Comparative Analysis of Machine Learning Algorithms for Sentiment Analysis on Instagram App Reviews: A Guideline for Enhancing User Experience

Nutthapat Kaewrattanapat[1], Mongkolchai Tiansoodeenon[2,*] and Jarumon Nookhong[3]

[1]Department of Digital Technology for Education, Faculty of Education, Suan Sunandha Rajabhat University

[2]Department of English and Linguistics, Faculty of Humanities, Ramkhamhaeng University

[3]Department of Digital Innovation Management, Faculty of Science and Technology, Suan Sunandha Rajabhat University

Email: nutthapat.ke@ssru.ac.th, mongkolchai.t@rumail.ru.ac.th, jarumon.no@ssru.ac.th

*Corresponding author

Abstract.

This study focuses on a comparative analysis of machine learning algorithms for classifying and clustering Instagram app reviews from the Google Play Store, aimed at understanding user sentiments and improving app performance. The dataset, obtained from Kaggle, consists of 210,542 reviews, with a randomly selected sample of 1,103 reviews used for analysis. Four algorithms were applied: Naïve Bayes, k-Nearest Neighbors (k-NN), k-Means, and k-Medoids. To evaluate their performance, 12 models were tested using different vectorization techniques such as TF-IDF, Term Occurrences, Binary Term Occurrences, and n-grams. The results show that Naïve Bayes, combined with TF-IDF and 2-grams, delivered the highest accuracy for sentiment classification at 99.46%. This high accuracy can be attributed to Naïve Bayes' effectiveness in handling probabilistic distributions in text data. In the clustering analysis, k-Means with 2-grams and stemming produced the most distinct groupings, with a Davies-Bouldin index of -7.279, successfully separating positive and negative sentiments. This research provides a robust framework for conducting Sentiment Analysis on large-scale user reviews. By applying these machine learning techniques, developers and app managers can gain valuable insights into user satisfaction and identify areas for improvement. Ultimately, these methods can help enhance user experience and drive app success through a deeper understanding of user feedback.

Keywords: Sentiment Analysis, Machine Learning, Naïve Bayes, k-Means Clustering, Instagram Reviews

1. Introduction

In an era where technology and social media play a pivotal role in sharing experiences, the concepts and utilization of social media applications profoundly impact contemporary culture and human behavior. They have become essential components in connecting and fostering diverse relationships among people worldwide. Widely recognized and highly popular applications such as Instagram (Amrina et al., 2024) —a platform for sharing photos and videos—exemplify this influence. To comprehend user engagement and satisfaction with this application, the present study aims to analyze and compare clustering and classification techniques applied to review datasets. Employing methods such as Naïve Bayes, k-Nearest Neighbors (k-NN), k-Means (Ahmad et al., n.d.) (Lallogo, 2024), and k-Medoids, the study analyzes the relationship between review texts and user satisfaction levels using datasets sourced from Kaggle. These datasets comprise review texts and five levels of ratings from Instagram users, facilitating an understanding of user trends and opinions toward the application.

The Kaggle dataset contains reviews of the Instagram application from the Google Play Store (Jhalani, 2023), totaling 210,542 samples. A probabilistic random sampling at a rate of 0.005 was conducted, maintaining the proportional representation of all five satisfaction rating levels. This resulted in a sample dataset of 1,103 instances used for analyzing and comparing the accuracy of clustering and classification techniques on the review data.

The outcomes of this study will showcase the analysis and comparative accuracy of various methods employed in clustering and classifying review data. This will aid in determining the most appropriate technique for analyzing Instagram application reviews. Additionally, the findings serve as a guideline for application developers and social media administrators to interpret review texts and ratings, thereby assisting in the enhancement of the application to provide users with the most appropriate and efficient services.

2. Research Objective

2.1 To analyze and explore samples from the dataset of Instagram application reviews on the Google Play Store.

2.2 To analyze and compare the accuracy of text classification techniques applied to sample data from Instagram application reviews on the Google Play Store, specifically between Naïve Bayes and the k-Nearest Neighbors (k-NN) algorithm.

2.3 To analyze and compare the Davies-Bouldin Index values obtained from clustering techniques applied to sample text data from Instagram application reviews on the Google Play Store, specifically between k-Means and k-Medoids algorithms.

3. Data and Methodology

The dataset utilized for this analysis consists of Instagram application reviews from the Google Play Store, obtained from Kaggle. This dataset compiles review texts from users of the Instagram application on smartphones, including user experiences and sentiments (review descriptions), review dates, and satisfaction levels across five ratings ranging from one to five stars, as shown in Table 1. This dataset aids in understanding user satisfaction, serving as a means to evaluate the application's performance and to identify emerging trends associated with the Instagram application.

Table 1: Structure of the Instagram Application Review Dataset on the Google Play Store

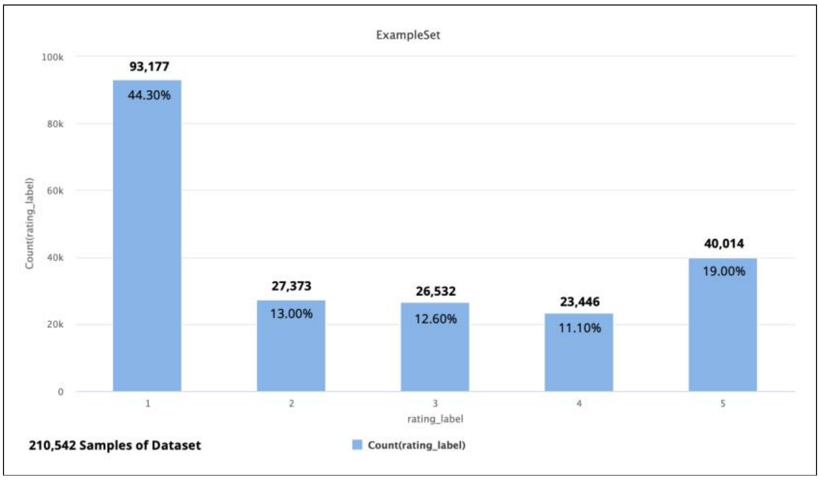
Attribute Name	Data Type	Example
review description	text	I like this app a lot, but the messaging part of the app is very frustrating. I sometimes won't get notifications when my friends respond to me. One time recently, I sent my friend a reel, and they responded, and then I went to see what they said, and the reel I sent was gone on my end, but they could still see them. Sometimes, I'll send messages, then leave the app, then go back, and it said I never sent the message. Wish the bugs would be fixed because I've dealt with this for a while...
rating	nominal	1, 2, 3, 4, 5
review date	datetime	2023-07-11 23:57:07

Source: Instagram Play Store. (2023). Retrieved from https://www.kaggle.com/datasets/saloni1712/instagram-play-store-reviews

The dataset contains a total of 210,542 samples, consisting of user review texts accompanied by star ratings, indicating five levels of satisfaction.

This breakdown shows that a significant portion of the reviews falls within the 1-star category, followed by the 5-star category, indicating a polarizing trend in user satisfaction. These proportions will be represented visually in Figure 1, providing a clearer understanding of the overall sentiment and feedback distribution among Instagram users.

Figure 1: Proportion of All Samples in the Review Dataset Categorized by 5 Rating Levels (Rating 1 - 5)



Source: Instagram Play Store. (2023). Retrieved from https://www.kaggle.com/datasets/saloni1712/instagram-play-store-reviews

3.1 Data Preprocessing

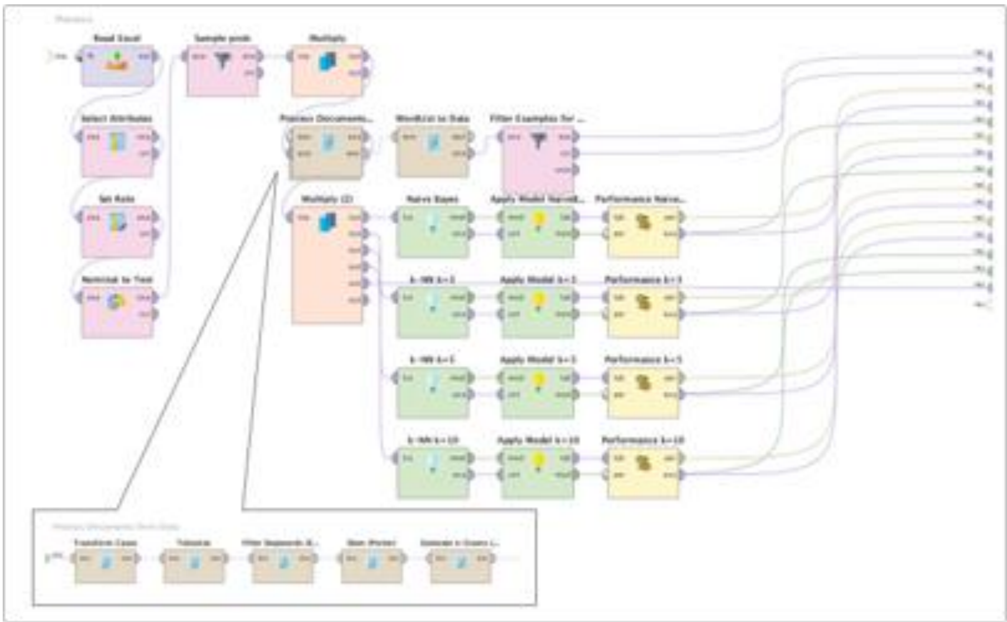
Text preprocessing steps include converting text to lowercase, removing stop words, and applying stemming. A variety of vectorization techniques were employed, such as TF-IDF, Term Frequency (TF), and Term Occurrence (TO), to prepare the text for analysis (Kularbphettong, 2019).

3.2 Classification Techniques

Naïve Bayes and k-NN (with varying k values) were applied to classify the text reviews. The accuracy of these techniques was compared based on various n-gram models (1-gram, 2-

gram), with Naïve Bayes achieving a maximum accuracy of 99.46% when used with 2-grams and TF-IDF vectorization as shown in Figure 2.

Figure 2: The Process of Classifying Sample Text Data from Instagram Application Reviews on the Play Store Using Naïve Bayes and k-Nearest Neighbors (k-NN)

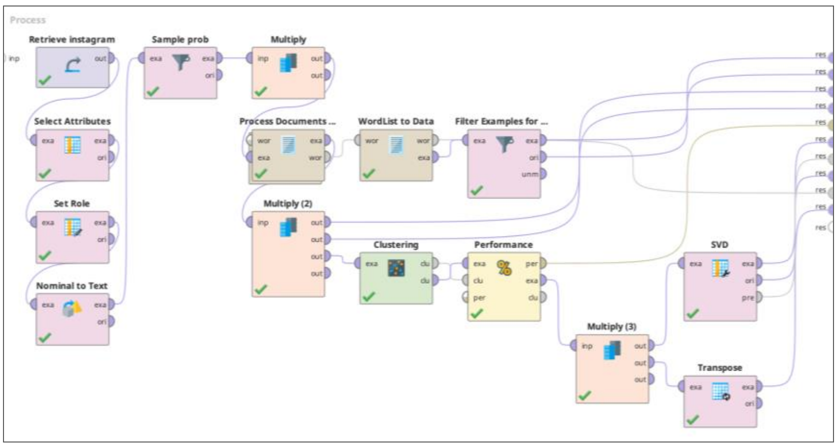


3.3 Clustering Techniques

Clustering was performed using k-Means and k-Medoids, with the number of clusters determined through the Davies-Bouldin Index. The results showed that k-Means with 2-grams and stemming provided the most distinct clusters of user reviews, particularly distinguishing between positive and negative sentiments.

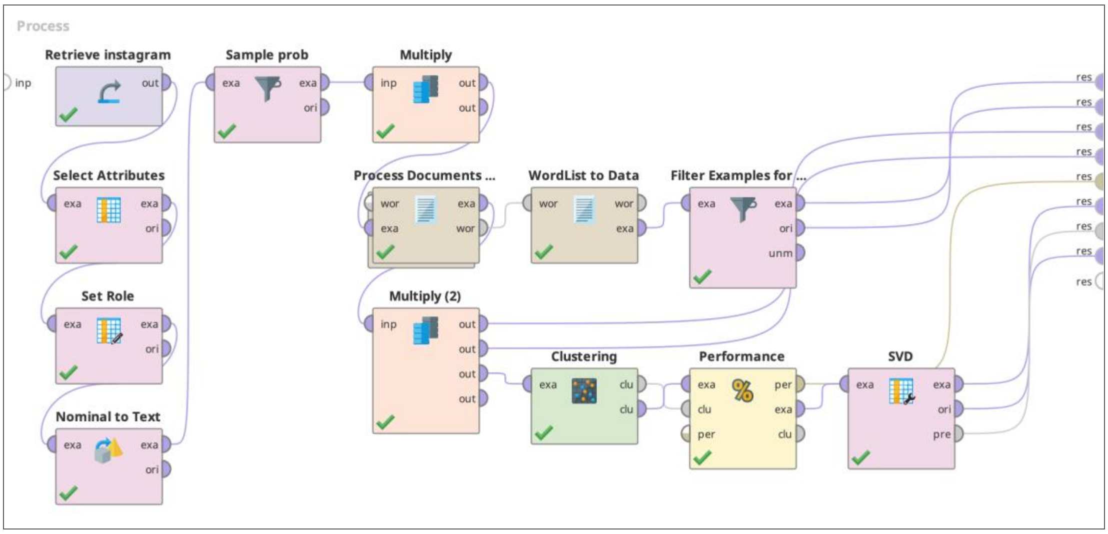
The process of analyzing and comparing the Davies-Bouldin Index from clustering techniques applied to sample text data from Instagram application reviews on the Play Store between k-Means and k-Medoids.

Figure 4: The Process of Cluster Analysis Using k-Means Clustering



Following the analysis and comparison of accuracy from text classification techniques applied to Instagram application reviews on the Play Store using Naïve Bayes and k-Nearest Neighbors (k-NN), we proceeded to analyze and compare the Davies-Bouldin Index from clustering techniques, specifically between k-Means and k-Medoids.

Figure 5: The Process of Determining the Optimal Number of Clusters for k-Medoids Clustering



4. Results and Discussion

4.1 Classification Results

Naïve Bayes outperformed k-NN in text classification, particularly when applied with TF-IDF vectorization. The accuracy of the Naïve Bayes model reached up to 99.46%, indicating its robustness in identifying user sentiments based on review text. The comparison of accuracy results is presented in the following table.

Table 2: Accuracy Values for Each Model from the Analysis of 1,103 Sample Groups

Model	**TF-IDF**		**Term Frequency (TF)**		**Term Occurrences (TO)**		**Binary Term Occurrences (Bi-TO)**	
	1-gram	**2-grams**	**1-gram**	**2-grams**	**1-gram**	**2-grams**	**1-gram**	**2-grams**
Naïve Bayes	92.11%	**99.46%*** **	91.66%	99.37%**	89.21%*	99.18%**	89.30%	99.18%**
k-NN (k=3)	95.47%*	95.47%	93.56%*	93.74%	89.12%	90.39%	89.85%*	89.21%
k-NN (k=5)	72.44%	71.89%	69.99%	71.26%	61.38%	48.59%	57.48%	36.63%
k-NN (k=10)	72.44%	71.89%	69.99%	71.26%	61.38%	48.59%	57.48%	36.63%

1,103 Samples (Probability = 0.005) from 210,542 Samples of Dataset

**** Maximum accuracy, ** Maximum accuracy value among n-grams groups,*

** Maximum accuracy value within a group of n-grams.*

Table 3: Accuracy Values for Each Model from the Analysis of 1,103 Sample Groups, Including Stemming

Model	TF-IDF		Term Frequency (TF)		Term Occurrences (TO)		Binary Term Occurrences (Bi-TO)	
	1-gram	2-grams	1-gram	2-grams	1-gram	2-grams	1-gram	2-grams
Naïve Bayes	86.85%	**99.37%*** **	86.04%	99.27%**	83.14%	99.00%**	82.96%	99.00%**
k-NN (k=3)	94.83%*	93.56%	92.84%*	94.83%	90.48%*	90.03%	88.40%*	89.30%
k-NN (k=5)	71.35%	69.99%	69.63%	70.17%	64.46%	50.86%	62.19%	38.71%
k-NN (k=10)	71.35%	69.99%	69.63%	70.17%	64.46%	50.86%	62.19%	38.71%

Table 4: Accuracy Values for Each Model from the Analysis of 1,103 Sample Groups, Including Term Importance Reduction Using the Percentual Method (Below 3% and Above 30%)

Model	TF-IDF		Term Frequency (TF)		Term Occurrences (TO)		Binary Term Occurrences (Bi-TO)	
	1-gram	2-grams	1-gram	2-grams	1-gram	2-grams	1-gram	2-grams
Naïve Bayes	39.71%	41.80%	39.53%	42.43%	38.17%	38.44%	37.17%	38.71%
k-NN (k=3)	88.03%***	86.67%*	85.77%**	84.41% *	84.41%**	84.41%**	84.41%*	84.59%**
k-NN (k=5)	67.09%	67.82%	65.46%	66.18%	65.55%	65.37%	66.00%	65.91%
k-NN (k=10)	67.09%	67.82%	65.46%	66.18%	65.55%	65.37%	66.00%	65.91%

Table 5: Accuracy Values for Each Model from the Analysis of 1,103 Sample Groups, Including Stemming and Term Importance Reduction Using the Percentual Method (Below 3% and Above 30%)

Model	TF-IDF		Term Frequency (TF)		Term Occurrences (TO)		Binary Term Occurrences (Bi-TO)	
	1-gram	2-grams	1-gram	2-grams	1-gram	2-grams	1-gram	2-grams
Naïve Bayes	40.89%	43.43%	40.44%	43.25%	39.98%	40.34%	38.17%	38.89%
k-NN (k=3)	90.75%*	**91.39%*** **	89.12%*	89.30%**	87.85%*	88.30%**	88.76%**	87.76%*
k-NN (k=5)	68.00%	67.09%	69.54%	68.00%	66.73%	66.46%	66.27%	66.09%
k-NN (k=10)	68.00%	67.09%	69.54%	68.00%	66.73%	66.46%	66.27%	66.09%

4.2 Clustering Results

The Davies-Bouldin Index revealed that both k-Means and k-Medoids formed effective clusters of user reviews. However, k-Means consistently showed better performance with lower Davies-Bouldin values, especially when stemming and 2-gram models were applied.

Table 6: Summary of the Number of Clusters

Number of clusters	No pruning		Pruning 3%-30%	
n-gram	No stemming	Stemming	No stemming	Stemming
1-gram	10	10	10	7
2-grams	6	6	6	9

Table 7: Davies-Bouldin Index Values and Number of Clusters from k-Means Clustering Analysis

Davies Bouldin	Pruning 3%-30%			
k (Number of clusters)	No stemming		Stemming	
	1-gram	2-grams	1-gram	2-grams
2	-6.544	-5.352	-7.255	-7.279*
3	-5.659	-5.764	-6.199	-6.412
4	-4.975	-5.584	-5.597	-5.839
5	-5.224	-4.528	-5.508	-5.674
6	-4.623	-4.468	-5.172	-5.000

7	-4.430	-4.735	-4.805	-5.097
8	-4.419	-4.362	-4.494	-4.848
9	-4.075	-3.867	-4.409	-4.476
10	-4.162	-3.862	-4.173	-4.233

Table 7 shows the Davies-Bouldin Index, which is used to evaluate the effectiveness of k-Means clustering. The lower the index value, the more efficient the clustering. The table presents different tests involving the use of term importance reduction (pruning) and stemming. The results indicate that using 2-grams, with stemming and term importance reduction, achieved the lowest Davies-Bouldin Index of -7.279, demonstrating the most effective clustering in this case.

Table 8: Davies-Bouldin Index Values and Number of Clusters from k-Medoids Clustering Analysis

Davies Bouldin	**Pruning 3%-30%**			
k (Number of clusters)	**No stemming**		**Stemming**	
	1-gram	**2-grams**	**1-gram**	**2-grams**
2	-2.216*	-1.901	-1.853	-1.862
3	-1.846	-1.860	-1.921	-2.000
4	-1.838	-1.836	-1.967	-1.956
5	-1.850	-1.812	-1.871	-1.905
6	-1.930	-1.859	-1.837	-1.874
7	-1.850	-1.796	-1.853	-1.858
8	-1.846	-1.822	-2.033	-1.833
9	-1.845	-1.826	-1.798	-1.902
10	-1.792	-1.777	-1.839	-1.883

Figure 8: Example of Clustering Results Using k-Medoids (k=2)

Figure 9: Word Groups from Clustering Analysis Using k-Medoids (k=2) with 2-grams

5. Conclusion

This article focuses on the analysis and comparison of text classification and clustering techniques applied to a dataset of user reviews, using methods such as Naïve Bayes, k-NN, k-Means, and k-Medoids. These techniques were implemented alongside text preprocessing steps, including word segmentation, stop word removal, stemming, n-gram generation, and four different vectorization methods (TF-IDF, Term Occurrences, Term Occurrences, and Binary Term Occurrences). The objective was to analyze the relationship between review texts and user satisfaction levels using a dataset sourced from Kaggle. This dataset consists of 210,542 user reviews and five levels of satisfaction ratings from Instagram users. A probabilistic sampling method with a probability of 0.005 was applied, maintaining the proportion of ratings, resulting in a sample of 1,103 reviews for analysis.

The classification comparison between Naïve Bayes and k-NN revealed that Naïve Bayes, combined with TF-IDF vectorization and 2-gram models, achieved the highest accuracy (Accuracy = 99.46%). In terms of clustering techniques, the comparison between k-Means and k-Medoids demonstrated that k-Means, combined with stemming and 2-grams, and k-Medoids, with 1-grams, were both effective at forming two distinct clusters (k-Means: Davies-Bouldin Index = -7.279, k-Medoids: Davies-Bouldin Index = -2.216). Upon examining the terms in these clusters, one group consisted of positive sentiment words such as "good," "love," "great," and "I love," while the other group contained negative sentiment words such as "fix," "problem," "please fix," and "issue."

References

Amrina, Nur, Aida., Primandani, Arsi., Ranggi, Praharaningtyas, Aji., Tarwoto, Tarwoto. (2024). 10. Analisis Sentimen Pengguna Aplikasi Instagram Pada Situs Google Play Menggunakan Metode Naïve Bayes. Jurnal media informatika Budidarma, doi: 10.30865/mib.v8i2.7388

Ahmad, M., Riadi, I., & Prayudi, Y. (n.d.). Cyberbullying Analysis on Instagram Using K-Means Clustering. [No publication details available]

Jhalani, S. (2023). Instagram Play Store reviews [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/saloni1712/instagram-play-store-reviews

Kularbphettong, K. (2019). Developing the Thai regional dialect based on semi-automatic technique. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(2), 2842-2846. https://doi.org/10.35940/ijrte.B1986.078219

LALLOGO, Lassané. (2024). 3. KMEANS.KNN: KMeans and KNN Clustering Package. doi: 10.32614/cran.package.kmeans.knn