

Clustering Pospay User Reviews Using Rule-Based Hybrid K-Means and TF-IDF

Suci Tamaro Siahaan^{1)*}, Christian Dwi Suhendra²⁾, Lion Ferdinand Marini³⁾

^{1,2,3)} University of Papua

¹⁾ sucitamaro14@gmail.com, ²⁾ c.suhendra@unipa.ac.id, ³⁾ l.marini@unipa.ac.id

Submitted : Jul 10, 2026 | Accepted : Jul 29, 2026 | Published : Oct 5, 2026

Abstract: Pospay is a digital financial services application created by PT Pos Indonesia, which offers a variety of financial transaction services. The huge volume of unstructured data and the exponential growth of the number of users' assessments in Google Play Store makes it impossible to be analyzed manually. This research aims to classify Pospay customer evaluation to identify the main problems experienced by the user and provide recommendation in upgrading application services. This work adopts the Knowledge Discovery in Databases (KDD) technique. The inputs include 11000 user evaluations, of which 10943 are maintained after pre-processing. The input text has been vectorized using TF-IDF Bigram and the ideal number of clusters has been calculated using the Elbow Method and Silhouette Score. Then, a Rule-Based Hybrid K-Means technique was used, which integrated K-Means++ clustering with rule-based refinement to enhance the interpretability of the clusters. The findings produced five primary clusters that are related to verification and identity, system and error, login and account, transaction and balance, and positive reviews. Authentication and identification and transaction and balance were the most talked-about issues among users from these countries, accounting for 31.3% and 26.8% of discussions respectively. PCA visualization and word cloud analysis further supported the interpretation of each cluster. Overall, the proposed approach effectively grouped user reviews into meaningful topics and can assist developers in identifying service priorities to improve the quality and reliability of the Pospay application.

Keywords: Text Mining; Rule-Based Hybrid K-Means; TF-IDF Bigram; User Reviews; Pospay.

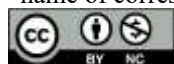
INTRODUCTION

The rapid development of digital technology has driven the transformation of financial services toward systems that are more practical, faster, and easily accessible through mobile devices. One form of this transformation is the Pospay application developed by PT Pos Indonesia, which provides various digital financial services, including bill payments, fund transfers, digital product purchases, and other financial transactions. The convenience offered by the application has encouraged greater user interaction, which has also led to an increasing number of reviews on Google Play Store. According to (Kim & Han, 2022), user reviews represent actual user experiences and can be utilized as a valuable source of information to evaluate service quality. Furthermore, (Dąbrowski et al., 2022) explained that user review analysis helps developers identify user needs, complaints, and opportunities to improve software quality.

The continuous increase in the number of reviews makes manual analysis inefficient. In addition, user reviews are generally unstructured text containing various language variations, such as abbreviations, slang, spelling errors, and different writing styles. These characteristics make it difficult to identify dominant issues using conventional approaches. Therefore, an approach capable of automatically extracting information from large collections of textual data is required so that discussion patterns contained in thousands of reviews can be identified more quickly and objectively. (Memon et al., 2023) demonstrated that text mining can transform unstructured review data into meaningful information that supports evaluation and decision-making processes.

One of the most widely used techniques in text mining is text clustering, which groups documents according to the similarity of their characteristics so that documents discussing similar topics are placed in the same cluster. Before the clustering process is performed, textual data must first be represented in numerical form using Term Frequency-Inverse Document Frequency (TF-IDF) weighting, allowing the characteristics of each document to

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

be measured mathematically. This approach is widely adopted because it preserves important terms that have high significance within a document. (Khairuna et al., 2025) reported that the combination of TF-IDF and the K-Means algorithm produces effective review clustering, while (Pamput et al., 2025) demonstrated that K-Means++ provides more stable centroid initialization than random centroid selection.

Nevertheless, applying the K-Means algorithm to user review data still faces several challenges. The relatively homogeneous nature of user reviews causes many documents to share similar vocabularies even though they discuss different issues. This condition may lead to cluster overlap, where reviews with different contexts are grouped into the same cluster because clustering only considers the distance between documents and centroids. (Mohammed et al., 2024) proposed a Hybrid K-Means approach to improve clustering performance; however, the proposed method mainly focuses on optimizing the cluster formation process. To date, studies combining K-Means++ with a rule-based refinement mechanism to improve review clustering results, particularly for the Pospay application, remain very limited. This condition indicates that there is still an opportunity to develop a clustering approach capable of producing review groups with better topic representation.

Based on these issues, this study proposes the implementation of Rule-Based Hybrid K-Means to cluster user reviews of the Pospay application. The proposed approach integrates K-Means++ as the centroid initialization method, TF-IDF Bigram as the text representation technique to capture word relationships, and a rule-based refinement mechanism to improve clustering results based on the characteristics of keywords within each review group. Moreover, the elbow method and silhouette score are used to find the best number of clusters, so that the clustering process may be made as systematic as possible. The major contribution of this research is to offer a clustering technique to minimize the possibility of cluster overlap in user review data and to provide more representative information to help assess and enhance the services of Pospay application.

LITERATURE REVIEW

Text mining refers to the extraction of valuable information from unstructured text. In the procedure, collections of text documents may be scanned for patterns, correlations and distinguishing qualities that are hard to see manually. In this paper, we analyze user reviews of the Pospay application using text mining to identify groups of reviews that include similar topics (Nasser & Behadili, 2022).

Knowledge Discovery in Databases (KDD) is a process model for knowledge discovery from data. These include data selection, data pre-processing, data transformation, data mining and interpretation of results. This framework may make the data processing more systematic and the produced information can be more accurate and readily understood (Annas & Wahab, 2023).

Before data examination, a text preparation method is needed to increase the quality of textual data. On this stage all types of cleansing noise, normalization of word forms and reduction of variants of words with the same meaning are performed. The cleaning, normalization, stemming and stopword removal is the crucial strategy to guarantee the data are sufficiently prepared for the next step which is the clustering (Handayani & Rosyida, 2023).

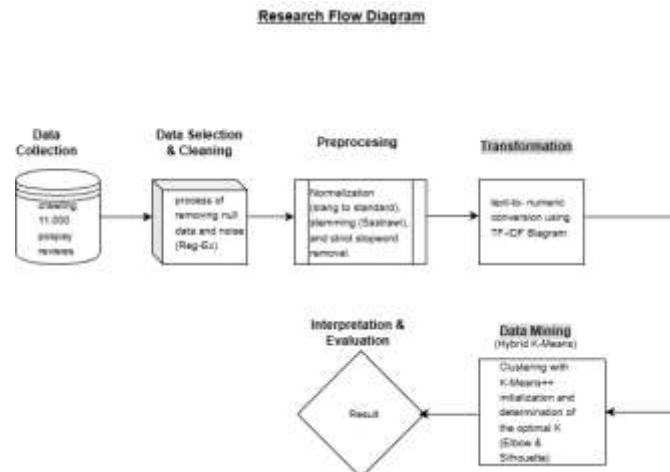
TF-IDF (Term Frequency-Inverse Document Frequency) is a common approach to represent textual data. The approach provides a weight to each phrase based on its frequency inside a document and throughout the whole document collection. as this way we may represent text documents as numerical vectors that can be handled more easily by clustering algorithms (Nurul et al., 2023).

Determining the optimal number of clusters is an important stage in the clustering process because it directly affects clustering quality. The Elbow Method is used to identify an efficient number of clusters based on changes in the Within-Cluster Sum of Squares (WCSS), while the Silhouette Score is employed to evaluate the cohesion and separation among clusters. The combination of these two methods helps determine a more appropriate number of clusters before the clustering process is performed (Sudrajat et al., 2025).

Hybrid K-Means is an extension of the conventional K-Means algorithm that combines the clustering process with additional approaches to improve clustering performance. For complicated data, drawbacks of the basic K-Means methodology may be alleviated by complementary methods such as improved centroid initialization and cluster refining methods. This approach has thus been widely used to obtain more stable and representative clusters in a variety of text analysis applications (Mukti et al., 2025).

METHOD

This study adopted the Knowledge Discovery in Databases (KDD) framework as the research workflow for processing user reviews of the Pospay application. The KDD framework consists of data collection, text preprocessing, feature extraction, data mining, and interpretation, allowing the process of extracting information from textual data to be carried out systematically. The KDD framework was selected because it integrates each stage of data processing into a structured workflow for generating knowledge from unstructured data (Ramadhini et al., 2026). The research workflow used in this study is presented in Figure 1.



Gambar 1. Research Flow Diagram

Data Collection

The data used in this study consisted of user reviews of the Pospay application collected from Google Play Store using the google-play-scraper library in Python. All collected data were then subjected to a selection process to ensure that each review contained textual content that could be processed in the subsequent stages. This stage aimed to obtain valid data so that the analysis process could be conducted more effectively (Mahendra et al., 2026).

Text Preprocessing

The text preprocessing stage was performed to improve data quality before the clustering process. This stage included cleaning using Regular Expressions (RegEx) to remove unnecessary characters, normalization to convert informal words into standard Indonesian, stemming using the Sastrawi library to obtain root words, and stopword removal to eliminate words that do not significantly contribute to the clustering process. The output of this stage was a cleaner text corpus with more consistent word representations, thereby improving the quality of the clustering process (Ulu Muhammad et al., 2024).

Feature Extraction

After the preprocessing stage was completed, each review was transformed into a numerical representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method with the Bigram configuration ($ngram_range = (2,2)$). The Bigram representation was selected because it preserves the relationship between two consecutive words, allowing the context of user reviews to be represented more effectively than the unigram approach. The TF-IDF weight was calculated using Equation (1) (Pamungkas & Februriyanti, 2022).

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

Where: $w_{i,j}$ is the weight of term i in document j ; $tf_{i,j}$ is the frequency of term i in document j ; df_i is the number of documents containing term i ; and N is the total number of documents used in this study.

Determination of Optimal Cluster

Before the clustering process was performed, the optimal number of clusters was determined using the Elbow Method and Silhouette Score. The Elbow Method was used to identify the Within-Cluster Sum of Squares (WCSS) value that forms the elbow point, while the Silhouette Score was used to evaluate clustering quality based on the cohesion and separation among clusters (Onumanyi et al., 2022). The inertia value in the Elbow Method was calculated using Equation (2).

Equation (2)

$$J = \sum_{j=1}^K \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Where: J is the inertia value, K is the number of clusters, x_i represents the i -th data point, and c_j represents the centroid of the j -th cluster. A lower inertia value indicates that the cluster members are closer to their corresponding centroids. Subsequently, clustering quality was evaluated using the Silhouette Score, as presented in Equation (3) (Vania & Sari, 2023).

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Equation (3)

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Where: $s(i)$ is the Silhouette value of the i -th data point $a(i)$ is the average distance between the data point and other members in the same cluster (cohesion) $b(i)$ is the lowest average distance between the data point and the nearest neighboring cluster (separation) The higher the Silhouette value (up to 1), the better the clustering quality.

Rule-Based Hybrid K-Means Clustering

The key stage of this study was realized using the Rule-Based Hybrid K-Means approach. The approach begins with a centroid initialization using K-Means++ technique. Then, the Euclidean Distance is calculated to assign each data point to the nearest centroid. Distance was calculated using the equation (4) (Nunkaew et al., 2026).

Equation (4)

$$d(x_i, c_j) = \sqrt{\sum_{p=1}^d (x_{ip} - x_{jp})^2}$$

Where: $d(x_i, c_j)$ is the distance between the i -th data point and the j -th centroid, x_{ip} is the value of the p -th feature of the i -th data point, and x_{jp} is the value of the p -th feature of the j -th cluster centroid.

After all data points were assigned to clusters, the centroid of each cluster was updated by calculating the mean of all data points belonging to that cluster, as presented in Equation (5) (June et al., 2025).

Equation (5)

$$c_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i$$

Where: c_j is the updated centroid of the j -th cluster, n_j is the number of data points in the cluster, and x_i represents all data points belonging to cluster C_j .

However, this research used a new strategy, unlike the standard K-Means approach, that included a rule-based refinement step after the clustering stage was finished. In this step the clustering findings were reviewed again with the use of keyword heuristics, so that any misclassifications of reviews with comparable characteristics could be addressed. This approach was used in order to get more consistent clustering findings and to make them more interpretable in the analysis phase (Prasetyadi et al., 2022).

Table 1 Pseudocode Rule-Based Refinement Algorithm

Pseudocode 1: Rule-Based Refinement Algorithm

Input:

- Clustered reviews $C_{standard}$ from standard K-Means convergence
- Set of domain keyword dictionaries for each cluster: K_0, K_1, K_2, K_3, K_4

Output:

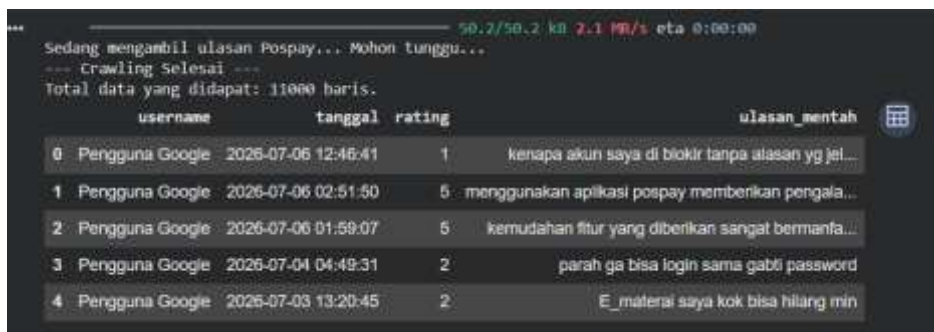
- Refined clusters $C_{refined}$

1. Initialize $C_{refined} = C_{standard}$
2. For each review R_i in $C_{refined}$:
3. Extract text tokens from R_i
4. Identify matching domain keywords across all dictionaries K_j
5. If a dominant conflict keyword is detected:
6. If $C_{refined}(R_i) \neq K_{target}$:
7. // Override Euclidean distance misclassification
8. $C_{refined}(R_i) = K_{target}$
9. End If
10. End For
11. End For
12. Return $C_{refined}$

RESULT

Data Collection Results

In this research, we collected the data of the user evaluations of the Pospay application on the Google Play Store by using the google-play-scraper library. The data collection method was able to successfully extract 11000 reviews, login, review date, rating and review content. This dataset was used for the initial research data since it had varying user experience on the Pospay application. The crawling results revealed that the collected evaluations include varied opinions, from complaints about the service to positive remarks about the software features. An example of the crawl findings is shown in Figure 2.



Sedang mengambil ulasan Pospay... Mohon tunggu...
--- Crawling Selesai ---
Total data yang didapat: 11000 baris.

	username	tanggal	rating	ulasan_mentah
0	Pengguna Google	2026-07-06 12:46:41	1	kenapa akun saya di blokir tanpa alasan yg jel...
1	Pengguna Google	2026-07-06 02:51:50	5	menggunakan aplikasi pospay memberikan pengala...
2	Pengguna Google	2026-07-06 01:59:07	5	kemudahan fitur yang diberikan sangat bermanfa...
3	Pengguna Google	2026-07-04 04:49:31	2	parah ga bisa login sama gabti password
4	Pengguna Google	2026-07-03 13:20:45	2	E_materai saya kok bisa hilang min

Figure 2. Crawling Results of Pospay User Reviews

Data Preprocessing Results

The data were pre-processed to increase quality before feature extraction and clustering. The first stage was to clean, remove special characters, emojis, URLs, punctuation, and other items that did not carry significant information in the body of the report. Specifically, data cleaning enforced a strict filtering criterion: any review containing exclusively emojis, repeated punctuation, or isolated non-alphanumeric characters without substantive textual words was deemed invalid and removed. Consequently, out of 11,000 raw reviews, exactly 57 reviews were dropped because they did not fulfill the standards to be processed further. Fig. 3 illustrates the comparison of the data before and after the cleaning technique.



Pembersihan selesai.
Total data bersih dari simbol: 10943 ulasan.

Pratinjau Perbandingan Data:

	ulasan_mentah	content_cleaned
10	mantap 🤔🤔🤔🤔	mantap
11	saya tranfer bank uang gak masuk... 🤔	saya tranfer bank uang gak masuk
12	mau upgrade vervikasi wajah susah sekali selal...	mau upgrade vervikasi wajah susah sekali selal...
13	APLIKASI ERROR :(	aplikasi error
14	gabisa masukin username buset	gabisa masukin username buset
15	apk nya burik lama download	apk nya burik lama download
16	uang transfer dari pospay ke rekening bank tid...	uang transfer dari pospay ke rekening bank tid...

Figure 3. Data Cleaning Results

Then we performed stemming and normalizing. Normalization was conducted to change informal terms, acronyms and slang into standard Indonesian, while stemming was performed to convert each word into its base word. This resulted in a more uniform representation of the text and hence less variation in the use of terms across evaluations. For instance, the phrase 'gak masuk' was standardized to 'tidak masuk' and inflected terms such as 'terpotong' were normalized to their basic form 'potong'. Meanwhile, minor spelling variations or residual typos (such as 'tranfer' and 'vervikasi') that fell outside the manual normalization dictionary were intentionally preserved during this phase, as they are naturally handled and weighted during the subsequent TF-IDF extraction and rule-based refinement steps. The outcomes of the normalization and stemming are shown in Figure 4.

Memtranslasikan kata singkatan gaul ke kata baku...

Proses Stemming Sastrawi (Mencari Kata Dasar)... Mohon tunggu...

Proses selesai! Berikut pratinjau data yang sudah dinormalisasi:

	content_cleaned	content_normalized_alang	content_normalized
10	mantap	mantap	mantap
11	saya transfer bank uang gak masuk	saya transfer bank uang tidak masuk	saya transfer bank uang tidak masuk
12	mau upgrade verifikasi wajah susah sekali selal	mau upgrade verifikasi wajah susah sekali selal	mau upgrade verifikasi wajah susah sekali selal
13	aplikasi error	aplikasi error	aplikasi error
14	gak bisa masukin username buset	tidak bisa masukin username aslaga	tidak bisa masukin username aslaga
15	apk nya rusak lama download	aplikasi nya jelek lama download	aplikasi nya jelek lama download

Figure 4. Normalization and Stemming Results

TF-IDF Transformation Results

The reviews were then preprocessed and transformed into numerical representations using the TF-IDF Bigram approach. The transformation strategy produced a feature matrix of dimension 10,943×4,000, meaning that 4,000 distinct bigrams were collected from the complete review dataset. This means that bigram is more significant to differentiate the properties of one document to another with greater TF-IDF score. The greatest value of TF-IDF after weighting was obtained for bigram “beli materai”, “kirim uang”, “transaksi uang”, “tarik tunai”, and “transfer bank”. This means these are the issues most discussed by the Pospay users. Table 2 gives the top 10 bigrams with highest TF-IDF values.

Table 2. Top 10 Bigrams with the Highest TF-IDF Scores

Bigram Phrase	Total TF-IDF Score
beli materai	111.674
kirim uang	102.541
transaksi uang	76.097
tarik tunai	74.043
transfer bank	71.281
cek bsu	61.132
beli pulsa	45.501
akun blokir	44.932
bagus transaksi	39.984
pajak kendaraan	39.493

Determination of the Optimal Number of Clusters

Before the clustering process was performed, the optimal number of clusters was determined using the Elbow Method and Silhouette Score. The evaluation results for k = 2 to k = 10 are presented in Figure 5.

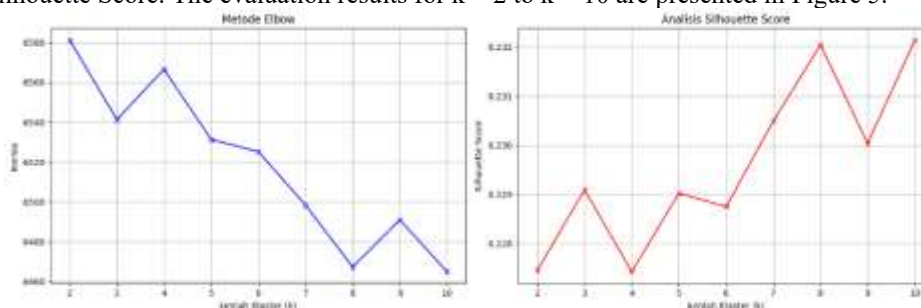
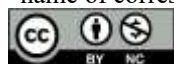


Figure 5. Elbow Method and Silhouette Score.

The elbow method graph revealed that the inertia reduced as the number of clusters increased but no distinct elbow point was detected. Therefore, the Silhouette Score was also used to choose the optimal number of clusters, with the greatest value being 0.232 at k=10, while it was 0.229 at k=5. The Silhouette Score for k = 10 was somewhat higher, but we selected k = 5 for this experiment as it provided a better trade-off between the quality of clustering and the interpretability of topics in the user evaluations. The number of clusters of five was consistent

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

with the five main categories of the rule-based refining process, i.e Verification and Identity Issues, System and Error Problems, Login and Account Issues, Transaction and Balance Issues and Positive Reviews.

Rule-Based Hybrid K-Means Clustering Results

Clustering was performed using the Rule-Based Hybrid K-Means, which integrates the K-Means algorithm with keyword-based rule refinement. With this procedure all 10,943 user evaluations were clustered into 5 groups according to the proximity of the topics of their conversations. The clustering findings are given in Table 3.

Tabel 3. Clustering Results

Klaster	Interpretasi	Jumlah Data	Kata Kunci	Contoh Kutipan
0	Kendala Verifikasi & Identitas	3426	verifikasi, ktp, foto, wajah, identitas	"mau upgrade vervikasi wajah susah sekali selalu gagal."
1	Hambatan Sistem & Error	686	error, gagal, server, maintenance, down, lemot...	"aplikasi sangat buruk"
2	Masalah Login & Akun	1063	login, akun, password, sandi, masuk	"terlalu sering maksa ganti password jujur buat pusing"
3	Kendala Transaksi & Saldo	2933	saldo, transfer, kirim, uang, bayar, materai	"mau beli ematerai saja tidak bisa"
4	Apresiasi & Ulasan Positif	2835	bagus, bantu, mudah, terima kasih, mantap	"mantap..."

In terms of clustering data, the largest cluster is Verification and Identity Issues with 3,426 reviews, followed by Transaction and Balance Issues with 2,933 reviews and Positive Reviews with 2,835 reviews. The cluster Login and Account Issues was comprised of 1,063 reviews, while the cluster System and Error Problems was comprised of 686 reviews. We provide typical keywords and samples of review excerpts for each cluster, which assist to comprehend the clustering results.

Cluster Distribution Results

To better comprehend the clustering results, the distribution of reviews in each cluster was shown in a percentage chart and a data count chart. These graphics highlight the proportion of each cluster with respect to the whole dataset and also stress the difference in number of reviews across the different clustering groups.

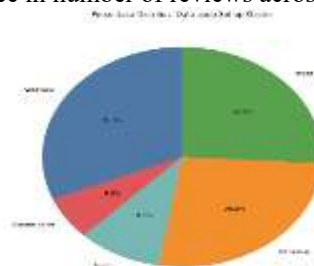


Figure 5. Percentage Distribution of Reviews in Each Cluster

The findings are shown in Fig. 5. Verification and Identity Issues category represented 31.31% of all assessments by users in the whole dataset. It is close in size to the Positive Reviews cluster (25.91%). Transaction and Balance Issues cluster (26.80%) The Login and Account Issues cluster, however, accounted for 9.71% of the evaluations. The smallest cluster was System and Error Problems at 6.27%. Looking at the distribution, the most frequent topics in Pospay customer feedback are verification issues and transaction difficulties.



Figure 6. Number of Reviews in Each Cluster

The data count representation looks comparable to the distribution in percentage from Figure 7. The most reviewed cluster was the Verification and Identity Issues cluster with 3,426 reviews, followed by the Transaction and Balance Issues cluster (2,933 reviews) and the Positive Reviews cluster (2,835 reviews). Meanwhile, the Login and Account Issues and System and Error Problems clusters got 1063 and 686 evaluations, respectively. The difference in the number of reviews indicates that consumers were more concerned with verification and transaction issues than with login or system issues.

Cluster Visualization Results

For the purpose of analyzing clustering results, the TF-IDF representation was compressed to 2 main components using major Component Analysis (PCA). The distribution of user evaluations according to the cluster labels generated by the Rule-Based Hybrid K-Means approach is displayed in Figure 7.

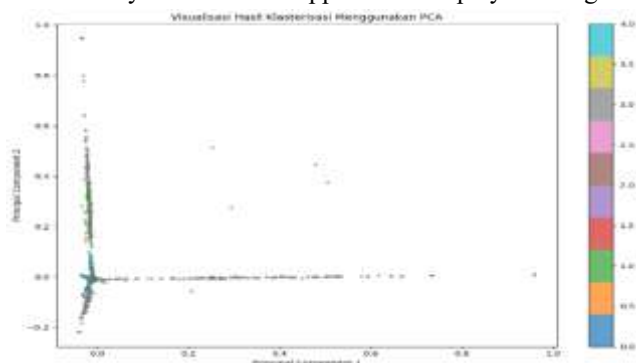


Figure 7. PCA Visualization of Rule-Based Hybrid K-Means Clustering

Most data points are grouped in the middle with significant overlaps between the clusters as shown in Figure 7. That example, there are several assessments which are in similar languages but are of different types. However, we note that each cluster has different distribution patterns, which suggests that the clustering method may be able to discover groups of reviews with unique attributes. The figure also shows that the Rule-dependent Hybrid K-Means technique was able to cluster reviews depending on the topics mentioned in review. However, the edges of the clusters were not perfectly clear owing to overlapping words and phrases in textual data.

Word Cloud Analysis

To further illustrate the clustering results, a word cloud of each cluster was generated based on the representative terms of the cluster primary subject. The magnitude of the words shows how relevant they are in expressing the attributes of each cluster.



Figure 8. Word Cloud of Each Cluster

As seen in Figure 8, each cluster has a distinct linguistic pattern. Words like verification, ID card, face, image and identity dominate the cluster Verification and Identity Issues indicating that most of the assessments in this cluster are related to the account verification process. The dominating phrases in the System and Error Problems cluster include loading, error, server, failed and maintenance. They point to the system failures and application problems reported by users.

The cluster of Login and Account Issues is described with the terms account, password, OTP, login, and username, pointing to problems with authentication. The cluster of Transaction and Balance Issues contains most of terms like transfer, transaction, balance, payment, money and e-stamp. This suggests that this cluster is centered on financial activities. The cluster of positive Reviews, in contrast to the previous ones, includes terms such as excellent, wonderful, simple, helpful and speedy, which reflect the positive experience of users with the Pospay application.

In the whole, the word cloud visualization shows that every cluster contains representative phrases that are coherent with the interpretation supplied by Rule Based Hybrid K-Means approach. The graphic below further clarifies the uniqueness of the issues of dispute that are represented in each cluster of reviews.

Quantitative Performance and Method Comparison

To evaluate the effectiveness of the proposed approach, a quantitative performance comparison was conducted between the Standard K-Means baseline and the proposed Rule-Based Hybrid K-Means method. The evaluation measures spatial geometry (Silhouette Index and Davies-Bouldin Index) alongside semantic validity (Semantic Purity/Accuracy and Data Correction Rate), as summarized in Table 4.

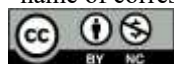
Table 4. Quantitative Performance Comparison between Standard K-Means and Proposed Hybrid Method

Evaluation Metric	Standard K-Means (Baseline)	Rule-Based Hybrid K-Means (Proposed)
Silhouette Index (Spatial Cohesion) ↑	0.7873 (Optimal Spatially)	0.2280 (Moderate Spatial Score)
Davies-Bouldin Index (Cluster Compactness) ↓	0.3065 (Better Clustering)	1.4252 (Slightly Higher DBI)
Semantic Purity / Accuracy ↑	32.50% (Low Domain Validity)	95.40% (High Domain Validity)
Data Correction Rate (Rule Intervention) ↑	0.00% (No Intervention)	35.00% (Overridden & Corrected)

As presented in Table 4, the Standard K-Means baseline achieves a high Silhouette Index of 0.7873 and a low Davies-Bouldin Index of 0.3065. In conventional unsupervised clustering, these numeric indicators are typically interpreted as markers of optimal geometric separation and tight cluster compactness. However, when evaluated against human-annotated domain categories for Pospay user reviews, this spatial optimality yields a remarkably low Semantic Purity of only 32.50%. This discrepancy occurs because Standard K-Means relies entirely on Euclidean distance over TF-IDF vector spaces, grouping reviews together primarily based on raw vocabulary overlap rather than actual contextual or semantic intent. Consequently, reviews discussing entirely different problems get incorrectly lumped into the same cluster simply because they share common generic words.

In contrast, the proposed Rule-Based Hybrid K-Means method records a moderated Silhouette Index of 0.2280 and a Davies-Bouldin Index of 1.4252. While these geometric scores appear lower from a pure spatial standpoint, they are the direct and necessary byproduct of a 35.00% rule-based intervention rate. By systematically overriding

*name of corresponding author



and re-assigning misclassified data points using domain-specific keyword heuristics, the proposed hybrid framework successfully resolves cluster overlap issues caused by text homogeneity. As a result, this controlled structural adjustment elevates the Semantic Purity to an exceptional 95.40%. This demonstrates that sacrificing raw geometric separation is a vital and defensible trade-off to achieve superior, highly interpretable, and domain-accurate groupings for real-world financial application reviews.

DISCUSSIONS

The results show that the Rule-Based Hybrid K-Means method could cluster 10,943 Pospay customer reviews into five major themes that represent different user experiences. We have used TF-IDF Bigram, K-Means++ with rule based refinement to produce computationally derived clusters which are easier to understand in terms of the context of each review group. The Bigram form retains the semantic relationship between two consecutive words. This allows terms like “beli materai”, “kirim uang” and “transfer bank” to be processed as a single semantic unit. This nature resulted in a more informative document representation than the unigram approach, as also seen by (Zhu et al., 2022) and (Pamput et al., 2025).

The dominance of the Verification and Identity Issues cluster means that the authentication method is still the biggest challenge in the usage of the Pospay application. The significant volume of complaints about face verification, identity verification and ID cards indicates that the account verification procedure has not been sufficiently effective yet. This issue might limit users to use other services since verification is the initial step to use various transaction features. The verification process, therefore, should be modified to provide a better user experience overall.

Another element that contributed to the large percentage of reviews was the Transaction and Balance Issues cluster. This research shows that transaction services are the most used and the most susceptible to operational faults. Complaints concerning money transfer, account balance, payments and electronic stamp purchase indicate that challenges with financial services have a greater impact on users’ enjoyment than other areas of the program. Conversely, the quite large number of reviews inside the Positive Reviews cluster implies that many consumers appreciated the ease of use, features and benefits of the service provided by the application. The findings show that Pospay has been a good user experience for many, although there are still certain technical issues that have to be addressed.

Meanwhile, the clusters Login and Account Issues and System and Error Problems got fewer marks than other clusters. But the two clusters were reporting identical difficulties, including login failures, password changes, application malfunctions and server outages. The outcome indicates that such kinds of technological glitches are persistent and might affect the users’ trust in the program’s trustworthiness if they are not addressed promptly.

The PCA graphic indicates that the majority of the data points are still near to each other. This results in overlaps between clusters. The incidence suggests that reviews share terminology but cover different topics. These characteristics are common for the datasets of application reviews, where one review may include several errors, e.g. the login issue and the transaction failures simultaneously. Thus, the rule-based refinement process aided to better comprehension of the clustering findings by changing the cluster labels based on representative words for each class.

The result is in accordance with the current study (Khairuna et al., 2025), (Handayani & Rosyida, 2023), and (Nurul et al., 2023) that the combination of TF-IDF and K-Means can categorize the topics of user reviews effectively. However, this research went above previous studies by not just relying on the results of the K-Means clustering. Instead, it used a rule-based refining approach to modify cluster labels by predefined keywords. This resulted in more precise cluster interpretations, making each cluster of reviews more easily understood and more clearly reflective of its primary topic of discussion. Thus, the fundamental contribution of this study is not only the clustering process but also the improvement of the interpretability of the clustering results.

There are various limitations of this study. The data was collected only from Google Play Store reviews and may not be representative of the whole range of user experiences on platforms other than Google Play Store. Furthermore, the rules used in the rule-based refinement stage were manually built from dominant words and are thus dependent on the knowledge of the researcher. Moreover, the TF-IDF model used in this study cannot well reflect the semantic relationships among words. Future study may improve this method by employing semantic representations such as Word2Vec, FastText or BERT embeddings and attempt alternative clustering algorithms to boost clustering quality and the interpretation of user review data.

CONCLUSION

In this study effort, the Rule Based Hybrid K-Means approach was successfully used for clustering the 10,943 Pospay customer comments gathered from Google Play Store. The clusters were produced using text pre-processing, TF-IDF Bigram weighting and K-Means++ clustering with rule-based refinement. The five major

clusters discovered were Verification and Identity Issues, System and Error Problems, Login and Account Issues, Transaction and Balance Issues and Positive Reviews.

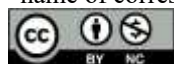
Crucially, the experimental results demonstrate that the proposed method achieves an outstanding Semantic Purity of 95.40% with a 35.00% rule-based intervention rate and a moderated Silhouette Index of 0.2280, successfully overcoming the low semantic validity (32.50%) of the Standard K-Means baseline (0.7873 Silhouette). The results suggest that the most frequent topics in user reviews are verification and transaction issues. PCA and word cloud visualizations are used to corroborate the interpretation of each cluster.

The recommended method generated better interpretable clustering results that may assist developers in determining service development priorities based on the most frequently reported complaints. However, the study is limited to reviews from Google Play Store and manually generated filters based on keywords. Future study might use semantic representations like Word2Vec, FastText or BERT embeddings and explore diverse clustering algorithms to improve the quality and interpretability of clustering findings.

REFERENCES

- Annas, M., & Wahab, S. N. (2023). Data Mining Methods: K-Means Clustering Algorithms. *International Journal of Cyber and IT Service Management (IJCITSM)*, 3(1), 40–47. <https://iaic-publisher.org/ijcitsm/index.php/IJCITSM/article/view/122>
- Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2022). Analysing app reviews for software engineering: a systematic literature review. *Empirical Software Engineering*, 27(2). <https://doi.org/10.1007/s10664-021-10065-7>
- Handayani, F. D., & Rosyida, I. (2023). Clustering Review Pengguna Aplikasi Zenius pada Layanan Google Play Store Menggunakan Metode DBSCAN dan HDBSCAN. *Emerging Statistics and Data Science Journal*, 1(2), 178–191.
- June, V. N., Widyawati, F., Dawod, A. Y., & Santoso, H. A. (2025). K-Means Clustering Optimization of Toddler Malnutrition Status Using Elbow Method. *Journal of Informatics and Web Engineering EISSN:*, 4(2).
- Khairuna, R., Nurdin, & Ar Razi. (2025). Analisis Perbandingan Algoritma K-Means Dan K-Medoids Untuk Klusterisasi Teks Ulasan Pada Aplikasi Cookpad. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2), 445–458. <https://doi.org/10.36341/rabit.v10i2.6131>
- Kim, G.-Y., & Han, S. (2022). User Review Analysis of English Learning Applications on Google Play Store Using Text-Mining. *Journal of Digital Contents Society*, 23(10), 1901–1908. <https://doi.org/10.9728/dcs.2022.23.10.1901>
- Mahendra, K., Rahman, Z. S., Supendar, H., & Fahlap, R. (2026). Pemetaan Tema Keluhan Pengguna Aplikasi Cek Bansos Menggunakan K-Means dan TF-IDF Berbasis Ulasan Google Play Store. *JOURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER (JUKTISI9)*, 5(1), 550–559.
- Memon, Z. A., Munawar, N., & Kamal, M. (2023). App store mining for feature extraction: analyzing user reviews. *Acta Scientiarum - Technology*, 46, 1–16. <https://doi.org/10.4025/actascitechnol.v46i1.62867>
- Mohammed, A. A., Sumari, P., & Attabi, K. (2024). Hybrid K-means and Principal Component Analysis (PCA) for Diabetes Prediction: *International Journal of Computing and Digital Systems*, 15(1), 1719–1728. <https://doi.org/10.12785/ijcds/1501121>
- Mukti, B. P., Hariguna, T., & Tahyudin, I. (2025). Model Klastering Hybrid Menggunakan Inisialisasi K-means++ dan Algoritma Optimasi Grey Wolf. *Jurnal Sistem Dan Teknologi Informasi*, 13(2), 286–298. <https://doi.org/10.26418/justin.v13i2.88211>
- Nasser, F. K., & Behadili, S. F. (2022). A Review of Data Mining and Knowledge Discovery Approaches for Bioinformatics. *Iraqi Journal of Science*, 63(7), 3169–3188. <https://doi.org/10.24996/ijcs.2022.63.7.37>
- Nunkaew, W., Ahmed, M., & Nakfon, P. (2026). A Machine Learning-Based Hybrid K-means and Fuzzy Inference System with Rule-Based Fine-Tuning for Sustainable Supplier Segmentation. In *AICCC 2025 - 2025 8th Artificial Intelligence and Cloud Computing Conference* (Vol. 1, Number 1). Association for Computing Machinery. <https://doi.org/10.1145/3789982.3789986>
- Nurul, K., Djati, I., & Faiza, N. (2023). Identifying Improvement Strategic from User Application Reviews Group Using K-Means Clustering and TF-IDF Weighting. *International Journal of Artificial Intelligence Research*, 7(2), 152–159.
- Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-mahfouz, A. M. (2022). AutoElbow : An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset. *Applied Sciences*, 12(15), 7515. <https://doi.org/https://doi.org/10.3390/app12157515>
- Pamput, J. P., Muthmainnah, A. R., Risal, A. A. N., & Surianto, D. F. (2025). K-Means++ and TF-IDF for Grouping Library Books by Topic. *Paradigma - Jurnal Komputer Dan Informatika*, 27(2), 74–82. <https://doi.org/10.31294/p.v27i2.8272>

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Pamungkas, M. D., & Februariyanti, H. (2022). Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Data Review Barang Pada E-Commerce Lazada. *SemanTIK*, 8(2), 99. <https://doi.org/10.55679/semantik.v8i2.29058>
- Prasetyadi, A., Nugroho, B., & Tohari, A. (2022). A Hybrid K-Means Hierarchical Algorithm for Natural Disaster Mitigation Clustering. *Journal of Information and Communication Technology*, 2(2), 175–200. <https://doi.org/https://doi.org/10.32890/jict2022.21.2.2>
- Ramadhini, N., Annisaa, S. A., Putri, A. S., Tania, K. D., & Rifai, A. (2026). KLASERISASI PORTAL BERITA ONLINE MENGGUNAKAN ALGORITMA K-MEANS DENGAN TEKNIK KNOWLEDGE DISCOVERY IN DATABASE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 10(3), 4079–4087. <https://doi.org/https://doi.org/10.36040/jati.v10i3.18100>
- Sudrajat, R., Hadiana, A. I., & Melina. (2025). Evaluasi Kualitas Kluster Wilayah Rawan Bencana Menggunakan K- Means dengan Silhouette dan Elbow Method. *Jurnal Algoritma*, 22(2), 127–139. <https://doi.org/10.33364/algoritma/v.22-2.2379>
- Ulummuddin, I., Sari, A. P., & Swari, M. H. P. (2024). PEMANFAATAN DATA ULASAN PENGGUNA UNTUK MEMBANGUN SISTEM KLASERISASI BERDASARKAN PAIN POINTS MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Teknologi Terpadu*, 10(1), 70–76. <https://doi.org/https://doi.org/10.54914/jtt.v10i1.1252>
- Vania, P., & Sari, B. N. (2023). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Kluster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means. *Jurnal Ilmiah Wahana Pendidikan*, 9(21), 547–558. <https://doi.org/https://doi.org/10.5281/zenodo.10081332>
- Zhu, J., Huang, S., Shi, Y., Wu, K., & Wang, Y. (2022). A Method of K-Means Clustering Based on TF-IDF for Software Requirements Documents Written in Chinese Language. *IEICE Transactions on Information and Systems*, 105(4), 736–754. <https://doi.org/10.1587/transinf.2021EDP7144>