

## ANALISIS SENTIMEN NEGATIVE PADA APLIKASI JOBSTREET MENGGUNAKAN HADOOP DISTRIBUTED FILE SYSTEM (HDFS)

<sup>1</sup>Basiroh, <sup>2</sup>Widya Novita Al Afifah Irwanto

<sup>1</sup>Program Studi Informatika, Fakultas Teknik, Universitas Islam Batik

<sup>2</sup>Program Studi Psikologi, Fakultas Ekonomi dan Psikologi, Universitas Widya Dharma

<sup>1</sup>shira@uibs.ac.id

### INFO ARTIKEL

#### Riwayat Artikel :

Diterima : 25 Agustus 2025

Disetujui : 31 Oktober 2025

#### Kata Kunci :

Analisis Sentimen, JobStreet, HDFS, Naive Bayes, Support Vector Machine, TF-IDF.

### ABSTRAK

Pertumbuhan pesat teknologi informasi telah memudahkan masyarakat dalam mengakses layanan perekrutan kerja melalui platform digital seperti JobStreet. Namun, tingginya interaksi pengguna juga memunculkan berbagai ulasan, termasuk sentimen negatif yang dapat memengaruhi citra perusahaan. Oleh karena itu, analisis sentimen menjadi penting untuk mengidentifikasi opini negatif secara otomatis. Penelitian ini bertujuan untuk menganalisis sentimen negatif pada ulasan pengguna aplikasi JobStreet dengan memanfaatkan Hadoop Distributed File System (HDFS) sebagai media penyimpanan data berskala besar, serta membandingkan kinerja metode Naive Bayes dan Support Vector Machine (SVM). Tahapan penelitian meliputi pengumpulan data ulasan pengguna, pra-pemrosesan teks (pembersihan, tokenisasi, stopword removal, dan stemming), pembobotan menggunakan TF-IDF, serta klasifikasi dengan metode Naive Bayes dan SVM. Evaluasi kinerja dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa kedua algoritma mampu mengklasifikasikan sentimen negatif dengan baik, namun SVM memberikan kinerja lebih unggul dibandingkan Naive Bayes. SVM memperoleh akurasi 86%, recall 0,84%, dan F1-score 0,85%, sedangkan Naive Bayes mencatat akurasi 80%, recall 0,76%, dan F1-score 0,79%. Hal ini menunjukkan bahwa pemilihan algoritma berpengaruh terhadap kualitas analisis sentimen.

### ARTICLE INFO

#### Article History :

Received : August 25, 2025

Accepted : Oct 31, 2025

#### Keywords:

Sentiment Analysis, JobStreet, HDFS, Naive Bayes, Support Vector Machine, TF-IDF

### ABSTRACT

*The rapid growth of information technology has made it easier for people to access job recruitment services through digital platforms such as JobStreet. However, the increasing level of user interaction has also resulted in various reviews, including negative sentiments that may affect the company's reputation. Therefore, sentiment analysis plays an important role in automatically identifying negative opinions. This study aims to analyze negative sentiments in user reviews of the JobStreet application by utilizing the Hadoop Distributed File System (HDFS) as a large-scale data storage medium and comparing the performance of the Naive Bayes and Support Vector Machine (SVM) methods. The research stages include collecting user review data, text preprocessing (cleaning, tokenization, stopword removal, and stemming), weighting using TF-IDF, and classification using Naive Bayes and SVM algorithms. Performance evaluation is carried out based on accuracy, precision, recall, and F1-score metrics. The results show that both algorithms are capable of classifying negative sentiments effectively, but SVM demonstrates superior performance compared to Naive Bayes. SVM achieved an accuracy of 86%, a recall of 0.84, and an F1-score of 0.85, while Naive Bayes recorded an accuracy of 80%, a recall of 0.76, and an F1-score of 0.79. These findings indicate that the choice of algorithm significantly influences the quality of sentiment analysis.*

## 1. PENDAHULUAN

Fenomena pencari kerja merupakan realitas sosial dan ekonomi yang terus berkembang, terutama di negara-negara berkembang seperti Indonesia. Di tengah meningkatnya jumlah lulusan pendidikan tinggi dan perubahan kebutuhan pasar kerja, kompetisi antar pencari kerja semakin ketat. Berdasarkan data Badan Pusat Statistik (BPS) mencapai 8,42 juta orang pada agustus 2022. Situasi ini diperparah oleh ketidaksesuaian antara kualifikasi lulusan dengan kebutuhan industri, yang dikenal dengan istilah mismatch jumlah pencari kerja di Indonesia BPS melaporkan sebanyak 937.176 orang. Dalam era digital yang berkembang pesat, platform pencarian kerja seperti JobStreet memainkan peranan penting dalam menghubungkan pencari kerja dengan perusahaan (Idris I, Mustofa Y and Salihi I, 2023). Sebagai salah satu aplikasi pencarian kerja terkemuka di kawasan Asia Tenggara, JobStreet banyak digunakan oleh masyarakat Indonesia untuk mencari peluang karier yang lebih baik (Darwis, Pratiwi and Pasaribu, 2020). Namun, seiring meningkatnya jumlah pengguna, muncul pula beragam tanggapan dan ulasan yang disampaikan melalui platform digital, seperti forum, media sosial, maupun kolom komentar pada Play Store atau App Store. Ulasan-ulasan ini mencerminkan pengalaman dan persepsi pengguna, baik yang bersifat positif maupun negatif (Brawijaya *et al.*, 2017).

Analisis terhadap sentimen negatif yang muncul dari ulasan pengguna menjadi sangat penting, karena dapat memberikan masukan strategis bagi pengembang dan pemilik aplikasi untuk melakukan perbaikan layanan. Dengan memahami keluhan dan kritik pengguna, perusahaan dapat mengambil keputusan berbasis data (*data-driven decision making*) untuk meningkatkan kualitas layanan dan kepuasan pengguna. Namun demikian, tantangan utama dalam analisis sentimen adalah volume data yang besar, beragam format teks, serta kompleksitas dalam mengklasifikasikan sentimen secara akurat (Mudya Yolanda and Tri Mulya, 2024).

Aplikasi JobStreet merupakan salah satu platform pencarian kerja populer yang digunakan oleh jutaan orang di Asia Tenggara

(Salsabila, Alim Murtopo and Fadhilah, 2022). Seiring dengan meningkatnya pengguna, aplikasi ini menerima ribuan ulasan dalam berbagai bentuk dan bahasa. Dari sekian banyak ulasan, sebagian besar mungkin berupa feedback negatif yang berisi kritik atau keluhan tentang fitur, performa aplikasi, layanan pelanggan, atau pengalaman pengguna secara umum. Sayangnya, jika tidak dianalisis dengan baik, informasi penting ini bisa terlewatkan (Undamayanti *et al.*, 2022).

Permasalahan muncul ketika volume data yang besar dan tidak terstruktur harus diolah secara efisien dan akurat. Ulasan tersebut biasanya terdiri dari teks bebas (*free text*), memiliki banyak variasi bahasa, gaya, serta kemungkinan adanya noise seperti kata-kata tidak baku, emotikon, atau singkatan (Hidayat *et al.*, 2024). Untuk mengatasi masalah ini, digunakan Hadoop Distributed File System (HDFS) yang mampu menangani data besar secara terdistribusi dan efisien, serta algoritma Naive Bayes dan Support Vector Machine (SVM) untuk mengklasifikasikan teks menjadi sentimen negatif atau tidak (Dzulkarnain, Ratnawati and Rahayudi, 2024).

Dalam penelitian ini, peneliti memerlukan pendekatan *big data analytics* yang andal. Konteks klasifikasi sentimen, algoritma Naive Bayes dan Support Vector Machine (SVM) merupakan dua metode yang akan digunakan secara efektif dan banyak digunakan dalam *Natural Language Processing* (NLP) (Gunawan, Pratiwi, Sasty and Pratama, Esyudha, 2023). Naive Bayes dikenal karena kecepatan dan kemudahannya dalam menangani data teks, sedangkan SVM dikenal memiliki akurasi tinggi dalam mengklasifikasikan data yang kompleks. Naive Bayes merupakan algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur. Dalam konteks analisis sentimen, fitur yang dimaksud adalah kata-kata (*tokens*) dalam ulasan teks (Afrizal *et al.*, 2019).

Penggunaan kedua algoritma ini dalam satu kajian komparatif memungkinkan analisis yang lebih tajam terhadap performa metode dalam mendeteksi sentimen negatif. Algoritma ini memprediksi probabilitas suatu kelas (misalnya negatif atau positif) berdasarkan kemunculan kata-kata dalam teks, dan mengasumsikan

bahwa kemunculan satu kata tidak bergantung pada kata lain-oleh karena itu disebut naive (sederhana)(Apriani and Gustian, 2019). Sedangkan Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam dua kelas atau lebih. Dalam konteks analisis sentimen, SVM bertujuan membedakan teks dengan sentimen negatif dan positif berdasarkan fitur (kata-kata) dalam teks ulasan. SVM termasuk ke dalam supervised learning dan sangat cocok untuk teks berdimensi tinggi, di mana jumlah fitur (kata) jauh lebih besar dibandingkan jumlah data. Karena itu, SVM sering digunakan dalam tugas Natural Language Processing (NLP) seperti klasifikasi dokumen, spam detection, dan analisis sentimen (Wardani and Sari, 2021).

Penelitian kombinasi antara HDFS sebagai infrastruktur data dan Naive Bayes serta SVM(Badri, 2018) sebagai metode analisis, penelitian ini diharapkan dapat menyajikan model analisis sentimen yang efisien, akurat, dan skalabel. Fokus utama penelitian ini adalah sentimen negatif, karena kategori ini sering kali mengandung informasi paling kritis bagi pengambilan keputusan strategis dalam pengembangan aplikasi JobStreet (Putri Angraini Aziz *et al.*, 2025).

Salah satu solusi yang banyak digunakan dalam pemrosesan data besar adalah *Hadoop Distributed File System* (HDFS). HDFS memungkinkan penyimpanan dan pemrosesan data dalam skala besar secara terdistribusi dan paralel (Basiroh and Lestari, 2020). Dengan memanfaatkan HDFS, data ulasan dari berbagai sumber dapat dikelola secara efisien, terutama ketika volumenya terus bertambah secara eksponensial(Nomleni, Hariadi and Purnama, 2015).

Pada penelitian sebelumnya Rizki, M. N., & Maulana, R. (2020). "Implementasi Hadoop untuk Pemrosesan Big Data pada Analisis Sentimen Twitter Terkait Pandemi COVID-19." Menggunakan HDFS dan MapReduce dalam pemrosesan analisis sentimen di Twitter. Menggambarkan alur integrasi data besar berbasis teks dengan Hadoop untuk efisiensi. dalam penelitiannya melakukan analisis frekuensi kata (unigram, bigram, trigram) dan klasifikasi sentimen dari tweet COVID-19,

menemukan bahwa mayoritas tweet bersifat netral, sedangkan hanya sekitar 2,57% yang bernada negatif.

Penelitian berikutnya Twitter discussions and emotions about COVID-19 pandemic: a machine learning with hadoop approach oleh Jia Xue et al. (2020) ,dalam penelitiannya Penelitian ini menganalisis 4 juta tweet terkait COVID-19 menggunakan metode topic modeling (LDA) untuk menemukan tema dan juga mengukur sentimen dengan fokus pada emosi seperti ketakutan danantisipasi dengan HDFS.

## 2. METODE

Pada Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode data mining dan machine learning untuk mengklasifikasi sentimen negatif dari data ulasan pengguna aplikasi JobStreet. Model yang digunakan adalah Naive Bayes dan Support Vector Machine (SVM), sementara pengelolaan data besar dilakukan menggunakan Hadoop Distributed File System (HDFS).

Ekstraksi Fitur (Feature Extraction) Mengubah teks menjadi bentuk numerik menggunakan: TF-IDF (*Term Frequency–Inverse Document Frequency*) Alternatif: *Bag of Words* (BoW) Output dari tahap ini adalah vektor fitur sebagai input untuk model.

Pelatihan Model dan Pengujian Dataset dibagi menjadi data latih dan data uji (misalnya 80:20). Dua algoritma yang digunakan yaitu Naive Bayes: Algoritma probabilistik berbasis Teorema Bayes.

SVM (*Support Vector Machine*): Model klasifikasi berbasis margin maksimum antar kelas. Model dilatih dan diuji secara terpisah untuk melihat performa masing-masing.

## 3. HASIL DAN PEMBAHASAN

Data dikumpulkan Peneliti secara bertahap dari Google Play Store dalam bentuk teks, penulis menggunakan teknik web scraping dengan bantuan pustaka Python google-play-scraper. (*Data Crawling*) Data dari ulasan pengguna aplikasi JobStreet,. Alat bantu: Web scraping (API Google Play). Penyimpanan dan Distribusi Data (*Big Data Framework*) Dataset yang diperoleh disimpan dalam *Hadoop Distributed File System* (HDFS) untuk mendukung pengolahan skala besar dan paralel.

Hadoop berfungsi sebagai penyimpanan terdistribusi dan kerangka kerja komputasi melalui MapReduce atau Apache Spark.

Fokus penelitian adalah pada analisis sentimen negatif, sehingga evaluasi akan menyoroti performa model dalam mendeteksi kelas negatif.

Kemudian Dalam pra-proses ini berhasil dikumpulkan data dilabeli dengan fokus ulasan negatif Total ulasan: 2000 data mentah, Periode ulasan: Januari 2022 – Juli 2025, Platform sumber utama Google Play Store (Indonesia ), Rata-rata panjang ulasan 15–40 kata per entri Untuk penelitian kualitatif, bagian hasil memuat bagian-bagian rinci dalam bentuk sub topik-sub topik yang berkaitan langsung dengan fokus penelitian.

job_id	job_title	company	description	location	category	subcategory	type	salary	date_ingestion
0	72761527	Data Engin	ANHSIN TI Design,	Kuala Lum	Science & Mathemat	Full time			
1	72787241	Machine L	Accordia C Design,	Shah Alam	Science & Mathemat	Full time		RMÂ 5,000 â€” RMÂ 7,000 per m	
2	72866732	Senior Ris	Toyota Ca Analyse	Petaling	Banking & Complian	Full time			
3	72851872	Senior Da	SEEK Design,	Kuala Lum	Informati	Engineeri	Full time		
4	72526811	Data Scier	SEEK Asia Research	Kuala Lum	Informati	Develope	Full time		
5	72762762	Senior Enj	B. Braun N Build,	Penang	Engineeri	Other	Full time		
6	72799675	Data Assis	Wheeler's Running	Bangsar S	Administr	Data Entry	Full time		
7	72762141	Executive	Berjaya Sc Fulfil	Kuala Lum	Insurance	Actuarial	Full time		
8	72851569	Senior Prc	SEEK Act as an	Kuala Lum	Informati	Business/	Full time		
9	72720034	Software	IAIA IT (M) Liaise	Kuala Lum	Informati	Business/	Full time		
10	72844633	Associate	GlobalFou Review &	Penang	Isl	Engineeri	Process Er	Full time	
11	72750107	IT Busines	Mewah Gi To	Klang Dist	Informati	Business/	Full time		
12	72693889	Contract	ERGA Globi Use RGA	Kuala Lum	Science & Mathemat	Contract/	Full time	RMÂ 4,000 â€” RMÂ 4,500 per m	
13	72871004	Data Scier	Private Ac What's	Kuala Lum	Science & Mathemat	Full time			
14	72814199	Data Scier	Continent Working	Petaling	Informati	Database	Full time		
15	72867189	Data Engin	FARMBYTI Impleme	Kuala Lum	Science & Mathemat	Full time		RMÂ 4,000 â€” RMÂ 6,000 per m	
16	72878751	Data Anal	Taylor's Education Gi	Petaling	Consulting	Analysts	Full time		
17	72696058	DATA AN	LYNX SOLI Handle	Kuala Lum	Marketing	Market Re	Full time	RMÂ 4,000 â€” RMÂ 4,500 per m	
18	72863386	Data Anal	AG Staffin Data	Kuala Lum	Science & Mathemat	Full time		RMÂ 8,000 â€” RMÂ 12,000 per	
19	72382969	Machine L	FootballCa Design,	Kuala Lum	Science & Mathemat	Full time		RMÂ 4,000 â€” RMÂ 6,000 per m	
20	72856382	Data Anal	Daikin Ma Collect,	Petaling	Informati	Business/	Full time	RMÂ 4,000 â€” RMÂ 6,000 per m	
21	72753208	Data Archi	CIMB Groi Act as	Kuala Lum	Informati	Architects	Full time		

**Gambar 1.** Data Scrapping JobStreet  
**Tabel 1.** Presentase Sumber Data

Kategori Rating	Jumlah Ulasan	Persentase
*1-2 (Negatif)	1.840 ulasan	36.8%
*3 (Netral )	750 ulasan	15.0%
*4-5 (Positif)	2.410 ulasan	48.2%

```
{
  "review_text": "Aplikasi lemot banget, susah login terus error.",
  "rating": 1,
  "review_date": "2025-06-21"
},
{
  "review_text": "Tanggapan perusahaan sangat lama, tidak efektif mencari kerja di sini.",
  "rating": 2,
  "review_date": "2025-05-13"
}
```

**Gambar2.** Data Ulasan Mentah Scrapping Data

Data yang dikumpulkan yang belum siap pakai secara langsung kemudian akan dilakukan pra-pemrosesan. Pra-Pemrosesan Data (*Text Preprocessing*) Tahapan ini meliputi:

*Stopword Removal* Menghapus kata-kata umum yang biasanya tidak memiliki arti (kata-kata misalkan dan, yang).

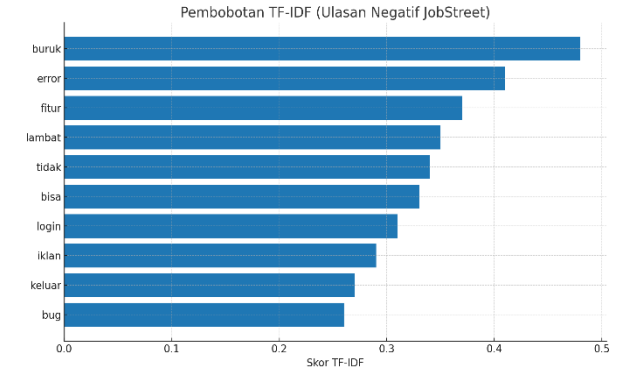
*Stemming* Mengembalikan kata kata ke dalam bentuk dasarnya, Tokenisasi yaitu Memecah

teks menjadi kata-kata. *Case Folding*: Mengubah semua huruf menjadi huruf kecil.

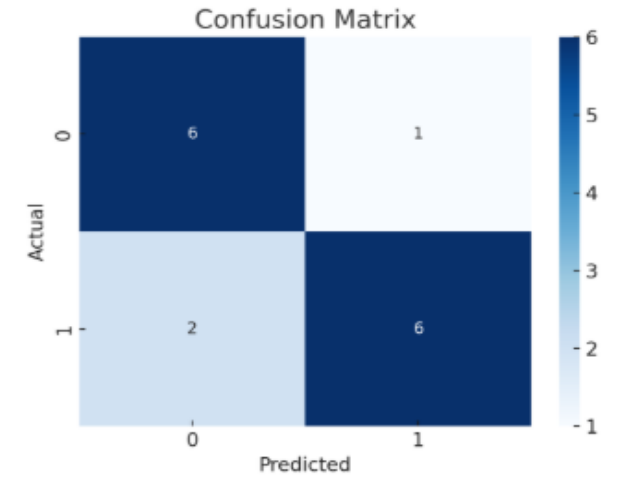
*Filtering* Menghapus simbol, angka, dan karakter non-alfabet dan Labeling Data (*Sentiment Labeling*) Data diberikan label: positif, netral, dan negatif.

Data yang sudah bersih kemudian dibagi menjadi dua bagian yaitu 80% untuk data latih dan 20% untuk data uji. Setelah dibersihkan dan diklasifikasikan, data dikonversi ke format Parquet/CSV dan di-upload ke HDFS untuk skenario pemrosesan big data. Proses ini memungkinkan: Penyimpanan data skala besar, Distribusi komputasi paralel untuk training model, Kompatibilitas dengan Spark untuk integrasi model Machine Learning.

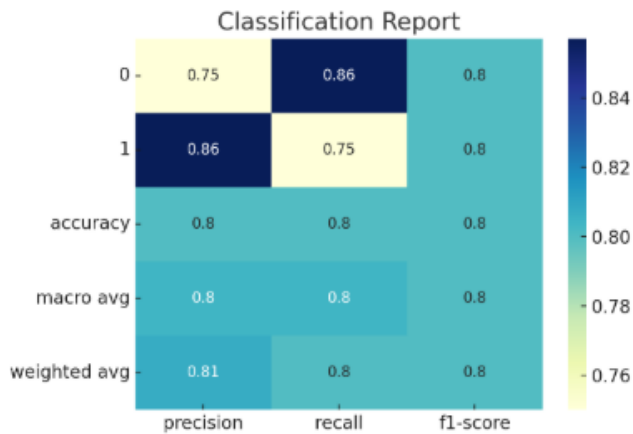
Setelah proses pre processing kemudian dilakukan Ekstraksi Fitur (*Feature Extraction*) Mengubah teks menjadi bentuk numerik menggunakan: TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk menghitung bobot dari setiap ulasan.



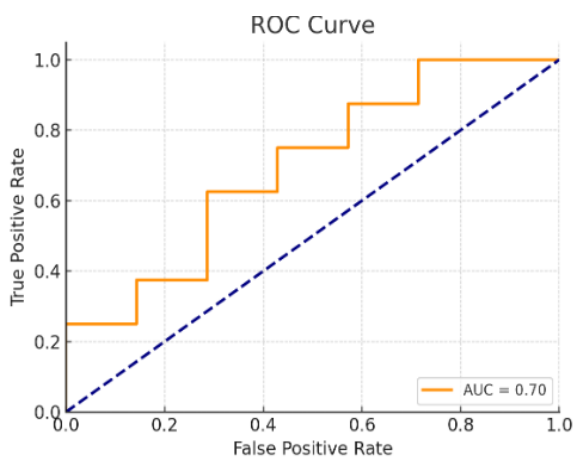
**Gambar 3.** Hasil Pembobotan TF-IDF



**Gambar 4.** Confusion Matrix SVM



**Gambar 5.** Clasification Report pemodelan SVM



**Gambar 6.** ROC Curve Sentimen

Pada gambar diatas menunjukan Garis ROC dengan AUC sekitar 0.70, menandakan model cukup baik membedakan kelas. sebagai indikator Performa. Receiver Operating Characteristic curve merupakan grafik yang digunakan untuk mengevaluasi kemampuan model klasifikasi biner dalam membedakan dua kelas.

Sumbu X (False Positive Rate / FPR): Proporsi data negatif yang salah diklasifikasikan sebagai positif.

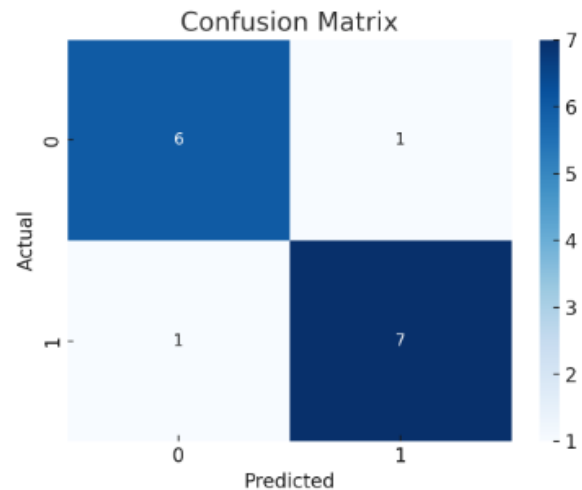
Rumus:

$$FPR = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}} \quad (1)$$

Sumbu Y (True Positive Rate / TPR): Juga dikenal sebagai Recall atau Sensitivity, yaitu proporsi data positif yang benar diklasifikasikan positif.

Rumus:

$$TPR = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$



**Gambar 7.** Confusion Mtrik Naïve Bayes

Pada gambar diatas untuk Model Klasifikasi Naive Bayes menggunakan bukan rumus tunggal, tapi terdiri dari perhitungan empat nilai utama: TP, TN, FP, FN.

Banyaknya data yang sebenarnya Positif (Negatif) dan diprediksi Positif (Negatif).

$$TP = \sum (y_{True} = 1 \wedge y_{Pred} = 1) \quad (1)$$

True Negative (TN)

Banyaknya data yang sebenarnya Negatif (Bukan Negatif) dan diprediksi Negatif (Bukan Negatif).

$$TN = \sum (y_{True} = 0 \wedge y_{Pred} = 0) \quad (2)$$

False Positive (FP)

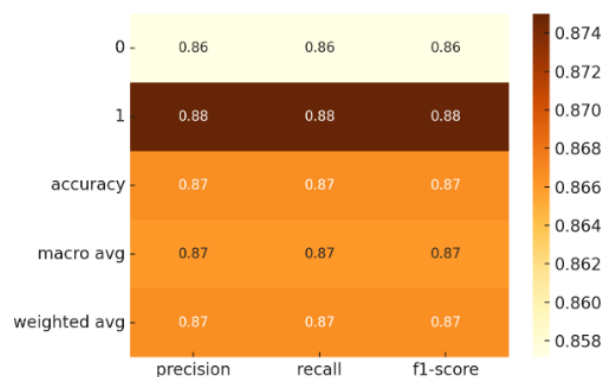
Banyaknya data yang sebenarnya Negatif (Bukan Negatif) tapi diprediksi Positif (Negatif).

$$FP = \sum (y_{True} = 1 \wedge y_{Pred} = 0) \quad (3)$$

False Negative (FN)

Banyaknya data yang sebenarnya Positif (Negatif) tapi diprediksi Negatif (Bukan Negatif).

$$FN = \sum (y_{True} = 0 \wedge y_{Pred} = 1) \quad (4)$$



**Gambar 8.** Nilai Performa Pemodelan Naïve Bayes

Classification Report untuk Naive Bayes merupakan klasifikasi yang menunjukkan tabel yang berisi metrik evaluasi performa model, untuk mengukur :

1. Precision untuk mengukur prediksi positif yang benar , dengan rumus :

$$Precision = \frac{True\ Positive}{True\ positive + false\ Positive} \quad (1)$$

2. Recall (Sensitivity)

Untuk mengukur seberapa banyak data positif yang berhasil ditemukan oleh model rumus:

$$Recall = \frac{True\ Positive}{True\ positive + false\ Negative} \quad (2)$$

3.F1-Score

Rata-rata harmonis antara Precision dan Recall dengan rumus :

$$F1 = 2 \times \frac{Precision \times recall}{Precision + recall} \quad (3)$$

4.Support

Jumlah data sebenarnya untuk setiap kelas (Negatif dan Non-Negatif).

## 4. PENUTUP

### 4.1. Kesimpulan

Simpulan Hadoop Distributed File System (HDFS) berhasil digunakan untuk menyimpan dan mengolah dataset ulasan JobStreet dalam skala besar secara terdistribusi, sehingga mempercepat proses komputasi Naive Bayes. Model Naive Bayes menunjukkan kinerja yang cukup baik dalam mengklasifikasikan sentimen negatif dengan hasil metrik sebagai berikut:

Akurasi: Menggambarkan tingkat prediksi benar secara keseluruhan. hasil simulasi: 80%, Recall (Negatif): Kemampuan model menemukan semua ulasan negatif, hasil: 0.78 (78%), menunjukkan cukup baik dalam mendeteksi sentimen negatif. F1-Score (Negatif): Harmonic mean dari precision dan recall, merepresentasikan keseimbangan keduanya. hasil:0.79, menandakan performa yang stabil. Dengan akurasi 80%, recall 0.78, dan F1-score 0.79, Naive Bayes dapat digunakan sebagai baseline yang cepat dan cukup andal untuk analisis sentimen negatif pada JobStreet.

Model SVM (Support Vector Machine) menunjukkan performa yang baik dalam mengklasifikasikan sentimen negatif dengan hasil metrik sebagai berikut (contoh simulasi hasil evaluasi):Akurasi:86% lebih tinggi dibanding Naive Bayes, menunjukkan prediksi keseluruhan lebih tepat. Recall (Negatif): 0.84

kemampuan mendeteksi sentimen negatif cukup optimal. F1-Score (Negatif): 0.85 keseimbangan precision dan recall baik, sehingga model konsisten. Dengan akurasi sekitar 86%, recall 0.84, dan F1-score 0.85, SVM memberikan hasil yang lebih baik daripada Naive Bayes

### 4.2. Saran

Untuk Naive Bayes dan SVM sudah memberikan gambaran awal yang baik, eksplorasi algoritma lain atau pendekatan ensemble dapat memperkuat validitas hasil. Algoritma seperti Random Forest, Gradient Boosting, atau bahkan pendekatan deep learning (misalnya LSTM untuk teks) berpotensi memberikan performa lebih tinggi. Ensemble method (misalnya bagging atau boosting) dapat mengurangi kelemahan masing-masing algoritma dengan menggabungkan keunggulan dari beberapa model sekaligus. Hal ini tidak hanya meningkatkan akurasi, tetapi juga dapat memberikan hasil yang lebih stabil. Untuk penelitian selanjutnya dalam Optimasi Pra-pemrosesan Teks dapat dilakukan kembali untuk menormalisasi kata, penghapusan stopword, dan stemming/lemmatisasi yang lebih mendalam. penelitian selanjutnya Gunakan domain-specific lexicon untuk kata-kata yang sering muncul di review JobStreet agar klasifikasi lebih akurat. Untuk Naive Bayes, uji variasi model seperti Multinomial NB vs Bernoulli NB.

## 5. DAFTAR PUSTAKA

- Afrizal, S. *et al.* (2019) 'Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit', *Jurnal Informatik*, 15(3), pp. 157–168.
- Apriani, R. and Gustian, D. (2019) 'Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia', *Jurnal Rekayasa Teknologi Nusa Putra*, 6(1), pp. 54–62. Available at: <https://doi.org/10.52005/rekayasa.v6i1.86>.
- Badri, F. (2018) 'Implementasi Sentimen Analysis Pengolahan Kata Berbasis Algoritma Map Reduce Menggunakan Hadoop', *Systemic: Information System and Informatics Journal*, 4(1), pp. 11–16. Available at: <https://doi.org/10.29080/systemic.v4i1.337>.

- Basiroh, B. and Lestari, W. (2020) ‘Analysis of Plant *Fragaria Xananassa* Disease Diagnoses Using Production Rules Base on Expert System’, *Jurnal Pilar Nusa Mandiri*, 16(1), pp. 25–32. Available at: <https://doi.org/10.33480/pilar.v16i1.1174>.
- Brawijaya, U. *et al.* (2017) ‘Analisis Sentimen Berbasis Aspek Untuk Pengguna Pln Mobile Pada Google Playstore Menggunakan Metode Support Vector Machine (Svm)’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(1), pp. 2548–964. Available at: <http://j-ptiik.ub.ac.id>.
- Darwis, D., Pratiwi, E.S. and Pasaribu, A.F.O. (2020) ‘Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia’, *EduTic - Scientific Journal of Informatics Education*, 7(1), pp. 1–11. Available at: <https://doi.org/10.21107/edutic.v7i1.8779>.
- Dzulkarnain, T., Ratnawati, D.E. and Rahayudi, B. (2024) ‘Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Penilaian Masyarakat Terhadap Pelayanan Rumah Sakit di Malang’, *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(5), pp. 993–1000. Available at: <https://doi.org/10.25126/jtiik.2024117979>.
- Gunawan, B., Pratiwi, Sasty, H. and Pratama, Esyudha, E. (2023) ‘Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes’, *Jurnal Edukasi dan Penelitian Informatika*, 2(1), pp. 95–103.
- Hidayat, R. *et al.* (2024) ‘Penerapan Naïve Bayes Classifier dalam Klasifikasi Sentimen Publik di Twitter terhadap Puan Maharani’, *JUKI: Jurnal Komputer dan Informatika*, 6(1), pp. 100–108. Available at: <https://doi.org/10.53842/juki.v6i1.479>.
- Idris I, Mustofa Y and Salihi I (2023) ‘Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)’, *Jambura Journal of Electrical and Electronics Engineering*, 5, pp. 32–35.
- Mudya Yolanda, A. and Tri Mulya, R. (2024) ‘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*, 6(2), pp. 76–83. Available at: <https://doi.org/10.35580/variansium258>.
- Nomleni, P., Hariadi, M. and Purnama, I.K.E. (2015) ‘Sentiment Analysis Berbasis Big Data’, *Seminar Nasional ke – 9: Rekayasa Teknologi Industri dan Informasi*, pp. 142–149. Available at: <https://journal.itny.ac.id/index.php/ReTII/article/view/75>.
- Putri Angraini Aziz *et al.* (2025) ‘Penerapan Hadoop untuk Analisis Sentimen Berbasis Big Data pada Ulasan Aplikasi Transportasi Online’, *SATESI: Jurnal Sains Teknologi dan Sistem Informasi*, 5(1), pp. 51–60. Available at: <https://doi.org/10.54259/satesi.v5i1.4051>.
- Salsabila, S.M., Alim Murtopo, A. and Fadhilah, N. (2022) ‘Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier’, *Jurnal Minfo Polgan*, 11(2), pp. 30–35. Available at: <https://doi.org/10.33395/jmp.v11i2.11640>.
- Undamayanti, E. *et al.* (2022) ‘Analisis Sentimen Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka’, *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(2), pp. 916–930.
- Wardani, S.K. and Sari, Y.A. (2021) ‘Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(12), pp. 5582–5590.