SETSCI Conference Proceedings
SETSCI Conference Proceedings 21, 1-5, 2024

International Trend of Tech Symposium

7-8 December 2024, İstanbul - Türkiye

https://doi.org/10.36287/setsci.21.1.001
2687-5527/

Multilingual Sentiment Analysis for Mobile Gaming: A Comparative Study of Machine Learning and Hybrid Deep Learning Approaches

Erol KINA[1*], Recep ÖZDAĞ[2]

[1]Özalp Vocational School, Van Yüzüncü Yıl University, Van, Türkiye
[2] Faculty of Engineering, Depatment of Computer Engineering, Van Yüzüncü Yıl University, Van, Türkiye

[*]erolkina@yyu.edu.tr

Abstract – The rapid expansion of the mobile gaming industry has underscored the need to understand user sentiment, with social media platforms like X (Twitter) providing key insights. This study applied sentiment analysis to English and Turkish tweets, utilizing the TEMSAP-CNNLSTM model a hybrid architecture combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for precise classification of complex textual data. The model's performance was benchmarked against traditional machine learning methods, including Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). Results revealed that TEMSAP-CNNLSTM consistently achieved superior performance, with the highest Accuracy, Precision, Recall, and F1-Score across both datasets. The model attained 96% accuracy on English and Turkish training datasets and 93% and 92% on English and Turkish test datasets, respectively. These findings highlighted the model's capability in handling sentiment data, surpassing traditional approaches while demonstrating robust generalizability across languages. The TEMSAP-CNNLSTM model offers valuable insights for mobile game developers and suggests broader applicability for other industries requiring accurate sentiment analysis in multilingual contexts.

Keywords – Machine learning, Sentiment analysis, Twitter, Text mining

I. INTRODUCTION

The mobile gaming sector has rapidly evolved into a global industry with the widespread adoption of smartphones. Mobile games play a significant role in contemporary society by fulfilling users' needs for entertainment, social interaction, and personal development [1]. The sector's appeal to a diverse audience spanning various age groups and demographic profiles highlights its broad impact. Mobile games have become integral to daily life, which has, in turn, contributed to the sector's growing economic importance. For instance, data from Newzoo [2] indicate that mobile games account for approximately half of the global gaming industry, reflecting a substantial market share. Beyond mere entertainment, mobile games offer considerable benefits in education and social engagement. Consequently, understanding the effects of mobile games on users holds strategic importance for the industry.

Social media has become deeply embedded in our daily lives, allowing people to stay connected with the outside world and share their experiences in a way that feels personal and convenient. Platforms such as Twitter (X), Facebook, Instagram, and WhatsApp have become central spaces where users can voice their opinions on a range of topics and interact with others. These dynamic communication channels make it crucial to share insights into people's perspectives, emotions, and attitudes toward various products, ideas, or policies. As users increasingly turn to these platforms for information through posts and tweets, the role of social media in shaping public opinion continues to expand [3, 4]. X (Twitter), in particular, is a prominent platform where users share real-time perspectives on mobile games, providing a substantial data source for analyzing emotional responses. Sentiment analysis plays a pivotal role in processing social media data and discerning positive and negative reactions within it [5]. This analytical method categorizes user opinions about mobile games, thereby offering game developers valuable insights into user feedback. The primary objective of sentiment analysis is to determine the emotional tone expressed within written texts, thereby enabling the measurement of users' sentiments and reactions.

In sentiment analysis, both traditional and advanced machine learning algorithms are commonly employed to determine whether texts contain positive, negative, or neutral content [6]. The literature frequently highlights methods such as Logistic Regression, Random Forest, Support Vector Machines (SVM), and Naïve Bayes algorithms. For example, in a study, conducted by Mantika et al. [7] utilized Twitter data to analyze public sentiment toward candidates in the 2024 Indonesian presidential election. The research compared the effectiveness of machine learning algorithms, specifically Naive Bayes and Logistic Regression, in classifying sentiment related to the candidates. Naive Bayes outperformed Logistic Regression, with accuracy rates of 63% for Anies Baswedan, 77% for Ganjar Pranowo, and 44% for Prabowo Subianto. These results suggested that Naive Bayes was a more effective approach for assessing public sentiment on the 2024 presidential candidates. Tabany and Gueffal, [8] sought to perform sentiment analysis on short and long Amazon reviews to assess their impact on an SVM model used for fake review classification. Initially, the SVM model's performance was evaluated in comparison with Naive Bayes, Logistic Regression, and Random Forest models. It demonstrated superior results, confirming the second assumption, with accuracy at 70%, precision at 63%, recall at 70%, and an F1-score of 62%. In another study Dharta et al. [9] aimed to analyze social media users' sentiment towards the election. Data was collected through a literature review and observation, and sentiment was analyzed using the Naïve Bayes Classifier and Support Vector Machine algorithms. TextBlob was used to classify over 15,000 preprocessed tweets, yielding 11,000 clean tweets labeled as either positive or negative. Results indicated that 4,000 tweets were positive, showing general support for the elections, while the rest reflected negative sentiments. The Naïve Bayes method achieved accuracy, precision, and recall of 85%, 80%, and 75%, respectively. The Support Vector Machine, with a poly kernel, produced the highest accuracy, precision, and recall at 90%, 90%, and 85%.

In addition to the aspects mentioned above methods, deep learning techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) facilitate a more nuanced understanding of the semantic and contextual structure of texts, thereby increasing analytical accuracy. CNNs are proficient at evaluating spatial relationships within data, whereas LSTMs are noted for capturing patterns within temporally dynamic data [10]. In recent years, the adoption of deep learning methods in sentiment analysis has enabled more in-depth analyses of social media data [11]. Based on this approach, the present study utilized these advanced techniques to conduct sentiment analysis on user feedback related to mobile games. In this study, sentiment analysis was applied to evaluate user feedback on mobile games, facilitating the identification of innovative trends within the sector. The study introduced an innovative approach to sentiment analysis on mobile gaming data obtained from Twitter. Specifically, the TEMSAP-CNNLSTM (Twitter Mobil Game Sentiment Analysis Approach) model was developed by integrating CNN and LSTM layers to achieve more precise classification of complex textual data.

II. Materials and Method

In this study, sentiment analysis on tweets related to mobile games was conducted through a comprehensive approach that involved data collection, preprocessing, feature extraction, and the application of various machine learning algorithms. The flowchart in Fig.1 illustrates the procedures implemented in this study

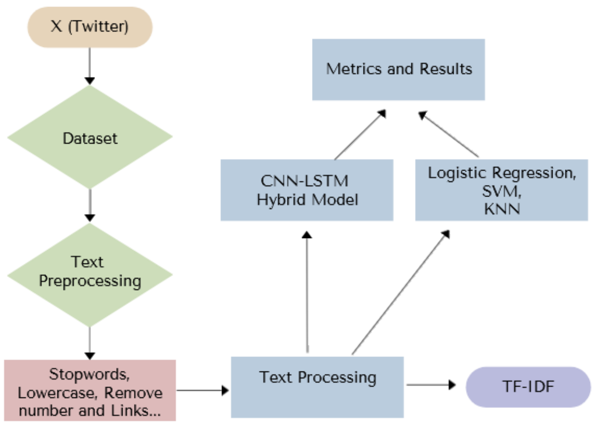

Fig. 1. Flowchart of the study

The methodology adopted for each phase is described in detail below:

A. Data Collection and Dataset

Data collection was performed using Twitter's API to retrieve real-time tweets concerning mobile games. This process resulted in a substantial dataset comprising 50.000 tweets, with 25.000 in Turkish and 25.000 in English. Tweets were sourced from multiple regions, including the Türkiye, United States, South Korea, and France, to ensure a broad spectrum of user perspectives. To prepare the data for analysis, tweets were filtered based on relevance to mobile games, ensuring that only pertinent content was included. Furthermore, each tweet was annotated with sentiment labels positive, negative, or neutral based on the expressed sentiment, thereby facilitating the subsequent classification tasks.

B. Data Preprocessing

The raw tweet data underwent a preprocessing phase to enhance its analytical suitability [12]. Initially, non-informative elements such as punctuation, emojis, and hyperlinks were removed to reduce noise [13]. Stop words were subsequently eliminated, which refined the dataset by retaining only meaningful terms [14]. Language filtering was then applied to ensure that only tweets in Turkish and English were included, following which tokenization and normalization were carried out. Tokenization divided the tweets into individual tokens [15], while normalization standardized text entries to improve coherence [16]. Stemming was finally employed to reduce words to their root forms, minimizing lexical variations and thereby improving the accuracy of sentiment classification [17].

C. Feature Extraction

After preprocessing, feature extraction techniques were applied to convert textual data into numerical representations that machine learning algorithms could process [18]. This was followed by the application of Term Frequency-Inverse Document Frequency (TF-IDF), which not only accounted for term frequency but also adjusted for the importance of each word across the dataset. TF-IDF assigns higher weights to words that are rare across the dataset but frequent in specific tweets, thus prioritizing informative terms for sentiment classification. This approach allowed the models to focus on the most salient words, facilitating a more accurate analysis.

D. Machine Learning Algorithms

For sentiment classification, several machine learning algorithms were deployed, including Logistic Regression, Random Forest, Support Vector Machines (SVM). Each algorithm was selected for its unique strengths in handling textual data:

Logistic Regression operates as a linear classifier, efficiently categorizing data into binary classes. Its ease of interpretation and relatively rapid training time made it suitable for this analysis. Support Vector Machines (SVM) maximize the margin between data classes by identifying a hyperplane that optimally separates them. This capability is especially advantageous for high-dimensional data, which is typical in text classification tasks. K-Nearest Neighbors (KNN) classifies instances based on the majority class among the nearest neighbors. It is a non-parametric, instance-based learning method that is especially effective in scenarios where decision boundaries are complex and non-linear. KNN was included due to its simplicity and robustness in handling noisy data. The study also introduced the TEMSAP-CNNLSTM model, a custom hybrid model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The CNN component captures spatial dependencies in text data, while the LSTM component excels at handling sequential information, making this model particularly effective for interpreting complex patterns within textual data.

E. Model Evaluation Criteria

The effectiveness of the machine learning models was assessed using several performance metrics, which provided a comprehensive evaluation of classification capabilities: Accuracy was used to measure the proportion of correctly classified instances relative to the total number of predictions. Precision and Recall evaluated the quality of positive and negative classifications, focusing on the proportion of true positives relative to the total predicted positives (Precision) and the actual positives (Recall), respectively. F1 Score offered a balanced metric by integrating Precision and Recall, making it particularly useful for handling imbalanced data where classes may not be equally represented. Through this robust methodology, the study provides valuable insights into the sentiment dynamics surrounding mobile games, contributing to a better understanding of user perspectives and informing the development of strategies for more effective user engagement.

III. EXPERIMENTAL RESULTS

As a result of the analysis, the performances of the machine learning algorithms were evaluated.The metrics demonstrating the performance of machine learning algorithms are presented in Tables 1, 2, 3 and 4. Table 1 presents the metrics calculated for the English training dataset for mobile games. Table 2 shows the metrics calculated for the English test dataset, Table 3 for the Turkish training dataset and Table 4 for the Turkish test dataset.

Table 1. Metrics calculated for the English training dataset in mobile games

Model Name	**Precision**	**Recall**	**F1-Score**	**Accuracy**
LR	0.89941	0.86145	0.88042	0.89
SVM	0.85456	0.86541	0.87551	0.86
KNN	0.87912	0.89215	0.88313	0.88
TEMSAP-CNNLSTM	0.96421	0.96545	0.94125	0.96

Table 2. Metrics calculated for the English test dataset in mobile games

Model Name	**Precision**	**Recall**	**F1-Score**	**Accuracy**
LR	0.83312	0.80089	0.82142	0.84
SVM	0.79216	0.80541	0.81698	0.80
KNN	0.82318	0.83562	0.84125	0.85
TEMSAP-CNNLSTM	0.96126	0.93260	0.93056	0.93

Table 3. Metrics calculated for the Turkish training dataset in mobile games

Model Name	**Precision**	**Recall**	**F1-Score**	**Accuracy**
LR	0.80644	0.96211	0.88333	0.87
SVM	0.81986	0.92318	0.88682	0.86
KNN	0.83714	0.96113	0.89116	0.88
TEMSAP-CNNLSTM	0.93289	0.98614	0.96391	0.96

Table 4. Metrics calculated for the Turkish Test dataset in mobile games

Model Name	**Precision**	**Recall**	**F1-Score**	**Accuracy**
LR	0.77395	0.94715	0.85271	0.85
SVM	0.76296	0.92417	0.88219	0.82
KNN	0.78521	0.91822	0.84819	0.84
TEMSAP-CNNLSTM	0.91833	0.94716	0.92943	0.92

This study evaluated various machine learning models for sentiment classification in mobile games, with an emphasis on English and Turkish datasets. The analysis was focused on four models: Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (KNN) and TEMSAP-CNNLSTM. These models were evaluated on metrics such as Precision, Recall, F1-Score and Accuracy applied to both training and test datasets for each language.

For the English dataset, TEMSAP-CNNLSTM demonstrated high precision and robust generalization, outperforming the other models with an Accuracy of 0.96 on the training data. The model also showed strong results on the test data, maintaining a high Accuracy of 0.93, significantly outperforming traditional models such as SVM and LR, which showed significant performance degradation.

On the Turkish dataset, TEMSAP-CNNLSTM once again demonstrated its effectiveness by achieving an accuracy of 0.96 on the training data and 0.92 on the test data. This consistency across languages demonstrated that TEMSAP-CNNLSTM was well suited for multilingual sentiment analysis. The combination of CNN for feature extraction and LSTM for capturing long-term dependencies provided a significant advantage over other models.

IV. DISCUSSION

The results presented in the tables indicated that the TEMSAP-CNNLSTM model consistently outperforms the other models-namely Logistic Regression, SVM, and KNN-across both English and Turkish datasets, for both training and testing phases. The TEMSAP-CNNLSTM model consistently achieved the highest values for Precision, Recall, F1-Score, and Accuracy, highlighting its superior performance in sentiment classification within the mobile gaming context. English Dataset: Across both training and test datasets, the TEMSAP-CNNLSTM model attained the highest accuracy rates (96% for training and 93% for testing) along with elevated F1-Scores (0.94125 for training and 0.93056 for testing). While Logistic Regression, SVM, and KNN models also performed commendably, they exhibited marginally lower Precision and Recall scores, suggesting that the TEMSAP-CNNLSTM model more effectively captured the nuances and complexities inherent in English-language sentiment data. Turkish Dataset: The TEMSAP-CNNLSTM model demonstrated similarly robust performance on the Turkish dataset, with accuracy rates of 96% in training and 92% in testing, alongside the highest F1-Scores among the models evaluated. The consistent superiority in both Precision and Recall metrics further underscored the model's robustness and its capacity for generalization across languages.

The results indicated that the TEMSAP-CNNLSTM model consistently exhibited superior performance in sentiment classification for mobile gaming data across both English and Turkish datasets. By attaining the highest scores in Accuracy, Precision, Recall, and F1-Score, this model significantly surpasses traditional machine learning approaches such as Logistic Regression, SVM, and KNN.

Hybrid models have consistently demonstrated their effectiveness in previous research, successfully combining the strengths of various machine learning techniques to achieve superior performance. For instance, in their study Rehman et al. [19] developed a Hybrid CNN-LSTM Model to tackle sentiment analysis by combining CNN's feature extraction with LSTM's ability to capture long-term dependencies. They used Word2Vec embeddings to convert text into vectors, enhancing grouping and analysis. The model, incorporating dropout and normalization for improved accuracy, significantly outperformed traditional methods on IMDB and Amazon review datasets, achieving higher precision, recall, F1-score, and accuracy.

Recently, inspired by the rise of social networks and the need to analyze public opinion, Ombabi et al. developed a new deep learning model for Arabic sentiment analysis. Recognizing the challenges posed by Arabic's complex morphology and dialect variations, they designed a model that combines CNN for local feature extraction and a two-layer LSTM to capture long-term dependencies. The extracted features were classified using an SVM powered by FastText word embeddings. After extensive testing on multi-domain datasets, the model achieved 90.75% accuracy, outperforming other state-of-the-art models. They also tested different embedding models and classifiers, confirming that the FastText skip-gram model and SVM are effective choices for Arabic sentiment analysis [20].

Similarly, our results highlighted the TEMSAP-CNNLSTM model's advanced capacity to manage the complexities of sentiment data, as well as its robust generalizability across different languages. The findings suggested that the TEMSAP-CNNLSTM model has substantial potential as a tool for multilingual sentiment analysis tasks, extending beyond the scope of mobile gaming to other domains where understanding user sentiment is essential. Given the model's demonstrated effectiveness, it is well-positioned for applications that demand precise and reliable sentiment detection in linguistically diverse settings. Future research could further examine the model's adaptability to additional languages and sentiment datasets, thereby enhancing its broader applicability and scalability across various industry contexts.

V. CONCLUSION

In this study, it was determined that, compared to other traditional machine learning methods, the TEMSAP-CNNLSTM model demonstrated superior accuracy, facilitating a more comprehensive understanding of users' emotional states regarding mobile games. This research provided mobile game developers with a strategic tool for gaining insights into user sentiments, whether positive or negative, and highlights the effectiveness of novel machine learning-based methodologies in analyzing social media data. Thus, the findings of this study can be considered a valuable resource for the mobile gaming industry, offering insights that may aid in developing superior products and meeting user needs more effectively.

ACKNOWLEDGMENT

The present study was conducted as part of Erol KINA's doctoral thesis, under the supervision of Assoc. Prof. Dr. Recep ÖZDAĞ.

REFERENCES

[1] S. Ngah, S. Rahi, F. Long, C. Gabarre, A. Rashid, and A. H. Ngah, "Future behavioural of console gamers and mobile gamers: are they differ?," *Qual. Quant.*, pp. 1–27, 2024.

[2] Newzoo, "Global esports market report." [Online]. Available: https://newzoo.com/products/reports/global-esports-market-report

[3] J. Anvar Shathik and K. Krishna Prasad, "A literature review on application of sentiment analysis using machine learning techniques," *Int J Appl Eng Manag Lett*, vol. 4, no. 2, pp. 41–67, 2020.

[4] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput.*, vol. 97, p. 106754, 2020.

[5] A. Modi, K. Shah, S. Shah, S. Patel, and M. Shah, "Sentiment analysis of Twitter feeds using flask environment: A superior application of data analysis," *Ann. Data Sci.*, vol. 11, no. 1, pp. 159–180, 2024.

[6] A. Alslaity and R. Orji, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," *Behav. Inf. Technol.*, vol. 43, no. 1, pp. 139–164, 2024.

[7] A. M. Mantika, A. Triayudi, and R. T. Aldisa, "Sentiment Analysis on Twitter Using Naïve Bayes and Logistic Regression for the 2024 Presidential Election," *SaNa J. Blockchain, NFTs Metaverse Technol.*, vol. 2, no. 1, pp. 44–55, 2024.

[8] M. Tabany and M. Gueffal, "Sentiment analysis and fake amazon reviews classification using SVM supervised machine learning model," *J. Adv. Inf. Technol.*, vol. 15, no. 1, pp. 49–58, 2024.

[9] F. Y. Dharta, A. J. Mahardhani, S. R. Yahya, A. Dirsa, and E. M. Usulu, "Application of Naive Bayes Classifier Method to Analyze

Social Media User Sentiment Towards the Presidential Election Phase," *J. Inf. dan Teknol.*, pp. 176–181, 2024.

[10] F. Alhaek, W. Liang, T. M. Rajeh, M. H. Javed, and T. Li, "Learning spatial patterns and temporal dependencies for traffic accident severity prediction: A deep learning approach," *Knowledge-Based Syst.*, vol. 286, p. 111406, 2024.

[11] I. K. S. Al-Tameemi, M.-R. Feizi-Derakhshi, S. Pashazadeh, and M. Asadpour, "A comprehensive review of visual–textual sentiment analysis from social media networks," *J. Comput. Soc. Sci.*, pp. 1–72, 2024.

[12] P. Chauhan, N. Sharma, and G. Sikka, "On the importance of pre-processing in small-scale analyses of twitter: a case study of the 2019 Indian general election," *Multimed. Tools Appl.*, vol. 83, no. 7, pp. 19219–19258, 2024.

[13] S. Jain and P. K. Roy, "E-commerce review sentiment score prediction considering misspelled words: A deep learning approach," *Electron. Commer. Res.*, vol. 24, no. 3, pp. 1737–1761, 2024.

[14] Y. S. Soekamto, I. Maryati, C. Christian, and E. Kurniawan, "Improving Publishing: Extracting Keywords and Clustering Topics," *J. Appl. Data Sci.*, vol. 5, no. 2, pp. 412–423, 2024.

[15] S. Wu and Y. Gao, "Machine learning approach to analyze the sentiment of airline passengers' tweets," *Transp. Res. Rec.*, vol. 2678, no. 2, pp. 48–56, 2024.

[16] A. Zadgaonkar and A. J. Agrawal, "An Approach for analyzing unstructured text data using topic modeling techniques for efficient information extraction," *New Gener. Comput.*, vol. 42, no. 1, pp. 109–134, 2024.

[17] N. A. Sharma, A. B. M. S. Ali, and M. A. Kabir, "A review of sentiment analysis: tasks, applications, and deep learning techniques," *Int. J. Data Sci. Anal.*, pp. 1–38, 2024.

[18] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J. Big Data*, vol. 11, no. 1, p. 87, 2024.

[19] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis," *Multimed. Tools Appl.*, vol. 78, pp. 26597–26613, 2019.

[20] A. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks," *Soc. Netw. Anal. Min.*, vol. 10, pp. 1–13, 2020.