Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Sentiment Analysis of MobileJKN App Reviews using Neural Network Algorithm

Muhammad Daffa Ayyasy [1] *, Rudi Kurniawan, Yudhistira Arie Wijaya [3]

[1,2,3]STMIK IKMI Cirebon
*mdaffaayyasy1102@gmail.com [1] **

Abstract

The advancement of information technology has encouraged the use of user data to improve digital services, particularly in health-related applications such as MobileJKN, developed by BPJS Kesehatan Indonesia. This research conducts sentiment analysis on user reviews of MobileJKN from the Google Play Store, aiming to identify key areas for improvement based on user perceptions. A Deep Learning approach is utilized, with Neural Networks as the primary model and Altair AI Studio as the main data processing tool. Following the Knowledge Discovery in Databases (KDD) methodology, the study involves various preprocessing stages including case folding, tokenization, filtering, stopword removal, and stemming, using the Kamus Besar Bahasa Indonesia (KBBI) to standardize local language terms. After preprocessing, clustering and classification are performed to extract sentiment patterns. The most frequently mentioned keywords "register," "app," "number," "sign in," and "verify" highlight common user concerns. The sentiment classification model achieved a 100% accuracy rate, with the Shuffled Sampling technique and a 90:10 training-testing ratio yielding optimal results. These findings demonstrate the effectiveness of Neural Networks in analyzing sentiment within health applications, providing valuable insights for developers seeking to enhance MobileJKN's performance and user satisfaction. The study also offers a practical reference for future sentiment analysis research in the Indonesian digital health context.

Keywords: *Sentiment Review, Neural Network, MobileJKN, Sampling Technique, Sentiment Classification*

1. Introduction

The development of information and communication technology has led to significant changes across various sectors, such as business, education, and healthcare. These changes have resulted in a massive and diverse volume of data from various digital platforms, such as social media, mobile applications, and e-commerce. In order to improve the quality of healthcare services, leveraging user data through sentiment analysis of health app reviews has become crucial. The MobileJKN app, released by BPJS Kesehatan Indonesia, has many reviews on the Google Play Store, offering an opportunity to gather useful information through sentiment analysis. Deep learning approaches, particularly neural networks with complex architectural structures, have shown great potential in classifying sentiments with higher accuracy, especially for Indonesian text rich in context [1].

One of the main challenges in sentiment analysis of MobileJKN app reviews is identifying the most frequently occurring words or phrases in specific sentiments. These words typically have unique characteristics, both in their usage patterns and context, which require special approaches for accurate identification. Furthermore, selecting the appropriate sampling technique to produce the best accuracy model is also a challenge that needs to be addressed. The comparison between the training and testing data ratio also affects the model's performance in sentiment classification, making it essential to find the most effective ratio. Several previous studies have developed deep learning models for sentiment analysis, such as multi-head attention mechanisms, RNN classification, and LSTM-CNN optimized with grid search [2].

While these models have demonstrated good performance in sentiment analysis, most of these studies focused on the English language or specific domains that may not be relevant to Indonesian and the context of local health apps. This study seeks to address this gap by adapting neural network models for Indonesian language reviews and integrating relevant dictionaries to improve classification performance [3].

The goal of this research is to enhance the sentiment classification model for MobileJKN app reviews using Neural Networks. By tailoring the deep learning model for Indonesian, particularly in health apps, this study aims to fill the gaps in previous research. The practical benefits of this research include improving the quality of app services through a deeper understanding of user reviews, as well as contributing to information and communication technology related to sentiment analysis. The approach used in this study is quantitative, using experiments based on a deep learning framework to build a Neural Network model, utilizing labeled review data through clustering techniques. The methodology applied refers to Knowledge Discovery in Database (KDD), which includes the stages of Data Selection, Pre-Processing, Transformation, Data Mining, and Evaluation. During preprocessing, case folding, tokenization, token filtering (by length), stopword filtering, and stemming are performed [4].

The Kamus Besar Bahasa Indonesia (KBBI) is used as a reference to ensure optimal data normalization. This stage also includes transforming local language into standard language and removing irrelevant foreign language terms. Local and foreign languages in this context refer to non-standard lexicons, abbreviations, or informal expressions commonly used on social media. The model's performance will be measured using metrics such as accuracy, precision, and recall, aiming to extract patterns or critical information from MobileJKN app user reviews [5].

The expected outcome of this research is to improve the accuracy and precision of sentiment classification models, which can be beneficial in understanding user feedback. These findings have the potential to form the basis for future research on sentiment analysis technologies specific to the healthcare domain and the Indonesian language [6].
Additionally, this study can provide practical insights for app developers, which are essential in ensuring that digital healthcare services are more responsive and data-driven [7].

2. Research Methodology

This study utilizes a quantitative research method with an experimental approach, focusing on sentiment analysis of user reviews for the MobileJKN application available on the Google Play Store. The objective is to explore user perceptions and experiences with the app through a structured classification of sentiment using Neural Network algorithms. The reviews are categorized into three sentiment classes: positive, negative, and neutral. The choice of Neural Networks is based on their ability to handle large, unstructured textual data and identify complex patterns, making them suitable for analyzing diverse and spontaneous user feedback. The experimental approach in this context is used to test hypotheses and validate theoretical assumptions related to user satisfaction and service performance in the MobileJKN app.

The data used in this study is secondary data, sourced from 1,400 user reviews collected through scraping techniques from the Google Play Store. A netnographic approach was applied in the data collection process to ensure authentic, user-generated content that reflects real-life experiences. The data was stored in CSV format to facilitate efficient preprocessing, which includes case folding, tokenization, filtering by word length, stopword removal, and stemming.

The Kamus Besar Bahasa Indonesia (KBBI) was used as the reference lexicon to standardize Indonesian terms, while foreign and non-standard language content was excluded. The analysis process follows the Knowledge Discovery in Databases (KDD) methodology, which consists of several stages: data selection, preprocessing, transformation, data mining, and evaluation. The sentiment classification process using the Neural Network algorithm achieved a 100% accuracy rate, with the Shuffled Sampling method and a 90:10 train-test data split yielding the most optimal results.
Key terms such as “register,” “app,” “number,” “sign in,” and “verify” emerged as the most frequently mentioned words in the reviews, highlighting user concerns related to account access and verification issues. These findings not only demonstrate the effectiveness of Neural Networks in sentiment analysis but also provide actionable insights for app developers to improve service delivery and user experience.
The study was conducted over a four-month period from September to December 2024, allowing for a structured and thorough implementation of all research stages, and offering practical contributions to both academic research and the development of health-related digital services in Indonesia.

3. Results and Discussion

3.1. Results

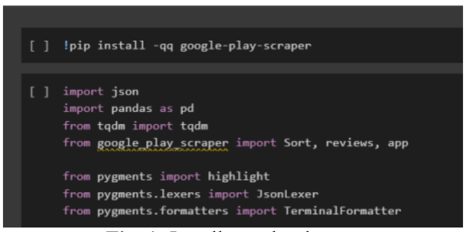

Fig. 1: Install google-play-scraper

```
[ ] app_packages = [
        'app.bpjs.mobile'
    ]

[ ] app_infus = []

    for application in tqdm(app_packages):
      info = app(application, lang='id', country='id')
      del info['comments']
      app_infus.append(info)

[ ] app_reviews = []

    for ap in tqdm(app_packages):
      for score in list(range(1,6)):
        for sort_order in [Sort.MOST_RELEVANT, Sort.NEWEST]:
          rvs, _ = reviews(
              ap,
              lang='id',
              country='id',
              sort=sort_order,
              count=300 if score == 3 else 100,
              filter_score_with= score
          )
          for r in rvs:
            r['sortOrder'] = 'most_relevant' if sort_order == Sort.MOST_RELEVANT else 'newest'
            r['appId'] = ap
          app_reviews.extend(rvs)
```

Fig 2: MobileJKN Data Scraping Process

Table 1: Python Scraping Data Results

No	Parameter	Content
1	Type	Include attributes
2	Attribute filter type	A subset
3	Select subset	ReviewId

Evaluation is a crucial step in assessing the performance of the generated model, as well as manually calculating accuracy, recall, and precision values. The calculation is based on the results obtained from the Split Data operator using the Shuffled Sampling technique with a ratio of 90% training data and 10% testing data. The table below shows the results of applying the 90% training data and 10% testing data ratio using the Split Data operator, demonstrating the Neural Network model's excellent performance with a 100% accuracy rate [8]. All data in each cluster (neutral, positive, and negative) were accurately predicted without any misclassification. Predictions for true cluster_0 (neutral), consisting of 10 data points, true cluster_2 (positive) with 8 data points, and true cluster_1 (negative) with 21 data points, each achieved perfect precision and recall, both at 100%.

This demonstrates the model's optimal performance in distinguishing data based on sentiment and highlights the Neural Network algorithm's capacity to effectively learn patterns and relationships within the dataset. Below are the manual calculations for evaluation: Precision and recall for each cluster were calculated by dividing the number of correctly predicted data points by the total number of data points in the cluster. For cluster_0 (neutral), the model correctly predicted all 10 data points, achieving 100% precision and recall. Similarly, for cluster_2 (positive) and cluster_1 (negative), the model also predicted all data points correctly, resulting in 100% precision and recall for both clusters. These perfect scores demonstrate the model's efficiency in classifying sentiment accurately across all clusters [9].

3.2. Discussion

At this stage, the goal is to identify the most frequently appearing words in user reviews of the MobileJKN application. This process is carried out to study language patterns and identify key words that are dominant in these reviews. This research aims to uncover important words that may reflect user sentiment toward the application, whether positive, negative, or neutral, through frequency data analysis. The identification results will be presented in a table, utilizing the Wordlist to Data operator, Sort, and Filter Example Range in Altair AI Studio [10].

The identification process can be seen in Figure 3.3 below.

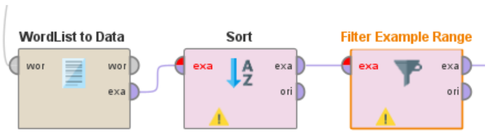

Fig 3: Identification Process

Figure 3 displays the WordList to Data operator used to convert a list of words into a structured dataset. This dataset contains information such as the list of words, the frequency of word occurrences in the document, and the total number of words. The results of WordList to Data are shown in Table 1 below.

Table 2: WordList to Data Results

No	Word	In Document	Total
1	abaikan	1	1
2	abis	7	7
3	aceh	1	1
4	ada	3	3
5	admin	19	25
...	...	...	...
250	entas	1	1
251	error	44	50
252	erti	5	5
...	...	...	...
951	ya	7	7
952	zaman	3	3

After obtaining the results from the WordList to Data operator, the process continues with the use of the Sort operator to sort the frequency of word occurrences from the most frequent to the least frequent (descending order). Next, the Filter Example Range operator is used to select the data to be displayed. In this study, the first row to the 25th row will be shown. Therefore, the results displayed are limited to rows 1 through 25. The process and results of the combination of the Sort and Filter Example Range operators [11].

Table 3: Sort Operator Parameters

No	Parameter	Content
1	sort by	attribute name
		sorting order
	total	descending

Table 4: Filter Example Range Operator Parameters

No	Parameter	Content
1	first example	1
2	last example	25

Table 5: Sort and Filter Example Range Results

No	Word	In Document	Total
1	daftar	431	656
2	aplikasi	467	646
3	nomor	252	420
4	masuk	259	367
5	verifikasi	253	353
6	kode	198	255
7	tolong	227	251
8	baik	193	200
9	pakai	146	197
10	coba	156	184
11	susah	146	169
12	buka	130	164
13	mohon	153	161
14	mudah	137	154
15	bantu	139	146
16	kirim	113	144
17	muncul	119	144
18	udah	117	144
19	kali	119	135
20	bpjs	106	133
21	data	111	128
22	antre	93	121
23	layan	98	115
24	ulang	85	112
25	akun	92	111

Table 5 displays the results of identifying the most frequently occurring words in user reviews of the MobileJKN application. The words "daftar" (register) and "aplikasi" (application) appear most frequently, with 656 and 646 occurrences respectively, followed by words such as "verifikasi" (verification), "masuk" (login), and "nomor" (number). These words indicate that users are primarily focused on the registration and verification processes as well as technical issues. In addition, words like "tolong" (please/help), "bantu" (assist), and "mohon" (request) reflect user appeals for help, while "mudah" (easy) and "baik" (good) suggest positive experiences. On the other hand, words such as "susah" (difficult) and "ulang" (repeat) indicate problems users have encountered [12].
These findings provide an overview of the key concerns and points of satisfaction expressed by users, which can serve as a basis for improving the quality of the MobileJKN application services. The following section presents a word cloud visualization featuring the 25 most frequently mentioned words in user reviews of the MobileJKN app. The size of each word corresponds to its frequency larger words indicate higher frequency [13].

These findings are supported by research conducted by Pal, which indicates that user experiences with digital healthcare applications are often influenced by technical aspects and the ease of use of the application. Additionally, Zhao emphasizes the importance of user experience in digital service environments to enhance satisfaction and trust in digital products.
At this stage, an analysis was conducted to determine which sampling technique yields the highest accuracy for the Neural Network model used in this study. The techniques assessed include Linear Sampling, Shuffled Sampling, and Stratified Sampling, with evaluation based on the model's accuracy for each respective method. The analysis is presented systematically to identify the most effective sampling technique in managing data distribution, thus ensuring optimal model performance. This analysis is displayed in Table 6 to facilitate understanding of the described process [14].

Table 6: Analysis of Sampling Techniques on Model Accuracy

No	Sampling Technique	Ratio	Accuracy (%)
1	Linear Sampling	90 : 10	97.44
		80 : 20	93.59
		70 : 30	89.74
2	Shuffled Sampling	90 : 10	100.00
		80 : 20	94.87
		70 : 30	96.58
3	Stratified Sampling	90 : 10	92.31

80 : 20	96.15
70 : 30	94.87

Based on Table 6 it is evident that the Shuffled Sampling technique with a training and testing data ratio of 90:10 yields the highest accuracy value of 100%, followed by the same technique at a 70:30 ratio with an accuracy of 96.58%. This technique outperforms both Linear Sampling and Stratified Sampling, which show lower accuracy across all data split ratios. These results indicate that Shuffled Sampling can better handle data variation, ensuring a random yet proportional distribution. Therefore, Shuffled Sampling is selected as the most effective sampling technique to support the performance of the Neural Network model in this study. This analysis highlights the importance of choosing the appropriate sampling technique to maximize the results of data classification. An effective sampling technique is crucial for dealing with imbalanced data and improving classification accuracy [15].

Although Stratified Sampling can provide a more proportional data distribution in cluster analysis, it may still face limitations in handling the complexity of textual data. In contrast, this study achieved the best accuracy using Shuffled Sampling [16].
This approach enables the Neural Network model to be trained with a better data distribution, resulting in optimal classification performance. Hence, these findings emphasize the significance of selecting the right sampling method to support the effectiveness of sentiment classification models.

This stage focuses on comparing different train-test data ratios to determine which yields the best accuracy for the Neural Network model. In this study, three sampling techniques were tested Linear Sampling, Shuffled Sampling, and Stratified Sampling using different data split ratios: 90:10, 80:20, and 70:30. The goal is to identify the data split ratio that delivers the best performance based on the accuracy results obtained from each sampling technique [17].

Based on Table 3.7, the 90:10 ratio delivers the best results for the Shuffled Sampling technique, achieving an accuracy of 100%. Meanwhile, for Linear Sampling and Stratified Sampling, although the 90:10 ratio also yields relatively high accuracy 97.44% and 92.31% respectively it still falls short compared to the results obtained using Shuffled Sampling at the same ratio [18].
Overall, the comparison of train-test data ratios shows that the 90:10 ratio provides optimal performance in enhancing the Neural Network model's accuracy, particularly when using Shuffled Sampling as the sampling technique. These results indicate that selecting the optimal train-test data ratio is crucial to ensuring a balanced model training and testing process [19].
Moreover, research by highlights that the 90:10 ratio is commonly used in various applications due to its ability to produce consistent and accurate results, especially in handling complex data such as sentiment analysis. This aligns with the current study's findings, where the best results were achieved using a 90% training and 10% testing ratio. However, the findings also show that despite generally good performance across techniques, Stratified Sampling produced the lowest accuracy in the experiment using the 90:10 ratio [20].

4. Conclusion

This study presents several key findings regarding the sentiment analysis of app reviews based on experimental results. In line with the research questions, the following conclusions can be drawn. First, the sentiment analysis revealed that the most frequently occurring words in user reviews are "daftar" (register), "aplikasi" (application), "nomor" (number), "masuk" (login), and "verifikasi" (verification), indicating that users primarily focus on technical aspects of the app usage process. Second, the application of the Neural Network algorithm, combined with Shuffled Sampling, successfully achieved a maximum accuracy of 100%, demonstrating the effectiveness of this combination in optimizing model performance and accurately classifying sentiment in reviews. Third, the optimal train-test data ratio identified in this study is 90% training data and 10% testing data, which significantly contributed to the model's high accuracy. Based on these findings, several recommendations are proposed to improve user experience, enhance algorithm performance, and refine training methodologies for better outcomes. Firstly, considering the frequent occurrence of technical-related keywords, it is recommended that developers focus on improving the user experience in processes such as registration, verification, and app access. Addressing these concerns through more intuitive features could help reduce the user difficulties frequently mentioned in the reviews. Secondly, while the Neural Network algorithm and Shuffled Sampling have proven highly effective, it is advised to continue testing with larger and more diverse datasets to maintain performance under more complex conditions. Future research may also explore the influence of additional algorithm parameters to enhance processing speed and model efficiency. Furthermore, while the 90:10 train-test data ratio yielded optimal results in this study, it is advisable to continue using this ratio in similar training scenarios. However, testing alternative ratios remains important to evaluate model stability under varying conditions and to understand potential accuracy shifts under different data configurations. Overall, these findings emphasize the significance of technical factors in user satisfaction and the importance of choosing effective sampling techniques and data ratios to support robust sentiment classification. Future work that builds on these insights may further improve the development of digital health applications and their reception among users. In conclusion, this research highlights the value of data-driven approaches in understanding user sentiment and improving app performance. By focusing on the most frequent pain points identified through text analysis and optimizing machine learning methods, developers and researchers alike can better address user needs and enhance the functionality of health-related digital platforms.

References

[1] Acito, F. (2023). *Neural Networks BT - Predictive Analytics with KNIME: Analytics for Citizen Data Scientists* (F. Acito (ed.); pp. 229–254). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45630-5_11

[2] Airikkala, E., Laaksonen, M., Halkoaho, A., & Kaunonen, M. (2023). Perception of inherited risk in type 2 diabetes: a systematic review. *Frontiers in Public Health*, *11*(December). https://doi.org/10.3389/fpubh.2023.1293874

[3] Al-Deen, H. S. S., Zeng, Z., Al-Sabri, R., & Hekmat, A. (2021). An improved model for analyzing textual sentiment based on a deep neural network using multi-head attention mechanism. *Applied System Innovation*, *4*(4). https://doi.org/10.3390/asi4040085

[4] Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R and D Management*, *50*(3), 329–351. https://doi.org/10.1111/radm.12408

[5] Aritonang, P. A., Johan, M. E., & Prasetiawan, I. (2022). Aspect-Based Sentiment Analysis on Application Review using CNN (Case Study : Peduli

Lindungi Application). *Ultima Infosys : Jurnal Ilmu Sistem Informasi*, *13*(1), 54–61.

[6] Astuti, W., Kurniawan, R., & Wijaya, A. Y. (2024). Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *7*(6), 3314–3319. https://doi.org/10.36040/jati.v7i6.8178

[7] Aufa, R. N., Prasetiyowati, S. S., & Sibaroni, Y. (2023). The Effect of Feature Weighting on Sentiment Analysis TikTok Application Using The RNN Classification. *Building of Informatics, Technology and Science (BITS)*, *5*(1), 345–353. https://doi.org/10.47065/bits.v5i1.3597

[8] Chahid, I., Elmiad, A. K., & Badaoui, M. (2023). Data Preprocessing For Machine Learning Applications in Healthcare: A Review. *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 1–6. https://doi.org/10.1109/SITA60746.2023.10373591

[9] Chakraborty, S., Islam, S. K. H., & Samanta, D. (2022). *Introduction to Data Mining and Knowledge Discovery BT - Data Classification and Incremental Clustering in Data Mining and Machine Learning* (S. Chakraborty, S. H. Islam, & D. Samanta (eds.); pp. 1–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-93088-2_1

[10] Hadwan, M., Al-Sarem, M., Saeed, F., & Al-Hagery, M. A. (2022). An Improved Sentiment Classification Approach for Measuring User Satisfaction toward Governmental Services' Mobile Apps Using Machine Learning Methods with Feature Engineering and SMOTE Technique. *Applied Sciences (Switzerland)*, *12*(11). https://doi.org/10.3390/app12115547

[11] Haron, N. H. bin. (2022). Stratified sampling using cluster analysis. *AIP Conference Proceedings*, *2472*(1), 50012. https://doi.org/10.1063/5.0092740

[12] Hasanah, K. (2024). Comparison of Sentiment Analysis Model for Shopee Comments on Google Play Store. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *13*(1), 21–30. https://doi.org/10.32736/sisfokom.v13i1.1916

[13] Heisler, G. V., & Assunção, J. V. C. (2024). *A Aplicação do Processo de KDD aos Dados da COVID-19: Um Estudo de Caso no Rio Grande do Sul, Brasil*. 91–100. https://doi.org/10.5753/erbd.2024.238871

[14] Irwandi, Santoso, S., Sakroni, Lukitasari, M., & Hasan, R. (2022). School-community Collaboration in Inquiry-based Learning to Strengthen Religious Character and Improve Learning Outcome of Students. *International Journal of Instruction*, *15*(3), 913–930. https://doi.org/10.29333/iji.2022.15349a

[15] Ito, T., Tsubouchi, K., Sakaji, H., Yamashita, T., & Izumi, K. (2020). Word-level contextual sentiment analysis with interpretability. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 4231–4238. https://doi.org/10.1609/aaai.v34i04.5845

[16] Jamil, R., Ashraf, I., Rustam, F., Saad, E., Mehmood, A., & Choi, G. S. (2021). Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/peerj-cs.645

[17] Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining*, *15*(4), 531–538. https://doi.org/10.1002/sam.11583

[18] Krosuri, L. R., & Aravapalli, R. S. (2023). Novel heuristic-based hybrid ResNeXt with recurrent neural network to handle multi class classification of sentiment analysis. *Machine Learning: Science and Technology*, *4*(1). https://doi.org/10.1088/2632-2153/acc0d5

[19] Priyadarshini, I., & Cotton, C. (2021). A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. *Journal of Supercomputing*, *77*(12), 13911–13932. https://doi.org/10.1007/s11227-021-03838-w

[20] Putra, R. A., Novita, R., Ahsyar, T. K., & Zarnelly. (2024). Implementation of Classification Algorithm for Sentiment Analysis: Measuring App User Satisfaction. *Teknika*, *13*(2), 204–212. https://doi.org/10.34148/teknika.v13i2.827

[21] Qureshi, A. A., Ahmad, M., Ullah, S., Yasir, M. N., Rustam, F., & Ashraf, I. (2023). Performance evaluation of machine learning models on large dataset of android applications reviews. *Multimedia Tools and Applications*, *82*(24), 37197–37219. https://doi.org/10.1007/s11042-023-14713-6

[22] Wahyudi, D., & Sibaroni, Y. (2022). Deep Learning for Multi-Aspect Sentiment Analysis of TikTok App using the RNN-LSTM Method. *Building of Informatics, Technology and Science (BITS)*, *4*(1), 169–177. https://doi.org/10.47065/bits.v4i1.1665