

BAB I

PENDAHULUAN

1.1 Latar Belakang

Review pengguna berperan sebagai sumber informasi yang sangat penting dalam meningkatkan kualitas sebuah aplikasi, berbagai metode dapat digunakan untuk menganalisis ulasan, salah satunya adalah metode *clustering* (Rafi Haidar Arsyad et al., 2024). Opini yang diberikan oleh pengguna mengenai sebuah aplikasi menjadi bahan pertimbangan bagi calon pengguna sebelum mereka memutuskan untuk menggunakan aplikasi tersebut (Diki Hendriyanto et al., 2022) Ulasan pengguna merupakan bentuk umpan balik yang sangat penting bagi pengembang aplikasi. Ulasan ini mencakup komentar, saran, kritik, dan pengalaman pengguna saat menggunakan aplikasi. Dengan menganalisis ulasan ini, pengembang dapat mengidentifikasi kekuatan dan kelemahan aplikasi mereka, sehingga memungkinkan mereka mengambil keputusan yang lebih baik ketika mengembangkan fitur baru dan meningkatkan layanan mereka (Danahiswari et al., 2022).

Cookpad adalah aplikasi yang dirancang untuk membantu pengguna menemukan dan berbagi resep masakan. Seiring meningkatnya popularitas *Cookpad*, jumlah ulasan yang diberikan oleh pengguna juga meningkat secara dramatis (Danahiswari et al., 2022). Berdasarkan pengamatan terhadap aplikasi *Cookpad* di *Google Play Store*, banyak ulasan pengguna yang memuji kemudahan penggunaan aplikasi, terutama pada fitur resep masakan dan komunitas. Namun, terdapat juga keluhan yang sering muncul terkait kesulitan dalam navigasi aplikasi dan pencarian resep, yang dianggap kurang efisien. Beberapa pengguna juga mengusulkan fitur interaksi sosial yang lebih mendalam. Fenomena ini menunjukkan bahwa meskipun aplikasi *Cookpad* memberikan kemudahan dalam beberapa aspek, terdapat kesulitan yang perlu diatasi untuk meningkatkan pengalaman pengguna secara keseluruhan.

Agar dapat mempertahankan kualitas dari aplikasi *Cookpad* dan tetap menjadi andalan bagi yang suka mencoba pengalaman memasak yang baru, serta

mendapatkan pengalaman yang berkesan dari pengguna aplikasi *Cookpad* di *Google Play Store*, dapat digunakan teknik *Clustering*. Ini dilakukan agar dapat mengetahui fitur teknis atau fungsionalitas dibagian manakah yang disukai oleh pengguna aplikasi *Cookpad* (Danahiswari et al., 2022).

Berdasarkan pengamatan terkait aplikasi *Cookpad*, banyak pengguna yang memberikan masukan terkait *user experience*. Di sisi lain, aplikasi ini memiliki ciri tampilan yang tenang, fitur pencarian yang cepat, dan petunjuk yang mudah diikuti. Namun, ada juga beberapa yang menimbulkan masalah, seperti perubahan tampilan akibat *update* yang kurang baik atau *bug* yang menyebabkan fitur tidak berfungsi dengan baik. Ulasan yang beragam ini belum dikategorikan secara sistematis, sehingga perlu mengkategorikannya menjadi dua kelompok, yaitu kemudahan pengguna dan kesulitan pengguna, agar informasi lebih mudah dipahami.

Salah satu teknik *data mining* yang digunakan untuk pengelompokan adalah *clustering*. *Clustering* merupakan proses pengelompokan sekumpulan data ke dalam dua kelompok atau lebih, sehingga data yang berada dalam kelompok yang sama memiliki kemiripan lebih besar dibandingkan data yang ada di kelompok lain. Pengelompokan ini hanya didasarkan pada informasi yang tersedia dari data tersebut (Danahiswari et al., 2022).

Salah satu metode *clustering* adalah *K-Means*, yang termasuk dalam pengelompokan *non-hierarki*. Metode ini mampu membagi data ke dalam dua atau lebih kelompok, di mana data dengan karakteristik yang sama akan dikelompokkan bersama, sedangkan data dengan karakteristik yang berbeda akan dimasukkan ke kelompok lain. Tujuan utama dari pengelompokan ini adalah meminimalkan fungsi objektif yang digunakan dalam proses pengelompokan, dengan berupaya untuk mengurangi variasi dalam satu kelompok dan meningkatkan perbedaan antara kelompok yang berbeda (Nugraha et al., 2022).

Proses *K-Means* dimulai dengan menentukan jumlah *cluster* yang diinginkan, yang disebut sebagai *k* selanjutnya, *centroid* atau pusat *cluster* dipilih secara acak dari titik data dalam dataset. Setelah *centroid* ditentukan, setiap titik data dikelompokkan ke dalam *cluster* berdasarkan kedekatannya dengan *centroid*, yang dihitung menggunakan jarak *Eulidean*. Setelah seluruh titik data dikelompokkan,

centroid baru dihitung dengan mengambil rata-rata posisi semua titik dalam *cluster* tersebut. Proses ini diulang data dikelompokkan kembali berdasarkan *centroid* yang baru, dan *centroid* diperbarui berdasarkan data yang telah dikelompokkan proses ini terus berlanjut hingga tidak ada perubahan yang signifikan dalam penugasan *cluster* atau sampai jumlah iterasi maksimum tercapai (Suraya et al., 2023).

Selain *K-Means*, ada juga algoritma *K-Medoids* yang menggunakan *medoid* (perwakilan data) untuk menentukan pusat klaster, menggantikan rata-rata (*mean*) pada *K-Means*. *Medoid* adalah objek yang paling representatif dalam klaster. Prosesnya dimulai dengan memilih beberapa objek secara acak sebagai *medoid*, lalu mengelompokkan data berdasarkan kedekatannya dengan *medoid* yang dipilih (Ayu et al., 2021).

Adapun cara kerja algoritma *K-Medoids* adalah dengan menggunakan *medoid* sebagai pusat klaster, yaitu objek yang ada dalam klaster dengan jarak terkecil ke semua objek lainnya, bukan rata-rata. Proses dimulai dengan menentukan jumlah klaster *k*, lalu memilih *k* objek secara acak sebagai *medoid*. Setiap data kemudian dikelompokkan berdasarkan kedekatannya dengan *medoid*, dihitung menggunakan jarak *Eulidean*. Selanjutnya, *medoid* baru dipilih berdasarkan objek dalam klaster yang meminimalkan jarak total ke semua objek lainnya. Proses ini diulang hingga posisi *medoid* tidak lagi berubah, menandakan klaster sudah stabil (Ena Tasia & Afdal, 2023).

Pengelompokan data bisa dilakukan dengan berbagai cara, salah satunya menggunakan algoritma *K-Means* dan *K-Medoids*. Kedua metode ini mudah digunakan dan efektif untuk mengelola data yang banyak. *K-Means* mengelompokkan data berdasarkan jarak antar objek, kemudian membaginya ke dalam beberapa kelompok. Sedangkan *K-Medoid* memilih objek yang paling mewakili kelompok, yang disebut *medoid*, sebagai pusat kelompok. Perbedaan utama antara keduanya adalah cara menentukan pusat kelompok. Pada *K-Medoids*, pusat kelompok dipilih dari objek yang ada dalam kelompok itu sendiri (*medoid*), sedangkan pada *K-Means*, pusat kelompok dihitung dari rata-rata nilai semua objek dalam kelompok (Hidayat et al., 2024).

Penelitian ini bertujuan untuk mengisi kekosongan dalam literatur yang ada terkait dengan analisis ulasan pengguna pada aplikasi *Cookpad*. Dengan menerapkan algoritma *K-Means* dan *K-Medoids*, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam membantu pengembang aplikasi untuk meningkatkan fitur-fitur yang ada, berdasarkan analisis ulasan pengguna yang lebih mendalam.

Salah satu penelitian terdahulu yang relevan mengenai perbandingan algoritma *K-Means* dan *K-Medoids* dalam pengelompokan data adalah studi yang dilakukan oleh (Ena Tasia & Afdal, 2023) yang dipublikasikan dalam *Indonesian Journal of Informatic Research and Software Engineering*. Penelitian ini membandingkan kedua algoritma tersebut untuk mengelompokkan daerah rawan banjir di Kabupaten Rokan Hilir pada periode 2019-2022. Data yang digunakan mencakup atribut seperti kecamatan, desa, dan jumlah jiwa yang terdampak banjir, yang diperoleh dari Dinas Sosial Kabupaten Rokan Hilir. Hasil eksperimen menunjukkan bahwa *K-Means* lebih optimal dibandingkan *K-Medoids* dalam hal validitas *cluster*. Penelitian ini juga menemukan bahwa klaster 3 merupakan klaster dengan jumlah daerah terbanyak, sedangkan klaster 2 berisi 3 daerah dengan jumlah korban terdampak banjir mencapai 13.463 jiwa. Daerah-daerah tersebut adalah Bagan Cempedak, Rantau Kopar, dan Melayu Besar, yang terletak di kecamatan Tanah Putih Tanjung Melawan dan Rantau Kopar. Faktor penyebab sering terjadinya banjir di daerah tersebut antara lain adalah aliran sungai (DAS), struktur tanah gambut, dan dataran rendah. Penelitian ini menunjukkan pentingnya penerapan algoritma *clustering* untuk mengidentifikasi daerah rawan bencana alam, yang dapat membantu dalam perencanaan penanganan bencana dan distribusi bantuan secara lebih efisien (Ena Tasia & Afdal, 2023).

Penelitian lain yang relevan mengenai perbandingan algoritma *K-Means* dan *K-Medoids* dalam pengelompokan data studi yang dilakukan oleh (Hidayat et al., 2024) yang dipublikasikan dalam Jurnal TEKINKOM. Penelitian ini bertujuan untuk mengelompokkan kabupaten dan kota di Jawa Barat berdasarkan masalah gizi buruk pada balita selama periode 2019-2022. Data yang digunakan mencakup atribut seperti nama kabupaten, kategori gizi, dan jumlah balita yang terdampak,

yang diperoleh dari Dinas Kesehatan. Hasil analisis dengan *K-Means* menghasilkan 3 klaster, yaitu Klaster 1 (tinggi) mencakup 16 kabupaten/kota, Klaster 2 (sedang) mencakup 10 kabupaten/kota, dan Klaster 3 (rendah) mencakup 1 kabupaten/kota. Sedangkan dengan *K-Medoids*, didapat klaster 1 (tinggi) terdiri dari 10 kabupaten/kota, Klaster 2 (sedang) terdiri dari 4 kabupaten/kota, dan Klaster 3 (rendah) terdiri dari 13 kabupaten/kota. Dari hasil perbandingan, *K-Means* menunjukkan performa lebih baik dengan nilai *Silhouette Coefficient* sebesar 0,617, menunjukkan bahwa *K-Means* lebih optimal dalam mengelompokkan data berdasarkan kasus gizi buruk di wilayah tersebut (Hidayat et al., 2024).

Dari beberapa penelitian terdahulu yang menggunakan proses *data mining* dengan metode *K-Means* dan *K-Medoids*, penelitian tersebut membuktikan bahwa kedua algoritma tersebut merupakan metode yang sesuai digunakan pada proses klasterisasi. Namun, penelitian tersebut penulis melihat tidak ada yang membahas mengenai sistem klasterisasi yang menyangkut ulasan pengguna aplikasi *Cookpad*. Oleh karena itu, penulis merasa penting untuk melakukan penelitian ini. Klasterisasi ini dilakukan karena beragamnya jenis ulasan yang diberikan oleh pengguna aplikasi *Cookpad*.

Berdasarkan pembahasan sebelumnya, dapat disimpulkan bahwa analisis ulasan pengguna aplikasi *Cookpad* menggunakan metode *clustering* dengan algoritma *K-Means* dan *K-Medoids* dapat memberikan gambaran yang jelas tentang kemudahan dan kesulitan yang dialami oleh pengguna. Dengan mengelompokkan ulasan pengguna, penelitian ini menemukan pola-pola yang dapat membantu pengembang dalam memperbaiki area yang membutuhkan peningkatan dan mempertahankan fitur yang disukai. Penelitian ini memberikan kontribusi dalam meningkatkan kualitas aplikasi *Cookpad* melalui pemahaman yang lebih baik terhadap pengalaman pengguna, serta menjadi referensi dalam pengembangan fitur dan perbaikan aplikasi di masa depan.

Berdasarkan penjelasan yang telah dijelaskan sebelumnya, analisis ulasan terhadap aplikasi *Cookpad* dengan metode *clustering* menggunakan algoritma *K-Means* dan *K-Medoids* dapat memberikan gambaran yang jelas tentang kesulitan dan kendala yang dihadapi oleh pengguna. Penelitian ini mengungkap pola yang

dapat membantu pengembang memperbaiki area yang perlu ditingkatkan dan mempertahankan fitur yang disukai, berkat kompilasi ulasan pengguna. Penelitian ini berkontribusi untuk meningkatkan kualitas aplikasi *Cookpad* melalui pemahaman yang lebih baik tentang pengalaman pengguna dan berfungsi sebagai panduan untuk pengembangan fitur dan peningkatan aplikasi di masa mendatang.

Berdasarkan latar belakang yang telah dijabarkan, maka judul proposal yang diangkat pada penelitian ini berjudul “Analisis Perbandingan Algoritma dan *K-Means* dan *K-Medoids* untuk Klasterisasi Teks Ulasan pada Aplikasi *Cookpad*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana perbandingan kualitas hasil klasterisasi dan ketahanan terhadap *outlier* antara algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan ulasan pengguna aplikasi *Cookpad*?
2. Bagaimana pengaruh parameter jumlah *cluster* (*k*) terhadap kualitas hasil klasterisasi ulasan pengguna aplikasi *Cookpad* menggunakan algoritma *K-Means* dan *K-Medoids*?
3. Bagaimana penelitian ini dapat memberikan panduan akademik yang berguna bagi peneliti di masa depan dalam memahami algoritma *K-Means* dan *K-Medoids* dalam klasterisasi?

1.3 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan sebagai berikut:

1. Membandingkan performa algoritma *K-Means* dan *K-Medoids* dalam klasterisasi teks ulasan pengguna aplikasi *Cookpad* berdasarkan kualitas hasil klaster dan ketahanan terhadap *outlier*.
2. Menganalisis pengaruh jumlah *cluster* (*k*) terhadap kualitas hasil klasterisasi ulasan pengguna aplikasi *Cookpad* menggunakan algoritma *K-Means* dan *K-Medoids*.

3. Memberikan panduan bagi peneliti atau praktisi untuk memahami tentang algoritma *K-Means* dan *K-Medoids* dalam klasterisasi.

1.4 Manfaat Penelitian

Penelitian ini memiliki beberapa manfaat yang dapat dijelaskan sebagai berikut:

1. Memberikan pemahaman tentang performa algoritma *K-Means* dan *K-Medoids* dalam klasterisasi teks ulasan pengguna aplikasi *Cookpad*, khususnya kualitas hasil klasterisasi, dan ketahanan terhadap data *outlier*.
2. Memberikan referensi tentang pengaruh parameter jumlah *cluster* (*k*) terhadap hasil klasterisasi, sehingga pengembang atau peneliti lain dapat menggunakan konfigurasi yang optimal untuk mengolah dataset serupa.
3. Penelitian ini dapat dijadikan acuan bagi peneliti lain yang ingin menggunakan metode klasterisasi *K-Means* dan *K-Medoids* dalam analisis data teks, khususnya ulasan pengguna.

1.5 Ruang Lingkup dan Batasan Penelitian

Agar penelitian ini lebih terarah, berikut adalah beberapa batasan yang digunakan:

1. Penelitian ini hanya mengambil data ulasan teks terbaru di *Google Playstore*, dan tidak mencakup *rating* atau ulasan yang berupa angka tanpa komentar.
2. Data yang di ambil berkisar 15000 data (2020-2025)
3. Perbandingan akan difokuskan hanya pada dua aspek utama, yaitu kualitas klaster dan ketahanan terhadap *outlier*, untuk memastikan analisis tetap mendalam tanpa meluas ke terlalu banyak aspek.
4. Fokus penelitian ini adalah pada analisis dan perbandingan kualitas hasil klasterisasi, khususnya dalam hal evaluasi kualitas klaster dan ketahanannya terhadap *outlier*. Setelah proses klasterisasi, dilakukan interpretasi isi tiap klaster untuk memahami kecenderungan topik dalam masing-masing klaster.
5. Metode yang digunakan untuk mengklasterisasi pada penelitian ini menggunakan algoritma *K-Means* dan *K-Medoids*.