

KLASIFIKASI SENTIMEN ULASAN MINECRAFT PADA GOOGLE PLAY STORE: STUDI KOMPARATIF NAÏVE BAYES DAN RANDOM FOREST

Oleh :

Luluk Asti Qomariah (221080200106)

Dosen Pembimbing :

Ade Eviyanti, S.Kom., M.Kom.

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

2025/2026



Pendahuluan

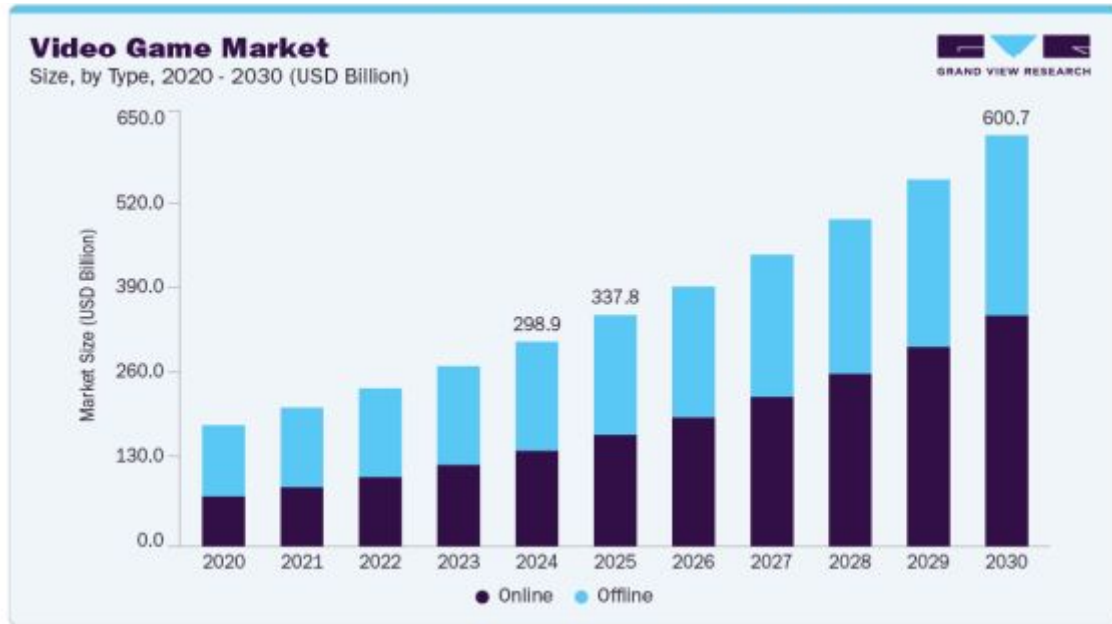


Diagram di atas menunjukkan total nilai pasar dalam miliar dolar AS (USD Billion) untuk tahun tersebut. Diagram ini memprediksi pertumbuhan pasar yang stabil hingga tahun 2030, dengan nilai pasar mencapai \$600,7 miliar pada tahun 2030.

Industri game telah menjadi salah satu industri hiburan terbesar di dunia dengan nilai pasar miliaran dolar.

Minecraft adalah game populer dengan jutaan pengguna aktif, menghasilkan jutaan ulasan di Google Play Store yang ideal untuk dianalisis.

google play store menyediakan fitur ulasan yang menjadi sumber feedback berharga dari pengguna aplikasi

Penelitian ini membandingkan algoritma Naïve Bayes dan Random Forest untuk menentukan mana yang lebih efektif dalam mengklasifikasikan sentimen ulasan game Minecraft.

Rumusan Masalah

- Bagaimana hasil klasifikasi sentiment menggunakan algoritma naïve bayes?
- Bagaimana hasil klasifikasi sentiment menggunakan algoritma random forest?
- Bagaimana perbandingan kinerja antara kedua algoritma tersebut?

Tujuan dan Batasan Masalah

Tujuan

- Menganalisis hasil klasifikasi sentiment menggunakan naïve bayes
- Menganalisis hasil klasifikasi sentiment menggunakan random forest
- Membandingkan kinerja kedua algoritma untuk menemukan model terbaik.

Batasan masalah

- Data berasal dari komentar game Minecraft di google play store
- Jumlah data sekitar 8.000 hingga 10.000 komentar.
- Bahasa yang di analisis hanya Bahasa Indonesia.
- Hanya menggunakan algoritma naïve bayes dan random forest.

- Metrik evaluasi yang digunakan adalah akurasi, presisi, recall, Dan F1-score

Penelitian Terdahulu

A. Miftahusalam, H. Pratiwi, dan I. Slamet(2023)

Judul : Perbandingan Metode Random Forest dan Naive Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile

D. Reinaldo and Y. F. Riti(2024)

Judul : Analisis Sentimen Ulasan Aplikasi Genshin Impact di Play Store Menggunakan Random Forest

R. P. Setiawan, B. Irawan, and W. P. Prihartono(2025)

Judul : Analisis Sentimen Ulasan Growtopia Di Google Play Store Menggunakan Naïve Bayes Classifier Untuk Identifikasi Kebutuhan Pengguna

Kesimpulan : Penelitian yang menggunakan kedua algoritma tersebut dalam analisis sentiment Sebagian besar berfokus pada aplikasi non-game, dan belum banyak yang membandingkan kedua metode secara langsung. Belum di temukan juga penelitian yang membahas komentar pengguna game Minecraft. Maka dari itu penelitian ini dilakukan untuk mengisi kekosongan tersebut.

Metodologi Penelitian

- **Jenis penelitian:** Penelitian kuantitatif dan komparatif
- **Populasi :** ulasan pengguna game Minecraft di google play store
- **Jumlah data :** 12.691 data
- **Teknik pengambilan data :** web scraping menggunakan library python google-play-scraper
- **Model :** naïve bayes dan random forest
- **Alat dan bahan :**
 - **Hardware:** Laptop dengan spesifikasi intel Celeron N4120, RAM 8 GB.

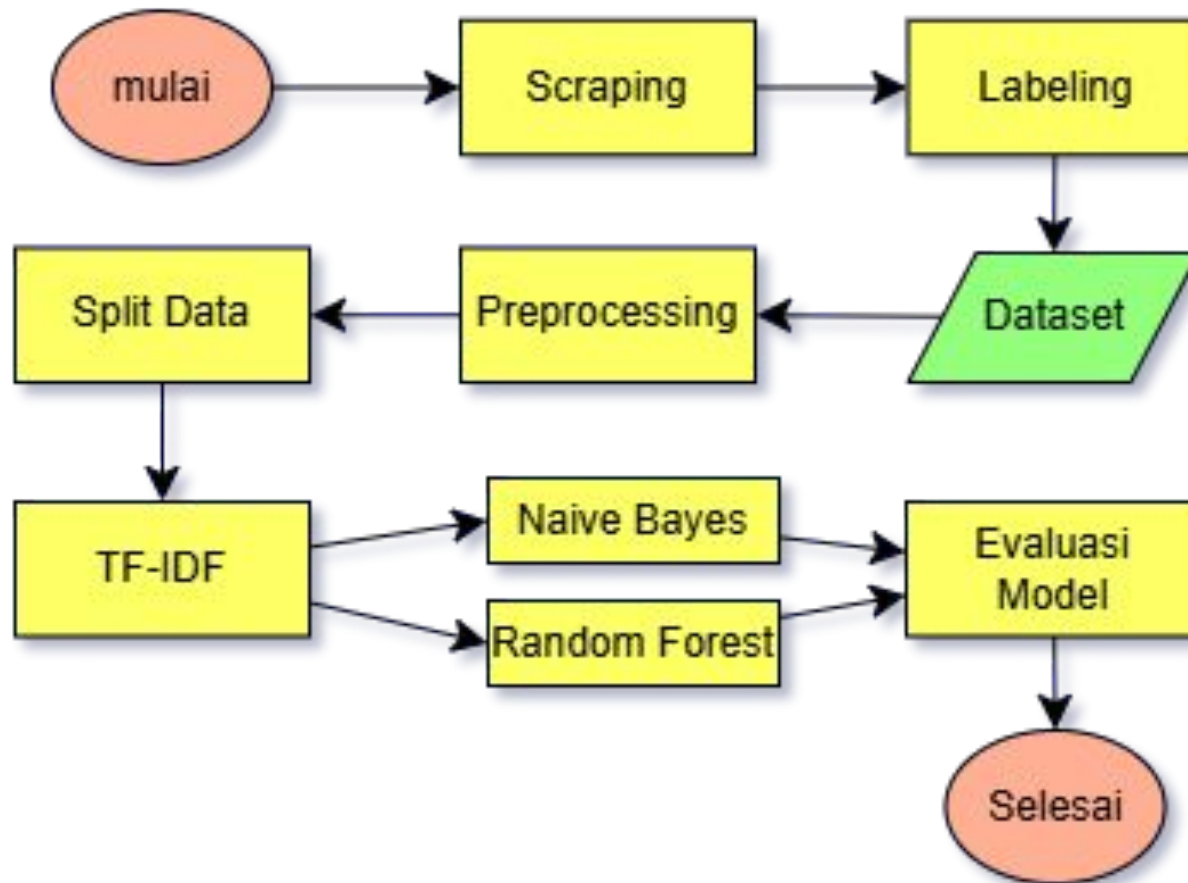
Variabel Penelitian

- **Variabel Independen** : naïve bayes dan random forest
- **Variabel Dependen** : akurasi, presisi, recall, dan F1-Score.
- **Variabel Kontrol** : dataset, tahapan preprocessing dan Teknik pembobotan kata

Variabel Penelitian

- **Variabel Independen** : naïve bayes dan random forest
- **Variabel Dependen** : akurasi, presisi, recall, dan F1-Score.
- **Variabel Kontrol** : dataset, tahapan preprocessing dan Teknik pembobotan kata

Alur Penelitian



Scarping Data

- **Sumber** : Google Play Store
- **Library** : google-play-scraper
- **Jumlah Data** : 12.691 ulasan

NO	Content	Score
1	Sgt seru dan menyenangkan dimainkan bersama keluarga.	5
2	Bisa bermain bersama teman, sgt menyenangkan.	5
3	Saya sudah bermain sejak 2019, game ini masih seru sampai sekarang.	5
4	Game ini sangat seru menurut saya, meskipun mungkin orang lain berbeda pendapat.	4
5	Bagus sekali, game ini cocok dimainkan anak-anak juga.	5

Pelebelan

- Proses pelebelan dilakukan secara manual
- Tool : Microsoft Excel
- Kelas : Positif & Negatif
- Alasan manual : - Menjaga akurasi ground truth

NO	Conten	Label
1	Sgt seru dan menyenangkan dimainkan bersama keluarga.	1
2	Bisa bermain bersama teman, sgt menyenangkan.	1
3	Saya sudah bermain sejak 2019, game ini masih seru sampai sekarang.	1
4	kapan mojang lu bakal update dimensi diend bosen bet gitu gitu aja	0
5	uang saya hilang saya tidak bisa mendownload game ini saya sudah membayar kembalikan uang saya	0

Preprocessing Data Teks

- Cleaning : Membersihkan teks dari karakter tidak penting seperti URL, emoji, angka, dan tanda baca.
- Case Folding : Mengubah seluruh Huruf Kecil agar konsisten.
- Tokenization : Memecah teks menjadi kata-kata (token).
- Normalization : Mengubah kata tidak baku atau singkatan menjadi baku.
- Stopword Removal : Menghapus kata umum yang tidak memiliki makna penting dalam analisis.

- Stemming

Before	game nya bagus banget aku kasih bintang 5 👍👍👍	mi.
After	bagus banget kasih bintang bagus bagus bagus	

Pembersihan Sesudah Preprocessing

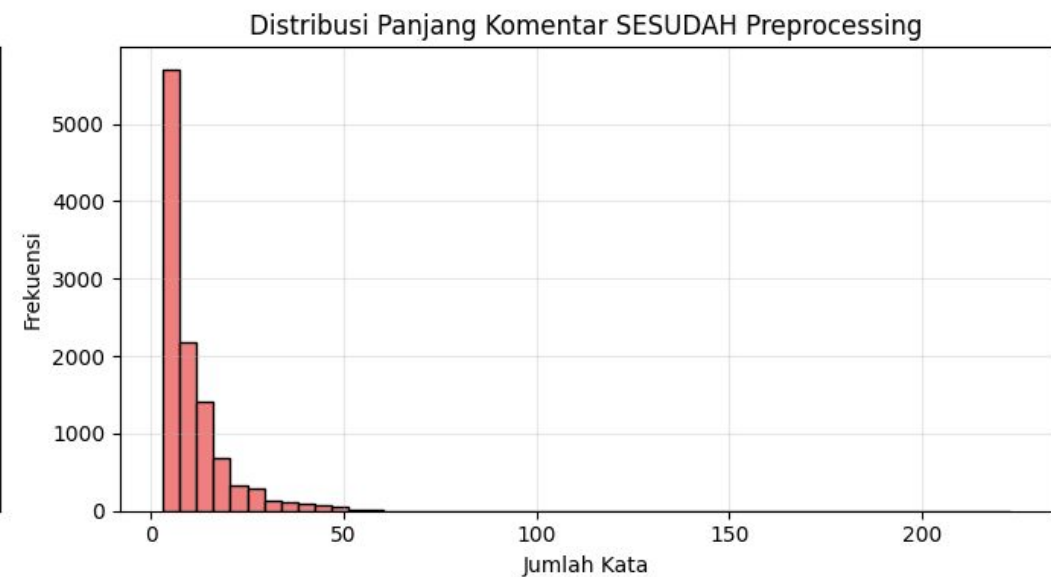
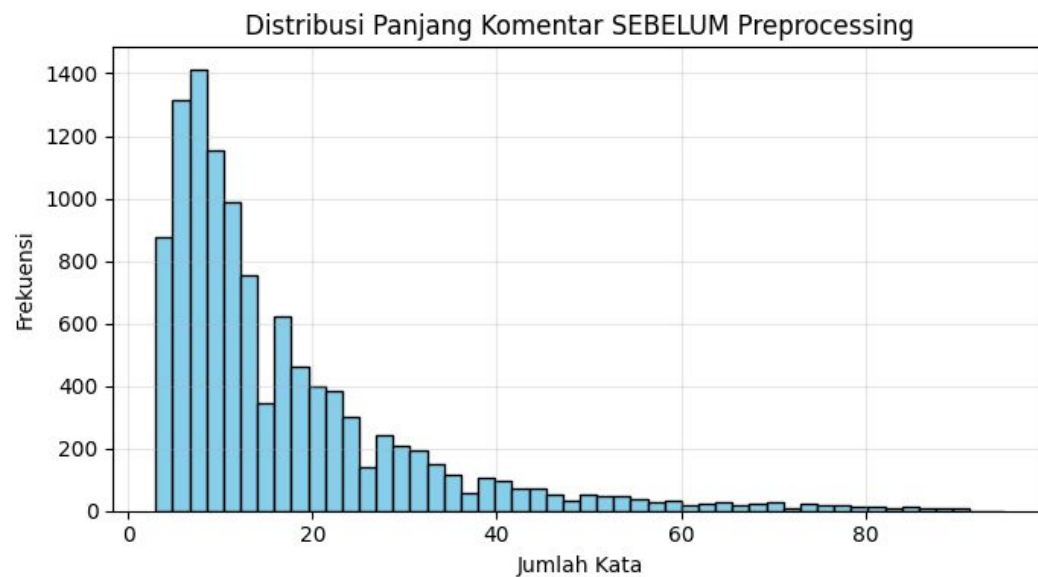
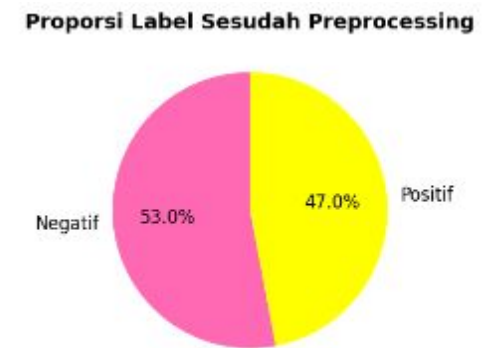
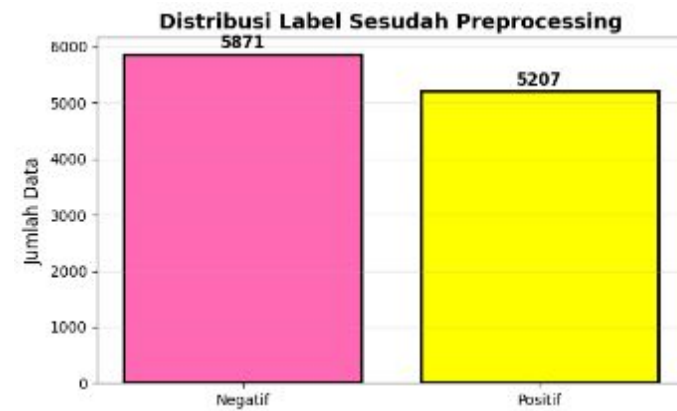
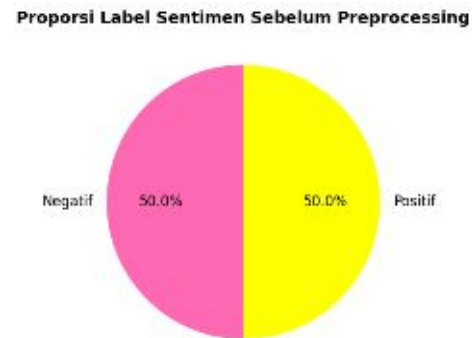
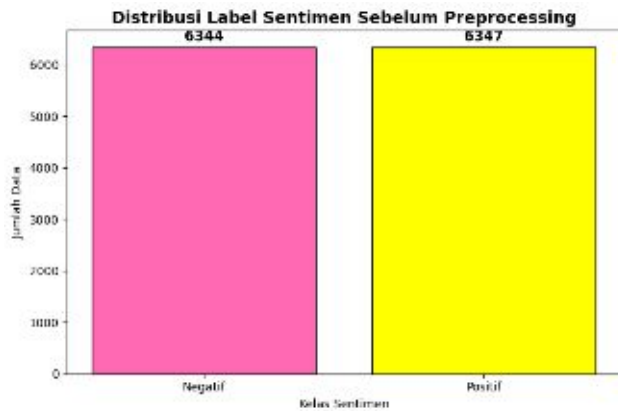
Penghapusan : Data duplikat, Data kosong (NaN), Komentar kurang dari 3 kata

```
=====:  
📊 RINGKASAN PEMBERSIHAN DATA  
=====:  
- Baris NaN dihapus: 4  
- Baris duplikat dihapus: 742  
- Baris < 3 kata dihapus: 867  
- Total data akhir: 11078 baris  
=====:
```

Total data sebelum preprocessing 12.691

Total data setelah preprocessing 11.078

EDA Before & After Preprocessing

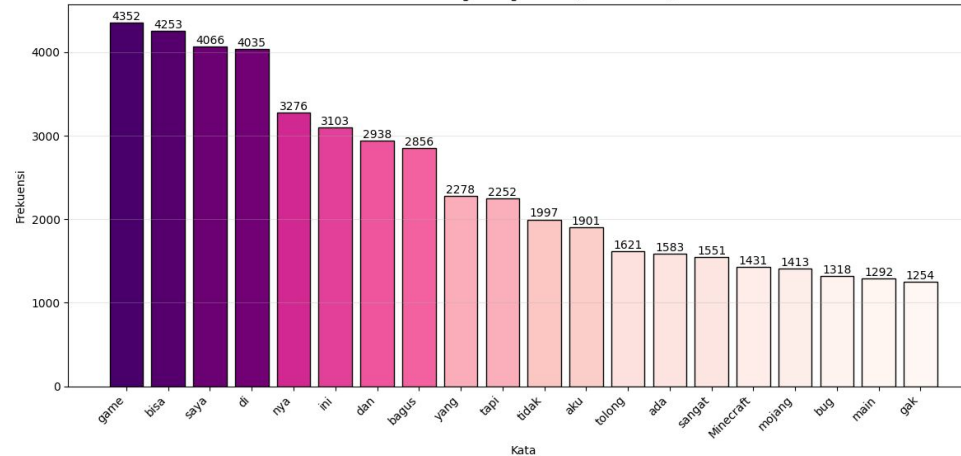


EDA Before & After Preprocessing

Wordcloud Setelah Preprocessing - Semua Data



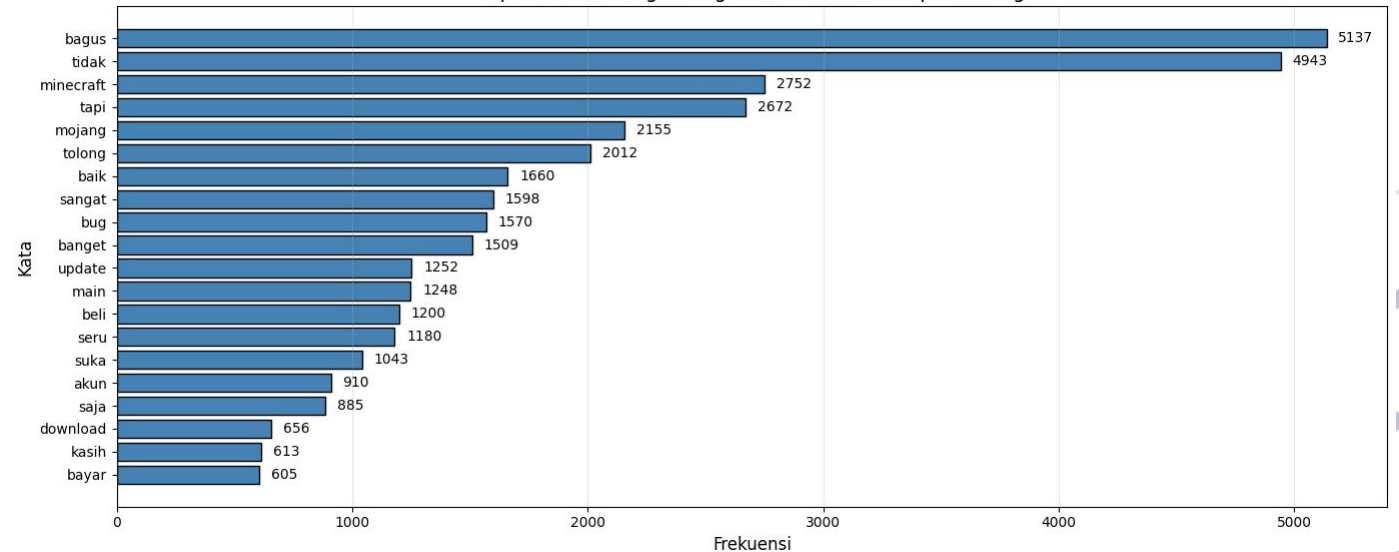
20 Kata Paling Sering Muncul (Data Mentah)



WordCloud Data Mentah (Sebelum Preprocessing)



Top 20 Kata Paling Sering Muncul Setelah Preprocessing



Pembagian Data

Pembagian Rasio : 70:30, 80:20, 90:10.

Rasio 70:30 :

Hasil split:

Data training: 7754 (69.99%)

Data testing: 3324 (30.01%)

Distribusi label TRAINING:

Label 0: 4109 (52.99%)

Label 1: 3645 (47.01%)

Distribusi label TESTING:

Label 0: 1762 (53.01%)

Label 1: 1562 (46.99%)

Rasio 80:20 :

Hasil split:

Data training: 8862 (80.00%)

Data testing: 2216 (20.00%)

Distribusi label TRAINING:

Label 0: 4697 (53.00%)

Label 1: 4165 (47.00%)

Distribusi label TESTING:

Label 0: 1174 (52.98%)

Label 1: 1042 (47.02%)

Rasio 90:10 :

Hasil split:

Data training: 9970 (90.00%)

Data testing: 1108 (10.00%)

Distribusi label TRAINING:

Label 0: 5284 (53.00%)

Label 1: 4686 (47.00%)

Distribusi label TESTING:

Label 0: 587 (52.98%)

Label 1: 521 (47.02%)

Pembobotan TF IDF

- TF-IDF (*Term Frequency – Inverse Document Frequency*) merupakan metode pembobotan kata yang digunakan untuk mengubah data teks menjadi representasi numerik.

Metode ini memberikan bobot pada setiap kata berdasarkan:

- **Seberapa sering kata muncul dalam satu dokumen (TF)**
- **Seberapa jarang kata tersebut muncul pada seluruh dokumen (IDF)**

Sehingga, kata yang sering muncul pada satu ulasan tetapi jarang muncul pada keseluruhan data akan memiliki bobot yang lebih tinggi dan dianggap lebih informatif.

Rasio	Train Size	Test Size	TF-IDF Features
90:10	9970	1108	5000
80:20	8862	2216	5000
70:30	7754	3324	5000

Pembobotan TF IDF

- TF-IDF (*Term Frequency – Inverse Document Frequency*) merupakan metode pembobotan kata yang digunakan untuk mengubah data teks menjadi representasi numerik.

Metode ini memberikan bobot pada setiap kata berdasarkan:

- **Seberapa sering kata muncul dalam satu dokumen (TF)**
- **Seberapa jarang kata tersebut muncul pada seluruh dokumen (IDF)**

Sehingga, kata yang sering muncul pada satu ulasan tetapi jarang muncul pada keseluruhan data akan memiliki bobot yang lebih tinggi dan dianggap lebih informatif.

Rasio	Train Size	Test Size	TF-IDF Features
90:10	9970	1108	5000
80:20	8862	2216	5000
70:30	7754	3324	5000

Pemodelan Klasifikasi

Hasil Algoritma Naïve Bayes :

Rasio 90:10 :

Rasio 80:20 :

Rasio 70:30 :

Hasil Naive Bayes (Rasio 90:10):

- Accuracy: 0.9052
- Precision: 0.8985
- Recall: 0.9002
- F1-Score: 0.8993
- ROC-AUC: 0.9537

Hasil Naive Bayes:

- Accuracy: 0.8935
- Precision: 0.8845
- Recall: 0.8896
- F1-Score: 0.8871
- ROC-AUC: 0.9534

Hasil Naive Bayes (Rasio 70:30):

- Accuracy: 0.8908
- Precision: 0.8865
- Recall: 0.8803
- F1-Score: 0.8834
- ROC-AUC: 0.9534

Confusion Matrix Naive Bayes:

```
[[534  53]
 [ 52 469]]
```

Confusion Matrix Naive Bayes:

```
[[1053  121]
 [ 115  927]]
```

Confusion Matrix Naive Bayes:

```
[[1586  176]
 [ 187 1375]]
```

Pemodelan Klasifikasi

Hasil Algoritma Random Forest:

Rasio 90:10 :

Rasio 80:20 :

Rasio 70:30 :

Hasil Random Forest (Rasio 90:10):

- Accuracy: 0.8944
- Precision: 0.8961
- Recall: 0.8772
- F1-Score: 0.8865
- ROC-AUC: 0.9544

Hasil Random Forest:

- Accuracy: 0.8840
- Precision: 0.8807
- Recall: 0.8714
- F1-Score: 0.8760
- ROC-AUC: 0.9506

Hasil Random Forest (Rasio 70:30):

- Accuracy: 0.8806
- Precision: 0.8741
- Recall: 0.8713
- F1-Score: 0.8727
- ROC-AUC: 0.9481

Confusion Matrix Random Forest:

```
[[534  53]
 [ 64 457]]
```

Confusion Matrix Random Forest:

```
[[1051 123]
 [ 134 908]]
```

Confusion Matrix Random Forest:

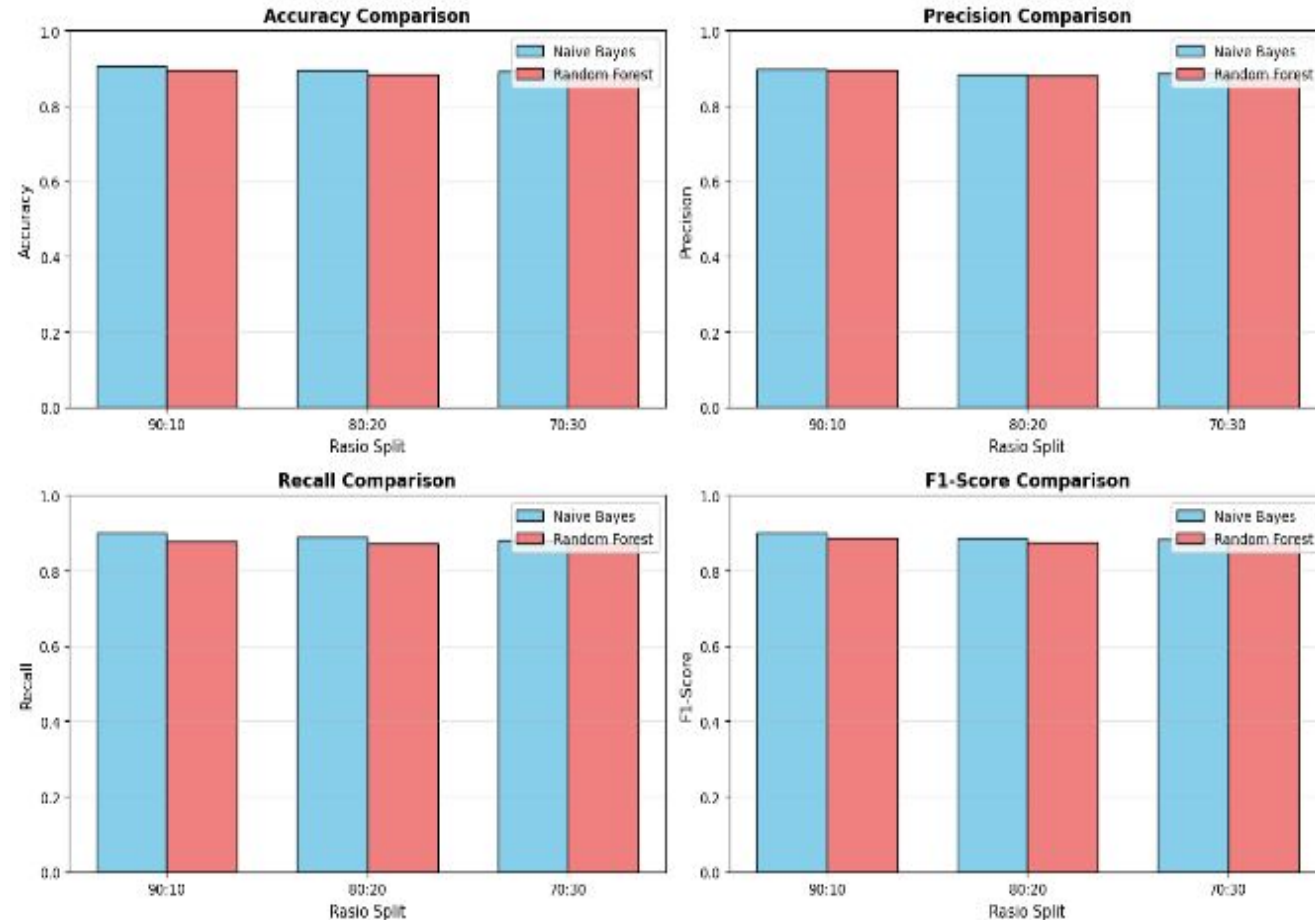
```
[[1566 196]
 [ 201 1361]]
```

Ringkasan Hasil Semua Eksperimen

Rasio	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
90:10	Naïve Bayes	0.905234657	0.898467433	0.9001919386	0.8993288591	0.9537434563
90:10	Random Forest	0.8944043321	0.8960784314	0.877159309	0.8865179437	0.9544431983
80:20	Naïve Bayes	0.8935018051	0.8845419847	0.8896353167	0.8870813397	0.9534156566
80:20	Random Forest	0.8840252708	0.8806983511	0.8714011516	0.8760250844	0.950579903
70:30	Naïve Bayes	0.8907942238	0.8865248227	0.8802816901	0.8833922261	0.9534392663
70:30	Random Forest	0.8805655836	0.8741168915	0.871318822	0.872715614	0.9480532976

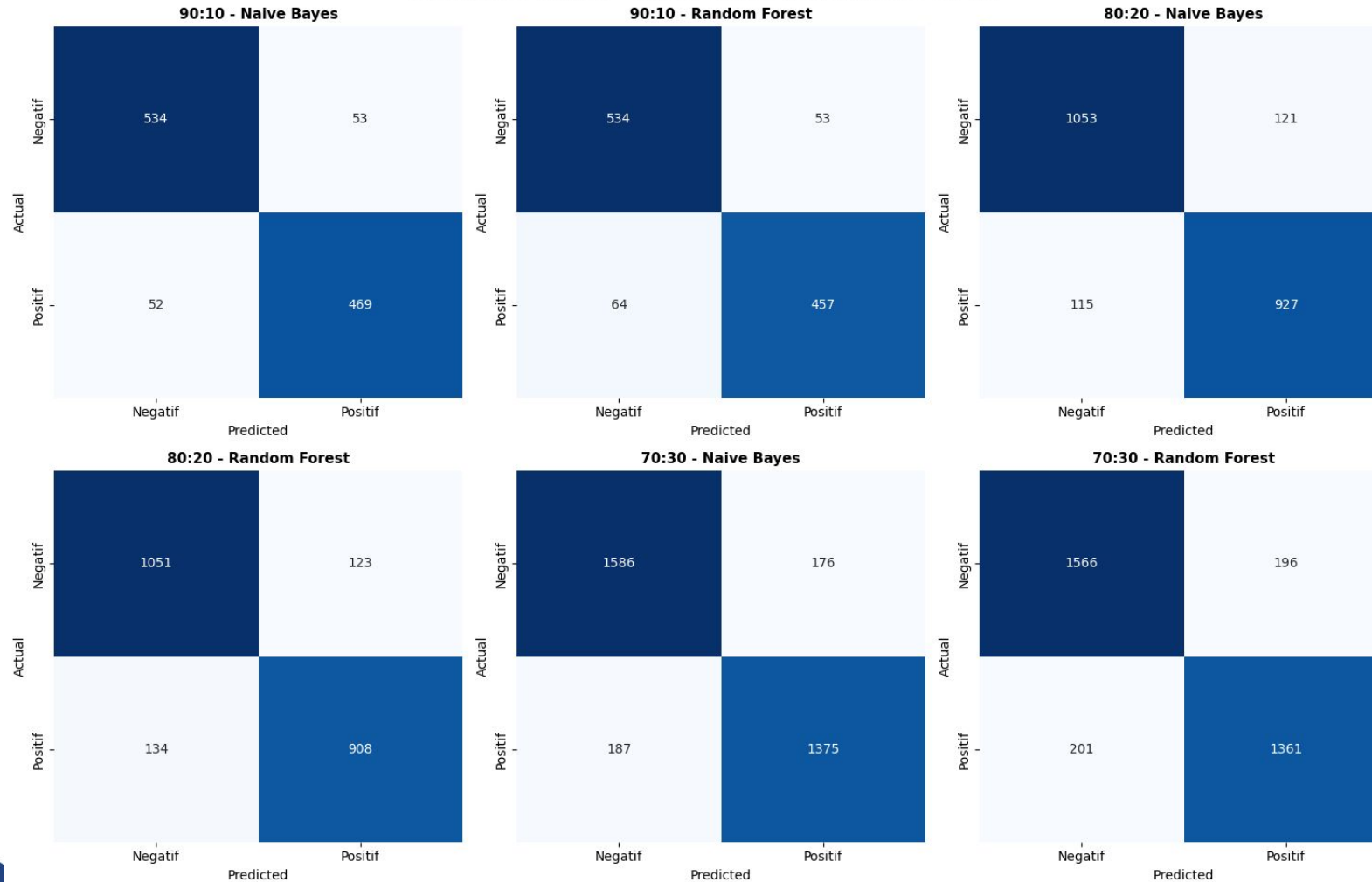
Visualisasi Perbandingan Model

Perbandingan Model Naive Bayes vs Random Forest



Visualisasi Perbandingan Model

Confusion Matrix - Naive Bayes vs Random Forest



Referensi

- V. Game *et al.*, "Video Game Market Summary," pp. 1–9, 2025.
- F. H. Hariatama, K. Nugraha Putra, G. Evelina Buji, and R. Alexandro, "Analisis Pengembangan Media Pembelajaran Berbasis Game Interaktif Edunomic Adventure Pada Pembelajaran Ekonomi," *Journal*, vol. 5, no. 2, pp. 148–159, 2024.
- A. Suryana, A. Irma Purnamasari, and I. Ali, "Mengoptimalkan Kepuasan Pengguna: Analisis Sentimen Review Aplikasi Grab Di Indonesia," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3396–3404, 2024, doi: 10.36040/jati.v8i3.9688.
- S. Mujilawati and M. Azarki Ubaydillah, "Analisis Sentimen Berbasis Aspek pada Ulasan Game Mobile Legends dengan Pendekatan Decision Tree untuk Evaluasi Pengalaman Pengguna," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 11, pp. 325–333, 2024, doi: 10.52436/1.jpti.465.
- M. S. Hamid, "Analisis Sentimen Tingkat Kepuasan Aplikasi WordPress Menggunakan Metode K-Nearest Neighbor dan Naive Bayes," vol. 15, no. 1, pp. 102–114, 2025.
- O. K. A. Azanda, G. Fdzilah Prayoganing, and W. Fredi, "Analisis Sentimen Ulasan Pengguna Aplikasi Mobile Legends Pada Google Playstore Menggunakan Naïve Bayes," *Teknol. Inf.*, vol. 4, no. 1, pp. 213–227, 2025, [Online]. Available: <https://journal.unilak.ac.id/index.php/Semaster/article/view/18526>

• D. Reinaldo and Y. F. Riti, "Analisis Sentimen Ulasan Aplikasi Genshin Impact di Play Store Menggunakan Random Forest," *Snestik*, 2024.

Referensi

- [9] C. G. Indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1131–1139, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] Fadhilah R, "Perbandingan algoritma Random Forest dan Naïve Bayes dalam analisis sentimen ulasan Spotify," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 78–90, 2024.
- [11] S. Nur Adhan, G. N. A. Wibawa, D. C. Arisona, I. Yahya, and R. Ruslan, "Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest," *AnoaTIK J. Teknol. Inf. dan Komput.*, vol. 2, no. 1, pp. 6–15, 2024, doi: 10.33772/anoatik.v2i1.32.
- [12] N. D. Kurniawan, P. R. Ferdian, and N. Hidayati, "Analisis Sentimen Algoritma Naïve Bayes, Support Vector Machine, dan Random Forest Pada Ulasan Aplikasi Ajaib," *J. Nas. Teknol. dan Sist. Inf.*, vol. 11, no. 1, pp. 87–97, 2025, doi: 10.25077/teknosi.v11i1.2025.87-97.
- [13] R. P. Setiawan, B. Irawan, and W. P. Prihartono, "Analisis Sentimen Ulasan Growtopia Di Google Play Store Menggunakan Naïve Bayes Classifier Untuk Identifikasi Kebutuhan Pengguna," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6415.
- [14] K. Kaeren and A. Andrianingsih, "Analisis Sentimen Aplikasi LinkAja di Google Play Store Menggunakan Algoritma Naïve Bayes dan Random Forest," *J. Ris. dan Apl. Mhs. Inform.*, vol. 6, no. 02, pp. 438–447, 2025, doi: 10.30998/jrami.v6i02.13821.
- [15] I. Deasiva, O. Nurdiawan, and M. F. Basysyar, "Model Sentimen Analisis Berdasarkan Ulasan Aplikasi Webtoon pada Goggle Play Store Ditingkatkan dengan Algoritma Random Forest," pp. 22–35.

Sekian,
Terimakasih

