

---

---

## ANALISIS SENTIMEN ULASAN GAME MOBILE FIRST-PERSON SHOOTER DI GOOGLE PLAY STORE MENGGUNAKAN METODE PEMBOBOTAN TF-IDF

Syahroni Wahyu Iriananda<sup>1\*</sup>, Rangga Pahlevi Putra<sup>1)</sup>, Anugrah Ahzul Raihan<sup>1)</sup>, Deni Adi Saputra<sup>1)</sup>, Egi Verdiansyah<sup>1)</sup>

<sup>1)</sup> Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas Widyagama Malang

---

---

INFORMASI ARTIKEL	ABSTRAK
<p><b>Data Artikel:</b> Naskah masuk, 10 Agustus 2023 Direvisi, 14 Agustus 2023 Diterima, 17 Agustus 2023</p> <p><b>*Email Korespondensi:</b> <a href="mailto:syahroni@widyagama.ac.id">syahroni@widyagama.ac.id</a></p>	<p>Paper ini membahas tentang analisis sentimen ulasan game mobile genre FPS menggunakan metode pembobotan TF-IDF. Dalam penelitian ini, penulis menggunakan 2180 ulasan yang telah divalidasi dan dibersihkan, di mana 1258 ulasan diklasifikasikan sebagai positif dan 922 ulasan sebagai negatif. Dengan menggunakan pembobotan TF-IDF dan pengujian model klasifikasi, penelitian ini mencapai tingkat akurasi sebesar 76%, dengan presisi 75%, recall 74%, dan F1-score 75%. Hasil ini menunjukkan bahwa metode pembobotan TF-IDF dapat menghasilkan analisis sentimen yang efektif dan otomatis untuk ulasan game mobile genre FPS, memberikan kontribusi penting dalam pengembangan metode analisis sentimen dalam konteks tersebut.</p> <p><b>Kata Kunci:</b> FPS; TF-IDF; Game Mobile; Ulasan; Analisis Sentimen</p>

---

---

### 1. PENDAHULUAN

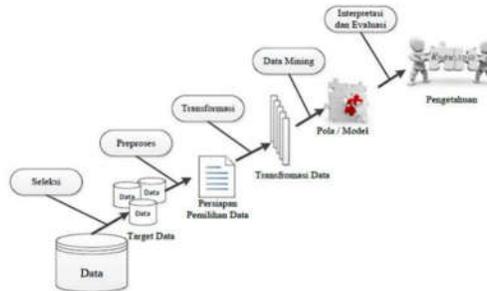
Analisis sentimen adalah salah satu cabang dari bidang pengolahan bahasa alami (natural language processing) yang bertujuan untuk mengidentifikasi dan mengekstraksi opini, emosi, dan sikap dari teks [1]. Analisis sentimen dapat diterapkan pada berbagai domain, salah satunya adalah ulasan game mobile. Ulasan game mobile merupakan sumber informasi yang berguna bagi pengembang game untuk mengetahui kepuasan dan masukan dari pengguna. Namun, ulasan game mobile seringkali berjumlah sangat banyak dan bervariasi, sehingga sulit untuk dianalisis secara manual. Oleh karena itu, dibutuhkan sebuah metode yang dapat mengotomatisasi proses analisis sentimen ulasan game mobile. [2]

Salah satu metode yang dapat digunakan untuk analisis sentimen adalah pembobotan TF-IDF (term frequency-inverse document frequency)[3]. Pembobotan TF-IDF adalah sebuah teknik yang menghitung bobot dari setiap kata dalam sebuah dokumen berdasarkan frekuensi kemunculan kata tersebut dalam dokumen dan dalam koleksi dokumen. Pembobotan TF-IDF dapat membantu mengurangi dampak dari kata-kata yang sering muncul tetapi tidak relevan dengan topik dokumen, seperti kata hubung atau kata sapaan. Dengan menggunakan pembobotan TF-IDF, kita dapat merepresentasikan setiap dokumen sebagai sebuah vektor yang terdiri dari bobot-bobot kata yang ada di dalamnya. [4]

Paper ini bertujuan untuk menganalisis sentimen ulasan game mobile genre FPS (first-person shooter) di playstore dengan menggunakan pembobotan TF-IDF. Genre FPS adalah salah satu genre game yang populer dan menantang, yang menuntut keterampilan, strategi, dan koordinasi dari pemainnya [5]. Ulasan game mobile genre FPS dapat mencerminkan aspek-aspek

seperti grafis, gameplay, kontrol, suara, cerita, dan lain-lain. Dengan menganalisis sentimen ulasan game mobile genre FPS, kita dapat mengetahui kelebihan dan kekurangan dari game-game tersebut, serta preferensi dan harapan dari penggunanya.

## 2. METODE PENELITIAN



Gambar 1 Metode Penelitian [6]

Metode penelitian yang digunakan dalam penelitian ini adalah Knowledge Discovery in Databases (KDD). Knowledge Database Discovery adalah suatu proses penggalian informasi yang terdapat pada database. Proses ini dilakukan dengan menggunakan teknik-teknik tertentu seperti clustering, data summarization, learning classification rules, finding dependency dan lain-lain [7]. Proses KDD memiliki 5 fase, yaitu:

- 1) Data Selection  
Proses seleksi data, juga dikenal sebagai data selection, melibatkan pengambilan data yang sudah ada untuk persiapan tahap awal dalam Knowledge Discovery in Database (KDD). [8]
- 2) Preprocessing  
Proses preprocessing merupakan proses deduplikasi data, selain itu proses ini juga memeriksa data yang tidak konsisten dan mengoreksi data yang ada jika terdapat kesalahan penulisan [8].
- 3) Transformation  
Data diubah atau digabungkan menjadi format yang cocok untuk pemrosesan data mining. Beberapa metode penambangan data memerlukan format data khusus untuk diterapkan. Misalnya, beberapa metode standar seperti analisis asosiasi dan pengelompokan hanya dapat menerima data input kategorikal. Oleh karena itu, perlu dilakukan pembagian data berupa nilai numerik kontinu menjadi beberapa interval. Proses ini sering disebut transformasi data. [9]
- 4) Data Mining  
Data mining merupakan salah satu tahapan dalam KDD yang bertujuan untuk menemukan pola-pola menarik dalam data yang telah dipilih, menggunakan berbagai teknik tertentu. Metode yang digunakan dalam data mining dapat bervariasi, dan pemilihan metode yang tepat sangat terkait dengan tujuan keseluruhan dan proses KDD. [10]
- 5) Evaluation  
Pada tahap ini, dilakukan identifikasi pola-pola menarik dalam knowledge base yang telah diidentifikasi. Selanjutnya, pola-pola khas dan model prediksi dihasilkan dan dievaluasi guna menilai apakah penelitian yang ada telah mencapai target yang diinginkan. [11]

### 3. HASIL DAN PEMBAHASAN

Hasil penelitian yang telah dilakukan ialah bagaimana melakukan sentimen analisis ulasan game mobile first-person shooter di google play store. Sentimen analisis tersebut menggunakan algoritma SVM serta seleksi fitur TF-IDF.

#### 3.1. Data Selection (pemilihan data)

Data ulasan yang digunakan berjumlah 3282 dataset dengan rentang score 1-5 akan diseleksi menggunakan perintah python untuk menghapus bagian yang kosong seperti pada Gambar 2.

```
data.dropna(subset=['content'], inplace=True)
data.dropna(subset=['score'], inplace=True)
Executed at 2023.07.06 12:16:41 in 211ms

data['content'].isnull().sum()
data['score'].isnull().sum()
Executed at 2023.07.06 12:16:41 in 187ms

0
```

Gambar 2 Menghapus baris kosong

Pada Gambar 2 setelah dilakukan penghapusan kolom yang kosong lalu dicek menggunakan perintah python “isnull”. Setelah itu menambahkan kolom baru untuk memudahkan dalam pemrosesan data seperti pada Gambar 3, dengan kondisi jika nilai ulasan antara 1-2 diberikan hasil negative, 3 diberikan hasil NaN dan 4-5 diberikan hasil positif.

```
data['sentimen'] = data['score'].apply(
    lambda x: 'negative' if x <= "2" else (
        'positive' if x >= "4" else np.nan
    )
)
```

Gambar 3 kolom baru pada tabel

Rincian dari ulasan yang telah di seleksi dan dibersihkan disajikan pada Tabel 1.

Tabel 1 Jumlah Ulasan yang Telah Diseleksi

Jeni Ulasan	Jumlah Ulasan
Positif	1258
Negaif	922

Tabel 1 menunjukkan jumlah ulasan yang telah diseleski dan dibersihkan yaitu 2180 ulasan , terdiri dari 1258 ulasan positif dan 922 ulasan negatif.

#### 3.2. Preprocessing

Setelah dilakukan seleksi dan pembersihan data, maka selanjutnya akan dilakukan tahap preprocessing untuk menghilangkan kata yang tidak diperlukan, symbol yang terdapat pada kalimat, mengubah semua huruf menjadi *lowercase*. Semua itu perlu dilakukan untuk mempermudah tahap kalsifikasi.

```
def preprocess_text(text):  
    # Menghilangkan simbol dan emotikon  
    text = re.sub(r'[^\w\s]', '', text)  
    text = re.sub(r'[\U0001F600-\U0001F6FF]', '', text)  
  
    # Mengubah huruf menjadi lowercase  
    text = text.lower()  
  
    # Menghilangkan stopwords  
    stop_words = set(stopwords.words('english'))  
    tokens = text.split()  
    filtered_tokens = [word for word in tokens if word not in stop_words]  
    text = ' '.join(filtered_tokens)  
  
    return text
```

Gambar 4 Preprocessing data

```
< < 5 rows > > 5 rows x 13 columns pd.DataFrame #  
content  
unfair matchmaking please add additional punis...  
hey game developers whats wrong many updates p...  
hate every time game lags game forces quit los...  
get higher rank put trash teammates plus snipe...  
bad experience got little bad network suddenly...
```

Gambar 5 Hasil preprocessing data

Gambar 4 adalah tahap preprocessing data yang dimulai dari menghilangkan simbol dan emoticon pada data, selanjutnya mengubah semua huruf menjadi *lowercase*, dan langkah terakhir adalah menghilangkan *stopwords*. Hasil dari tahap preprocessing data dapat dilihat pada Gambar 5.

### 3.3. Transformasi

```
# Pembobotan TF-IDF  
tfidf = TfidfVectorizer()  
X_train_tfidf = tfidf.fit_transform(X_train)  
X_test_tfidf = tfidf.transform(X_test)  
Executed at 2023.07.14 00:26:31 in 63ms
```

Gambar 6 Pembobotan TF-IDF

```
Feature: crimson TF-IDF value: 0.7598809481004654  
Feature: pakistan TF-IDF value: 0.36796269985209124  
Feature: spin TF-IDF value: 0.4171468390378313  
Feature: suit TF-IDF value: 0.3364118174384533
```

Gambar 7 Hasil Transformasi TF-IDF

Seperti yang ditunjukkan pada Gambar 6 terlihat nilai dari TF-IDF dari kata yang terdapat pada contoh kalimat ulasan yang telah ditransformasi.

### 3.4. Data Mining

Pada tahap sebelumnya telah dilakukan transformasi, hasil dari transformasi digunakan dalam tahap data mining yaitu tahap klasifikasi data menggunakan algoritma SVM (*support virtual machine*) yang akan ditampilkan dalam bentuk tabel *Confusion Matrix*. Penelitian ini akan menggunakan rasio pembagian data 80:20, lalu akan dibandingkan dengan model yang menerapkan TF-IDF, *Count Vectorizer*, dan *Hashing Vectorizer*.

Tabel 2 Hasil Klasifikasi TF-IDF

Prediksi	Aktual	
	Positif	Negatif
Positif	120	64
Negatif	42	210

Berdasarkan Tabel 2 terdapat 120 data diprediksi benar sebagai kelas positif, sedangkan 42 data positif diprediksi negatif. Lalu sebanyak 210 data negatif benar diprediksi sebagai kelas negatif, sedangkan 64 data kelas negatif diprediksi sebagai kelas positif.

Tabel 3 Hasil Klasifikasi Count Vectorizer

Prediksi	Aktual	
	Positif	Negatif
Positif	113	71
Negatif	37	215

Berdasarkan Tabel 3 terdapat 113 data diprediksi benar sebagai kelas positif, sedangkan 37 data positif diprediksi negatif. Lalu sebanyak 215 data negatif benar diprediksi sebagai kelas negatif, sedangkan 71 data kelas negatif diprediksi sebagai kelas positif.

Tabel 4 Hasil Klasifikasi Hashing Vectorizer

Prediksi	Aktual	
	Positif	Negatif
Positif	115	69
Negatif	42	210

Berdasarkan Tabel 4 terdapat 115 data diprediksi benar sebagai kelas positif, sedangkan 69 data positif diprediksi negatif. Lalu sebanyak 210 data negatif benar diprediksi sebagai kelas negatif, sedangkan 42 data kelas negatif diprediksi sebagai kelas positif.

### 3.5. Evaluasi

Setelah dilakukan seluruh pengujian maka tahap selanjutnya adalah perbandingan untuk mencari model terbaik dalam proses klasifikasi Gambar 7 menunjukkan hasil dari evaluasi

TF-IDF

```
Classification Report:
              precision    recall  f1-score   support

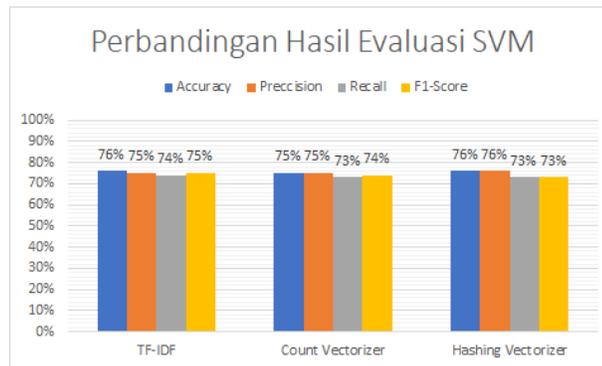
   negative      0.74      0.65      0.69      184
   positive      0.77      0.83      0.80      252

 accuracy              0.76      436
 macro avg              0.75      0.74      0.75      436
weighted avg              0.76      0.76      0.75      436
```

Gambar 8 Evaluasi klasifikasi TF-IDF

Pada Gambar 7 menunjukkan bahwa hasil akurasi menggunakan algoritma SVM dengan menerapkan TF-IDF menghasilkan akurasi sebesar 76%, *precision* sebesar 75%, *recall* sebesar 74% dan *F1-score* sebesar 75%. Hasil akurasi dengan menerapkan *Count Vectorizer* ditunjukkan pada Gambar 8.





Gambar 12 Grafik Perbandingan Hasil Evaluasi

Dapat disimpulkan berdasarkan hasil perbandingan tersebut model pembobotan yang terbaik untuk klasifikasi SVM pada rasio 80:20 berada pada pembobotan TF-IDF dengan tingkat presentasi akurasi tertinggi sebesar 76%, dengan *precision* sebesar 75%, *recall* sebesar 74% dan *F1-score* sebesar 75%. Pemilihan model TF-IDF dilakukan karena meskipun hanya ada perbedaan 1-2% pada pembobotan lainnya, sangat berpengaruh pada hasil confusion matrik yang ada. TF-IDF tidak hanya menghitung bobot dari kata yang sering muncul, namun juga kata unik atau kata yang tidak sering muncul. Hal tersebut membuat TF-IDF memiliki *balance* atau keseimbangan dalam menghitung setiap kata pada dataset.

#### 4. KESIMPULAN

Berdasarkan dari penelitian yang telah dilakukan dapat diambil kesimpulan proses dari analisis sentimen ulasan game mobile fps dari data ulasan yang berjumlah 3282 lalu di validasi dan dibersihkan sehingga berjumlah 2180 ulasan, terdiri dari 1258 ulasan positif dan 922 ulasan negatif. Menghasilkan sentimen ulasan positif diakibatkan presentase klasifikasi lebih besar ke arah ulasan positif. Pengujian terhadap model klasifikasi menggunakan rasio 80:20 dengan menggunakan pembobotan TF-IDF yang menghasilkan tingkat presentase akurasi tertinggi sebesar 76%, dengan *precision* sebesar 75%, *recall* sebesar 74% dan *F1-score* sebesar 75%. Hasil tersebut dikarenakan TF-IDF tidak hanya menghitung bobot dari kata yang sering muncul, namun juga kata unik atau kata yang tidak sering muncul. Hal tersebut membuat TF-IDF memiliki *balance* atau keseimbangan dalam menghitung setiap kata pada dataset. Hasil ini menunjukkan bahwa metode pembobotan TF-IDF dapat menghasilkan analisis sentimen yang efektif dan otomatis untuk ulasan game mobile genre FPS, memberikan kontribusi penting dalam pengembangan metode analisis sentimen dalam konteks tersebut.

#### 6. DAFTAR PUSTAKA

- [1] M. Furini and M. Montanero, "Sentiment analysis and Twitter: a game proposal," *Pers. Ubiquitous Comput.*, vol. 22, no. 4, pp. 771-785, 2018, doi: 10.1007/s00779-018-1142-5.
- [2] Risda, F. J. Septriwinti, and F. Nasution, "Pendekatan Pemrosesan Informasi," *Mudabbir (Journal Res. Educ. Stud.)*, vol. 3, no. 1, pp. 49-59, 2023.

- 
- [3] R. Kosasih and A. Alberto, "Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier," *Ilk. J. Ilm.*, vol. 13, no. 2, pp. 101-109, 2021, doi: 10.33096/ilkom.v13i2.721.101-109.
- [4] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43-49, 2019, doi: 10.52985/insyst.v1i1.36.
- [5] R. Weber, K.-M. Behr, R. Tamborini, U. Ritterfeld, and K. Mathiak, "What Do We Really Know about First-Person-Shooter Games? an Event-Related, High-Resolution Content Analysis," *J. Comput. Commun.*, vol. 14, no. 4, pp. 1016-1037, Jul. 2009, doi: 10.1111/j.1083-6101.2009.01479.x.
- [6] G. F. Mandias, G. A. Sandag, A. G. Takalumbide, and C. Wahongan, "Analisa Pola Peminjaman Buku di Perpustakaan Universitas Klabat Menggunakan Algoritma Apriori," *Konf. Nas. Sist. Inf.*, pp. 8-9, 2018.
- [7] G. Piatesky-Shapiro, "Knowledge discovery in databases: Progress report," *Knowl. Eng. Rev.*, vol. 9, no. 1, pp. 57-60, 1994, doi: 10.1017/S0269888900006573.
- [8] R. P. Sidiq, B. A. Dermawan, and Y. Umaidah, "Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes," *J. Inform. Univ. Pamulang*, vol. 5, no. 3, p. 356, 2020, doi: 10.32493/informatika.v5i3.6571.
- [9] J. Eska, "Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksianan," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, pp. 9-13, 2016.
- [10] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46-63, 2022, doi: 10.34010/jati.v12i1.6733.
- [11] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17-24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24