

Analisis Ulasan Pengguna Aplikasi Gojek Berdasarkan Versi Aplikasi Menggunakan Metode Clustering untuk Evaluasi Kepuasan

Kayyisu Willyani*

Universitas Muhammadiyah Banjarmasin

kayyisu123@gmail.com*

Abstrak

Penelitian ini bertujuan untuk menganalisis data ulasan pengguna aplikasi Gojek dengan menggunakan metode clustering untuk mengelompokkan pengguna berdasarkan versi aplikasi dan tingkat kepuasan mereka. Dataset yang digunakan diperoleh dari Kaggle dan berisi kolom-kolom seperti "username", "content", "score", "at", dan "appversion". Dalam penelitian ini, dilakukan pengelompokan data ke dalam tiga cluster menggunakan algoritma K-Means. Setiap versi aplikasi dikonversi menjadi nilai numerik, dan data dipangkas menjadi maksimal 1.000 entri per versi untuk menjaga keseimbangan. Hasil clustering menunjukkan adanya pola kepuasan pengguna yang bervariasi pada setiap versi aplikasi. Cluster pertama merepresentasikan pengguna dengan tingkat kepuasan rendah, cluster kedua menunjukkan tingkat kepuasan tinggi pada versi aplikasi terbaru, dan cluster ketiga menampilkan tingkat kepuasan tinggi pada versi aplikasi awal. Visualisasi hasil clustering memberikan wawasan yang berharga bagi pengembang aplikasi untuk meningkatkan pengalaman pengguna.

Kata Kunci: Clustering, K-Means, Analisis Ulasan, Gojek, Kepuasan Pengguna

Abstract

This study aims to analyze user review data from the Gojek app using clustering methods to group users based on the app version and their satisfaction levels. The dataset used was obtained from Kaggle and contains columns such as "username," "content," "score," "at," and "appversion." In this study, the data is grouped into three clusters using the K-Means algorithm. Each app version is converted into a numeric value, and the data is limited to a maximum of 1,000 entries per version to maintain balance. The clustering results show varying patterns of user satisfaction across different app versions. The first cluster represents users with low satisfaction, the second cluster shows high satisfaction for the latest app version, and the third cluster displays high satisfaction for the initial app version. The visualization of the clustering results provides valuable insights for app developers to improve user experience.

Keywords: Clustering, K-Means, Review Analysis, Gojek, User Satisfaction

1. Pendahuluan

Perkembangan teknologi saat ini semakin pesat, memberikan banyak kemudahan untuk mendukung aktivitas manusia. Salah satu inovasi tersebut adalah hadirnya layanan transportasi berbasis ojek online. Kehadiran ojek online menjadi solusi yang mampu menjawab berbagai kekhawatiran masyarakat terkait kualitas layanan transportasi umum. Salah satu layanan ojek online yang banyak diminati di Indonesia adalah GoJek(1). Gojek adalah perusahaan yang menghadirkan inovasi baru di bidang transportasi umum berbasis teknologi informasi. Dengan semakin pesatnya pertumbuhan pengguna internet, masyarakat tidak lagi perlu membuang waktu dan tenaga untuk mencari transportasi umum. Gojek menawarkan kemudahan melalui sistem reservasi online yang dapat diakses melalui aplikasi di smartphone, sehingga mempermudah pengguna dalam melakukan transaksi(2).

Dengan jutaan pengguna aktif setiap hari, evaluasi kepuasan pengguna terhadap berbagai versi aplikasi menjadi langkah penting bagi pengembang untuk terus meningkatkan kualitas layanan. Dalam hal ini, data ulasan pengguna yang tersedia secara publik menjadi sumber informasi berharga untuk menganalisis tingkat kepuasan dan pengalaman mereka. Analisis data

ulasan seringkali menghadapi tantangan, terutama karena volume data yang besar, Dalam mengatasi tantangan ini, data mining menjadi pendekatan penting yang digunakan untuk menggali pola, tren, dan informasi berharga dari data yang kompleks dan besar. Data mining adalah proses yang memanfaatkan kecerdasan buatan, statistik, matematika, dan pembelajaran mesin untuk mengidentifikasi serta mengekstrak informasi berguna dan pengetahuan terkait dari database berukuran besar(3). Salah satu teknik data mining yang efektif dalam memahami pola dari data besar adalah clustering. Clustering adalah proses mengelompokkan data ke dalam kelas atau kelompok, di mana objek dalam satu cluster memiliki tingkat kesamaan yang tinggi antara satu sama lain, namun sangat berbeda dibandingkan dengan objek yang berada di cluster lain(4). Tujuan clustering adalah untuk mengklasifikasikan data dengan mengidentifikasi kelompok-kelompok dalam sebuah dataset yang sebelumnya belum terdefinisi(5). Dalam konteks ini, clustering digunakan untuk mengelompokkan pengguna aplikasi Gojek berdasarkan versi aplikasi yang mereka gunakan dan skor kepuasan yang mereka berikan.

Penelitian ini menggunakan algoritma K-Means sebagai metode clustering. Algoritma K-Means berfungsi membagi data ke dalam beberapa cluster, dengan mengelompokkan data yang memiliki kemiripan ke dalam satu cluster yang sama, sementara data yang berbeda ditempatkan di cluster lainnya(6). Algoritma ini dipilih karena memiliki kelebihan yaitu kemampuannya untuk mengelompokkan objek dalam jumlah besar dan menangani data yang tidak sesuai atau berbeda jauh dari kelompok utama (outlier) dengan sangat cepat, sehingga mempercepat proses pengelompokan (7).

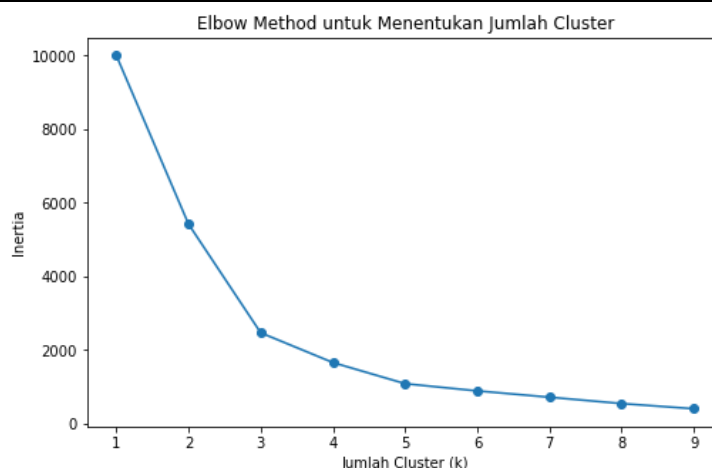
Tujuan penelitian ini adalah untuk memberikan wawasan kepada pengembang aplikasi Gojek terkait tingkat kepuasan pengguna pada setiap versi aplikasi, sehingga dapat digunakan sebagai dasar untuk perbaikan dan inovasi di masa mendatang. Artikel ini akan memaparkan langkah-langkah preprocessing data, implementasi algoritma K-Means, analisis hasil clustering, dan rekomendasi berdasarkan temuan penelitian.

2. Metode Penelitian

Penelitian ini dilakukan secara kuantitatif dengan tahapan yang meliputi pengumpulan dataset ulasan pengguna aplikasi Gojek yang diperoleh dari Kaggle. Penelitian kuantitatif adalah jenis penelitian yang didasarkan pada filsafat positivisme, berfokus pada studi terhadap populasi atau sampel tertentu. Sampel diambil secara acak, data dikumpulkan menggunakan instrumen penelitian, dan analisis datanya dilakukan secara statistik(8).

Dataset yang diperoleh dari Kaggle tersebut terdiri dari lima kolom utama, yaitu "username", "content", "score", "at", dan "appversion". Dari data ini, dipilih lima versi aplikasi dengan ulasan terbanyak, yaitu 4.31.1, 4.32.1, 4.34.1, 4.36.1, dan 4.40.1, yang kemudian dianalisis lebih lanjut. Proses analisis data dilakukan menggunakan aplikasi Jupyter Anaconda, dimulai dengan pembersihan data, termasuk konversi kolom "appversion" menjadi nilai numerik agar dapat diproses secara lebih efisien. Untuk menjaga keseimbangan, data dipangkas dengan membatasi maksimal 1.000 entri per versi menggunakan teknik sampling.

Setelah pembersihan, fitur "appversion_numeric" dan "score" distandarkan untuk memastikan keseragaman skala, sehingga analisis menjadi lebih akurat. Clustering dilakukan dengan algoritma K-Means, sementara jumlah cluster optimal ditentukan melalui metode Elbow. Data kemudian dibagi menjadi tiga cluster berdasarkan kemiripan skor kepuasan dan versi aplikasi. Visualisasi scatter plot digunakan untuk memahami distribusi data antar-cluster. Analisis tambahan berupa perhitungan rata-rata skor dan versi aplikasi pada setiap cluster dilakukan untuk menggali lebih dalam karakteristik masing-masing cluster. Langkah-langkah ini dirancang untuk menghasilkan wawasan yang relevan terkait pola kepuasan pengguna pada setiap versi aplikasi.



Gambar 1. Elbow Untuk Menentukan Jumlah Cluster

3. Hasil Penelitian dan Pembahasan

3.1 Data Mining

Perkembangan data mining (DM) yang pesat tidak dapat dilepaskan dari kemajuan teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi. Sebagai contoh, platform aplikasi layanan digital seperti Gojek mencatat setiap aktivitas pengguna, mulai dari pemesanan layanan hingga ulasan yang diberikan pengguna. Database dari aktivitas tersebut dapat mencapai ukuran yang sangat besar, terutama dengan jumlah pengguna yang terus meningkat setiap harinya. Perkembangan internet juga berkontribusi signifikan dalam akumulasi data ini. Namun, pertumbuhan pesat dari data yang terakumulasi sering menciptakan kondisi yang disebut sebagai “rich of data but poor of information”, di mana data yang terkumpul tidak dimanfaatkan secara optimal untuk menghasilkan informasi yang berguna. Dalam banyak kasus, data ini hanya dibiarkan begitu saja seperti “kuburan data” (data tombs).

Data mining (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang sebelumnya tidak diketahui secara manual(9). Dalam konteks aplikasi seperti Gojek, DM dapat digunakan untuk menganalisis pola kepuasan pengguna berdasarkan ulasan mereka terhadap berbagai versi aplikasi. Kata mining sendiri berarti upaya untuk mendapatkan sesuatu yang berharga dari sejumlah besar material dasar. Oleh karena itu, DM memiliki akar yang kuat pada bidang ilmu seperti kecerdasan buatan (artificial intelligence), machine learning, statistik, dan database.

Beberapa teknik yang sering digunakan dalam DM, seperti clustering, classification, dan association rule mining, menjadi sangat relevan untuk aplikasi skala besar(10). Misalnya, dalam penelitian ini, teknik clustering diterapkan untuk mengelompokkan ulasan pengguna aplikasi Gojek berdasarkan tingkat kepuasan dan versi aplikasi yang digunakan. Dengan menggunakan algoritma seperti K-Means, DM memungkinkan identifikasi pola-pola penting dalam kumpulan data besar yang sebelumnya sulit diakses atau dipahami secara manual. Hal ini memberikan wawasan yang signifikan bagi pengembang aplikasi untuk meningkatkan kualitas layanan berdasarkan kebutuhan dan pengalaman pengguna.

3.2 Proses Data Mining

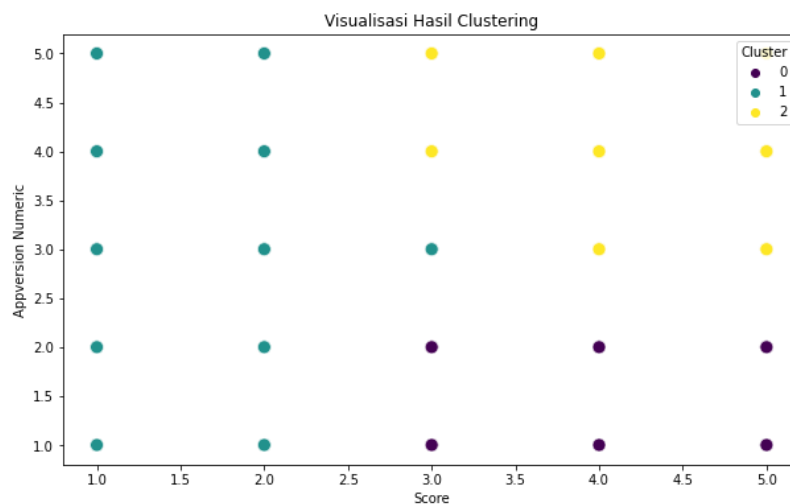
Tahap tahap Data Mining, karena Data Mining adalah suatu proses.

1. Pembersihan Data: Menghilangkan data yang tidak konsisten dan mengurangi noise pada dataset. Pada penelitian ini, proses ini dilakukan dengan menghapus data duplikat atau yang tidak relevan dari ulasan pengguna aplikasi Gojek.
2. Integrasi Data: Menggabungkan data dari berbagai sumber jika diperlukan. Namun, dalam penelitian ini, data ulasan sudah berasal dari satu sumber dataset aplikasi Gojek.
3. Transformasi Data: Data diubah menjadi format yang sesuai untuk proses mining. Pada penelitian ini, kolom appversion dikonversi menjadi format numerik untuk mempermudah analisis clustering. Selain itu, nilai score dan appversion_numeric distandarkan agar memiliki skala yang seragam.

4. Aplikasi Teknik Data Mining: Penerapan algoritma K-Means untuk clustering data ulasan berdasarkan tingkat kepuasan pengguna dan versi aplikasi yang digunakan.
5. Evaluasi Pola yang Ditemukan: Pola dari hasil clustering dianalisis untuk mengidentifikasi insight yang menarik, seperti cluster dengan kepuasan rendah pada versi aplikasi tertentu.
6. Presentasi Pengetahuan: Hasil clustering divisualisasikan menggunakan grafik scatter plot untuk mempermudah interpretasi dan pemahaman pola yang ditemukan.

3.3 Hasil

Hasil analisis clustering menunjukkan bahwa data pengguna aplikasi Gojek terbagi menjadi tiga cluster utama, masing-masing dengan karakteristik yang berbeda. Cluster 0 mencakup 1.609 data pengguna dengan rata-rata skor kepuasan 4,83 dan rata-rata versi aplikasi berada di sekitar 1,5, menunjukkan bahwa cluster ini didominasi oleh pengguna versi awal dengan tingkat kepuasan tinggi. Cluster 1 mencakup 1.019 data dengan rata-rata skor kepuasan 1,20 dan rata-rata versi aplikasi di sekitar 3,0, menggambarkan pengguna pada versi menengah dengan tingkat kepuasan rendah. Cluster 2, yang mencakup 2.372 data, memiliki rata-rata skor kepuasan 4,83 dan rata-rata versi aplikasi sekitar 4,0, merepresentasikan pengguna pada versi terbaru dengan tingkat kepuasan tinggi.



Gambar 2. Visualisasi Menggunakan Scatter Plot

Visualisasi hasil clustering menunjukkan distribusi data yang jelas di mana cluster dengan skor kepuasan tinggi (Cluster 0 dan Cluster 2) terpisah dari cluster dengan skor kepuasan rendah (Cluster 1). Hal ini mengindikasikan bahwa pembaruan pada versi aplikasi memiliki dampak signifikan terhadap kepuasan pengguna. Versi aplikasi terbaru cenderung mendapatkan ulasan yang lebih positif, sementara versi menengah menunjukkan adanya ketidakpuasan yang mungkin disebabkan oleh masalah performa atau fitur yang belum sempurna. Pada grafik scatter plot tersebut, distribusi data menunjukkan bahwa cluster dengan skor rendah (Cluster 1) sangat terkonsentrasi pada satu area, mencerminkan masalah yang mungkin terjadi pada versi menengah aplikasi. Di sisi lain, persebaran data pada Cluster 0 dan Cluster 2 lebih merata, menunjukkan bahwa pengguna pada versi awal dan terbaru memiliki pengalaman yang konsisten dan lebih memuaskan.

Dari distribusi data per cluster, terlihat bahwa jumlah data terbesar berada pada Cluster 2, yang merepresentasikan pengguna versi aplikasi terbaru dengan kepuasan tinggi. Hal ini menunjukkan bahwa pembaruan yang dilakukan oleh pengembang aplikasi pada versi terbaru berhasil meningkatkan pengalaman pengguna. Sebaliknya, Cluster 1 yang memiliki jumlah data terkecil menunjukkan adanya tantangan yang perlu diperhatikan pada pengguna versi menengah.

Hasil ini memberikan implikasi penting bagi pengembang aplikasi, yaitu perlunya fokus pada peningkatan fitur dan stabilitas di setiap versi untuk memastikan tingkat kepuasan pengguna yang konsisten. Selain itu, analisis ini juga dapat digunakan sebagai acuan untuk mengidentifikasi masalah spesifik pada versi aplikasi tertentu sehingga dapat segera diatasi.

4. Kesimpulan

Hasil analisis clustering pada data ulasan pengguna aplikasi Gojek menunjukkan bahwa versi aplikasi memiliki pengaruh signifikan terhadap tingkat kepuasan pengguna. Cluster 0 terdiri dari pengguna versi awal aplikasi (rata-rata versi 1,5) dengan tingkat kepuasan tinggi (rata-rata skor 4,83), Cluster 1 mewakili pengguna pada versi menengah (rata-rata versi 3,0) dengan tingkat kepuasan rendah (rata-rata skor 1,20), dan Cluster 2 mencakup pengguna versi terbaru aplikasi (rata-rata versi 4,0) dengan tingkat kepuasan tinggi (rata-rata skor 4,83). Visualisasi scatter plot memperlihatkan distribusi data yang tersegmentasi secara jelas berdasarkan skor kepuasan dan versi aplikasi. Versi aplikasi terbaru secara konsisten mendapatkan ulasan yang lebih positif, sementara versi menengah menghadapi beberapa tantangan yang menurunkan kepuasan pengguna. Oleh karena itu, pembaruan aplikasi yang berfokus pada stabilitas dan penyempurnaan fitur di setiap versi menjadi penting. Temuan ini memberikan acuan strategis bagi pengembangan aplikasi untuk terus meningkatkan pengalaman pengguna melalui pendekatan berbasis data.

Referensi

- [1] Riadi S, Ulum F. Analisis Penerapan Algoritma First Come First Served (Fcfs) Dalam Proses Pesanan Pada Aplikasi Gojek. *J Inform dan Rekayasa Perangkat Lunak* [Internet]. 2021;2(2):268–75. Available from: <http://jim.teknokrat.ac.id/index.php/informatika>
- [2] Muflihadi I, Rubiyanti RN. Pengaruh Perceived Usefulness , Perceived Ease of Use , Dan Trust Terhadap Kepuasan Konsumen (Studi Pada Gojek Bandung). *e-Proceeding Manag* [Internet]. 2016;3(2):2026–33. Available from: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/management/article/view/1945/1846>
- [3] Herlinda V, Darwis D. Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. Darwis, Dartono [Internet]. 2021;2(2):94–9. Available from: <http://jim.teknokrat.ac.id/index.php/JTSl>
- [4] Bakri M. Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means. *J Teknoinfo* [Internet]. 2017;11(1):6. Available from: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/article/view/3/3>
- [5] Wakhidah N. Clustering Menggunakan K-Means Algorithm. *J Transform* [Internet]. 2010;8(1):33. Available from: <https://journals.usm.ac.id/index.php/transformatika/article/view/45/45>
- [6] nurul rohmawati, sofi defiyanti mohamad jajuli. Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jitter 2015* [Internet]. 2015;1(2):62–8. Available from: <https://journal.widyatama.ac.id/index.php/jitter/article/view/56>
- [7] JASMINE K. 濟無No Title No Title No Title. Penambahan Natrium Benzoat Dan Kalium Sorbat Dan Kecepatan Pengadukan Sebagai Upaya Penghambatan Reaksi Inversi Pada Nira Tebu [Internet]. 2014;(1):26–32. Available from: https://d1wqtxts1xzle7.cloudfront.net/96758486/238400_penerapan_algoritma_k_means_clustering_a_97bef344-libre.pdf?1672778339=&response-content-disposition=inline%3B+filename%3DPenerapan_Algoritma_K_Means_Clustering_A.pdf&Expires=1735821117&Signature=Ay2Pa
- [8] Balaka MY. Metode penelitian Kuantitatif. *Metodol Penelit Pendidik Kualitatif* [Internet]. 2022;1:130. Available from: <https://repository.penerbitwidina.com/media/publications/464453-metodologi-penelitian-kuantitatif-10d6b58a.pdf>
- [9] Kusuma WA, Husniah L. Skeletonization using thinning method for human motion system. In: 2015 International Seminar on Intelligent Technology and Its Applications, ISITIA 2015 - Proceeding. Institute of Electrical and Electronics Engineers Inc.; 2015. p. 103–6.
- [10] Nur Khormarudin A, Kevin Synagogue Panjaitan OC, Septianingsih A, Faisal M, Utami WS, Novia Ningsi L, et al. Teknik Data Mining: Algoritma K-Means Clustering. *J Lebesgue J Ilm Pendidik Mat Mat dan Stat* [Internet]. 2022;1(2):116–23. Available from: <https://djournals.com/klik%0Ahttps://ilmukomputer.org/category/datamining/>

