

Topic Modelling Pengalaman Pengguna Game Roblox Menggunakan Algoritma Non-Negative Matrix Factorization (NMF)

Kamarudin¹, Nuri Cahyono²

^{1,2}Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Sleman, Indonesia

Author Email: kamarudin@amikom.ac.id¹, nuricahyono@amikom.ac.id²

Abstrak. Roblox merupakan salah satu platform permainan daring berbasis user-generated content yang memiliki jutaan pengguna aktif dan menghasilkan ulasan dalam jumlah besar di Google Play Store. Ulasan tersebut memuat berbagai pengalaman, kepuasan, maupun keluhan pengguna yang dapat dimanfaatkan sebagai masukan bagi pengembang. Namun, banyaknya data ulasan yang bersifat tidak terstruktur menyebabkan proses analisis secara manual menjadi tidak efektif. Penelitian ini bertujuan mengidentifikasi topik-topik utama yang terdapat pada ulasan pengguna aplikasi Roblox menggunakan metode topic modelling berbasis Non-Negative Matrix Factorization (NMF). Data penelitian diperoleh melalui proses scraping terhadap 10.000 ulasan terbaru berbahasa Indonesia di Google Play Store. Tahapan penelitian meliputi preprocessing berupa cleaning, case folding, normalisasi kata slang, stopword removal, dan penghapusan dokumen kosong sehingga diperoleh 9.835 ulasan yang siap dimodelkan. Selanjutnya dilakukan representasi teks menggunakan TF-IDF, evaluasi variasi jumlah topik menggunakan Coherence Score, serta pemodelan NMF dengan jumlah topik terbaik. Hasil penelitian menunjukkan bahwa nilai K=6 menghasilkan Coherence Score tertinggi sebesar 0,5111, sehingga dipilih sebagai jumlah topik optimal. Enam topik yang dihasilkan menggambarkan pengalaman pengguna terkait kualitas permainan, keseruan bermain, kepuasan pengguna, performa aplikasi, serta berbagai kendala teknis seperti fitur chat, map, dan jaringan. Temuan ini menunjukkan bahwa kombinasi algoritma NMF dan evaluasi Coherence Score mampu menghasilkan topik yang koheren dan mudah diinterpretasikan, sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan untuk meningkatkan kualitas layanan dan pengalaman pengguna aplikasi Roblox.

Kata kunci: Topic Modelling, Non-Negative Matrix Factorization, Roblox, Coherence Score

1. PENDAHULUAN

Perkembangan industri permainan digital telah mengalami pertumbuhan yang sangat pesat seiring meningkatnya penetrasi internet, perangkat mobile, serta teknologi komputasi berbasis cloud. Salah satu platform permainan yang mengalami pertumbuhan signifikan adalah Roblox yang merupakan platform game online yang mengusung konsep user-generated content (UGC), sehingga memungkinkan pengguna tidak hanya memainkan permainan, tetapi juga merancang, mengembangkan, dan mendistribusikan game mereka sendiri melalui Roblox Studio [1]. Model ekosistem tersebut menjadikan Roblox sebagai salah satu platform metaverse terbesar dengan jutaan pengalaman (experiences) yang diciptakan oleh komunitas global. Fleksibilitas dalam pengembangan game, sistem ekonomi virtual berbasis Robux, serta dukungan lintas platform telah menjadikan Roblox populer di berbagai kelompok usia, khususnya generasi muda [2]. Tingginya jumlah pengguna aktif menghasilkan interaksi yang masif, termasuk munculnya berbagai ulasan pengguna pada Google Play Store yang merepresentasikan persepsi, kepuasan, maupun keluhan terhadap kualitas layanan dan pengalaman bermain.

Ulasan pengguna pada Google Play Store merupakan sumber informasi yang sangat berharga untuk memahami pengalaman pengguna (user experience) terhadap aplikasi Roblox. Berbagai aspek seperti performa aplikasi, kualitas grafis, stabilitas server, keamanan akun, mekanisme monetisasi, pembaruan sistem, hingga interaksi sosial sering kali menjadi topik utama yang dibahas oleh pengguna. Jumlah ulasan yang terus bertambah setiap hari menyebabkan proses identifikasi isu secara manual menjadi tidak efisien dan rentan terhadap subjektivitas. Selain itu, karakteristik data ulasan yang bersifat tidak terstruktur (unstructured text) menyulitkan pengembang dalam memperoleh gambaran menyeluruh mengenai kebutuhan, harapan, maupun permasalahan yang dialami pengguna [3]. Akibatnya, informasi penting yang terkandung dalam ribuan ulasan berpotensi tidak dimanfaatkan secara optimal sebagai dasar pengambilan keputusan untuk peningkatan kualitas aplikasi.

Salah satu pendekatan yang dapat digunakan untuk mengekstraksi informasi dari kumpulan dokumen teks dalam jumlah besar adalah topic modelling [4]. Topic modelling merupakan teknik unsupervised machine learning yang bertujuan mengidentifikasi sekumpulan topik laten berdasarkan pola kemunculan kata pada dokumen tanpa

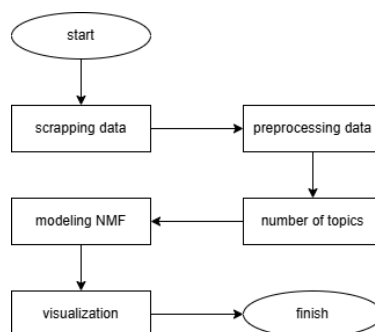
memerlukan proses pelabelan sebelumnya [5]. Melalui pendekatan ini, ribuan ulasan pengguna dapat dikelompokkan ke dalam beberapa topik utama yang merepresentasikan isu, opini, maupun tema pembahasan yang dominan. Hasil identifikasi topik tersebut dapat membantu pengembang dalam memahami kebutuhan pengguna secara lebih sistematis, mengidentifikasi prioritas perbaikan aplikasi, serta mendukung proses evaluasi kualitas layanan berbasis data. Topic modelling menjadi solusi yang efektif untuk mengatasi kompleksitas analisis data ulasan pengguna yang berskala besar [6].

Algoritma topic modelling yang memiliki performa baik pada data teks adalah Non-Negative Matrix Factorization (NMF). Algoritma ini bekerja dengan melakukan dekomposisi matriks dokumen-term menjadi dua matriks non-negatif yang merepresentasikan hubungan antara dokumen dan topik serta hubungan antara topik dan kata [7]. Berbeda dengan pendekatan probabilistik, NMF menghasilkan representasi topik yang lebih mudah diinterpretasikan karena seluruh bobot bernilai positif sehingga menghasilkan kumpulan kata yang lebih koheren. Dalam penelitian topic modelling, kualitas topik yang dihasilkan perlu dievaluasi menggunakan metrik kuantitatif, salah satunya adalah Coherence Score [8]. Metrik ini mengukur tingkat keterkaitan semantik antar kata penyusun suatu topik sehingga dapat digunakan untuk menentukan jumlah topik yang optimal sekaligus mengevaluasi kualitas interpretasi topik yang dihasilkan oleh model. Semakin tinggi nilai Coherence Score menunjukkan bahwa kata-kata dalam suatu topik memiliki hubungan semantik yang semakin kuat sehingga topik lebih mudah dipahami [9].

Penelitian ini bertujuan untuk mengidentifikasi topik-topik utama yang muncul pada ulasan pengguna aplikasi Roblox di Google Play Store menggunakan algoritma Non-Negative Matrix Factorization (NMF). Penelitian ini diharapkan mampu menghasilkan informasi mengenai pengalaman pengguna yang mencakup aspek kepuasan, keluhan, maupun kebutuhan pengguna terhadap aplikasi Roblox secara lebih komprehensif. Kebaruan (novelty) penelitian ini terletak pada penerapan algoritma NMF untuk analisis topic modelling pada ulasan pengguna game Roblox, penentuan jumlah topik terbaik berdasarkan evaluasi Coherence Score, serta penyajian interpretasi topik sebagai dasar rekomendasi bagi pengembang dalam meningkatkan kualitas layanan dan pengalaman pengguna. Hasil penelitian ini diharapkan dapat memperkaya kajian text mining pada domain permainan digital sekaligus memberikan kontribusi praktis bagi pengembang dalam memanfaatkan ulasan pengguna sebagai sumber informasi strategis untuk pengembangan aplikasi berbasis data.

2. METODOLOGI

Penelitian ini menggunakan pendekatan text mining dengan metode topic modelling menggunakan algoritma Non-Negative Matrix Factorization (NMF) untuk mengidentifikasi topik utama pada ulasan pengguna aplikasi Roblox di Google Play Store. Tahapan penelitian terdiri atas lima proses utama, yaitu pengumpulan data (scraping), preprocessing data, evaluasi variasi jumlah topik menggunakan Coherence Score, pemodelan topic modelling menggunakan NMF, serta visualisasi hasil topic modelling. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Scraping Data

Tahap pertama penelitian adalah pengumpulan data ulasan pengguna aplikasi Roblox yang berasal dari Google Play Store. Proses pengambilan data dilakukan menggunakan bahasa pemrograman Python dengan pustaka Google Play Scraper [10]. Data diambil berdasarkan ulasan terbaru (newest reviews) yang ditulis dalam Bahasa Indonesia sebanyak 10.000 ulasan. Pembatasan bahasa dilakukan agar topik yang dihasilkan

merepresentasikan pengalaman pengguna Indonesia secara lebih spesifik. Data hasil scraping terdiri atas beberapa atribut yang digunakan dalam penelitian, yaitu: Username, sebagai identitas pengguna, Rating, berupa penilaian pengguna dalam skala 1–5 bintang, Tanggal, yaitu waktu pengguna memberikan ulasan dan Ulasan (Review), yaitu komentar atau opini pengguna terhadap aplikasi Roblox. Pada penelitian ini, atribut ulasan menjadi data utama yang digunakan dalam proses topic modelling, sedangkan atribut lainnya dimanfaatkan sebagai informasi pendukung dalam proses analisis.

2.2 Preprocessing Data

Data hasil scraping masih mengandung berbagai karakter yang tidak relevan sehingga perlu dilakukan proses preprocessing untuk meningkatkan kualitas data sebelum dilakukan pemodelan topik [11]. Tahapan preprocessing yang diterapkan pada penelitian ini meliputi sebagai berikut.

- Case Folding : Case folding bertujuan mengubah seluruh huruf menjadi huruf kecil (lowercase) sehingga tidak terjadi perbedaan representasi antara kata yang memiliki penulisan huruf kapital dan huruf kecil.
- Cleaning : Tahap cleaning dilakukan untuk menghapus karakter yang tidak diperlukan dalam proses analisis teks. Pembersihan data meliputi: penghapusan URL, penghapusan mention dan hashtag, penghapusan tanda baca, penghapusan simbol, penghapusan angka, penghapusan emoji, penghapusan karakter berulang, serta menyisakan karakter alfabet (a–z) sebagai representasi kata.
- Normalisasi Slang : Bahasa yang digunakan pada media sosial maupun Google Play Store umumnya mengandung kata tidak baku atau slang word, seperti "gk", "ga", "bgt", "udh", dan sebagainya. Oleh karena itu dilakukan proses normalisasi menggunakan kamus slang Bahasa Indonesia sehingga kata tidak baku diubah menjadi bentuk baku sesuai Kamus Besar Bahasa Indonesia (KBBI).
- Stopword Removal : Penghapusan kata-kata umum yang tidak memiliki makna penting dalam pembentukan topik menggunakan stopwords Bahasa Indonesia yang dikombinasikan dengan custom stopwords. Custom stopwords disusun berdasarkan kata-kata yang sering muncul pada ulasan Roblox tetapi tidak memberikan informasi tematik, misalnya kata "game", "roblox", "aplikasi", maupun kata umum lainnya yang terlalu dominan.
- Penghapusan Data Kosong : Setelah seluruh proses preprocessing selesai dilakukan, beberapa ulasan dapat berubah menjadi kosong karena seluruh katanya terhapus pada tahap cleaning maupun stopwords removal. Dokumen seperti ini tidak memiliki informasi yang dapat dianalisis sehingga dihapus dari dataset sebelum memasuki tahap topic modelling.

2.3 Evaluasi Variasi Jumlah Topik Menggunakan Coherence Score

Penentuan jumlah topik (number of topics) merupakan salah satu parameter penting pada topic modelling karena akan memengaruhi kualitas topik yang dihasilkan. Oleh karena itu penelitian ini melakukan evaluasi terhadap beberapa variasi nilai K menggunakan metrik Coherence Score. Tahapan evaluasi dilakukan sebagai berikut.

- Representasi Teks Menggunakan TF-IDF : Dokumen hasil preprocessing direpresentasikan ke dalam bentuk matriks Term Frequency–Inverse Document Frequency (TF-IDF). Representasi TF-IDF digunakan karena mampu memberikan bobot yang lebih tinggi pada kata-kata yang bersifat informatif serta mengurangi pengaruh kata yang terlalu sering muncul pada seluruh dokumen [12].
- Tokenisasi dan Dictionary Gensim : Selain membangun matriks TF-IDF, seluruh dokumen juga ditokenisasi untuk membentuk dictionary dan corpus menggunakan pustaka Gensim. Dictionary ini digunakan pada proses evaluasi Coherence Score karena menghitung tingkat keterkaitan semantik antar kata penyusun topik [13].
- Evaluasi Variasi Nilai K : Penelitian melakukan pengujian terhadap beberapa variasi jumlah topik (K). Pada setiap nilai K dilakukan proses pelatihan model NMF, kemudian dihitung nilai Coherence Score menggunakan metode c_v coherence. Nilai K yang menghasilkan Coherence Score tertinggi dipilih sebagai jumlah topik optimal dan digunakan pada proses pemodelan akhir [14].

2.4 Topic Modelling Menggunakan Non-Negative Matrix Factorization (NMF)

Setelah diperoleh jumlah topik terbaik, tahap berikutnya adalah membangun model Non-Negative Matrix Factorization (NMF) menggunakan matriks TF-IDF. NMF bekerja dengan melakukan faktorisasi matriks dokumen-term V menjadi dua matriks non-negatif, yaitu matriks dokumen-topik (W) dan matriks topik-kata (H), sebagaimana ditunjukkan pada Persamaan (1).

$$V = WH \quad (1)$$

dengan:

V = matriks TF-IDF berukuran $m \times n$,

W = matriks distribusi topik pada dokumen,

H = matriks distribusi kata pada setiap topik.

Hasil dekomposisi menghasilkan sejumlah topik yang direpresentasikan oleh kata-kata dengan bobot tertinggi. Selanjutnya dilakukan proses interpretasi untuk memberikan label pada setiap topik berdasarkan kata-kata dominan yang muncul.

2.5 Visualisasi Hasil Topic Modelling

Tahap terakhir adalah visualisasi hasil topic modelling untuk mempermudah interpretasi terhadap topik yang dihasilkan. Visualisasi yang digunakan terdiri atas dua bentuk. Visualisasi distribusi dominansi topik bertujuan menunjukkan jumlah dokumen yang termasuk ke dalam masing-masing topik hasil pemodelan NMF [16]. Setiap ulasan diklasifikasikan berdasarkan topik dengan probabilitas atau bobot tertinggi sehingga dapat diketahui topik yang paling banyak dibahas oleh pengguna Roblox. Selain distribusi topik, penelitian juga menghasilkan Word Cloud pada setiap topik. Word Cloud menampilkan kata-kata dominan berdasarkan bobot tertinggi yang diperoleh dari matriks topik-kata hasil NMF. Semakin besar ukuran kata pada Word Cloud menunjukkan semakin tinggi kontribusi kata tersebut dalam merepresentasikan topik yang bersangkutan [17]. Visualisasi ini membantu proses interpretasi dan pemberian label terhadap topik yang dihasilkan.

3. HASIL DAN DISKUSI

Tahap awal penelitian dilakukan dengan mengumpulkan data ulasan pengguna aplikasi Roblox dari Google Play Store menggunakan teknik web scraping. Data yang diperoleh selanjutnya digunakan sebagai sumber utama dalam proses topic modelling untuk mengidentifikasi topik-topik yang paling banyak dibahas oleh pengguna. Seluruh tahapan pengolahan dilakukan secara bertahap mulai dari pengumpulan data, preprocessing, evaluasi jumlah topik optimal, hingga pemodelan menggunakan algoritma Non-Negative Matrix Factorization (NMF). Pada bagian ini disajikan hasil setiap tahapan penelitian beserta pembahasannya.

Proses scraping dilakukan terhadap ulasan terbaru (newest reviews) yang ditulis dalam Bahasa Indonesia pada aplikasi Roblox di Google Play Store. Sebanyak 10.000 ulasan berhasil dikumpulkan dengan atribut yang meliputi username, score (rating), dan content (isi ulasan). Atribut username digunakan untuk mengidentifikasi pengguna, score menunjukkan penilaian pengguna terhadap aplikasi dalam rentang 1 hingga 5 bintang, sedangkan content berisi opini atau pengalaman pengguna selama menggunakan aplikasi Roblox. Contoh hasil proses scraping ditampilkan pada Tabel 1, yang memperlihatkan beberapa ulasan asli pengguna sebelum dilakukan proses preprocessing. Data mentah tersebut masih mengandung berbagai karakter yang tidak baku, singkatan, kata slang, serta kata-kata yang tidak memiliki makna penting untuk proses pemodelan topik sehingga memerlukan tahap pembersihan data.

Tabel 1. Dataset Ulasan Pengguna Roblox

Username	Score	Content
Pengguna google	1	game dengan sistem kacau pls lah kalo mau game
Pengguna google	5	aku suka
Pengguna google	5	bagus end enak
Pengguna google	5	game nya seru seru bangetttt
Pengguna google	1	map ny bug semua GK jelas

Tahap preprocessing dilakukan melalui beberapa proses, yaitu cleaning, slang normalization, dan stopword removal. Pada tahap cleaning, sistem menghapus URL, simbol, tanda baca, angka, emoji, serta karakter selain huruf alfabet sehingga hanya menyisakan teks yang relevan. Selanjutnya dilakukan slang normalization menggunakan kamus slang Bahasa Indonesia untuk mengubah kata-kata tidak baku menjadi bentuk baku, misalnya gk menjadi tidak dan pls menjadi tolong. Setelah itu diterapkan stopword removal menggunakan kombinasi stopword Bahasa Indonesia dan custom stopword untuk menghilangkan kata-kata umum yang tidak berkontribusi terhadap pembentukan topik. Hasil dari proses tersebut menunjukkan bahwa sebagian ulasan menjadi kosong karena seluruh katanya terhapus selama proses pembersihan. Ulasan kosong dihapus dari dataset

sehingga diperoleh 9.835 ulasan bersih yang siap digunakan pada tahap representasi TF-IDF dan pemodelan Non-Negative Matrix Factorization (NMF). Jumlah data yang tetap tinggi setelah proses preprocessing menunjukkan bahwa sebagian besar ulasan masih mengandung informasi yang relevan untuk mengidentifikasi pengalaman dan permasalahan pengguna aplikasi Roblox.

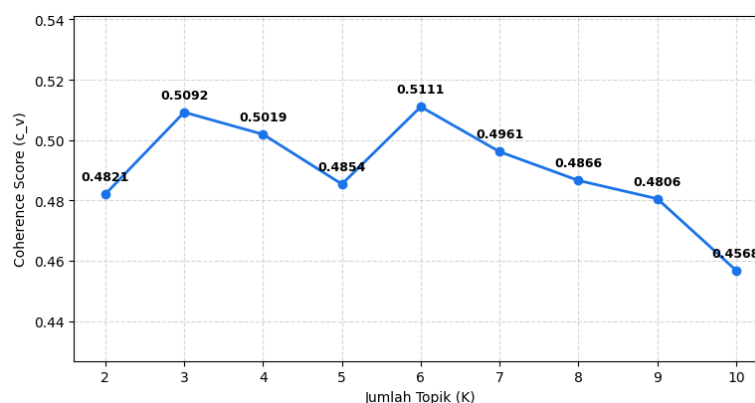
Teks utama Penentuan jumlah topik (number of topics) merupakan salah satu tahapan penting dalam proses topic modelling menggunakan algoritma Non-Negative Matrix Factorization (NMF). Jumlah topik yang terlalu sedikit berpotensi menggabungkan beberapa tema yang berbeda ke dalam satu topik, sedangkan jumlah topik yang terlalu banyak dapat menghasilkan topik yang saling tumpang tindih dan sulit diinterpretasikan. Oleh karena itu, penelitian ini melakukan evaluasi terhadap variasi nilai K mulai dari 2 hingga 10 topik menggunakan metrik Coherence Score untuk memperoleh jumlah topik yang paling optimal. Hasil evaluasi ditunjukkan pada Tabel 2.

Berdasarkan Tabel 2, nilai Coherence Score mengalami fluktuasi pada setiap variasi jumlah topik yang diuji. Nilai coherence berturut-turut untuk K=2 hingga K=10 adalah 0,4821, 0,5092, 0,5019, 0,4854, 0,5111, 0,4961, 0,4866, 0,4806, dan 0,4568. Hasil tersebut menunjukkan bahwa nilai K=6 menghasilkan Coherence Score tertinggi sebesar 0,5111, diikuti oleh K=3 dengan nilai 0,5092, sedangkan nilai terendah diperoleh pada K=10 sebesar 0,4568. Semakin tinggi nilai Coherence Score menunjukkan bahwa kata-kata penyusun suatu topik memiliki keterkaitan semantik yang lebih kuat sehingga topik yang dihasilkan lebih mudah dipahami dan diinterpretasikan.

Tabel 2. Performa Nilai K (Jumlah Topik)

K (Jumlah Topik)	Coherence Score
2	0.4821
3	0.5092
4	0.5019
5	0.4854
6	0.5111
7	0.4961
8	0.4866
9	0.4806
10	0.4568

Berdasarkan hasil evaluasi tersebut, penelitian ini menetapkan K=6 sebagai jumlah topik optimal untuk proses pemodelan Non-Negative Matrix Factorization (NMF). Pemilihan nilai tersebut didasarkan pada pencapaian Coherence Score tertinggi dibandingkan variasi jumlah topik lainnya, sehingga diharapkan mampu menghasilkan representasi topik yang lebih koheren dan menggambarkan pengalaman pengguna Roblox secara lebih akurat. Dengan menggunakan enam topik, proses interpretasi menjadi lebih seimbang karena mampu mengidentifikasi tema-tema utama dalam ulasan pengguna tanpa menghasilkan topik yang terlalu umum maupun terlalu spesifik.



Gambar 2. Grafik Penentuan K Optimal

Perubahan Coherence Score terhadap variasi jumlah topik (K) yang diuji pada gambar 2. Grafik menunjukkan bahwa nilai coherence mengalami fluktuasi seiring bertambahnya jumlah topik. Pada K=2, nilai coherence sebesar 0,4821 kemudian meningkat menjadi 0,5092 pada K=3, sedikit menurun pada K=4 dan K=5, lalu kembali meningkat hingga mencapai nilai tertinggi sebesar 0,5111 pada K=6. Setelah itu, nilai coherence

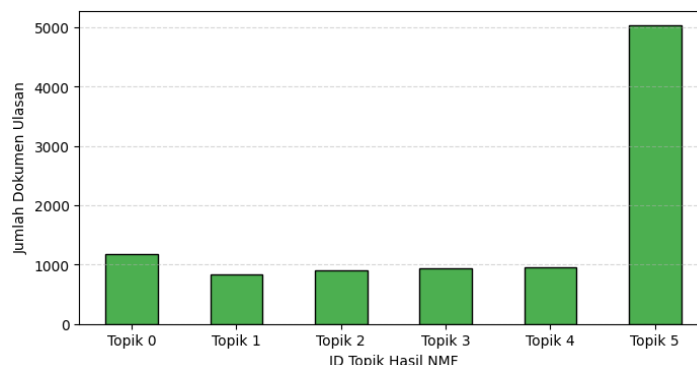
menunjukkan tren penurunan secara bertahap pada $K=7$ hingga $K=10$, dengan nilai terendah sebesar 0,4568 pada $K=10$. Pola tersebut menunjukkan bahwa penambahan jumlah topik tidak selalu meningkatkan kualitas model, karena terlalu banyak topik justru menyebabkan topik menjadi semakin spesifik dan mengurangi keterkaitan semantik antar kata. Oleh karena itu, berdasarkan grafik evaluasi tersebut, $K=6$ dipilih sebagai jumlah topik yang paling optimal karena menghasilkan Coherence Score tertinggi, sehingga mampu memberikan keseimbangan antara jumlah topik dan kualitas interpretasi topik yang dihasilkan oleh algoritma Non-Negative Matrix Factorization (NMF).

Tabel 3. Interpretasi Topik

Topik	Kata Kunci
Topik 0	bagus, sekali, gamenya, gem, lumayan, game, grafik, anak, lag, jaringan
Topik 1	seru, banyak, sekali, gamenya, permainan, gem, juga, bersama, teman, karna
Topik 2	sangat, menyenangkan, sekali, permainan, baik, mainkan, cocok, menarik, gem, menyukai
Topik 3	suka, aku, sama, sekali, kasih, bintang, mau, gem, karna, jadi
Topik 4	banget, keren, sama, gamenya, enak, jelek, pokoknya, sumpah, asik, main
Topik 5	banyak, tidak, chat, ada, gak, lagi, jadi, map, tolong, juga

Setelah diperoleh jumlah topik optimal sebanyak enam topik ($K=6$), proses selanjutnya adalah melakukan pemodelan menggunakan algoritma Non-Negative Matrix Factorization (NMF). Hasil pemodelan menghasilkan enam kelompok topik yang masing-masing direpresentasikan oleh sepuluh kata kunci dengan bobot tertinggi sebagaimana ditunjukkan pada Tabel 3. Kata-kata dominan pada setiap topik menggambarkan tema utama yang sering dibahas oleh pengguna Roblox dalam ulasan mereka. Sebagai contoh, Topik 0 didominasi kata bagus, gamenya, grafik, lag, dan jaringan yang mengindikasikan pembahasan mengenai kualitas permainan dan performa jaringan. Topik 1 memuat kata seru, permainan, bersama, dan teman yang merepresentasikan pengalaman bermain dan aspek sosial dalam permainan. Sementara itu, Topik 2 didominasi kata menyenangkan, baik, menarik, dan menyukai yang menunjukkan tingkat kepuasan pengguna terhadap pengalaman bermain Roblox.

Beberapa topik juga mencerminkan adanya keluhan maupun masukan dari pengguna. Topik 3 memuat kata-kata seperti kasih, bintang, dan mau yang mengindikasikan ekspresi penilaian atau harapan pengguna terhadap aplikasi. Topik 4 didominasi oleh kata keren, enak, jelek, dan sumpah yang menunjukkan adanya kombinasi opini positif maupun negatif mengenai kualitas permainan. Adapun Topik 5 berisi kata chat, gak, map, dan tolong yang mengarah pada pembahasan mengenai fitur komunikasi, permasalahan pada peta (map), serta permintaan bantuan akibat kendala yang dialami pengguna. Secara keseluruhan, hasil interpretasi topik menunjukkan bahwa ulasan pengguna Roblox tidak hanya berfokus pada aspek kesenangan bermain, tetapi juga banyak membahas performa aplikasi, kualitas fitur, stabilitas permainan, serta berbagai kendala teknis yang dapat menjadi masukan bagi pengembang dalam meningkatkan pengalaman pengguna.



Gambar 3. Grafik Distribusi Topik

Pada Gambar 3 menunjukkan distribusi dominansi topik hasil pemodelan Non-Negative Matrix Factorization (NMF) terhadap 9.835 ulasan pengguna Roblox. Berdasarkan grafik tersebut, Topik 5 merupakan topik yang paling dominan dengan jumlah lebih dari 5.000 dokumen ulasan, jauh melampaui lima topik lainnya

yang masing-masing hanya berada pada kisaran 800–1.200 dokumen. Dominannya Topik 5 menunjukkan bahwa sebagian besar pengguna membahas isu-isu yang berkaitan dengan kata kunci seperti chat, map, tidak, gak, dan tolong, yang mengindikasikan adanya pembahasan mengenai kendala fitur komunikasi, permasalahan pada peta permainan, serta berbagai keluhan teknis yang dialami pengguna. Topik 0 menjadi topik terbesar kedua dengan pembahasan mengenai kualitas permainan, grafik, dan jaringan, sedangkan Topik 1 hingga Topik 4 memiliki distribusi yang relatif seimbang sehingga mencerminkan keberagaman pengalaman pengguna, mulai dari keseruan bermain, tingkat kepuasan, penilaian terhadap kualitas permainan, hingga berbagai masukan untuk pengembangan aplikasi. Hasil distribusi ini menunjukkan bahwa meskipun pengguna memberikan banyak apresiasi terhadap Roblox, isu-isu teknis masih menjadi tema yang paling sering muncul dalam ulasan sehingga dapat dijadikan prioritas oleh pengembang untuk meningkatkan kualitas layanan dan pengalaman bermain pengguna.



Gambar 4. Wordcloud Pertopik

Visualisasi word cloud untuk masing-masing topik hasil pemodelan Non-Negative Matrix Factorization (NMF) pada gambar 4. Word cloud digunakan untuk mempermudah interpretasi topik dengan menampilkan kata-kata yang memiliki bobot tertinggi pada setiap topik, di mana ukuran kata berbanding lurus dengan tingkat kontribusinya dalam merepresentasikan topik tersebut. Berdasarkan visualisasi, setiap topik memiliki karakteristik kata kunci yang berbeda, seperti Topik 0 yang didominasi kata *bagus* dan *gamenya*, Topik 1 dengan kata *seru* dan *banyak*, Topik 2 dengan kata *sangat* dan *menyenangkan*, Topik 3 dengan kata *suka* dan *sama*, Topik 4 dengan kata *banget* dan *keren*, serta Topik 5 dengan kata *banyak*, *tidak*, *chat*, dan *lagi*. Visualisasi ini memperkuat hasil interpretasi pada Tabel 3 bahwa setiap topik merepresentasikan tema pembahasan yang berbeda, mulai dari apresiasi terhadap kualitas permainan, pengalaman bermain yang menyenangkan, hingga berbagai keluhan terkait fitur dan kendala teknis pada aplikasi Roblox. Dengan demikian, word cloud tidak hanya memudahkan identifikasi kata-kata dominan pada setiap topik, tetapi juga memberikan gambaran visual mengenai fokus utama pembahasan pengguna dalam ulasan Google Play Store.

4. KESIMPULAN

Algoritma Non-Negative Matrix Factorization (NMF) berhasil diterapkan untuk mengidentifikasi topik utama pada 9.835 ulasan pengguna aplikasi Roblox yang diperoleh dari 10.000 ulasan Google Play Store setelah melalui proses preprocessing. Evaluasi menggunakan Coherence Score menunjukkan bahwa jumlah topik yang paling optimal adalah enam topik ($K=6$) dengan nilai 0,5111, yang menghasilkan topik-topik dengan tingkat keterkaitan semantik yang paling baik. Hasil pemodelan menunjukkan bahwa pengalaman pengguna Roblox didominasi oleh pembahasan mengenai kualitas permainan, keseruan bermain, kepuasan pengguna, serta berbagai kendala teknis seperti fitur chat, map, dan performa aplikasi. Analisis distribusi topik memperlihatkan bahwa topik yang berkaitan dengan keluhan dan permasalahan teknis merupakan topik yang paling banyak dibahas oleh pengguna, sehingga menjadi aspek yang perlu mendapatkan perhatian lebih dari pengembang. Selain itu, visualisasi word cloud berhasil memperjelas karakteristik setiap topik melalui kata-kata dominan yang merepresentasikan tema pembahasan. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan topic modelling menggunakan algoritma NMF yang dikombinasikan dengan evaluasi Coherence Score mampu menghasilkan topik yang koheren dan mudah diinterpretasikan, sehingga dapat dimanfaatkan sebagai dasar dalam

memahami pengalaman pengguna serta mendukung pengambilan keputusan untuk pengembangan dan peningkatan kualitas aplikasi Roblox.

Referensi

- [1] D. Kurniawan, Roblox: Dunia Permainan Ciptaan Anda. Afdan Rojabi Publisher, 2025.
- [2] A. H. Sutopo, Pengembangan Bahan Ajar Berbasis Metaverse. Topazart, 2022.
- [3] D. D. Alfaridzi, T. H. Pudjiantoro, and I. Santikarama, "Implementation of Aspect-Based Sentiment Analysis for Outdoor Gear Product Reviews on the Tokopedia E-Commerce Platform (Case Study: Eiger Adventure)," *Holistic Science*, vol. 5, no. 2, pp. 173–184, 2025.
- [4] N. Cahyono, *Fundamental Topic Modelling: Memahami Tren & Topik dalam Text Mining*. Mega Press Nusantara, 2024.
- [5] U. N. Khadijah and N. Cahyono, "Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan Latent Dirichlet Allocation (LDA)," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, 2024.
- [6] N. Cahyono, N. Nurcahyo, A. Fauzan, and R. Agung, "Analisa Perbandingan Latent Semantic Indexing (LSI) dan Latent Dirichlet Allocation (LDA) untuk Topic Modelling Aplikasi Identitas Kependudukan Digital (IKD)," *Building of Informatics, Technology and Science*, vol. 6, no. 3, pp. 1638–1647, 2024.
- [7] S. Saikin, S. Fadli, J. Akbar, and H. Fahmi, "Topic Modeling Judul Penelitian Menggunakan Metode Latent Dirichlet Allocation (LDA) dan Faktorisasi Matriks Non-Negatif (NMF)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3439–3445, 2025.
- [8] Y. M. Yolanda, D. Swanjaya, and D. P. Pamungkas, "Sistem Pencarian Dongeng Berbasis Topik Menggunakan Latent Dirichlet Allocation (LDA) pada Aplikasi Nusantara Panji Kediri," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 394–407, 2025.
- [9] R. A. Kurniawan, P. Purwanto, and I. Maulana, "Pemodelan Topik Ulasan Pengguna Honkai Star Rail Menggunakan BERTopic Berbasis IndoBERT," *Blend Sains Jurnal Teknik*, vol. 4, no. 4, pp. 872–881, 2026.
- [10] R. Hasanah, S. Sulistiani, N. Nurhikmayani, Z. Hasanah, S. A. Wijaya, D. Abdennasser, and A. N. Sharkawy, "Play Store Data Scrapping and Preprocessing Done as Sentiment Analysis Material," *Indonesian Journal of Modern Science and Technology*, vol. 1, no. 1, pp. 16–21, 2025.
- [11] M. Faza, M. R. Sanjaya, A. Ibrahim, and M. H. Syahbani, "Analisis Ulasan Sentimen pada Aplikasi Mobile BSB Menggunakan Algoritma Convolutional Neural Networks dan Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 6, no. 2, pp. 829–838, 2026.
- [12] A. T. Nasirah, H. T. Surbakti, and R. Kaban, "Klasifikasi Sentimen Tayangan Televisi dengan TF-IDF dan Naive Bayes," *Jurnal Media Akademik (JMA)*, vol. 4, no. 7, 2026.
- [13] F. M. Hidayat, H. Sanjaya, D. Purnomo, and H. Wiranto, "Pemodelan Topik Berita Nasional Indonesia Menggunakan Latent Dirichlet Allocation," *INFOTECH Journal*, vol. 12, no. 1, pp. 33–39, 2026.
- [14] M. Hasanah and W. Yustanti, "Pengembangan Standarisasi Kategori Keluhan Aplikasi MyPelindo Berbasis Latent Dirichlet Allocation dan Klasifikasi," *Jurnal Ilmu Teknologi Informasi Indonesia*, vol. 2, no. 2, pp. 205–212, 2026.
- [15] D. Wang, P. Zhang, P. Deng, Q. Wu, W. Chen, T. Jiang, et al., "An Autoencoder-Like Deep NMF Representation Learning Algorithm for Clustering," *Knowledge-Based Systems*, vol. 305, Art. no. 112597, 2024.
- [16] N. Ramadhani, D. A. E. Aritonang, K. D. Tania, and A. Rifai, "Knowledge Discovery terhadap Informasi Kesehatan di Media X Menggunakan Topic Modeling Berbasis Non-Negative Matrix Factorization (NMF)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 10, no. 3, pp. 4040–4047, 2026.
- [17] G. I. Ratumbusang and I. R. Tangkawarow, "Studi Literatur: Analisis Topik Otomatis Isu Pemilu Indonesia Menggunakan Algoritma LDA dan NMF," *Edutik: Jurnal Pendidikan Teknologi Informasi dan Komunikasi*, vol. 6, no. 1, pp. 382–393, 2026.