

ANALISIS SENTIMEN KELUHAN PENGGUNA PADA GAME SUPREMACY 1914 MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Jery Cevin^{1*}, David Gultom², Nikolas Orlando³, Hendrik Fery Herdiatmoko⁴

^{1,2,3,4}Universitas Katolik Musi Charitas Palembang, Indonesia

¹Jerylevin22@gmail.com, ²davidgultom669@gmail.com

³nikolasorlando00@gmail.com, ⁴hendrik@ukmc.ac.id

Received: 30-10-2024

Revised: 15-11-2024

Approved: 27-11-2024

ABSTRACT

Penelitian ini bertujuan menganalisis keluhan pengguna untuk pengembangan game Supremacy 1914 dengan memanfaatkan 1000 ulasan berbahasa Indonesia di Google Playstore. Ulasan ini diproses melalui tahapan preprocessing, seperti pemisahan kata (tokenizing), penyaringan (filtering), penghapusan kata tidak penting (stop word), dan perubahan kata ke bentuk dasar (stemming). Data yang telah diproses dikelompokkan menggunakan metode K-Means menjadi tiga kelompok (cluster), yang divisualisasikan dalam bentuk wordcloud. Cluster 0 mencakup ulasan umum dengan kata-kata seperti "bagus", "seru", dan "mantap", sedangkan Cluster 1 fokus pada masukan terkait kualitas game dan pengalaman pengguna. Penelitian ini diharapkan membantu pengembang mengidentifikasi kebutuhan pengguna lebih cepat, meningkatkan kepuasan, dan mendukung pengembangan game secara berkelanjutan.

Kata Kunci: *Supremacy 1914, K-Means, WorldCloud.*

PENDAHULUAN

Game online merupakan salah satu jenis hiburan yang paling populer di berbagai masyarakat di era internet. Supremacy 1914, sebuah permainan strategi yang sangat disukai, menuntut pemain untuk mengatur sumber daya, membentuk aliansi, dan membuat taktik dalam simulasi perang. Pengalaman bermain pengguna tidak selalu lancar meskipun populer. Banyak pengguna menyampaikan keluhan mereka melalui berbagai platform, termasuk forum, ulasan aplikasi, dan media sosial. Berbagai elemen termasuk masalah ini, termasuk bug dalam permainan, sistem pembayaran, dan aspek interaksi antarpemain.

Pengembang game harus memahami keluhan untuk meningkatkan pengalaman pengguna. Analisis Sentimen merupakan metode efektif untuk mengekstraksi informasi bernilai dari data tidak terstruktur [1], ini efektif untuk menemukan pola emosi, baik positif, negatif, maupun netral, yang ditemukan dalam ulasan pengguna. Analisis sentimen adalah proses otomatis untuk menemukan dan mengolah kata-kata yang menunjukkan perasaan dalam sebuah teks. Teknik ini secara sistematis mengidentifikasi dan memproses informasi teks untuk mengungkap emosi yang tersembunyi di dalamnya [2]. Namun, pengelompokan data yang besar dan berbagai tema keluhan sering menjadi masalah. Dalam menganalisis perasaan atau pendapat orang-orang dari tulisan mereka, metode K-Means bisa secara otomatis memilah-milah tulisan tersebut ke dalam kelompok-kelompok yang berbeda berdasarkan sentimen yang terkandung di dalamnya [3].

Beberapa penelitian dilakukan terkait sentimen analisis menggunakan algoritma K-Means. Penerapan metode K-Means pada toko Helai, yaitu dengan cara mengelompokkan data stok baju. Kemudian memilih 3 cluster secara acak sebagai centroid awal. Setelah data pada setiap cluster tidak berubah-ubah, maka dapat diketahui hasil akhirnya yaitu yang sangat laris ada 97 artikel, yang laris ada 55 artikel, dan 34 artikel untuk yang kurang laris [4]. Studi penelitian tentang penerapan algoritma

K-Means Clustering pada data penjualan produk makana dingin (frozen food) menunjukkan pengelompokan minat konsumen. Dari 45 produk yang dianalisis, 3 produk tergolong dalam Cluster 1 dengan minat konsumen yang rendah, sedangkan 42 produk termasuk dalam Cluster 2 yang memiliki minat konsumen tinggi [5].

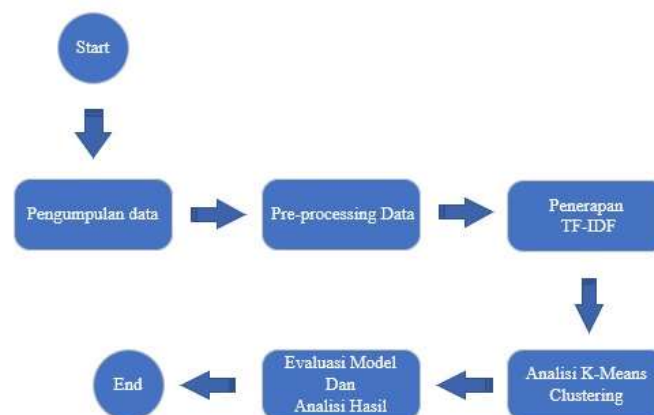
Penelitian menunjukkan bahwa statistik tiap pemain FPS bervariasi berdasarkan variabel K/D (Kill/Death ratio), KAST, dan Rounds Played. Performanya diwakili oleh K/D dan Rounds Played. Melalui algoritma K-Means serta metode elbow dan silhouette, clustering optimal tercapai saat $k = 3$ [6]. Analisis data peminjam buku menggunakan metode K-Means Clustering di RapidMiner menunjukkan hasil yang sama dengan hitungan manual. Buku yang paling sering dipinjam berada di Cluster 1 (9 item), yang paling sedikit dipinjam di Cluster 2 (15 item), dan yang cukup sering dipinjam di Cluster 0 (12 item). Penerapan data mining ini membantu perpustakaan dalam mengidentifikasi buku yang paling sering dipinjam [7].

Penelitian ini melakukan clustering pada 2000 universitas teratas di dunia dalam 10 iterasi, menghasilkan tiga cluster: rendah, sedang, dan tinggi. Profiling menunjukkan 667 universitas di cluster tinggi, 667 di cluster sedang, dan 666 di cluster rendah. Di antara universitas tersebut, Harvard University berada di cluster tertinggi (Cluster 2), University of Haifa di cluster sedang (Cluster 1), dan National Chung Cheng University di cluster rendah (Cluster 0) [8].

METODE PENELITIAN

Data yang digunakan untuk penelitian digunakan untuk melakukan analisis sentimen pada keluhan pengguna game Supremacy 1914 di Google Play Store. Untuk membuat dataset penelitian ini siap untuk dianalisis, proses preprocessing diperlukan untuk menghilangkan data yang tidak penting [9].

Tahapan preprocessing kalimat seperti case folding, tokenization, filtering, stopword, stemming, dan labeling dilakukan dengan metode TF-IDF dalam penelitian ini. Selanjutnya, clustering digunakan menggunakan K-Means Clustering untuk mendapatkan wordcloud. Gambar 1 menunjukkan jalur proses preprocessing kalimat.



Gambar 1. Alur Penelitian

1. Pengumpulan Data

Data dikumpulkan secara manual dengan mengambil data komentar pada Google Play Store untuk game Supremacy 1914 melalui Scraping dengan Python. Python adalah bahasa pemrograman interpretatif yang mudah

dipelajari dan dapat dijalankan pada berbagai platform dengan fokus utama pada keterbacaan kode[10]. Ulasan yang diambil sebagai data didasarkan pada opini pengguna terkait pengalaman mereka dalam menggunakan game Supremacy 1914. Jumlah data yang dikumpulkan adalah 1000 ulasan, yang mencakup ulasan positif dan negatif, sebagaimana terlihat pada Gambar 2.



Gambar 2. Sumber data pada komentar

2. Preprocessing

Penelitian ini dilalui dengan beberapa tahap Preprocessing sebagai berikut :

- **Case Folding**
Pada titik ini, data akan diproses untuk mengubah semua huruf kapital menjadi huruf kecil.
- **Tokenizing**
Proses tokenisasi akan membagi data menjadi kumpulan beberapa kata, dengan setiap kata memiliki nilai dan arti yang berbeda.
- **Filtering dan Stopword**
Filtering adalah Langkah ketiga dalam proses prapemrosesan data adalah filtering dan filtering titik akhir. Pada titik ini, tanda baca dan hashtag yang dianggap tidak penting dihapus. Selain itu, daftar stopwords yang telah ditentukan akan digunakan untuk menghapus kata-kata yang tidak banyak digunakan dalam data komentar.
- **Stemming**
Stemming merupakan tahap lanjutan setelah filtering dan stopwords removal. Pada tahap ini, kata-kata yang telah melalui langkah sebelumnya akan dikonversi menjadi kata dasar dalam Bahasa Indonesia yang sesuai dengan kaidah bahasa baku.

3. Pembobotan term weighting TF-IDF

Setelah data diproses awal, tahap selanjutnya yaitu menghitung nilai penting setiap kata dalam kumpulan dokumen. Nilai ini akan digunakan untuk menentukan seberapa relevan suatu kata dalam memprediksi sentimen.

3.1 Empat langkah utama digunakan dalam proses pembobotan kata. Pertama, Anda harus menemukan jumlah frekuensi kata (TF) untuk setiap kata. Setelah prapemrosesan selesai, langkah ini dilakukan dengan memotong kalimat menjadi kata-kata individu, seperti yang ditunjukkan pada tabel di bawah..

3.2 Langkah kedua adalah menghitung jumlah Document Frequency (DF) untuk

setiap kata, yaitu menghitung berapa banyak dokumen yang memuat kata tertentu.

3.3 Pada langkah ketiga, Inverse Document Frequency (IDF) dihitung dengan menggunakan rumus berikut:

$$idf = \log\left(\frac{n}{df}\right) \tag{1}$$

Tabel 1. *Ilustrasi Perhitungan Term Frequency pada Kalimat.*

TF	
bakal	1
offline	1
mengajarkan	1
mohon	1

bakal
offline
mengajarkan
mohon

Contoh, untuk kata "offline" yang hanya ditemukan pada satu dokumen, perhitungannya yaitu sebagai berikut :

$$idf = \log\left(\frac{5}{1}\right) = \log 2,5 = 0,3979400087$$

Sementara itu, untuk kata "game" yang muncul pada dua dokumen berbeda, perhitungannya menjadi :

$$idf = \log\left(\frac{5}{2}\right) = \log 1,6 = 0,2218487505$$

3.4 Mengalikan nilai Frekuensi Terma (tf) dan Frekuensi Inverse Dokumen (idf) untuk menghitung bobot atau berat (w). Proses ini dimulai dengan menghitung hasil kali antara tf dan idf untuk setiap kata dalam setiap kalimat. Kemudian, seperti yang dijelaskan dalam Persamaan ke-2, nilai-nilai ini dijumlahkan untuk mendapatkan bobot total (w) dari setiap kalimat.

$$tfidf = tf(t, d) * idf \tag{2}$$

Dengan menggunakan rumus berikut, persamaan ke-3 dapat menghitung nilai berat atau bobot kalimat secara keseluruhan:

$$w = (tfidf_1 + tfidf_2 + \dots + tfidf_n) \tag{3}$$

Tabel 2 menunjukkan hasil perhitungan tf-idf untuk kalimat yang memiliki kata negatif dan nilai bobot untuk setiap kalimat.

Tabel 2. *Ilustrasi Perhitungan bobot kalimat*

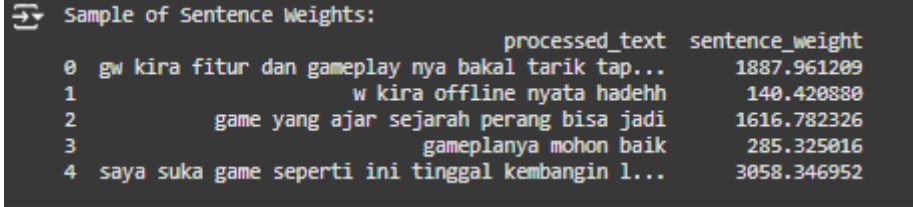
No	kata	tf	idf	Tf-idf
1	bakal	1	0.3979400087	0.3979400087
2	offline	1	0.3979400087	0.3979400087
3	mengajarkan	1	0.3979400087	0.3979400087
4	mohon	1	0.3979400087	0.3979400087
Bobot atau Weight (w)				1,59176003

4. Analisis Metode K-Means Clustering

Tahapan yang dijalankan untuk menganalisis menggunakan metode K-Means yaitu :

4.1 Menentukan jumlah cluster (k). cluster yang digunakan ialah 2 cluster.

4.2 Pada penelitian ini, jumlah cluster yang dipilih adalah Menentukan centroid awal (C_i). Setelah tahap pre-processing dan pembobotan selesai (menghasilkan nilai bobot atau (w)), bobot dari setiap kalimat akan digunakan dalam proses perhitungan untuk metode K-Means clustering. Data hasil pembobotan (w) yang diperoleh digunakan sebagai input untuk proses clustering bisa dilihat pada gambar 3.



	processed_text	sentence_weight
0	gw kira fitur dan gameplay nya bakal tarik tap...	1887.961209
1	w kira offline nyata hadehh	140.420880
2	game yang ajar sejarah perang bisa jadi	1616.782326
3	gameplanya mohon baik	285.325016
4	saya suka game seperti ini tinggal kembangin l..	3058.346952

Gambar 3. Perhitungan bobot w

Berdasarkan hasil pembobotan kalimat yang telah dihitung, nilai terkecil (kalimat 5, ($K 1$)= 1.40420880 dan nilai terbesar (kalimat 1, ($K 5$)= 3.058346952) digunakan untuk menentukan centroid.

4.3 Tahap berikutnya yaitu mengelompokkan data menjadi beberapa cluster berdasarkan kesamaan karakteristiknya. Pusat dari setiap cluster disebut centroid. Pengelompokan ini didasarkan pada nilai bobot yang diberikan pada setiap kalimat, sehingga kita bisa menentukan apakah suatu kalimat masuk ke cluster pertama atau kedua :

- 1) $C1 = k0, k1, k2$
- 2) $C2 = k3, k4$

4.4 Gunakan Persamaan ke-4 untuk menghitung jarak setiap anggota cluster untuk menghasilkan centroid baru.

$$C(i) = \frac{(x1+ x2+\dots+ xn)}{\sum x} \quad (4)$$

Keterangan :

$C(i)$ = centroid baru.

$x1, x2, xn$ = nilai yang ada pada masing-masing anggota cluster.

$\sum x$ = jumlah anggota pada setiap cluster.

Sampai pusat kelompok dan anggota kelompok mengalami perubahan atau tetap konsisten, langkah ketiga dan keempat diulangi.

HASIL DAN PEMBAHASAN

Gambar 4 menunjukkan hasil pengelompokan data metode K-Means. Algoritma K-Means menggunakan data yang telah diproses (Hasil Stemming) sebagai input, dan hasilnya adalah pembagian data menjadi beberapa kelompok, dengan setiap kalimat ditempatkan di kelompok yang tepat.

"saya," dan "untuk." Wordcloud Cluster 2 menampilkan ulasan seputar game dengan cakupan yang lebih beragam, dengan kata-kata seperti "yang," "di," "saya," "untuk," dan "ada" yang mencerminkan komentar bersifat personal dan spesifik terhadap berbagai aspek game. Penelitian ini dapat dikembangkan dengan menggabungkan metode K-Means dengan teknik lain untuk meningkatkan akurasi hasil, sehingga diharapkan dapat membantu perusahaan memahami keluhan pelanggan dan meningkatkan kualitas layanan mereka.

DAFTAR PUSTAKA

- [1] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [2] E. Sinduningrum, F. N. Hasan, A. R. Dzikrillah, A. B. Rossianiz, D. Febriawan, and I. Irawati, "Peningkatan Kemampuan Data Analytic Melalui Pelatihan Asean Data Science Explorers Menggunakan Sap Analytic Cloud," *SELAPARANG J. Pengabd. Masy. Berkemajuan*, vol. 6, no. 4, p. 1708, 2022, doi: 10.31764/jpmb.v6i4.10734.
- [3] R. Wati and S. Ernawati, "Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python," *J. Tek. Inform. UNIKA St. Thomas*, vol. 06, pp. 240–247, 2021, doi: 10.54367/jtiust.v6i2.1465.
- [4] N. I. R, A. Dahlia, and P. Ningum, "Sentiment Analysis of Public Opinion on Films Taylor Swift Eras Tour on the Twitter Platform Using the Machine Learning," vol. 1, no. August, pp. 85–92, 2024.
- [5] Tukino, S. Masrurroh, and D. Herdiana, "Applying the Prototype Model into the Electronic Reporting System for the Elementary School Student Base on Android," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, vol. 2, no. 1, pp. 17–21, 2021, doi: 10.36805/bit-cs.v2i1.1241.
- [6] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [7] L. Azzahra and Amru Yasir, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilmu Komput. dan Sist. Inf.*, vol. 3, no. 1, pp. 1–10, 2024, doi: 10.70340/jirsi.v3i1.88.
- [8] M. I. Muharizki and D. B. Arianto, "... Clustering Dengan Metode K-Means Terhadap Statistik Permainan Pro-Player Valorant Pada Kompetisi Valorant Champions 2022," *Serunai J. Ilm. Ilmu ...*, vol. 9, pp. 40–47, 2023, [Online]. Available: <https://ejournal.stkipbudidaya.ac.id/index.php/ja/article/view/846%0Ahttps://ejournal.stkipbudidaya.ac.id/index.php/ja/article/download/846/571>
- [9] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- [10] F. Dikarya and S. Muharni, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Universitas Terbaik Di Dunia," *J. Inform.*, vol. 22, no. 2, pp. 124–131, 2022, doi: 10.30873/ji.v22i2.3324.