsj/index

EXPERIMENTAL ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR SENTIMENT ANALYSIS IN THE UZBEK LANGUAGE

Suyunova Malika Odil qizi
PhD Researcher at the Department of Computational Linguistics and Digital Technology Alisher Navoi Tashkent State University of Uzbek Language and Literature
malikasuyunova0@gmail.com

Abstract

This paper investigates the problem of sentiment detection in Uzbek social media texts through a comparative evaluation of classical and deep learning models. Within the scope of the study, the performance of Naive Bayes, Support Vector Machine (SVM), LSTM recurrent neural network, and transformer-based models such as BERT/RoBERTa was experimentally assessed. The models were trained on a balanced corpus and evaluated using accuracy and F1-score metrics. The results demonstrate that transformer-based models, which effectively capture contextual dependencies, achieved the highest performance, while the LSTM model produced slightly lower but still competitive results. Among traditional approaches, SVM proved to be a stable and effective baseline classifier. The findings confirm the superiority of deep contextual models for Uzbek sentiment analysis, while also indicating that classical or hybrid approaches remain practically relevant for real-time applications.

Keywords: Sentiment analysis, Uzbek language, social media texts, machine learning, deep learning, SVM, LSTM, BERT, transformer models, text classification.

Introduction

Over the past decade, social media has become one of the primary platforms for expressing public opinions. On platforms such as Telegram, Instagram, Facebook, and others, users share their views on products, services, public policy, education, culture, and various social processes. The rapid growth in the volume of such textual data has created a need for its automated analysis.

Sentiment analysis (also referred to as Opinion Mining) is the process of determining whether the opinion expressed in a text is positive, negative, or

ISSN (E): 2980-4612
Volume 5, Issue 3, March - 2026
Website: intentresearch.org/index.php/irsj/index

neutral. This field is considered one of the key tasks in Natural Language Processing (NLP). Through sentiment analysis, it is possible to systematically and efficiently assess public mood, user attitudes, brand reputation, and service quality. Traditional sentiment analysis evaluates a text at the overall level, classifying an entire post or document as positive, negative, or neutral. However, in practice, user opinions often contain multiple aspects. For example, in a post about a restaurant, the service quality may be evaluated positively, while the prices may be assessed negatively. Therefore, in recent years, Aspect-Based Sentiment Analysis (ABSA) has emerged as a distinct research direction. ABSA identifies specific objects or attributes (aspects) within a text and separately determines the sentiment expressed toward each of them.

In recent years, scientific research on sentiment analysis in the Uzbek language has been developing steadily. However, Uzbek is considered a low-resource language. Large-scale annotated corpora, comprehensive sentiment lexicons, and specialized NLP tools are not yet sufficiently developed. Moreover, the agglutinative morphological structure of the Uzbek language, the extensive use of affixes, relatively flexible word order, and the informal style of social media texts further complicate the task of sentiment analysis.

The following characteristics are commonly observed in social media texts:

- spelling and grammatical errors;
- abbreviations and dialectal elements;
- emojis and special symbols;
- mixed words borrowed from Russian and English;
- irony and implicit meanings.

These factors complicate not only the identification of overall sentiment but also the process of aspect extraction. Therefore, developing ABSA approaches for the Uzbek language holds particular scientific significance.

The main studies conducted on sentiment analysis in the Uzbek language are presented in Table 2. While some studies are directly focused on Uzbek social media texts, others address different domains (such as product reviews); however, their approaches and methodologies are also relevant to the problem addressed in this study.

Intent Research Scientific Journal-(IRSJ)
ISSN (E): 2980-4612
Volume 5, Issue 3, March - 2026
Website: intentresearch.org/index.php/irsj/index

Table 2. Scientific research on sentiment analysis in the Uzbek language (recent years)

Research source (year)	Topic	Approaches	Best result
Kuriyozov and Matlatipov (2019)	Google Play app reviews (4.3k manually annotated + 20k machine-translated)	LR, SVM, RNN, CNN (using fastText embeddings)	89.6% accuracy – achieved by LR with word + character n-grams. The deep learning model did not show a significant advantage (classical and DL results were very close).
Matlatipov et al. (2022)	Local restaurant reviews (~8k, positive/negative)	LR, SVM, RNN, CNN	91% accuracy – achieved by LR (word + char n-grams). Introducing stemming beforehand significantly improved the result. SVM ~88%.
Allanazarova and Elova (2023)	Social network comments (various domains)	Lexicon-based (with manually built emotive lexicon)	No specific numerical result provided. Developed an emotive lexicon of emotionally rich words in Uzbek and proposed a lexicon-based analysis method. This approach relies on the emotional semantics of words to determine the sentiment in the text.
Mengliev et al. (2023)	General (various texts)	Rule-based (using morphological lexicons: 300+ suffixes, exception words, roots)	No specific numerical result. Proposed a rule-based method for determining sentence-level sentiment using word suffixes and negation rules in the agglutinative Uzbek language. This morphology-based approach is noted to be effective for sentiment detection.
Kuriyozov et al. (2024)	App reviews (positive/negative, 4k) and movie reviews (translated, 20k)	BERTbek (Uzbek monolingual BERT model), mBERT, LR, RNN, CNN	92.3% F1 – BERTbek (pretrained on a large corpus). The monolingual BERT model outperformed multilingual BERT as well as traditional models and RNN/CNN. It was observed that increasing the size of the pretraining corpus further improved the BERT model's performance.

The studies reviewed above indicate that the creation of an annotated corpus has been a decisive task in the early stages of sentiment analysis research for the Uzbek language (in 2019, the first dataset consisting of 4,300 manually labeled entries was developed). Traditional machine learning methods (such as Logistic Regression (LR) and Support Vector Machines (SVM)) have demonstrated considerable performance even with relatively small annotated datasets, and in

Intent Research Scientific Journal-(IRSJ)
ISSN (E): 2980-4612
Volume 5, Issue 3, March - 2026
Website: intentresearch.org/index.php/irsj/index

some cases have achieved results close to those of deep learning models. For example, in the study by Kuriyozov and Matlatipov (2019), a logistic regression model using word- and character-level n-gram features achieved an accuracy of 88–89% on 4,300 manually annotated reviews, while RNN and CNN models produced comparable results [1]. Similarly, in a 2022 study on restaurant reviews, the best performance was achieved by the logistic regression model (91%), whereas RNN and CNN models reached approximately 88–90% accuracy [2]. These findings suggest that in the Uzbek language context, the advantages of deep learning models can be fully realized only when larger and higher-quality datasets are available.

Overall, the literature review indicates that identifying sentiment in social media texts is more challenging due to linguistic ambiguity, stylistic variation, abbreviations, and dialectal elements. Social media posts frequently contain features of spoken language, emotionally charged expressions, and deviations from standard grammatical norms. Therefore, in the application of the models described below, particular attention was paid to text preprocessing, normalization, and the extraction of features in a manner appropriate for each model.

Methodology

This study aims to develop and evaluate an Aspect-Based Sentiment Analysis (ABSA) model based on Uzbek social media texts. Unlike traditional sentiment analysis, which determines the overall polarity of a text, the ABSA approach focuses on identifying specific aspects mentioned in the text and determining the sentiment expressed toward each of them.

In sentiment analysis, the choice of model directly depends on the following factors:

1. Data size and quality: When the annotated corpus is small, classical models (such as Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machines (SVM)) often provide strong baseline performance. As the dataset grows, deep learning and transformer-based models tend to demonstrate superior performance.

2. Language characteristics: Since Uzbek is an agglutinative language, words can take numerous affixes. Additionally, spelling variations, abbreviations, dialectal forms, and code-mixing frequently appear in social media texts. These factors increase the likelihood of model errors.

3. Context and negation: Negation forms such as "emas," "yo'q," "hech," and suffixes like "-ma/-mas," as well as intensifiers such as "juda," "rosa," and "o'ta," can significantly alter sentiment polarity. Some models capture these contextual cues effectively, while others process words independently and may lose contextual information.

4. Computational resources: Transformer-based models may achieve higher performance but require substantial memory and GPU resources, whereas SVM and NB models are computationally efficient and cost-effective.

Naïve Bayes Classifier

Naïve Bayes (NB) is one of the simplest and fastest text classification methods based on the probabilistic distribution of words within a document [3]. Using Bayes' theorem, the model calculates the probability that each word (feature) belongs to a particular sentiment class and combines the contributions of all words to determine the final classification. Although the independence assumption does not fully hold in real-world texts, NB is often considered a strong linear baseline model.

The main advantage of the NB model lies in its computational efficiency and relatively stable performance even with limited data. For example, on the Sentiment140 dataset, which contains 1.6 million English tweets, NB reportedly achieved approximately 85% accuracy. However, NB does not account for dependencies between words; therefore, misclassification may occur in sentences with complex contextual structures. For instance, in the sentence "Film unchalik yomon emas" ("The movie is not that bad"), the two-part negation ("unchalik … emas") conveys a positive meaning. However, since NB evaluates words independently, it may interpret the word "yomon" ("bad") as a negative signal. In such cases, more advanced models, described below, tend to demonstrate superior performance.

Support Vector Machine Classifier

Support Vector Machine (SVM) is a linear classification method that represents data in a high-dimensional space and seeks to determine the optimal separating hyperplane. By employing specialized kernel functions, SVM can also model nonlinear decision boundaries by transforming the problem into a linearly

separable form within a higher-dimensional feature space [4]. In text classification tasks, a linear kernel is typically sufficient, since in feature spaces such as Bag-of-Words (BoW), sentiment-related words are often linearly separable.
SVM generally achieves higher accuracy than NB, particularly when the features are well-balanced and properly preprocessed. For example, in sentiment analysis of Arabic tweets, SVM achieved an accuracy of 76.56%, outperforming NB (approximately 70–75%). Another advantage of SVM is its lower tendency toward overfitting and its stable performance even in high-dimensional feature spaces. This can be explained by its objective of maximizing the decision margin and relying only on support vectors (data points closest to the decision boundary).

CRF (Conditional Random Field)

Conditional Random Field (CRF) is a model originally developed for sequence labeling tasks and is widely used in natural language processing to label token sequences (e.g., Named Entity Recognition (NER) and Part-of-Speech (POS) tagging). CRF determines the optimal sequence of labels for a given sequence of tokens by maximizing the global conditional probability, while explicitly modeling dependencies between neighboring elements [5].
In the scientific literature, there have also been attempts to address sentiment classification as a sequence labeling task. For example, in the work of Lafferty et al. (2001), the CRF model was enriched with additional features and adapted to label sentences as “positive” or “negative” [6]. In this approach, CRF assigns a sentiment label to each word, and the final decision is derived from the resulting sequence of labels. However, in general, CRF is not widely applied at the document level. Nevertheless, within our research framework, CRF was considered as a component model, particularly for the following purposes:

- **Negation handling with CRF:** When negation words or affixes appear, their scope of influence is taken into account. For example, in the phrase “yaxshi emas” (“not good”), the negation marker “emas” changes the meaning of the nearby word “yaxshi” from positive to negative. CRF can model such dependencies and assign a negative label to “yaxshi.”

- **Intensifiers and context:** In expressions such as "juda yaxshi" ("very good"), the word "juda" intensifies the sentiment of "yaxshi." Through a sequential CRF model, the presence of "juda" can signal that the following word should be interpreted with intensified polarity.

LSTM Recurrent Neural Network

Among deep learning approaches, one of the most widely used models in sentiment analysis is the Long Short-Term Memory (LSTM) recurrent neural network. LSTM is designed to learn sequential data over time, such as text sequences, and is capable of preserving long-range contextual information through its memory mechanism. Standard RNNs often struggle with the problems of vanishing and exploding gradients, which limit their ability to retain context in long sequences. In contrast, LSTM addresses these issues through specialized memory cells and gating mechanisms [6]. As a result, LSTMs are able to capture long-distance dependencies that influence the sentiment of a text and summarize them within a final hidden state vector.

In our architecture, pre-trained fastText embeddings were loaded into the LSTM model (each word represented as a 300-dimensional vector). A single LSTM layer with 128 hidden units was employed. The LSTM processes the input text sequence, and its final hidden state is taken as the aggregated context vector representing the entire post. This vector is then passed to a fully connected output layer, where softmax probabilities are computed for three classes (positive, neutral, negative). However, one of the limitations of LSTM is the required training time and the need for careful hyperparameter tuning. As the corpus size increases, training the LSTM typically requires GPU resources, and in some cases, it may be more practical to use a pre-trained transformer model instead.

BERT and RoBERTa (Transformer Models)

In recent years, the models achieving the highest performance in sentiment analysis have been those based on the BERT transformer architecture. BERT (Bidirectional Encoder Representations from Transformers) was introduced by Google AI in 2018 and marked a significant breakthrough in language modeling. BERT is pre-trained on large-scale corpora using self-supervised learning objectives, such as masked language modeling and next sentence prediction [7].

As a result, the model learns contextual representations of words, generating hidden vectors that capture the full sentence context for each token.

BERT's transformer architecture consists solely of the encoder component (encoder-only). Its structure is composed of the following modules:

- **Tokenizer:** Splits the input text into tokens, typically using WordPiece subword tokenization.
- **Embedding layer:** Converts each token into a 768-dimensional vector representation. This layer incorporates three types of embeddings: (a) token embedding (representing the token itself), (b) positional embedding (indicating the position of the token in the sequence), and (c) segment embedding (indicating whether the token belongs to the first or second sentence).
- **Encoder layers:** A stack of 12 Transformer encoder blocks performs deep contextual transformations layer by layer. Each block includes a multi-head self-attention mechanism, which computes weighted relationships between tokens, allowing each word to attend to other words in the sequence.
- **Task head (output layer):** During pretraining, BERT uses a vocabulary-sized softmax layer to predict masked tokens. During fine-tuning for downstream tasks, the final hidden state of the [CLS] token is typically used, followed by a task-specific classifier layer. In our case, a three-dimensional output layer (positive, neutral, negative) was added on top of the [CLS] representation.

Interpreting transformer-based model outputs is also important, as they produce probability scores for each class. In our approach, the class with the highest softmax probability was selected as the final prediction. Additionally, by visualizing the attention mechanism, it is possible to analyze which words the model relies on most when making its decision.

Results and Analysis

In this study, several classical and deep learning models were evaluated to determine the sentiment of Uzbek social media posts. The experiments were conducted on a balanced corpus consisting of posts labeled as positive, negative, and neutral. The dataset was divided into 80% training and 20% testing sets. To ensure the stability and reliability of the results, 5-fold cross-validation was also applied.

Before conducting the experiments, the following preprocessing steps were performed:

- removal of URLs and unnecessary symbols;
- conversion of text to lowercase;
- normalization of emojis and abbreviations;
- tokenization;
- application of TF-IDF features (for classical models);
- use of pretrained embeddings (fastText for the LSTM model).

For the transformer-based models, a specialized tokenizer (WordPiece) was applied. Model performance was evaluated using the following metrics: accuracy, precision, recall, and F1-score. Since the corpus was balanced, accuracy was used as the primary evaluation metric; however, class-wise F1-scores were also analyzed separately.

The experimental results indicate that model performance increased with the level of model complexity.

Table 2. Performance of Classical and Advanced Models

Model	**Accuracy (%)**	**Average F1**
Naive Bayes	~78	0.77
SVM	~88	0.87
LSTM	~90	0.89
BERT/RoBERTa	~92	0.91–0.92

According to the results, the transformer-based model achieved the highest performance. The LSTM model produced slightly lower results than the transformer model but outperformed the classical approaches.

Analysis of Classical Models

Despite its computational efficiency, the Naïve Bayes model encountered difficulties in handling complex syntactic constructions. In particular, misclassifications were observed in sentences containing negation forms (such as “emas”) or contrastive conjunctions (such as “lekin” and “ammo”).

The SVM model, based on n-gram features, was able to separate the classes more accurately and achieved the best performance among the classical models. This confirms that SVM serves as an effective baseline model for Uzbek-language text classification tasks.

Analysis of Deep Learning Models

The LSTM model performed better on complex and mixed-polarity sentences, as it accounts for sequential context. For example, in sentences containing both positive and negative elements, LSTM was relatively successful in producing accurate predictions.

The transformer-based model (BERT/RoBERTa) achieved the highest accuracy overall. Owing to its self-attention mechanism, the model was able to deeply analyze the entire sentence context and correctly interpret complex constructions.

While the positive and negative classes were identified with high accuracy, relatively more errors were observed in the neutral class. Neutral posts often contain both positive and negative elements, which may lead the model to misclassify them. The main confusion was observed between the neutral and negative classes. The inference speed of the NB and SVM models was very high, making them suitable for real-time monitoring systems. In contrast, the transformer model provided higher accuracy but required substantial computational resources. Therefore, in practical applications, a hybrid approach (combining a fast model with a deep model) may be more effective.

Conclusion

The results of this study indicate that sufficient resources and methodological capabilities for sentiment analysis in the Uzbek language are gradually being developed; however, several limitations still remain. In particular:

- **Language resources:** Currently, large-scale text corpora, annotated sentiment lexicons, and specialized tools for the Uzbek language remain limited. In the future, it is necessary to expand the corpus, particularly by increasing the number of neutral examples and covering a wider range of topics. In addition, developing a lexicon of emotion-expressing synonyms and idiomatic expressions (e.g., “kayfiyati chog‘” – positive, “qoni qaynamoqda” – contextually negative) could further enhance model performance.
- **Model integration:** Ensemble approaches that combine classical and deep learning models may be considered. For example, integrating the outputs of LSTM and SVM through a majority voting mechanism could improve reliability in certain cases (some studies report a ~1–2% performance improvement with ensemble methods).

•**Contextual analysis:** Social media posts often derive their full meaning from comment threads or preceding messages. In this study, individual posts were analyzed independently. Future research could focus on contextual sentiment analysis, such as models capable of evaluating sentiment within comment networks or conversational threads.

•**Aspect-based analysis:** This study focused on overall sentiment classification. However, user posts frequently contain opinions about multiple aspects (e.g., service quality, price, environment). Future work could involve developing models that automatically detect which specific aspects are described positively or negatively within each post (e.g., fine-tuning BERT in an ABSA setting).

•**Real-time monitoring:** For practical deployment (e.g., real-time monitoring of opinions in Telegram channels), model optimization is necessary. Utilizing lightweight models such as DistilBERT or accelerating SVM with techniques such as the hashing trick may be appropriate.

This study presents an experimentally grounded comparative analysis aimed at advancing sentiment analysis in the Uzbek language and serves as a methodological foundation for future research in this field.

References

1. Kuriyozov, E., Matlatipov, S., Alonso, M. A., & Gómez-Rodríguez, C. (2019, May). Construction and evaluation of sentiment datasets for low-resource languages: The case of Uzbek. In Language and Technology Conference (pp. 232–243). Cham: Springer International Publishing.

2. Matlatipov, S., Rahimboeva, H., Rajabov, J., & Kuriyozov, E. (2022). Uzbek sentiment analysis based on local restaurant reviews. arXiv preprint arXiv:2205.15930.

3. Boltayevich, E. B., Turapovna, I. S., & Ibragimovna, T. G. (2024, November). Tagging Units in the Text and the Bayes Algorithm. In 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE) (pp. 1840–1843). IEEE.

4. Rahat, A. M., Kahir, A., & Masum, A. K. M. (2019, November). Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (pp. 266–270). IEEE.

ISSN (E): 2980-4612
Volume 5, Issue 3, March - 2026
Website: intentresearch.org/index.php/irsj/index

5. Wang, W., Xin, G., Wang, B., Huang, J., & Liu, Y. (2017). Sentiment information extraction of comparative sentences based on CRF model. Computer Science and Information Systems, 14(3), 823–837.

6. Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proceedings of ICML 2001.

7. Nowak, J., Taspinar, A., & Scherer, R. (2017, May). LSTM recurrent neural networks for short text and sentiment classification. In International conference on artificial intelligence and soft computing (pp. 553–562). Cham: Springer International Publishing.

8. Liao, W., Zeng, B., Yin, X., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Applied Intelligence, 51(6), 3522–3533.