

# Penerapan Metode Non-Negative Matrix Factorization dan Support Vector Machine pada Sentimen Pengguna terhadap Update Minecraft 1.21 berbasis Website

Tengku Syafiq Ali Syahputra<sup>1</sup>, Ardian Prima Atmaja<sup>2</sup>, Susilo Veri Yulianto<sup>3</sup>

Program Studi Teknologi Informasi,  
Fakultas Teknik,  
Politeknik Negeri Madiun, Madiun

<sup>1</sup>syafiqali253@gmail.com, <sup>2</sup>atmaja@pnm.ac.id, <sup>3</sup>susilo@pnm.ac.id

## Abstrak

Perkembangan teknologi informasi telah mengubah cara manusia menyampaikan pendapat melalui media sosial dan platform ulasan daring. Dalam konteks permainan digital, komunitas pemain memiliki peran penting dalam membentuk persepsi terhadap kualitas game melalui komentar dan ulasan yang mereka berikan. Sebagai salah satu *game sandbox* terpopuler, setiap pembaruan (*update*) *Minecraft* sering kali memunculkan reaksi beragam dari pemain, namun komentar yang sangat banyak dan tidak terstruktur sering kali tidak dianalisis secara menyeluruh. Kondisi ini menunjukkan pentingnya sistem yang mampu mengolah ulasan pengguna secara otomatis untuk membantu pengembang memahami persepsi pemain. Penelitian ini membahas penerapan metode *Non-Negative Matrix Factorization* (NMF) dan *Support Vector Machine* (SVM) pada analisis sentimen pengguna terhadap *update Minecraft 1.21* berbasis website. Tujuan penelitian ini adalah mengembangkan aplikasi web yang dapat mengekstrak topik, mengklasifikasikan sentimen pengguna, dan menampilkan hasil dalam bentuk visualisasi yang informatif. Hasil penelitian menunjukkan tingkat akurasi sebesar 90,95%, dengan distribusi sentimen netral (36,7%), negatif (32,5%), dan positif (30,8%). Analisis topik menggunakan NMF mengungkapkan tema dominan terkait masalah teknis, pengalaman positif, dan fitur baru. Secara keseluruhan, kombinasi metode NMF dan SVM memberikan pemahaman yang komprehensif terhadap persepsi pengguna terhadap pembaruan *Minecraft*.

**Kata kunci:** *Non-Negative Matrix Factorization* (NMF), *Support Vector Machine* (SVM), Sentimen, *Minecraft*, *Update*, Aplikasi Web.

## 1. Pendahuluan

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam cara manusia berinteraksi, menyampaikan pendapat, dan berbagi pengalaman. Salah satu dampak terbesar dari revolusi digital ini adalah meningkatnya peran media sosial, forum komunitas, dan platform ulasan online sebagai sarana utama dalam menyuarakan opini terhadap suatu produk atau layanan. Fenomena ini juga terlihat dalam dunia permainan digital, di mana komunitas pemain berperan aktif dalam membentuk persepsi terhadap kualitas *game* melalui komentar, diskusi, maupun evaluasi publik.

Sebagai salah satu *game sandbox* terpopuler di dunia, *Minecraft* memiliki lebih dari 170 juta pemain aktif setiap bulannya. *Game* ini secara rutin merilis pembaruan (*update*) besar yang memperkenalkan fitur-fitur baru, peningkatan *gameplay*, maupun perbaikan bug. Setiap *update* tersebut biasanya diikuti oleh gelombang reaksi dari komunitas pemain, yang disampaikan melalui forum resmi, media sosial seperti *Reddit* dan *Twitter*, serta ulasan di platform distribusi aplikasi seperti *Google play store* [1]. Komentar-komentar tersebut mencerminkan penilaian subjektif pemain terhadap kualitas pembaruan, namun dalam banyak kasus tidak dikompilasi atau dianalisis secara menyeluruh oleh pengembang. Akibatnya, banyak masukan yang sebenarnya bernilai tinggi justru terlewatkan, yang kemudian berdampak pada turunnya kepuasan pengguna serta reputasi pengembang di mata komunitas.

Tingginya volume data teks yang bersifat tidak terstruktur dari komentar pengguna menjadi tantangan tersendiri dalam proses pengambilan keputusan. Oleh karena itu, diperlukan suatu sistem yang mampu mengolah data ulasan tersebut secara otomatis dan menyajikan informasi yang berguna bagi tim pengembang. Salah satu pendekatan yang dapat digunakan adalah analisis sentimen, yaitu proses untuk mengidentifikasi dan mengklasifikasikan emosi atau opini seseorang berdasarkan data teks. Teknik ini menjadi bagian penting dari bidang *machine learning* dan *Natural Language Processing* (NLP), dan telah banyak digunakan dalam berbagai studi untuk memahami persepsi publik terhadap suatu isu atau produk [2].

Dalam konteks pengembangan *game*, analisis sentimen memiliki potensi besar untuk mendukung evaluasi fitur baru berdasarkan tanggapan komunitas secara luas. Dibandingkan dengan metode survei manual yang memakan waktu dan sumber daya, analisis sentimen menawarkan pendekatan yang cepat, akurat, dan bersifat *data-driven*. Namun demikian, kompleksitas bahasa alami seperti ambiguitas makna, penggunaan bahasa informal, serta ekspresi sarkastik membuat analisis teks menjadi tantangan tersendiri. Oleh karena itu, diperlukan metode klasifikasi yang tangguh dan mampu menangani data berdimensi tinggi.

Salah satu algoritma yang populer digunakan dalam klasifikasi sentimen adalah *Support Vector Machine* (SVM). SVM bekerja dengan membangun *hyperplane* optimal yang memisahkan data berdasarkan kelas sentimennya, dan telah terbukti efektif dalam berbagai penelitian dengan performa yang kompetitif [3][4]. Studi-studi sebelumnya menunjukkan bahwa SVM mampu menghasilkan akurasi yang tinggi pada data ulasan pengguna aplikasi seperti *TikTok* [5], *Sayurbox* [6], dan *MPStore* [7]. Keunggulan SVM terletak pada kemampuannya menangani data teks yang memiliki dimensi besar dan pola distribusi yang kompleks.

Agar kinerja klasifikasi optimal, proses *preprocessing* menjadi tahap yang sangat penting. Tahap ini mencakup normalisasi teks, penghapusan karakter khusus, *stopword removal* yaitu penghapusan kata-kata umum yang tidak memiliki makna, *tokenization* yaitu proses memecah kalimat menjadi unit kata atau token, serta *lemmatization* yaitu pengubahan setiap kata ke bentuk dasarnya [8][9]. Tujuan *preprocessing* adalah menyederhanakan teks mentah menjadi representasi numerik yang bermakna dan mengurangi noise dalam data.

Setelah proses *preprocessing*, teks perlu diubah menjadi representasi numerik agar dapat diproses oleh algoritma pembelajaran mesin. Salah satu teknik yang paling umum digunakan untuk tujuan ini adalah *Term Frequency–Inverse Document Frequency* (TF-IDF). TF-IDF bekerja dengan memberikan bobot pada setiap kata berdasarkan seberapa sering

kata tersebut muncul dalam satu dokumen dibandingkan dengan kemunculannya di seluruh korpus. Dengan cara ini, TF-IDF dapat menonjolkan kata-kata yang benar-benar penting dalam konteks analisis sentimen dan mengurangi pengaruh kata-kata umum yang tidak relevan. Metode ini telah banyak digunakan dalam berbagai penelitian karena kemampuannya meningkatkan akurasi klasifikasi teks dengan cara menyeimbangkan frekuensi lokal dan global dari setiap kata [10].

Selain klasifikasi, dibutuhkan pula pendekatan untuk mengelompokkan komentar berdasarkan tema atau topik yang sering muncul agar pengembang dapat mengetahui isu-isu dominan yang diangkat oleh komunitas. Untuk kebutuhan tersebut, metode *Non-Negative Matrix Factorization* (NMF) menjadi pilihan yang efektif. NMF adalah teknik dekomposisi matriks yang merepresentasikan data dalam bentuk non-negatif sehingga hasilnya lebih mudah diinterpretasikan secara manusiawi. Dalam konteks teks, NMF digunakan untuk mengekstraksi topik atau fitur utama dari kumpulan dokumen, memungkinkan sistem untuk memahami konteks pembicaraan secara lebih luas [11][12]. Pendekatan ini dapat mempermudah pengembang dalam meninjau fokus utama keluhan atau pujian pemain berdasarkan kelompok tema yang terbentuk.

Dalam literatur NLP modern, integrasi antara NMF dan SVM sering digunakan secara bersamaan dalam *pipeline* analisis teks. NMF berfungsi untuk mereduksi dimensi dan mengekstrak fitur penting, sementara SVM melakukan klasifikasi berdasarkan fitur tersebut [13]. Beberapa penelitian juga menunjukkan bahwa kombinasi ini memberikan hasil yang lebih baik dalam menangani data ulasan yang sangat bervariasi baik dalam gaya bahasa, struktur kalimat, maupun intensitas sentimen [14][15]. Selain itu, proses pengumpulan data dari *Google play store* dapat dilakukan secara otomatis menggunakan teknik *web scraping*, yang telah terbukti efisien dan banyak diterapkan dalam penelitian berbasis opini pengguna [16].

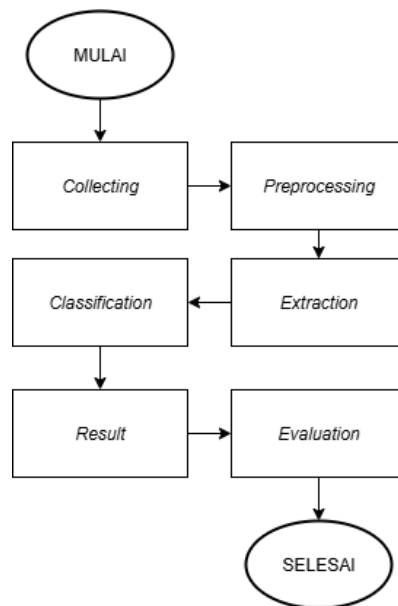
Dalam penelitian ini digunakan metode penelitian eksperimental, yaitu pendekatan penelitian yang dilakukan dengan cara melakukan percobaan secara langsung untuk menguji hipotesis yang telah dirumuskan [17]. Metode ini memungkinkan peneliti untuk mengamati hubungan sebab-akibat antara variabel yang diuji melalui serangkaian tahap yang terukur, sehingga hasilnya dapat dijadikan dasar dalam pengambilan keputusan berbasis data. Untuk mendukung pengembangan sistem analisis sentimen ini, digunakan pendekatan *Agile Development*, yaitu metodologi pengembangan perangkat lunak yang berfokus pada iterasi cepat, fleksibilitas terhadap perubahan kebutuhan, serta kolaborasi aktif antara pengembang dan pengguna [18]. *Agile* memungkinkan proses pengembangan dilakukan secara bertahap sehingga hasil setiap iterasi dapat dievaluasi dan disesuaikan sebelum melanjutkan tahap berikutnya. Pendekatan ini dinilai efektif dalam pengembangan sistem berbasis web karena mampu meminimalkan risiko kesalahan besar dan memastikan bahwa produk akhir sesuai dengan kebutuhan fungsional pengguna.

Dengan demikian, pengembangan sistem berbasis website yang mengintegrasikan metode NMF dan SVM menjadi pendekatan yang menjanjikan dalam pengolahan komentar pengguna secara lebih sistematis. Sistem ini dirancang untuk menganalisis ulasan pengguna dalam skala besar, mengekstraksi topik utama, dan mengklasifikasikan sentimen pengguna terhadap pembaruan tertentu. Pendekatan ini tidak hanya dapat digunakan untuk evaluasi pasca-rilis (*post-release review*), tetapi juga menjadi alat pendukung dalam perencanaan pembaruan selanjutnya. Informasi yang diperoleh dari sistem ini diharapkan dapat memperkuat komunikasi antara pengembang dan komunitas,

mendorong pengambilan keputusan berbasis data, serta menciptakan ekosistem pengembangan *game* yang lebih responsif dan berkelanjutan.

## 2. Metode

Metode penelitian yang digunakan adalah metode penelitian eksperimental. Metode ini dipilih karena mengaju dari tujuan penulis yaitu untuk menghasilkan aplikasi website yang menerapkan metode *Non-Negative Matrix Factorization* dan *Support Vector Machine*. Berikut adalah tahapan dari penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan dari penelitian. Tahapan pada metode penelitian eksperimental dalam pengembangan sistem ini dimulai dengan tahap *collecting*, yaitu pengumpulan data komentar pengguna mengenai *update Minecraft 1.21* dari platform *Google play store* menggunakan teknik *web scraping*. Selanjutnya dilakukan *preprocessing* untuk membersihkan dan menyiapkan data teks, melalui proses seperti *cleaning*, *tokenization*, *case folding*, *stopword removal*, dan *lemmatization*. Tahapan ini bertujuan untuk mengubah data mentah menjadi teks yang siap dianalisis. Proses *cleaning* menghapus karakter non-alfabet seperti angka, emoji, dan simbol khusus. *Case folding* mengubah seluruh huruf menjadi huruf kecil agar konsisten. *Tokenization* memecah kalimat menjadi unit kata atau token, sedangkan *stopword removal* menghapus kata umum yang tidak memiliki makna analitis, dan *lemmatization* mengembalikan kata ke bentuk dasarnya. Setelah data siap, dilakukan tahap *extraction* menggunakan metode *Non-Negative Matrix Factorization* (NMF) untuk mengidentifikasi topik utama dari kumpulan komentar. Data hasil ekstraksi kemudian masuk ke tahap *classification*, dimana metode *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan sentimen menjadi positif, netral, atau negatif. Hasil dari proses ini dianalisis pada tahap *result*, yang menampilkan distribusi sentimen dan topik yang muncul. Terakhir, dilakukan evaluasi terhadap performa sistem, termasuk pengukuran akurasi dan analisis *confusion matrix*, untuk menilai efektivitas metode yang diterapkan sebelum keseluruhan proses dinyatakan selesai.

Pada tahap implementasi metode, data ulasan yang telah melalui *preprocessing* direpresentasikan dalam bentuk matriks TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk mengekstraksi bobot kata yang signifikan. Matriks ini kemudian diolah menggunakan *Non-Negative Matrix Factorization* (NMF) untuk mereduksi dimensi dan mengidentifikasi lima topik utama berdasarkan kumpulan kata dominan. Representasi hasil NMF selanjutnya digunakan sebagai input pada algoritma *Support Vector Machine* (SVM) untuk proses klasifikasi sentimen. *Training* model SVM dilakukan pada tahap *classification* dengan beberapa langkah sistematis. Pertama, dataset dibagi menjadi 70% data latih dan 30% data uji menggunakan fungsi *train\_test\_split()* dari *Scikit-learn* dengan parameter *random\_state=42* untuk memastikan reproduibilitas hasil. Kemudian, model SVM diinisialisasi menggunakan *SVC(kernel='linear', C=1.0, random\_state=42)* dari *library Scikit-learn*, dimana kernel linear dipilih karena efektif untuk data teks yang telah ditransformasi ke dalam ruang fitur berdimensi tinggi melalui TF-IDF. Selanjutnya, model dilatih menggunakan fungsi *fit()* dengan input berupa matriks hasil NMF dari data latih (*X\_train*) dan label sentimen yang telah ditentukan (*y\_train*). Proses *training* ini membangun *hyperplane* optimal yang memaksimalkan margin pemisahan antar kelas sentimen (positif, netral, negatif) dalam ruang fitur multidimensi. Setelah model selesai dilatih, dilakukan prediksi pada data uji menggunakan fungsi *predict()* untuk memperoleh label sentimen hasil klasifikasi.

Penentuan kriteria sentimen dilakukan menggunakan analisis polaritas dari *library TextBlob*. Komentar dengan nilai polaritas lebih dari 0,05 dikategorikan sebagai positif, nilai antara -0,05 dan 0,05 dikategorikan sebagai netral, sedangkan nilai kurang dari -0,05 dikategorikan sebagai negatif. Pendekatan ini digunakan karena *TextBlob* menyediakan perhitungan polaritas berbasis leksikon yang efisien dan mudah diintegrasikan dalam *pipeline* NLP. Untuk memastikan hasil pengujian benar dan tidak bias, dilakukan evaluasi performa menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score* yang dihitung dari *confusion matrix*.

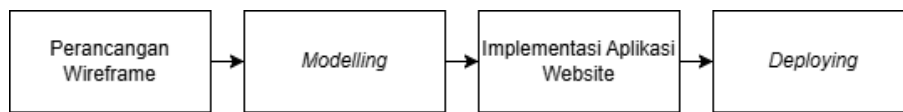
Seluruh proses implementasi dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan beberapa *library open-source*. Proses *tokenization* dan *stopword removal* dilakukan menggunakan *library NLTK*, sedangkan penentuan polaritas sentimen dilakukan dengan *TextBlob*. Untuk proses representasi teks, ekstraksi topik, dan klasifikasi digunakan *library Scikit-learn* melalui komponen *TF-IDF Vectorizer*, *Non-Negative Matrix Factorization* (NMF), serta *Support Vector Machine* (SVM). Selain itu, *Pandas* dan *NumPy* dimanfaatkan dalam pengolahan serta manipulasi data agar proses analisis menjadi lebih efisien dan terstruktur.

Pemilihan *Scikit-learn* dilakukan karena *library* ini memiliki dokumentasi yang lengkap, integrasi *pipeline* yang mudah, serta stabilitas tinggi dibandingkan *library* lain seperti *TensorFlow* atau *PyTorch* untuk kasus teks klasik. Selain itu, *Scikit-learn* telah terbukti efisien dalam menangani dataset berukuran menengah dengan performa yang kompetitif pada algoritma *machine learning* tradisional seperti SVM dan NMF.

Selanjutnya, metode pengembangan yang diimplementasikan pada proyek sistem analisis sentimen ini menggunakan pendekatan *Agile*, dipilih karena fleksibilitas, iterasi cepat, dan kemampuannya untuk merespons perubahan kebutuhan selama proses pengembangan. Berikut tahapan-tahapan dari metode pengembangan aplikasi.

Gambar 2 menunjukkan tahap dari pengembangan aplikasi. Tahapan pengembangan aplikasi ini terdiri dari empat tahap utama yang saling terintegrasi. Pertama, perancangan wireframe dilakukan untuk menentukan struktur tampilan antarmuka pengguna serta

alur navigasi dari aplikasi berbasis website yang akan dibangun. Tahap ini bertujuan memberikan gambaran visual awal mengenai desain halaman-halaman penting, seperti halaman *homepage*, *scraping*, *preprocessing*, dan visualisasi. Kedua, tahap *modelling* mencakup pengumpulan data, *preprocessing*, ekstraksi topik menggunakan metode Non-Negative Matrix Factorization (NMF), serta klasifikasi sentimen menggunakan Support Vector Machine (SVM). Model yang dihasilkan pada tahap ini menjadi inti dari analisis sentimen yang ditampilkan pada website. Ketiga, tahap implementasi aplikasi website melibatkan pengembangan *frontend* untuk menampilkan hasil analisis secara interaktif, termasuk halaman visualisasi topik dan distribusi sentimen. Terakhir, tahap *deploying* dilakukan untuk menerbitkan aplikasi ke server berbasis *cloud*, dalam hal ini menggunakan layanan *AWS EC2*, sehingga aplikasi dapat diakses secara publik dan digunakan oleh pengguna sesuai kebutuhan.



Gambar 2. Tahapan Pengembangan Aplikasi

### 3. Hasil

#### 3.1. Hasil Modelling

Hasil pemodelan diperoleh dari serangkaian tahapan, mulai dari pengumpulan data, *preprocessing*, hingga penerapan model ekstraksi topik dan klasifikasi sentimen. Berikut hasil dari tahapan-tahapan tersebut.

##### 3.1.1. Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses *web scraping* dari platform *Google Play Store* pada halaman ulasan aplikasi *Minecraft*. Proses *scraping* dilakukan untuk mengumpulkan komentar pengguna yang berkaitan dengan pengalaman dan pendapat mereka terhadap update aplikasi. Hasil dari pengumpulan data dapat dilihat pada Gambar 3 berikut ini.

	Author	Comment	Date
0	A Google user	Good Multiplayer but Windows 10 Edition runs f...	2024-12-02 23:41:49
1	Zelda	There's something absolutely amazing and beaut...	2024-12-02 23:10:04
2	Lois Billings	I love this game, and all the mods I have down...	2024-12-02 22:51:41
3	Stuart Willie	Sorry for 2 stars cause i cant play with my fr...	2024-12-02 21:42:26
4	Phoenix_Rising 86	Fun time waster just to what you can build wit...	2024-12-02 20:24:41
...	...	...	...
10211	Shubham Rakshit	Normal boy IQ = 100 Minecraft player IQ = 1000 🤡	2024-09-18 15:10:53
10212	Iman Nav	ldk, its just attractive	2024-09-18 14:59:45
10213	Sapna Kumari	Bhai mere me server nahi chel reha hai	2024-09-18 14:59:45
10214	Debjit Adhikari	The game is great. It would be better if you m...	2024-09-18 14:53:36
10215	Tecnoh Mafia	Best game 🏆🏆🏆🏆	2024-09-18 14:51:49

10216 rows × 3 columns

Gambar 3. Hasil Pengumpulan Data

Hasil dari pengumpulan data ini menghasilkan sebanyak 10.216 baris data yang terdiri dari tiga kolom utama, yaitu *Author* (nama pengguna), *Comment* (isi ulasan), dan *Date*

(waktu ulasan diberikan). Data tersebut kemudian disimpan dalam format *csv* untuk selanjutnya dilakukan proses *preprocessing* dan *modelling*.

### 3.1.2. Hasil Preprocessing

*Preprocessing* dilakukan agar data yang akan diproses lebih siap diolah agar hasil lebih akurat dan efisien. Hasil dari *preprocessing* dapat dilihat pada Gambar 4 berikut ini.

	processed_text	label
0	gud gud gud gud first game	negative
1	f game cant connect server	negative
2	game greatbut problem installed game month sho...	negative
3	long download	negative
4	video game visit comeback play year first vide...	negative
...	...	...
5848	good	positive
5849	thats good still wish add thing could choose l...	positive
5850	giving fps phone awesome	positive
5851	year think best game game love im undating yea...	positive
5852	best game world	positive

5853 rows × 2 columns

Gambar 4. Hasil Preprocessing

Gambar 4 menunjukkan hasil akhir dari proses *preprocessing* data komentar pengguna dari *Google Play Store*. Proses ini dimulai dengan pembersihan teks (*cleaning*) yang menghapus emoji, simbol, dan karakter non-alfabet. Setelah itu, dilakukan tokenisasi untuk memecah teks menjadi kata-kata, *case folding* untuk mengubah semua huruf menjadi huruf kecil, *stopword removal* untuk menghapus kata-kata umum yang tidak bermakna secara analitis, serta *lemmatization* untuk mengembalikan kata ke bentuk dasarnya. Selanjutnya, dilakukan penentuan label sentimen menggunakan analisis polaritas dari *library TextBlob*. Komentar dengan nilai polaritas  $> 0,05$  dikategorikan sebagai positif, nilai antara  $-0,05$  dan  $0,05$  sebagai netral, dan nilai  $< -0,05$  sebagai negatif. Pendekatan ini digunakan karena *TextBlob* memberikan hasil cepat dan stabil dalam pengukuran polaritas berbasis leksikon. Untuk menghindari ketidakseimbangan kelas, dilakukan *undersampling* menggunakan *RandomUnderSampler* agar distribusi data antar kelas seimbang. Hasil akhir ditampilkan dalam dua kolom utama, yaitu *processed\_text* berisi teks yang telah diproses dan label berisi kategori sentimen, dengan total 5.853 data yang siap digunakan pada tahap pemodelan.

Selain itu, dalam proses pengumpulan data dilakukan penyaringan bahasa awal dengan menetapkan parameter *lang="en"* pada fungsi *reviews()* dari *library google\_play\_scraper*. Langkah ini bertujuan agar komentar yang dikumpulkan didominasi oleh ulasan berbahasa Inggris. Namun demikian, hasil pengumpulan menunjukkan bahwa masih terdapat sejumlah kecil komentar dalam bahasa lain yang lolos karena keterbatasan filter bawaan *library* tersebut. Oleh karena itu, dilakukan penyaringan tambahan sederhana pada tahap *preprocessing*, seperti menghapus manual komentar yang berisi karakter non-Latin, kata dari bahasa lain, atau teks yang tidak dapat dikenali secara linguistik. Pendekatan ini memastikan sebagian besar data yang digunakan tetap konsisten berbahasa Inggris, sehingga analisis sentimen menggunakan *TextBlob*, TF-IDF, dan SVM dapat berjalan secara valid dan stabil.

### 3.1.3. Hasil Penerapan Model

Proses penerapan model dalam penelitian ini terdiri dari dua tahapan utama, yaitu ekstraksi topik menggunakan metode Non-Negative Matrix Factorization (NMF) dan klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM).

Tahap pertama, NMF diterapkan pada hasil representasi teks berbasis TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk mengekstraksi lima topik utama berdasarkan kata-kata dominan yang muncul dalam ulasan pengguna. Representasi ini membantu mereduksi dimensi dan menyoroti kata kunci yang paling relevan dengan pembahasan pengguna.

```
Memproses file: playstore_preprocessed.csv
Topic 1: game nice like favourite fun play love amazing playing great
Topic 2: good game really friend app creative play please server add
Topic 3: best ever world game played one life survival app seen
Topic 4: op game minecraft ever app gaming hai bro ha gem
Topic 5: minecraft love play please like cant friend server fix problem

Distribusi Topik untuk playstore_preprocessed.csv:
dominant_topic
1    1932
2     413
3     457
4     177
5    2874
Name: count, dtype: int64
```

Gambar 5 Hasil Extraction

Gambar 5 menunjukkan hasil ekstraksi topik dari data ulasan pengguna, yang menghasilkan lima topik utama dengan kumpulan kata-kata kunci seperti *game*, *love*, *server*, *play*, dan lainnya. Dari distribusi topik tersebut, topik kelima mendominasi dengan jumlah data sebanyak 2.874.

Tahap kedua, klasifikasi sentimen dilakukan dengan menggunakan algoritma *Support Vector Machine* (SVM) berbasis linear kernel. Data dibagi menjadi 70% data latih (4.097 data) dan 30% data uji (1.756 data). Hasil klasifikasi dapat dilihat pada Gambar 6 berikut ini.

```
Distribusi Label:
- negative: 1951 (33.33%)
- neutral: 1951 (33.33%)
- positive: 1951 (33.33%)
Pie chart distribusi label aktual disimpan ke: playstore_actual_label_distribution.png

Training Data: 4097, Testing Data: 1756

Memulai pelatihan model SVM...

Akurasi Model SVM: 0.9095

Laporan Klasifikasi:

      precision  recall  f1-score  support
negative    0.90    0.88    0.89     585
neutral     0.89    0.97    0.93     585
positive    0.95    0.87    0.91     586

accuracy                0.91    1756
macro avg             0.91    0.91    0.91    1756
weighted avg          0.91    0.91    0.91    1756
```

Gambar 6 Hasil Classification

Model SVM yang digunakan berhasil mencapai akurasi sebesar 90,95% dengan performa terbaik pada kelas positif, yang memiliki nilai *f1-score* sebesar 0,91. Evaluasi dilakukan menggunakan *confusion matrix* dan metrik *precision*, *recall*, serta *f1-score* untuk

memastikan validitas hasil klasifikasi. Distribusi label yang seimbang (masing-masing kelas berjumlah 1.951 data) serta hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan sentimen secara optimal dan memiliki kemampuan generalisasi yang baik.

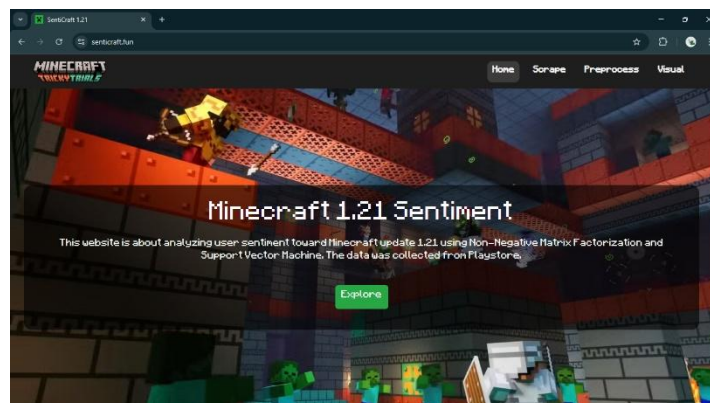
Seluruh proses pemodelan diimplementasikan menggunakan bahasa pemrograman *Python* dengan *library* utama *Scikit-learn* untuk TF-IDF, NMF, dan SVM, serta *TextBlob* untuk pelabelan sentimen. Pemilihan *Scikit-learn* dilakukan karena memiliki integrasi *pipeline* yang baik, dokumentasi lengkap, dan performa tinggi dalam menangani data teks berdimensi besar.

### 3.2. Hasil Implementasi Website

Sistem hasil penelitian diimplementasikan dalam bentuk aplikasi berbasis website. Website ini memiliki beberapa halaman utama, yaitu halaman homepage, halaman scraping, halaman preprocessing, dan halaman visualisasi.

#### 3.2.1. Halaman Homepage

Halaman *homepage* menampilkan deskripsi umum aplikasi dan navigasi ke fitur utama. Tampilan dirancang responsif dan informatif. Hasil halaman *homepage* dapat dilihat pada Gambar 7.



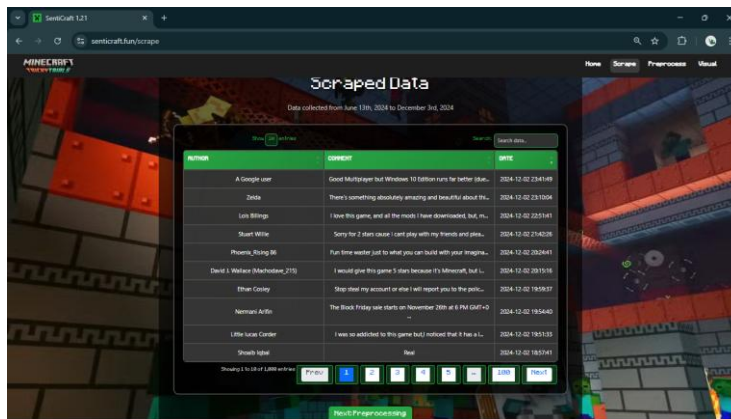
Gambar 7. Halaman Homepage

Gambar 7 menunjukkan *homepage* antarmuka untuk analisis sentimen terhadap *update game Minecraft 1.21*. Halaman ini menampilkan tema *Minecraft* dan menjelaskan tujuan website, yaitu menganalisis opini pengguna menggunakan metode *Non-Negative Matrix Factorization* (NMF) untuk ekstraksi topik dan *Support Vector Machine* (SVM) untuk klasifikasi sentimen. Data yang digunakan berasal dari ulasan di *Playstore*, dengan tombol "*Explore*" yang mengarahkan pengguna ke halaman *scraping*.

#### 3.2.2. Halaman Scraping

Halaman *scraping* menampilkan hasil pengumpulan data dari *Google Play Store*. Jumlah komentar, tanggal, dan format data ditampilkan secara ringkas. Hasil dari halaman *scraping* dapat dilihat pada Gambar 8.

Halaman *scraping* pada website ini menampilkan data yang dikumpulkan dari 13 Juni 2024 hingga 3 Desember 2024, seperti terlihat pada Gambar 4.6. Data berupa komentar pengguna ditampilkan dalam tabel interaktif dengan kolom *Author*, *Comment*, dan *Date*, serta dilengkapi fungsi pencarian dan paginasi untuk memudahkan eksplorasi.

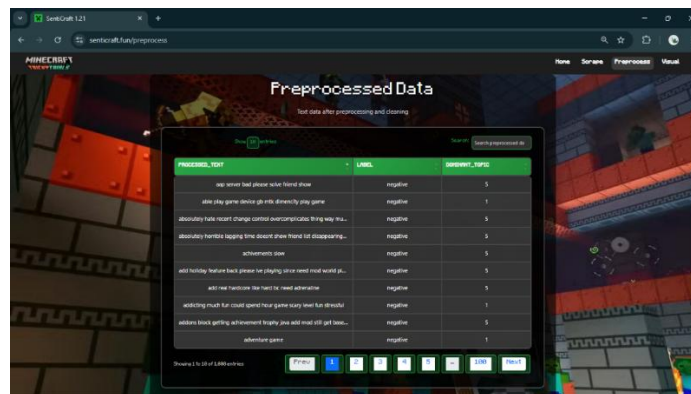


Gambar 8 Halaman Scraping

### 3.2.3. Halaman Preprocessing

Halaman *preprocessing* menampilkan hasil dari *preprocessing*. Hasil dari halaman *preprocessing* dapat dilihat pada Gambar 9.

Halaman *Preprocessing* menampilkan informasi tentang data teks setelah pembersihan dan pengolahan. Terdapat tabel dengan kolom "*PROCESSED\_TEXT*" yang berisi komentar pemain, kolom "*LABEL*" untuk sentimen (positif, negatif, atau netral), dan kolom "*DOMINANT\_TOPIC*" yang menunjukkan topik utama komentar.



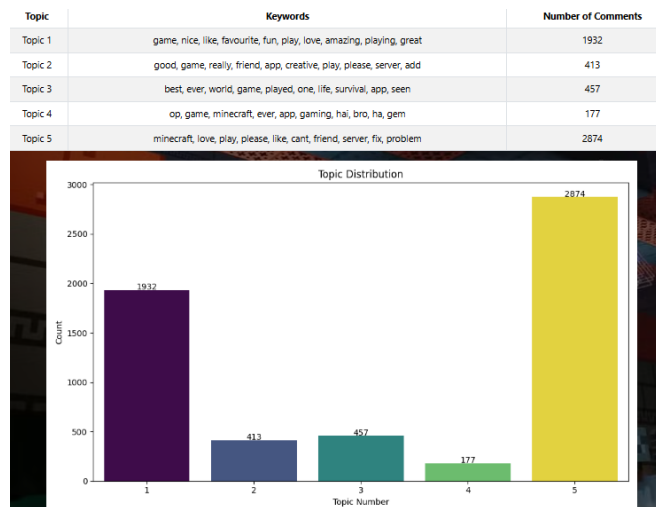
Gambar 9. Halaman Preprocessing

### 3.2.4. Halaman Visualization

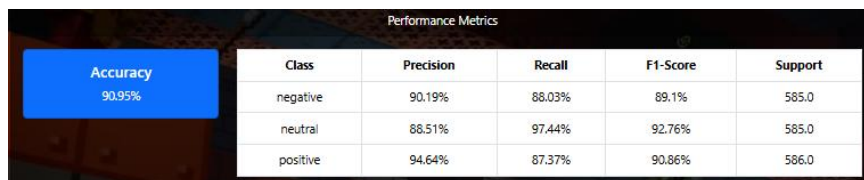
Pada halaman *visualization* terdapat 4 menu. Pada halaman *visualization* terdapat 4 menu yaitu, *Topics*, *Classifications*, *Sentiment*, dan *Summary*. Berikut adalah hasil dari 4 menu tersebut.

Pada menu *topics* akan ditampilkan visual dari hasil ekstraksi topik menggunakan *Non-Negative Matrix Factorization*. Proses ekstraksi topik menggunakan NMF menghasilkan lima topik utama yang teridentifikasi dari 5.853 komentar. Topik 1 mencerminkan ekspresi positif seperti "nice", "fun", dan "amazing" yang berjumlah 1.932 komentar. Topik 2 menyoroti permintaan pengguna terkait fitur dan server dengan 413 komentar. Topik 3 menggambarkan pengalaman pengguna dalam mode survival dan dunia game sebanyak 457 komentar. Topik 4 berisi kata-kata terkait pujian dengan 177 komentar, dan Topik 5 lebih banyak berisi permintaan perbaikan dan kritik terkait server dengan total 2.874 komentar, menjadikannya topik terbanyak.

Penerapan Metode Non-Negative Matrix Factorization dan Support Vector Machine pada Sentimen Pengguna terhadap Update Minecraft 1.21 berbasis Website  
 Tengku Syafiq Ali Syahputra, Ardian Prima Atmaja, Susilo Veri Yulianto



Gambar 10. Menu Topics

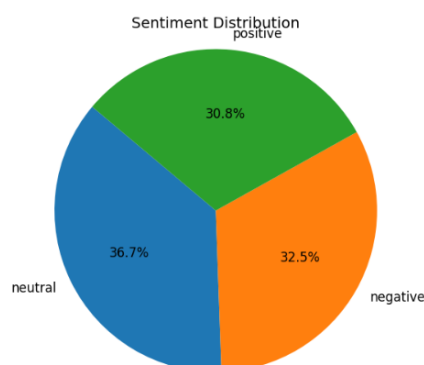


Gambar 11. Menu Classifications

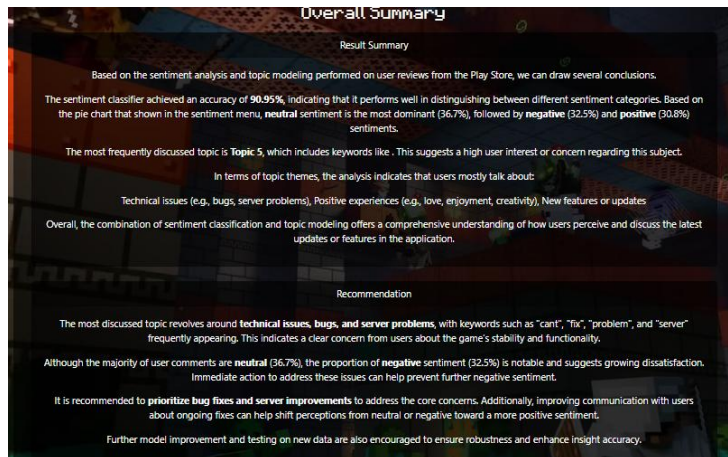
Pada menu *classification* akan ditampilkan performa dari model menggunakan metode *Support Vector Machine*. Hasil performa menunjukkan tingkat akurasi sebesar 90,95%. Selain itu, performa juga menunjukkan metrik performa yang tinggi dengan nilai *precision* tertinggi pada kelas positif (94,64%) dan *recall* tertinggi pada kelas netral (97,44%).

Pada menu *sentiment* akan ditampilkan distribusi sentimen berupa diagram *pie*. Berdasarkan hasil klasifikasi menunjukkan dominasi sentimen netral sebesar 36,7%, disusul oleh negatif sebesar 32,5% dan positif sebesar 30,8%. Diagram *pie* dapat dilihat pada Gambar 12.

Pada menu *summary* berisi interpretasi hasil analisis dan rekomendasi bagi pengembang. Halaman ini menjadi ringkasan dari seluruh proses dan hasil analisis.



Gambar 12. Menu Sentiment



Gambar 13 Menu Summary

### 3.3. Hasil Deploying

Hasil *deploying* merupakan tahap akhir dari proses implementasi aplikasi, dimana aplikasi yang telah dikembangkan di-*deploy* pada lingkungan *hosting* agar dapat diakses oleh pengguna. *Deploying* dilakukan menggunakan layanan *Amazon EC2 (Elastic Compute Cloud)* sebagai server dan domain publik untuk mengakses website. Tahapan ini memastikan bahwa seluruh fungsi aplikasi, mulai dari lihat data hingga visualisasi hasil analisis, dapat berjalan dengan baik secara online.

## 4. Kesimpulan

Berdasarkan hasil penelitian dan implementasi sistem, dapat disimpulkan bahwa metode *Non-Negative Matrix Factorization (NMF)* dan *Support Vector Machine (SVM)* berhasil diterapkan dalam aplikasi berbasis web untuk mengekstraksi topik dan mengklasifikasikan sentimen pengguna terhadap *update Minecraft 1.21*. Melalui proses pengolahan data yang dilakukan secara bertahap, sistem mampu mengelompokkan komentar ke dalam lima topik utama dengan bantuan NMF, di mana topik dominan berisi keluhan teknis seperti *bug* dan gangguan server. Sementara itu, klasifikasi sentimen menggunakan SVM menghasilkan akurasi sebesar 90,95%, yang menunjukkan performa model cukup baik dalam membedakan komentar positif, negatif, dan netral. Distribusi sentimen menunjukkan bahwa komentar netral paling mendominasi dengan 36,7%, diikuti oleh negatif sebesar 32,5%, dan positif sebesar 30,8%.

Meskipun hasil yang diperoleh menunjukkan kinerja model yang baik, penelitian ini belum melakukan perbandingan dengan algoritma lain seperti *Logistic Regression*, *Naïve Bayes*, atau *Random Forest*. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan komparasi antar-algoritma klasifikasi guna memastikan model terbaik dalam konteks analisis sentimen pengguna game. Selain itu, penelitian lanjutan juga dapat memperluas cakupan data dengan memasukkan ulasan dari berbagai platform dan bahasa lain, serta mengeksplorasi model berbasis *deep learning* untuk meningkatkan akurasi dan pemahaman konteks sentimen yang lebih kompleks..

## Daftar Pustaka

- [1] E. M. Cigognini and A. Nardi, "Minecraft as an Educational Tool Before, During, and After the Pandemic: A Case Study Research Project," *European Journal of Open, Distance*

- and *E-Learning*, vol. 26, no. 1, pp. 87–103, 2024. [Online]. Available: <https://doi.org/10.2478/eurodl-2024-0011>
- [2] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, 2022. [Online]. Available: <https://doi.org/10.1007/s10462-022-10144-1>
- [3] B. AlBadani, R. Shi, and J. Dong, "A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM," *Applied System Innovation*, vol. 5, no. 1, 2022. [Online]. Available: <https://doi.org/10.3390/asi5010013>
- [4] S. R. Hakim, M. A. Rizki, N. Fitri, Y. A. Rizkie, and R. Nooraeni, "Analisis Sentimen Pengguna Instagram Terhadap Kebijakan Kemdikbud Mengenai Bantuan Kuota Internet dengan Metode Support Vector Machine (SVM)," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 8, no. 2, pp. 15-24, 2020. [Online]. Available: <https://doi.org/10.24252/msa.v8i2.16795>
- [5] S. Fide, S. Suparti, and S. Sudarno, "Analisis sentimen ulasan aplikasi TikTok di Google Play menggunakan metode Support Vector Machine (SVM) dan asosiasi," *Jurnal Gaussian*, vol. 10, no. 3, pp. 346–358, 2021. [Online]. Available: <https://doi.org/10.14710/j.gauss.10.3.346-358>
- [6] M. Yolanda and R. T. Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 6, no. 2, pp. 76–83, 2024. [Online]. Available: <https://doi.org/10.35580/variasiunm258>
- [7] R. B. Afandi, T. F. Nurdiansyah, A. N. Ramadhani, and A. P. Sari, "Implementasi Support Vector Machine untuk Analisis Sentimen Aplikasi 'MPStore-Super App UMKM'," *Jurnal Informatika Polinema (JIP)*, vol. 10, no. 4, pp. 565-569, 2024.
- [8] A. Tabassum and R. R. Patil, "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing," *International Research Journal of Engineering and Technology*, vol. 7, no. 6, pp. 4864-4867, 2020. [Online]. Available: [www.irjet.net](http://www.irjet.net)
- [9] A. Petukhova and N. Fachada, "TextCL: A Python package for NLP preprocessing tasks," *SoftwareX*, vol. 19, 2022. [Online]. Available: <https://doi.org/10.1016/j.softx.2022.101122>
- [10] M. Das, K. Selvakumar, and P. J. A. Alphonse, "A Comparative Study on TF-IDF Feature Weighting Method and Its Analysis Using Unstructured Dataset," 2023. [Online]. Available: <https://arxiv.org/abs/2308.04037>
- [11] X. Lin and P. C. Boutros, "Optimization and expansion of non-negative matrix factorization," *BMC Bioinformatics*, vol. 21, no. 1, 2020. [Online]. Available: <https://doi.org/10.1186/s12859-019-3312-5>
- [12] S. Mifrah, "Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID-19 Corpus," *International Journal of Advanced Trends in Computer Science and Engineering*, 2020. [Online]. Available: <https://doi.org/10.30534/ijatcse/2020/231942020>
- [13] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*, 3rd ed., online manuscript, 2020. [Online]. Available: [https://web.stanford.edu/~jurafsky/slp3/old\\_dec20/ed3book\\_dec302020.pdf](https://web.stanford.edu/~jurafsky/slp3/old_dec20/ed3book_dec302020.pdf)

- [14]N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes," *Jurnal Informatika: Jurnal Pengembangan IT*, Vol. 10, no. 1, pp. 37-49, 2025. [Online]. Available: <https://doi.org/10.30591/jpit.v10i1.7106>
- [15]G. R. Ditami, E. F. Ripanti, and H. Sujaini, "Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 3, pp. 508–516, 2022. [Online]. Available: <https://doi.org/10.26418/jp.v8i3.56478>
- [16]E. Sukmana and L. Nilawati, "Penerapan Machine Learning Pada Sistem Informasi Klasifikasi Informasi Penggalan Potensi Pajak," *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 13, no. 3, 2024. [Online]. Available: <http://dx.doi.org/10.35889/jutisi.v13i3.2323> .
- [17]M. Ramdhan, *Metode Penelitian*, Surabaya: Cipta Media Nusantara, 2021.
- [18]M. Opalski, *A case study: A detailed look at real-world machine learning project using Agile*, AI-Agile.org, 2023. [Online]. Available: <https://www.ai-agile.org/2023/10/case-study-detailed-look-at-real-world.html>