Identification of Primary Complaints from Bank Saqu Users in Google Play Reviews Using the K-Means Clustering Algorithm

Ari Anshari Yuliansyah Siregar, Muhammad Iqbal

Abstract

Bank Saqu is a digital bank owned by Astra Financial and WeLab, with all of its services accessible via a mobile app; as a result, user reviews on the Google Play Store serve as the primary source of feedback. These reviews consist of unstructured text written in informal language, and their number continues to grow, while sentiment analysis only indicates polarity without specifying the nature of the complaints. This study aims to identify the main complaints of Bank Saqu users using the K-Means clustering algorithm. The research data consists of 2,000 reviews from April 29 to November 7, 2026, and 535 reviews (26.75%) with ratings of 1 and 2 were selected as complaint reviews. Following the Knowledge Discovery in Databases (KDD) workflow, the reviews were processed through case folding, cleaning, tokenization, normalization, negation merging, stopword removal, Sastrawi stemming, and filtering; they were then weighted using TF-IDF for unigrams and bigrams, resulting in a matrix of 477 reviews and 483 features. The number of clusters was tested for K = 2 through 10 using the Elbow Method, Silhouette Score, and Davies-Bouldin Index (DBI), and K = 8 was selected as the value with the highest Silhouette Score within the range of K = 3 to 8. The results showed that the most common complaints were related to account balances, payments, and transactions (182 reviews, 38.2%), followed by login and facial verification (24.7%), loans and Saku Kredit (20.3%), and the app failing to open (16.8%). The Silhouette Score of 0.0263 and DBI of 4.7568 indicate overlapping clusters, as is common in short review texts; therefore, the labels for each cluster were validated using centroid keywords and representative reviews. These findings provide a data-driven basis for prioritizing service improvements at Bank Saqu.

***Keywords**:* Bank Saqu, K-Means, TF-IDF, reviews, clustering

Ari Anshari Yuliansyah Siregar[1]
[1]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: arianshari32@gmail.com[1]

Muhammad Iqbal[2]
[2]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: muhammadiqbal@dosen.pancabudi.ac.id [2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

A digital bank is a bank that conducts its business primarily through electronic channels, with no physical branches other than its headquarters or with only a limited number of physical branches [1]. One of the rapidly growing digital banks in Indonesia is Bank Saqu, which was launched on November 20, 2023, by Astra Financial and WeLab through PT Bank Jasa Jakarta. The entity changed its name to PT Bank Saqu Indonesia on May 21, 2025, and had 3.2 million customers as of November 2025, approximately 40% of whom were solopreneurs [2].

For digital banks, mobile apps are the primary point of contact between the bank and its customers, so the quality of the app is crucial to the customer experience. The Google Play Store features a reviews and ratings section that serves as a direct channel for user feedback[3]. Reviews are written spontaneously and contain real-life experiences, making them a valuable source of information for developers to identify issues and prioritize improvements[4]. However, reviews consist of unstructured text that uses informal language, abbreviations, and typos, and their volume continues to grow, making it impossible to read and summarize them manually[5]

The most commonly used approach for analyzing banking app reviews is sentiment analysis[6]. Pratiwi applied the Naive Bayes algorithm to reviews of Bank Saqu and found that 24.5% of the reviews had negative sentiment[7], while Purnamasari and Thaha found that app stability and service quality were key factors in user satisfaction with five digital banks in Indonesia[8]. Sentiment analysis provides an overview of satisfaction levels, but it does not explain the specific nature of the complaints. App managers still need to read negative reviews one by one to determine whether the complaints relate to the login process, facial verification, transfers, or app performance[9].

Clustering is an unsupervised learning technique that groups data based on similarity without requiring labels [10]. This technique is suitable for identifying complaint themes because it does not require manual labeling, which is costly and subjective. K-Means [11] is widely used because it is simple, efficient, and its results are easy to interpret through the centroids of each cluster [12]. Several studies have applied K-Means with TF-IDF weighting to Google Play reviews—including those for the BCA Mobile [13], LINE Bank [14], and Tokopedia [15] apps—and have shown that K-Means is capable of grouping reviews into clusters based on dominant issues.

Based on the author's review, no study has yet been found that specifically applies the K-Means algorithm to Bank Saqu user complaints. A preliminary analysis of 2,000 Bank Saqu reviews on the Google Play Store revealed that 535 reviews (26.75%) gave a rating of 1 or 2, with diverse content ranging from the login and facial verification processes, delayed transactions, loan applications, to app performance. This diversity highlights the need for automatic clustering to objectively and measurably identify the main complaints. Identifying Key Complaints from Bank Saqu Users in Google Play Reviews Using the K-Means Clustering Algorithm by specifically clustering Bank Saqu user complaint reviews using K-Means, supplemented by normalization of non-standard banking terms, determination of the number of clusters based on several metrics, and validation of cluster labels.

Literature Review

K-Means is a partition-based clustering algorithm that divides data into K clusters by assigning each data point to the cluster with the nearest centroid [16]. This algorithm minimizes the variance within clusters and maximizes the variance between clusters [17]. Aulia et al. applied K-Means using Python and Jupyter Notebook to cluster patient medical record data based on age, blood pressure, and diagnosis, and identified clusters of high-intensity (49 members), moderate-intensity (28 members), and low-intensity (23 members) diseases [18]. . clustered 1,005 household data points from Bina Baru Village and used the Davies-Bouldin Index to determine the optimal value of k as 3, representing middle-, high-, and low-income households [19]. applied

K-Means to students' socioeconomic, behavioral, and psychological factors, then recommended intervention programs for students in specific clusters [20]. These studies demonstrate that K-Means results can be directly translated into a basis for decision-making.

In review text data, K-Means is generally combined with TF-IDF weighting. Kharisma and Aesyi applied TF-IDF and K-Means to Twitter data regarding the MyPertamina app and identified a single cluster containing public complaints about the app [13]. Sunarko et al. clustered 662 BCA Mobile reviews from Google Play into 10 clusters with a Silhouette Score of 0.1027 [21]. Oktaviani clustered Gojek and Grab reviews into three clusters with maximum Silhouette Scores of 0.5059 and 0.4704, respectively [22]. For non-banking apps, Subagja clustered 8,335 Tokopedia reviews using TF-IDF and the Elbow method into four complaint clusters: customer service, payments and refunds, shipping, and promotions and shipping costs; the payments and refunds cluster was the most dominant, accounting for 49.88% of the reviews, with a Silhouette Score of 0.0500 [23]. A combination of TF-IDF and K-Means was also applied to reviews of LINE Bank and online lending apps as evaluation material for service providers[24]

Several studies have compared K-Means with other algorithms or examined methods for determining the number of clusters. Khairuna and Razi compared K-Means and K-Medoids for clustering review texts from the Cookpad app, while Sudrajat et al. compared the Elbow and Silhouette methods to determine the optimal number of clusters and assess cluster quality [25]

K-Means with TF-IDF has proven effective at mapping app reviews into interpretable problem clusters. However, there are three points to note. First, the Silhouette score for short reviews tends to be low (0.05 to 0.10), so determining the value of K requires combining several metrics and considering the clarity of the themes. Second, studies on banking apps generally group all reviews together, so clusters may be mixed with positive reviews. Third, research on Bank Saqu has so far been limited to sentiment analysis.

Based on the discussion above, K-Means combined with TF-IDF has proven effective in mapping app reviews into interpretable problem clusters. However, research on banking apps generally groups all reviews together, resulting in clusters that may be mixed with positive reviews; furthermore, based on the author's literature review, no previous study has specifically examined complaints from Bank Saqu users. This study addresses that gap. This study clusters complaint reviews using K-Means with TF-IDF representation, determines the optimal number of clusters, and interprets each cluster as a complaint theme to serve as a basis for prioritizing service improvements.

Research Methodology

This is descriptive-exploratory quantitative research based on text mining. It is exploratory because it aims to discover complaint patterns that are not known in advance, and it uses unsupervised learning, so no class labels or statistical hypothesis tests are required. The object is the user reviews of the Bank Saqu application on the Google Play Store, as shown in Figure 1.

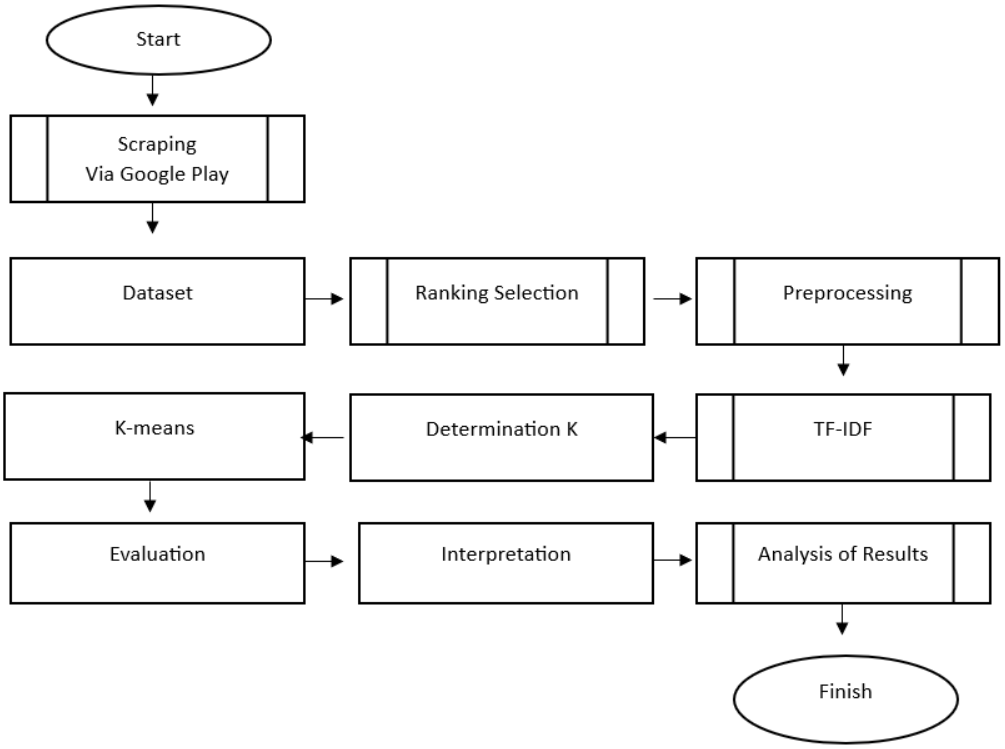


Figure 1. Research Flow

In this stage, reviews of the Bank Saqu app were collected from the Google Play Store and analyzed to understand their characteristics. The rating selection stage identified reviews with ratings of 1 and 2 as complaint reviews. The preprocessing stage cleaned and standardized the review text. The TF-IDF stage converts the cleaned text into numerical vectors. The K-determination stage tests the number of clusters, ranging from K = 2 to 10, and the K-Means stage forms clusters based on the selected K. The evaluation stage measures cluster quality; the interpretation stage assigns theme labels to each cluster; and the results analysis stage ranks the main complaints as the basis for recommendations. The entire process was implemented in Python on Jupyter Notebook using the pandas, NumPy, Sastrawi, scikit-learn, matplotlib, seaborn, and wordcloud libraries.

3.1 Scraping

Scraping is an automated technique for retrieving and extracting data from web pages. In this study, data was scraped using the Saqu Bank app from the Google Play Store.

3.2 Dataset

A total of 2,000 recent reviews of the Bank Saqu app from April 29 to November 7, 2026, are stored in the file banksaqu-reviews.xlsx. Preprocessing

	reviewId	content	score	thumbsUpCount	at	appVersion
0	71ec305d-057d...	mantaplah..pokoknya the best..bunganya rendah..tenornya panjan...	5	0	2026-11-07 02:28:50	26.10.1
1	2b1f2e9b-2bfd...	saya uninstall aplikasi ini ya, jika suatu saat terjadi masala...	2	0	2026-11-07 00:13:04	26.10.2
2	29b2b126-73c4...	oke baik , mantap	5	0	2026-11-07 00:04:39	26.10.2
3	fba91856-f525...	sangat baik	5	0	2026-11-06 17:36:53	26.10.2
4	f88cc6f7-82c7...	gajelas tiba-tiba keluar dari akun lawak lo	1	0	2026-11-06 15:25:58	26.10.2
5	4bf40494-55d6...	ini kok gk bisa login d jam9 malem?	1	0	2026-11-06 15:11:28	26.10.2
6	4042a1b5-9e4d...	Saldo deposito tidak bisa ditarik, terlalu lama untuk pemecaha...	1	0	2026-11-06 11:31:02	26.10.2
7	cf5213d1-fe4e...	Menurut saya sebagai nasabah dan pengguna aplikasinya Bank Saq...	5	0	2026-11-06 07:40:39	
8	1fa9c70c-2ada...	untuk sementara kasih bintang 1 karna belum ada fitur saku pin...	1	0	2026-11-06 07:25:15	26.10.2
9	8de4f7b3-d6aa...	baik	5	0	2026-11-06 07:24:37	26.10.2

Figure 2. Dataset

3.3 Rating Selection

Reviews were specifically filtered for 1- and 2-star ratings so that the analysis would focus on user complaints, issues, and negative feedback. 3-star reviews were excluded because they often contain a mix of praise and criticism. This stage yielded 535 complaint reviews, which served as input for the preprocessing stage.

3.4 Preprocessing

Preprocessing is a technique for improving algorithm performance by converting raw text into a format ready for processing [19]. The steps involved are as follows.

1. **Case folding**, which involves converting all letters to lowercase so that identical words are not treated as different due to capitalization.
2. **Cleaning**, which involves removing URLs, email addresses, numbers, punctuation, symbols, and emojis using regular expressions, as well as reducing letters that repeat three or more times ("bgttt" becomes "bgt").
3. **Tokenization**, which involves splitting sentences into tokens based on spaces.
4. **Normalization**, which involves replacing non-standard words, abbreviations, and slang terms using a normalization dictionary compiled by the researchers from frequently occurring terms—for example, "gk" and "gak" become "tidak," "gabisa" becomes "tidak bisa," "tf" becomes "transfer," "vermuk" becomes "verifikasi wajah," and "ngendap" becomes "tertahan"; profanity is removed.
5. **Negation merging**, which involves combining the words "tidak," "bukan," "belum," and "jangan" with the following word ("tidak bisa" becomes "tidak_bisa") so that the negation is not removed during the stopword removal stage.
6. **Stopword removal**: This involves removing common Indonesian stopwords, domain-specific stopwords that appear in nearly all reviews ("bank," "saqu," "astrapay," "app"), general evaluative terms ("good," "bad"), and tokens two letters or shorter.
7. **Stemming**, which involves converting inflected words back to their base forms using Sastrawi, which implements a confix-stripping algorithm [20]—for example, "kejelasan" becomes "jelas"—after which stopword removal is repeated. At this stage, Sastrawi removes underscores, so that negated phrases revert to two consecutive words ("tidak bisa"), which are then captured as bigrams by TF-IDF.
8. **Filtering**: this involves discarding reviews that contain fewer than two tokens and reviews that are identical after cleaning.

3.5 TF – IDF

The cleaned text was converted into a TF-IDF matrix using the formula with unigrams and bigrams, a minimum document frequency threshold of 3% and a maximum of 80%, sublinear frequency scaling, and L2 normalization. The keywords obtained from the centroid weights were then used to interpret the clusters.

3.6 Determination K

The value of K is not determined arbitrarily—to ensure that the resulting clusters are neither too general nor too fragmented—but is mathematically tested using evaluation methods such as the Elbow Method and the Silhouette Score. In the Elbow Method, the researcher runs the K - Means algorithm with various values of K in sequence (for example, from K=2 to K=10) to calculate the Within-Cluster Sum of Squares (WCSS)—the total squared distance between each review and its cluster's centroid—and then select the value of K that forms a sharp break resembling an elbow, at which point the decline in the error value begins to level off significantly

3.7 K Means

The process begins with a text preprocessing stage to remove noise and extract features using the TF-IDF method, so that all reviews are converted into numerical vector representations. Once the data is in numerical form, the K-Means algorithm begins by determining the desired number

of clusters (K) and randomly selecting a number of initial centroids. Next, the system calculates the mathematical distance between each review vector and each

3.8 Evaluation

The evaluation phase was conducted to measure the validity and quality of the complaint groups formed by the K-Means algorithm at the optimal number of clusters (K). Since the Bank Saqu Google Play review data lacked initial category labels, the testing was performed using internal validation methods, namely the Silhouette Coefficient and the Davies-Bouldin Index (DBI). The test using the Silhouette Coefficient aimed to assess the degree of clustering of reviews within a single complaint topic as well as the degree of separation from other complaint topics. Meanwhile, the Davies-Bouldin Index was used to ensure that the clusters were sufficiently distant from one another and did not overlap.

3.9 Interpretation

The final labels were assigned by researchers after reviewing keywords and sample reviews, so the dictionary serves only as a guide. The clusters were then sorted by the number of reviews, and clusters with similar themes were combined to determine the ranking of complaints by theme

3.10 Analysis of Results

Each cluster is interpreted based on the ten keywords with the highest centroid weights, the five original reviews closest to the centroid, and 3 the average rating and total number of "thumbs up." The top 15 keywords are used to automatically suggest labels.

Results

4.1 Dataset Characteristics

The rating distribution of 2,000 reviews is shown in Table V and Figure 3. Reviews rated 1 and 2 totaled 535 (26.75%), consisting of 460 reviews (23.00%) and 75 reviews (3.75%), and were retained as complaint data. The remainder consisted of neutral (rating 3) and positive (ratings 4 and 5) reviews, which were not analyzed. The predominance of 5-star reviews (62.35%) illustrates why complaint reviews must be separated prior to clustering, so that positive reviews do not form their own cluster.

Rating	number of reviews	Percentage	Category
1	460	23,00%	Complaints
2	75	3,75%	Complaints
3	96	4,80%	Neutral (not analyzed)
4	122	6,10%	Positive (not analyzed)
5	1.247	62,35%	Positive (not analyzed)
Total	2.000	100%	

Table 1. Rating distribution of the dataset.

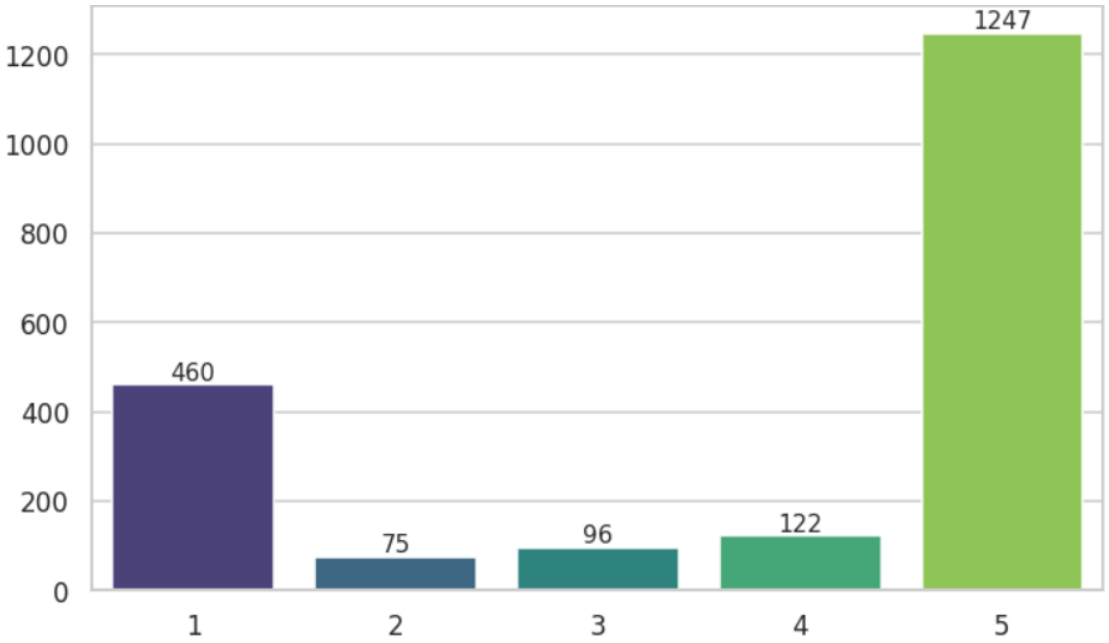


Figure 3. Distribution of review ratings.

4.2 TF IDF

The amount of data at each processing stage is shown in Table 1. Of the 535 complaint reviews, 56 were discarded because they contained fewer than two tokens, and 1 was discarded because it was a duplicate, leaving 478 reviews. After TF-IDF weighting, 1 review had no features, so the total number of grouped reviews was 477. The phrase "can't" was the most dominant (155 times), followed by "verification" (83), "face" (75), "difficult" (70), "pay" (69), "money" (57), "login" (54), "transaction" (51), and "balance" (47). From this stage onward, two problem categories became apparent: access issues (face verification, login, unable to log in) and transaction issues (pay, money, balance, transfer).

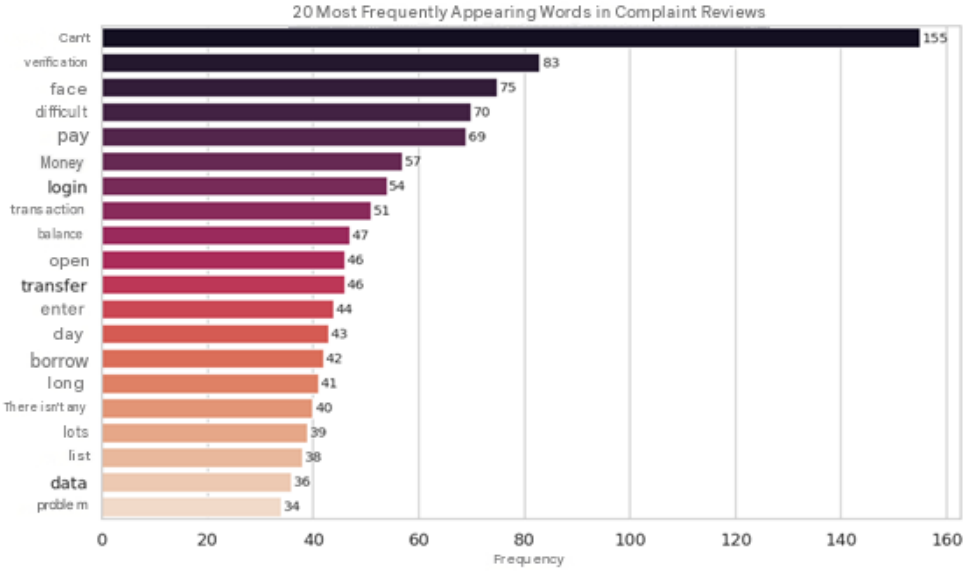


Figure 4: Most Frequently Used Words in Complaint Reviews



Figure 5. Word Cloud of Complaint Reviews

The TF-IDF weighting produced a matrix of 478 reviews × 483 features with a density of 1.86%. This means that, on average, each review contains only about 9 of the 483 terms, making the review vectors very sparse. The terms with the highest average weights are shown in Figure 8: "no," "can," "can't," "difficult," "verification," "face," "face verification," and "login." The appearance of bigrams such as "can't" and "face verification" indicates that the use of bigrams successfully captured complaint phrases.

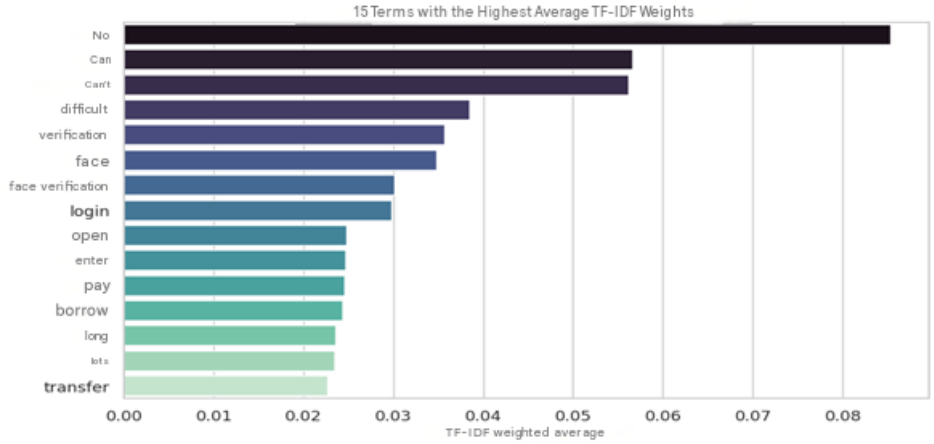


Figure 6 Terms with the Highest TF-IDF Scores

4.3 Determine K

Based on a comparative analysis of these three evaluation methods, the researchers obtained quantitative guidance for selecting the optimal number of clusters K to be used in the subsequent K-Means modeling stage, where K = 9 performed best in terms of the Silhouette Score and K = 6 performed best in terms of the Davies-Bouldin Index efficiency

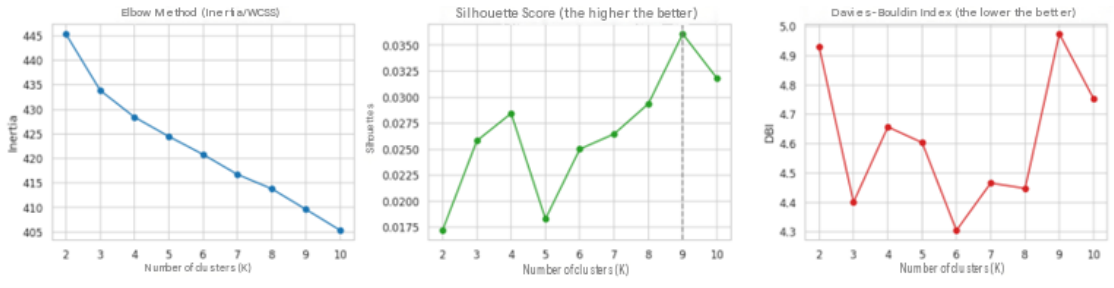


Figure 7. Comparative Model Evaluation

This stage includes a summary of the clusters, in which the clusters are divided into seven

groups, with each cluster having a different grouping and its own keywords; the percentages for each cluster can be seen in Figure 8 .

CLUSTER SUMMARY

Cluster	Amount	Percentage	Average ratings	Label suggestions
0	19	4.0%	1.05	Login & verification (face/data)
1	48	10.1%	1.12	Login & verification (face/data)
2	55	11.5%	1.25	Loans, credit pockets, interest & fees
3	20	4.2%	1.10	Login & verification (face/data)
4	80	16.8%	1.11	Application error, slow & difficult to open
5	42	8.8%	1.10	Loans, credit pockets, interest & fees
6	31	6.5%	1.13	Login & verification (face/data)
7	182	38.2%	1.14	Transfer/transaction failed or delayed

Figure 8. the number of clusters and the percentage in each cluster

4.4 Analisys of Result

This stage involves evaluating the results to determine the level of satisfaction and identify any issues or complaints from reviews by users of the Banksaqu app; therefore, the image can be viewed at

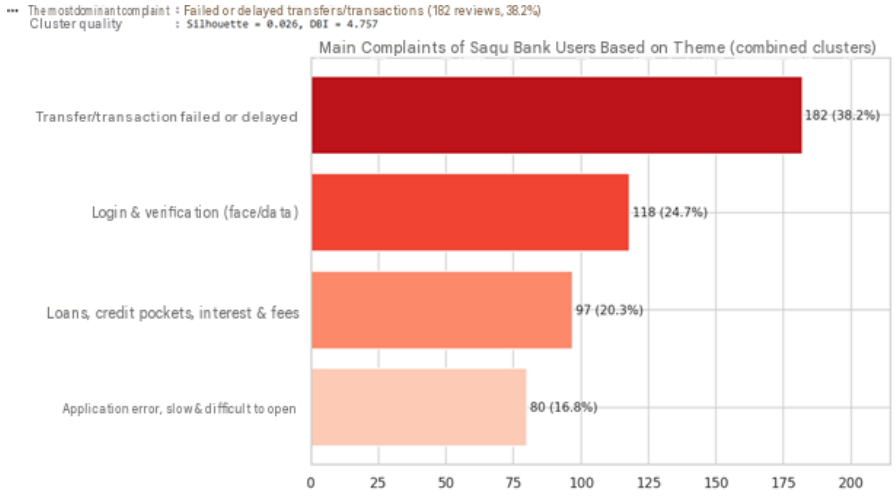


Figure 9. Chart of user complaints regarding Bank Saqu

Based on this visual data, user complaints are dominated by four main categories, with varying numbers and percentages. The most common complaint category is “Failed or delayed transfers/transactions,” which ranks first with a total of 182 reviews, accounting for 38.2% of all complaints. In second place is the “Login & verification (face/data)” category, which recorded 118 reviews, or about 24.7%. Next, in third place, is the “Loans, credit limit, interest & fees” category, with 97 reviews, or 20.3%. Finally, the complaint category with the fewest reviews among these four themes is “App Errors, Slow Performance, & Difficulty Opening,” which garnered 80 reviews, accounting for 16.8%. Overall, this chart clearly illustrates that issues related to transaction difficulties, as well as the login and verification processes, are the primary sources of user dissatisfaction—problems that require immediate attention and resolution from Bank Saqu’s developers.

Conclusion

This study successfully applied K-Means with TF-IDF weighting to identify the main complaints in 535 reviews rated 1 and 2 out of 2,000 Bank Saqu reviews on the Google Play Store from April 29 to November 7, 2026. The preprocessing stage retained negative phrases and normalized non-standard language, resulting in 477 reviews with 483 TF-IDF features. Testing K values from 2 to 10 using the Elbow Method, Silhouette Score, and Davies-Bouldin Index yielded 8 complaint clusters. The most common complaints were regarding account balances, payments, and transactions (182 reviews; 38.2%), followed by login and facial verification

(20.8%), loans and Saku Kredit (20.3%), the app failing to open (16.8%), and requests to delete data (4.0%). The cluster quality was classified as weak (Silhouette Score 0.0263; DBI 4.7568), as is common with short reviews; therefore, the labels for each cluster were determined after validation using centroid keywords and representative reviews.

This complaint cluster ranking provides a data-driven basis for Bank Saqu to prioritize improvements, particularly regarding transaction status clarity, facial verification reliability, and app access stability. This research methodology can also be replicated for other banking apps or digital wallets. Future research is recommended to perform stemming before negation merging to preserve negation tokens, use embedding-based text representations such as IndoBERT to improve cluster separation in short texts, compare K-Means with K-Medoids or HDBSCAN, incorporate data from the App Store, analyze monthly complaint trends, and utilize the generated cluster labels to build a complaint classification model.

References

[1] D. Irawan, G. Wijaya, and T. T. Warisaji, "Application of the K-Means Clustering Algorithm for Bank Customer Segmentation," vol. 6, no. 1, 2025.

[2] R. Radiatussiva et al., "Sentiment Clustering in LINE Bank App Reviews Using the K-Means Clustering Algorithm 1,2," vol. 2, no. 2, pp. 1–8, 2026.

[3] J. M. Polgan et al., "Implementation of the K-Means Clustering Algorithm in Student Grouping: A Case Study (Information Management Program)," vol. 12, pp. 777–783, 2023.

[4] W. Aulia, A. Putera, U. Siahaan, L. Marlina, and M. Iqbal, "K -Means Clustering Algorithm Analysis for Grouping Patient Medical Record Data Based on Disease Type," vol. 14, no. 04, pp. 832–843, 2024, doi: 10.54209/infosains.v14i04.

[5] R. Sudrajat et al., "Evaluation of the Quality of Clusters in Disaster-Prone Areas Using K-Means with the Silhouette and Elbow Methods," pp. 127–139, 2025, doi: 10.33364/algoritma/v.22-2.2379.

[6] W. Mahendra and T. M. Al-qoyyim, "Mapping the Global Landscape of Political Culture Research: A Bibliometric Analysis of Scopus-Indexed Literature," vol. 4, no. 2, pp. 81–92, 2026, doi: 10.56566/mandalika.v4i2.483.

[7] R. Khairuna and A. Razi, "A Comparative Analysis of the k-means and k-medoids Algorithms for Clustering Review Texts on the Cookpad App (Parts 1, 2, and 3)," vol. 10, no. 2, pp. 445–458, 2025.

[8] M. Iqbal, S. P. Sipayung, A. R. Sinaga, and P. M. Hasugian, "Analysis of Student Achievement Using K -Means on Socioeconomic, Behavioral, and Psychological Factors," vol. 14, no. 04, pp. 715–728, 2024, doi: 10.54209/infosains.v14i04.

[9] T. Ikhsan, E. Haerani, F. Wulandari, and F. Syafria, "Clustering Population Data Using the K -Means Algorithm in TIN: Applied Informatics of the Archipelago," vol. 5, no. 12, pp. 679–687, 2025, doi: 10.47065/tin.v5i12.7328.

[10] J. Ilmiah and W. Pendidikan, "The K-Means Clustering Method for Grouping Frozen Food Sales by Tifani Amalina 1, Danendra Bima Adhi Pramana 2, and Betha Nurina Sari 3, Department of Computer Science, Singaperbangsa Karawang University," vol. 8, no. September, pp. 574–583, 2022.

[11] U. S. Aesyi, S. Pengisian, B. Bakar, U. Spb, and K. Kunci, "ANALYSIS OF THE USEFULNESS LEVEL OF MYPERTAMINA USING K-MEANS CLUSTERING: Abstract, Keywords: Introduction, Research Methodology, Literature Review," vol. 4, no. 2, 2023.

[12] P. M. A. K-means, "Data Mining Clustering in the Categorization of Library Books Using the K-Means Algorithm," vol. 8, no. 3, pp. 802–814, 2023.

[13] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementation of the K-Means Clustering Algorithm for Selecting High-Achieving Students Based on Their Active Participation in

the Learning Process," vol. 7, no. 2, pp. 111–121, 2022.
[14] A. Rahmadani, "Implementation of the K-Means Algorithm for Inventory Data Clustering," vol. 5, no. 1, pp. 1–11, 2025.
[15] G. Sanko, S. Wasino, and T. Sutrisno, "Sentiment Clustering of User Reviews for the BCA Mobile App on the Google Play Store Using the K-Means Clustering Algorithm."
[16] R. T. Pangestu and R. Muliono, "Application of K-Means Clustering for Grouping the Competency Profiles of Internship Students at LKP Utama Jaya," vol. 6, no. 2, pp. 98–109, 2026, -Means Clustering for Grouping the Competency Profiles of Internship Students at LKP Utama Jaya," vol. 6, no. 2, pp. 98–109, 2026, doi: 10.34007/incoding.v6i2.1221.
[17] N. Hendrastuty, "Application of Data Mining Using the K-Means Clustering Algorithm in Evaluating Student Learning Outcomes," vol. 3, pp. 46–56, 2024.
[18] R. Purnamasari and A. R. Thaha, "Understanding User Satisfaction in Digital Banking: Sentiment Analysis and Topic Modeling of Indonesian Bank Reviews," vol. 19, no. 3, pp. 101–112, 2025, doi: 10.17323/2587-814X.2025. 3.101.112.
[19] A. Yahya and R. Kurniawan, "Implementation of the K-Means Algorithm for Clustering Sales Data Based on Sales Patterns," vol. 5, no. January, pp. 350–358, 2025.
[20] A. Nugraha et al., "APPLICATION OF THE K-MEANS CLUSTERING METHOD IN DATA MINING," vol. 6, no. 2, pp. 849–855, 2022.
[21] F. Pangestu, N. Yasin, and R. C. Hasugian, "Application of the K-Means Algorithm to Classify Drug Data," vol. 12, pp. 53–62, 2023.
[22] F. Sembiring and M. Fauzi, "Analysis of Chess Opening Patterns Using Data Mining Methods to Determine Players' Strategic Tendencies," vol. 6, no. 2, pp. 233–250, 2026, doi: 10.30811/jaise.v6i2.9781.
[23] V. Alvianatinova et al., "APPLICATION OF THE K-MEANS CLUSTERING ALGORITHM IN CLUSTERING SUPERMARKET SALES DATA BY BRANCH," vol. 8, no. 2, pp. 1529–1535, 2024.
[24] A. Pangestu, "APPLICATION OF DATA MINING USING THE K-MEANS ALGORITHM TO CLUSTER CUSTOMERS BASED ON WATER SALES VOLUME USING WEKA," vol. 11, no. 3, pp. 67–71, 2021.
[25] T. A. Aria, M. Julkarnain, and F. Hamdani, "Application of the K-Means Clustering Algorithm to Drug Data," vol. 4, no. 1, pp. 649–657, 2023, doi: 10.30865/klik.v4i1.1117