JSRCS Journal of Students' Research in Computer Science

e-ISSN: 2722 – 290X
Vol. 6 No. 1 (Mei 2025), Page: 113 – 124
https://doi.org/10.31599/96sfw544
Available Online at http://ejurnal.ubharajaya.ac.id/index.php/JSRCS

Analisis Sentimen Masyarakat Terhadap PHK di Indonesia Pada Twitter Menggunakan *Naïve Bayes* dan *Support Vector Machine (SVM)*

Abdu Malik AlHakim [1], Prima Dina Atika [1,*] , Herlawati Herlawati [1]

* Korespondensi: e-mail: prima.dina@dsn.ubharajaya.ac.id

[1] Informatika; Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya; Jl. Perjuangan No. 81, Marga Mulya, Bekasi Utara, Bekasi, Jawa Barat, 17143, Telp/Fax: (021) 88955882; e-mail: 202110715209@mhs.ubharajaya.ac.id, prima.dina@dsn.ubharajaya.ac.id herlawati@ubharajaya.ac.id

Submitted : **26 Maret 2025**
Revised : **17 April 2025**
Accepted : **13 Mei 2025**
Published : **30 Mei 2025**

Abstract

The phenomenon of layoffs in Indonesia has led to various public opinions, especially on social media. This research aims to analyze public sentiment on the layoff issue using data from Twitter, and compare the performance of two text classification algorithms, namely Naïve Bayes and Support Vector Machine. The Knowledge Discovery in Databases approach is used as the research framework, which includes the stages of data selection, text cleaning, transformation, classification, and evaluation. A total of 3,458 tweets were collected and processed through the pre-processing stage, then classified into positive and negative sentiments. Performance assessment was conducted with three scenarios of training and test data sharing: 80:20, 70:30, and 90:10. The results showed that Support Vector Machine gave the highest accuracy of 84.93% in the 90:10 scenario, compared to Naïve Bayes with 82.61% accuracy in the same scenario. Visualization through wordcloud was also used to strengthen the interpretation of dominant words in public opinion. The findings show that classification algorithms can be utilized to understand public perceptions of employment issues and support social data-based decision-making. This research can be further developed by expanding data coverage and evaluating more complex methods to improve classification accuracy.

Keywords*: Layoff, Naïve Bayes, Sentiment, Support Vector Machine (SVM), Twitter*

Abstrak

Fenomena pemutusan hubungan kerja (PHK) di Indonesia telah memunculkan berbagai opini publik, khususnya di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu PHK menggunakan data dari *Twitter*, serta membandingkan performa dua algoritma klasifikasi teks, yaitu *Naïve Bayes* dan *Support Vector Machine*. Pendekatan *Knowledge Discovery in Databases* digunakan sebagai kerangka kerja penelitian, yang mencakup tahapan pemilihan data, pembersihan teks, transformasi, klasifikasi, dan evaluasi. Sebanyak 3.458 tweet dikumpulkan dan diproses melalui tahapan pra-pemrosesan, kemudian diklasifikasikan ke dalam sentimen positif dan negatif. Penilaian performa dilakukan dengan tiga skenario pembagian data latih dan uji: 80:20, 70:30, dan 90:10. Hasil menunjukkan bahwa *Support Vector Machine* memberikan akurasi tertinggi sebesar 84,93% pada skenario 90:10, dibandingkan *Naïve Bayes* dengan akurasi 82,61% pada skenario yang sama. Visualisasi melalui *wordcloud* turut digunakan untuk memperkuat interpretasi terhadap kata-kata dominan dalam opini publik. Temuan ini menunjukkan bahwa algoritma klasifikasi dapat dimanfaatkan untuk memahami persepsi masyarakat

JSRCS status is accredited by the Directorate General of Research Strengthening and Development No. 225/E/KPT/2022 with Indonesian Scientific Index (SINTA) journal-level of S5, starting from Volume 1 (2) 2020 to Volume 6 (1) 2025

terhadap isu ketenagakerjaan dan mendukung pengambilan keputusan berbasis data sosial. Penelitian ini dapat dikembangkan lebih lanjut dengan memperluas cakupan data dan mengevaluasi metode yang lebih kompleks untuk meningkatkan akurasi klasifikasi.

Kata kunci: *Naïve Bayes*, Pemutusan Hubungan Kerja (PHK), Sentimen, *Support Vector Machine (SVM), Twitter*

1. Pendahuluan

Indonesia, sebagai negara berkembang, menghadapi tantangan besar dalam menjaga stabilitas ketenagakerjaan. Dalam beberapa tahun terakhir, tingkat pemutusan hubungan kerja (PHK) berfluktuasi secara drastis, terutama sejak pandemi COVID-19 memicu gelombang PHK massal di berbagai sektor industri. Berdasarkan data Kementerian Ketenagakerjaan RI, sebanyak 3,6 juta pekerja mengalami PHK pada tahun 2020, dan meskipun telah mengalami penurunan, prediksi pada tahun 2025 menunjukkan lebih dari 280.000 pekerja, terutama di industri tekstil, berpotensi terdampak (Widianto, 2024).

PHK tidak hanya berdampak secara finansial, tetapi juga menimbulkan tekanan psikologis dan keresahan sosial di tengah masyarakat (Retnaningsih, 2025). Seiring dengan perkembangan teknologi dan digitalisasi informasi, masyarakat Indonesia kini mulai aktif menyuarakan pendapatnya melalui media sosial, khususnya *twitter*. Dengan lebih dari 24 juta pengguna aktif di Indonesia (Duarte, 2025), *twitter* merupakan platform yang terbuka dan real-time untuk mengekspresikan opini publik, dengan fitur seperti tagar, topik yang sedang tren, dan pencarian kata kunci, membuat platform ini menjadi sumber data yang dapat digunakan untuk mengumpulkan data tentang pandangan dan opini masyarakat terhadap isu ketenagakerjaan.

Untuk memahami berbagai pandangan dan opini tersebut, maka analisis sentimen menjadi sangat penting. Analisis sentimen dapat mengklasifikasikan opini publik ke dalam kategori positif, negatif dengan menggunakan metode *machine learning*. Metode yang sering digunakan adalah *Naïve Bayes* dan *Support Vector Machine (SVM)*. *SVM* mencatat akurasi klasifikasi sebesar 84% dalam mengelompokkan sentimen terhadap PHK massal, sedangkan *Naïve Bayes* mencapai 74,1% (Saddam et al., 2023). Komparasi antara algoritma *Naïve Bayes*, Random Forest, dan SVM pada topik penghapusan tenaga honorer menunjukkan performa terbaik pada Random Forest, namun SVM dan *Naïve Bayes* masih menunjukkan akurasi di atas 64% (Miftahusalam et al., 2022). *Naïve Bayes* juga berkinerja baik dalam klasifikasi sentimen pada kebijakan Jaminan Hari Tua (JHT) dengan akurasi mencapai lebih dari 87% pada berbagai proporsi data latih dan data uji (Wijaya & Hakim, 2023). Komentar publik terhadap PHK massal oleh BT *Group* berhasil diklasifikasikan dengan akurasi 93,67% menggunakan *Naïve Bayes* (Charisma et al., 2024). Namun, sebagian besar penelitian hanya fokus pada satu metode atau jumlah data terbatas.

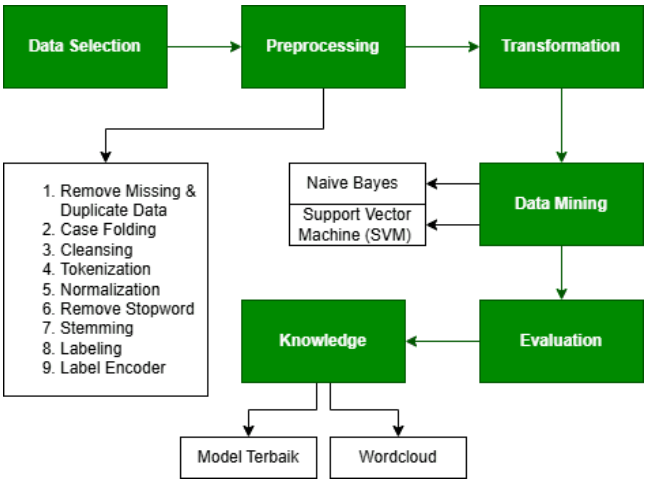
Penelitian ini ditujukan untuk mengetahui lebih dalam mengenai respon masyarakat terhadap isu PHK melalui klasifikasi sentimen berdasarkan data twitter, serta membandingkan akurasi algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* untuk menentukan metode yang paling baik dalam proses pemetaan opini publik. Hasilnya diharapkan mendukung pengembangan sistem analisis opini media sosial yang akurat, khususnya pada isu ketenagakerjaan di Indonesia yang masih jarang dianalisis secara komparatif.

2. Metode Penelitian

Penelitian ini mengacu pada tahapan *Knowledge Discovery in Databases (KDD)* sebagai kerangka utama dalam proses pengolahan dan analisis data. Struktur lengkap tahapan *KDD* yang digunakan dalam penelitian ini dijelaskan pada Gambar 1 yang merupakan kerangka penelitian.

2.1. *Knowledge Discovery in Databases (KDD)*

Knowledge Discovery in Databases (KDD) adalah proses yang terstruktur dan tersistematis dalam mengidentifikasi suatu pola dalam data yang valid, baru, berguna, dan dapat dimengerti. Proses ini bertujuan untuk mengubah bentuk data baku menjadi sebuah informasi yang bermakna sehingga dapat digunakan untuk pengambilan keputusan. *KDD* terdiri dari enam tahapan utama, yaitu *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, *Interpretation*/*Evaluation* dan *Knowledge* (Jansevskis & Osis, 2023).



Sumber: Hasil Penelitian (2025)

Gambar 1. Kerangka Penelitian

2.2. *Naïve Bayes*

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilistik yang memprediksi kelas data baru berdasarkan data pelatihan, dengan menghitung probabilitas bersyarat dari fitur yang ada (Herlawati et al., 2024). *Naïve Bayes* dirumuskan pada Rumus 1.

$$P(K \mid H) = \frac{P(H \mid K) \cdot P(K)}{P(H)} \quad (1)$$

Dengan keterangan K : hipotesis bahwa data termasuk dalam kelas tertentu; H : data atau fitur

yang diamati; P(K|H) : *posterior probability* (probabilitas kelas berdasarkan data); P(H|K) : *likelihood* (probabilitas data diberikan kelas); P(K) : *prior probability* dari kelas; P(H) = probabilitas H *(evidence)*.

2.3. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah algoritma *supervised learning* yang digunakan untuk tugas klasifikasi dengan cara mencari batas keputusan *(hyperplane)* yang secara optimal memisahkan data dari dua kelas. *SVM* bekerja dengan memproyeksikan data ke dalam ruang berdimensi tinggi, lalu menentukan *hyperplane* yang memaksimalkan margin antara dua kelas. Titik data yang berada paling dekat dengan *hyperplane* disebut sebagai support *vectors*, dan menjadi penentu utama dalam pembentukan batas klasifikasi (Ovirianti et al., 2022). *Hyperplane* pada *SVM* dinyatakan pada (Rumus 2).

$$WX + b = 0 \tag{2}$$

Dengan keterangan W : vektor normal terhadap hyperplane; X : vektor fitur dari data input; B : bias skalar yang menggeser *hyperplane*.

2.4. *Confusion Matrix*

Confusion Matrix adalah tabel yang digunakan untuk mengukur performa algoritma klasifikasi dengan membandingkan hasil prediksi model terhadap data aktual. Matriks ini menyajikan prediksi dalam empat nilai, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) (Krstinić et al., 2020).

3. Hasil dan Pembahasan

Bagian ini memaparkan hasil implementasi tiap tahapan dalam proses *KDD*, mulai dari pemilihan data hingga evaluasi model. Fokus utama berada pada analisis sentimen terhadap isu PHK menggunakan algoritma *Naïve Bayes* dan *SVM*.

3.1. Data Selection

Penelitian ini menganalisis sentimen dari 3.458 *tweet* yang dikumpulkan melalui proses *web scraping* di *Twitter* menggunakan bahasa pemrograman *Python* pada platform *Google Colaboratory*. Pengambilan data dilakukan dengan kata kunci “PHK”, berbahasa Indonesia, dalam rentang waktu 2020 hingga 2024.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3458 entries, 0 to 3457
Data columns (total 15 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   conversation_id_str       3458 non-null   float64
 1   created_at                3458 non-null   object 
 2   favorite_count            3458 non-null   int64  
 3   full_text                 3458 non-null   object 
 4   id_str                    3458 non-null   float64
 5   image_url                 426 non-null    object 
 6   in_reply_to_screen_name   2072 non-null   object 
 7   lang                      3458 non-null   object 
 8   location                  1971 non-null   object 
 9   quote_count               3458 non-null   int64  
 10  reply_count               3458 non-null   int64  
 11  retweet_count             3458 non-null   int64  
 12  tweet_url                 3458 non-null   object 
 13  user_id_str               3458 non-null   float64
 14  username                  3458 non-null   object 
dtypes: float64(3), int64(4), object(8)
```

Sumber: Hasil Penelitian (2025)

Gambar 2. Struktur Metadata *Dataset* Hasil *Scraping*

Dataset yang diperoleh memiliki 15 atribut untuk setiap entri *tweet*, namun hanya kolom *full_text* yang digunakan karena berisi opini pengguna dan menjadi fokus analisis sentimen. Struktur *metadata* lengkap dari hasil *scraping* terlihat pada Gambar 2, sedangkan hasil seleksi kolom *full_text* yang akan digunakan dalam proses berikutnya terlihat pada Gambar 3.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3458 entries, 0 to 3457
Data columns (total 1 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   full_text  3458 non-null   object
dtypes: object(1)
memory usage: 27.1+ KB
```

Sumber: Hasil Penelitian (2025)

Gambar 3. Struktur Metadata Hasil Seleksi Kolom *full_text*

3.2. Preprocessing

Tahap *preprocessing* bertujuan untuk membersihkan dan menyiapkan data mentah hasil *scraping* agar layak untuk dianalisis lebih lanjut. Proses diawali dengan penghapusan data duplikat dan entri kosong *(missing values)*, kemudian dilakukan pemrosesan pada teks secara bertahap melalui beberapa tahap seperti mengubah seluruh huruf kapital atau besar menjadi huruf kecil, *cleansing*, *tokenization, normalization, remove stopword, stemming.* Teks yang sudah bersih akan melalui proses *labeling* dan *label encoder.*

a. *Case Folding*

Tahap ini merupakan tahap pengubahan semua huruf kapital atau huruf besar pada teks menjadi huruf kecil. Gambar 4 menunjukan hasil penerapan *Case Folding* pada teks.

	full_text	case_folding
0	sampai saat ini kami karyawan korban Divestasi saham PT. FREEPORT INDONESIA yang di FOURLOGH dan...	sampai saat ini kami karyawan korban divestasi saham pt. freeport indonesia yang di fourlogh dan...
1	@Dagdigduq Kecuali klo ada ujian misal suami kena PHK itupun klo pasangannya sabar ya bisa dilal...	@dagdigduq kecuali klo ada ujian misal suami kena phk itupun klo pasangannya sabar ya bisa dilal...
2	1 kata buat temen lu yang Cepuin lu sampe lu di PHK dari kerjaan lu dan lu sendiri punya istri s...	1 kata buat temen lu yang cepuin lu sampe lu di phk dari kerjaan lu dan lu sendiri punya istri s...
3	Kasihan gw sama perusahaan yg bayar dia. Kalo ga salah manager kan yaa? Buang2 duit gaji gede or...	kasihan gw sama perusahaan yg bayar dia. kalo ga salah manager kan yaa? buang2 duit gaji gede or...
4	Omongan manusia pengangguran itu bebas ..krn tidak takut di PHK aplg dicopot dr Jabatanya...	omongan manusia pengangguran itu bebas ..krn tidak takut di phk aplg dicopot dr jabatanya...

Sumber: Hasil Penelitian (2025)

Gambar 4. Hasil Penerapan *Case Folding* Pada Teks

b. *Cleansing*

Teks dibersihkan dari elemen tidak penting seperti nama pengguna, tagar, URL, karakter emotikon, dan tanda baca. Tahap ini dilakukan untuk meminimalisir (*noise*) dalam analisis sentimen. Gambar 5 menunjukan penerapan hasil *cleansing*.

case_folding	cleansing
sampai saat ini kami karyawan korban divestasi saham pt. freeport indonesia yang di fourlogh dan...	sampai saat ini kami karyawan korban divestasi saham pt freeport indonesia yang di fourlogh dan ...
@dagdigduq kecuali klo ada ujian misal suami kena phk itupun klo pasangannya sabar ya bisa dilal...	kecuali klo ada ujian misal suami kena phk itupun klo pasangannya sabar ya bisa dilalui bareng b...
1 kata buat temen lu yang cepuin lu sampe lu di phk dari kerjaan lu dan lu sendiri punya istri s...	kata buat temen lu yang cepuin lu sampe lu di phk dari kerjaan lu dan lu sendiri punya istri ser...
kasihan gw sama perusahaan yg bayar dia. kalo ga salah manager kan yaa? buang2 duit gaji gede or...	kasihan gw sama perusahaan yg bayar dia kalo ga salah manager kan yaa buang duit gaji gede orang...
omongan manusia pengangguran itu bebas ..krn tidak takut di phk aplg dicopot dr jabatanya...	omongan manusia pengangguran itu bebas krn tidak takut di phk aplg dicopot dr jabatanya

Sumber: Hasil Penelitian (2025)

Gambar 5. Hasil Penerapan *Cleansing* Pada Teks

c. *Tokenization*

Tokenization merupakan pembagian kalimat menjadi beberapa bagian yang disebut token, Gambar 6 menunjukan penerapan hasil *tokenization* dari teks yang sudah dibersihkan.

cleansing	tokenization
sampai saat ini kami karyawan korban divestasi saham pt freeport indonesia yang di fourlogh dan ...	[sampai, saat, ini, kami, karyawan, korban, divestasi, saham, pt, freeport, indonesia, yang, di,...
kecuali klo ada ujian misal suami kena phk itupun klo pasangannya sabar ya bisa dilalui bareng b...	[kecuali, klo, ada, ujian, misal, suami, kena, phk, itupun, klo, pasangannya, sabar, ya, bisa, d...
kata buat temen lu yang cepuin lu sampe lu di phk dari kerjaan lu dan lu sendiri punya istri ser...	[kata, buat, temen, lu, yang, cepuin, lu, sampe, lu, di, phk, dari, kerjaan, lu, dan, lu, sendir...
kasihan gw sama perusahaan yg bayar dia kalo ga salah manager kan yaa buang duit gaji gede orang...	[kasihan, gw, sama, perusahaan, yg, bayar, dia, kalo, ga, salah, manager, kan, yaa, buang, duit,...
omongan manusia pengangguran itu bebas krn tidak takut di phk aplg dicopot dr jabatanya	[omongan, manusia, pengangguran, itu, bebas, krn, tidak, takut, di, phk, aplg, dicopot, dr, jaba...

Sumber: Hasil Penelitian (2025)

Gambar 6. Hasil Penerapan *Tokenization* Pada Teks

d. *Normalization*

Normalization yaitu proses mengubah kata-kata, bahasa gaul, atau singkatan yang tidak lazim menjadi bentuk baku dan konsisten, penerapan dari *normalization* terlihat pada Gambar 7.

tokenization	normalization
[sampai, saat, ini, kami, karyawan, korban, divestasi, saham, pt, freeport, indonesia, yang, di,...	[sampai, saat, ini, kami, karyawan, korban, divestasi, saham, pt, freeport, indonesia, yang, di,...
[kecuali, klo, ada, ujian, misal, suami, kena, phk, itupun, klo, pasangannya, sabar, ya, bisa, d...	[kecuali, jika, ada, ujian, misal, suami, kena, phk, itupun, jika, pasangannya, sabar, ya, bisa,...
[kata, buat, temen, lu, yang, cepuin, lu, sampe, lu, di, phk, dari, kerjaan, lu, dan, lu, sendir...	[kata, buat, temen, lu, yang, cepuin, lu, sampai, lu, di, phk, dari, kerjaan, lu, dan, lu, sendi...
[kasihan, gw, sama, perusahaan, yg, bayar, dia, kalo, ga, salah, manager, kan, yaa, buang, duit,...	[kasihan, gw, sama, perusahaan, yang, bayar, dia, kalau, tidak, salah, manager, kan, yaa, buang,...
[omongan, manusia, pengangguran, itu, bebas, krn, tidak, takut, di, phk, aplg, dicopot, dr, jaba...	[omongan, manusia, pengangguran, itu, bebas, karena, tidak, takut, di, phk, aplg, dicopot, dr, j...

Sumber: Hasil Penelitian (2025)

Gambar 7. Hasil Penerapan *Normalization* Pada Teks

e. *Remove Stopword*

Merupakan tahapan penghapusan kata-kata yang sangat sering digunakan yang memiliki makna rendah atau tidak signifikan dibandingkan kata lain dalam teks, Gambar 8 menunjukkan penerapan penghapusan kata yang sering digunakan pada teks.

normalization	remove_stopword
[sampai, saat, ini, kami, karyawan, korban, divestasi, saham, pt, freeport, indonesia, yang, di,...	[karyawan, korban, divestasi, saham, pt, freeport, indonesia, fourlogh, terselesaikan]
[kecuali, jika, ada, ujian, misal, suami, kena, phk, itupun, jika, pasangannya, sabar, ya, bisa,...	[kecuali, ujian, suami, kena, itupun, pasangannya, sabar, bareng, berusaha, jgn, biarin, merembet]
[kata, buat, temen, lu, yang, cepuin, lu, sampai, lu, di, phk, dari, kerjaan, lu, dan, lu, sendi...	[temen, cepuin, kerjaan, istri, anak, berumur, pandemi, bener, butuh, tambahan, ekonomi]
[kasihan, gw, sama, perusahaan, yang, bayar, dia, kalau, tidak, salah, manager, kan, yaa, buang,...	[kasihan, perusahaan, bayar, salah, manager, yaa, buang, duit, gaji, gede, bego, model, gini, di...
[omongan, manusia, pengangguran, itu, bebas, karena, tidak, takut, di, phk, aplg, dicopot, dr, j...	[omongan, manusia, pengangguran, bebas, takut, aplg, dicopot, dr, jabatanya]

Sumber: Hasil Penelitian (2025)

Gambar 8. Hasil Penerapan *Remove Stopword* Pada Teks

f. *Stemming*

Stemming merupakan proses mengubah bentuk kata menjadi kata dasar, yang merupakan tahap menemukan kata dasar dari setiap kata yang tersaring, Gambar 9 menunjukkan penerapan *stemming* pada teks.

	remove_stopword	stemming
0	[karyawan, korban, divestasi, saham, pt, freeport, indonesia, fourlogh, terselesaikan]	[karyawan, korban, divestasi, saham, pt, freeport, indonesia, fourlogh, selesai]
1	[kecuali, ujian, suami, kena, itupun, pasangannya, sabar, bareng, berusaha, jgn, biarin, merembet]	[kecuali, uji, suami, kena, itu, pasang, sabar, bareng, usaha, jgn, biarin, rembet]
2	[temen, cepuin, kerjaan, istri, anak, berumur, pandemi, bener, butuh, tambahan, ekonomi]	[temen, cepuin, kerja, istri, anak, umur, pandemi, bener, butuh, tambah, ekonomi]
3	[kasihan, perusahaan, bayar, salah, manager, yaa, buang, duit, gaji, gede, bego, model, gini, di...	[kasihan, usaha, bayar, salah, manager, yaa, buang, duit, gaji, gede, bego, model, gin, dipeliha...
4	[omongan, manusia, pengangguran, bebas, takut, aplg, dicopot, dr, jabatanya]	[omong, manusia, anggur, bebas, takut, aplg, copot, dr, jabatanya]

Sumber: Hasil Penelitian (2025)

Gambar 9. Hasil Penerapan *Stemming* Pada Teks

g. *Labeling*

Teks yang telah melalui proses pembersihan kemudian diberi label sentimen secara otomatis berdasarkan kamus sentimen *InSet Lexicon*, yang mengklasifikasikan kata-kata ke dalam kategori positif dan negatif pada Gambar 10.

	tokens	positive_words	negative_words	positive_score	negative_score	total_score	label
0	['maju', 'th', 'hitung', 'rizal', 'hitung', 'th', 'sektor', 'macet', 'tambah', 'tenaga', 'kerja', 'daya', 'beli', 'lambat']	hitung(1.0), hitung(1.0), tambah(3.0), tenaga(3.0), kerja(2.0), beli(2.0), lambat(1.0)	beli(-3.0)	+13	-3	+10	positive
1	['kue', 'sayur', 'ala', 'expegawai', 'hotel', 'mati', 'sumber', 'rezeki', 'sumber', 'rezeki', 'hempas', 'simak', 'video', 'yuk']	hotel(1.0), mati(4.0), rezeki(3.0), rezeki(3.0), simak(3.0), yuk(4.0)	mati(-5.0), hempas(-4.0), simak(-5.0)	+18	-14	+4	positive
2	['perintah', 'bikin', 'bijak', 'upah', 'rp', 'juta', 'usaha', 'sanggup', 'massal', 'kritisi']	bijak(3.0), upah(1.0), juta(2.0), usaha(1.0), sanggup(2.0)	bijak(-4.0), upah(-3.0), usaha(-4.0)	+9	-11	-2	negative
3	['capek', 'ham', 'ciptakeja', 'polemik', 'korupsi', 'bansos', 'ekonomi', 'covid', 'parah']	capek(4.0), polemik(3.0)	capek(-4.0), korupsi(-4.0), parah(-5.0)	+7	-13	-6	negative

Sumber: Hasil Penelitian (2025)

Gambar 10. Penerapan *Labeling* Pada Teks

Labeling dilakukan dengan menjumlahkan skor polaritas kata dalam setiap *tweet*, lalu menentukan klasifikasinya sebagai sentimen positif atau negatif. Hasil labeling otomatis pada teks terlihat pada Gambar 11.

	cleansing	sentiment
0	sampai saat ini kami karyawan korban divestasi saham pt freeport indonesia yang di fourlogh dan ...	positive
1	kecuali klo ada ujian misal suami kena phk itupun klo pasangannya sabar ya bisa dilalui bareng b...	positive
2	kata buat temen lu yang cepuin lu sampe lu di phk dari kerjaan lu dan lu sendiri punya istri ser...	negative
3	kasihan gw sama perusahaan yg bayar dia kalo ga salah manager kan yaa buang duit gaji gede orang...	negative
4	omongan manusia pengangguran itu bebas krn tidak takut di phk aplg dicopot dr jabatanya	negative

Sumber: Hasil Penelitian (2025)

Gambar 11. Hasil *Labeling* Otomatis Pada Teks

Label tersebut kemudian dikonversi ke bentuk numerik menggunakan teknik *label encoder* untuk kebutuhan pemodelan, terlihat pada Gambar 12 hasil penerapan dari *label encoder*.

	full_text	sentiment	sentiment_encoded
0	sampai saat ini kami karyawan korban Divestasi saham PT. FREEPORT INDONESIA yang di FOURLOGH dan...	positive	1
1	@Dagdigduq Kecuali klo ada ujian misal suami kena PHK itupun klo pasangannya sabar ya bisa dilal...	positive	1
2	1 kata buat temen lu yang Cepuin lu sampe lu di PHK dari kerjaan lu dan lu sendiri punya istri s...	negative	0
3	Kasihan gw sama perusahaan yg bayar dia. Kalo ga salah manager kan yaa? Buang2 duit gaji gede or...	negative	0
4	Omongan manusia pengangguran itu bebas ..krn tidak takut di PHK aplg dicopot dr Jabatanya...	negative	0

Sumber: Hasil Penelitian (2025)

Gambar 12. Hasil penerapan *Label Encoder* Pada Teks

3.3. *Transformation*

Pada tahap ini dilakukan splitting data (pembagian data) dan pembobotan *TF-IDF* unutk kebutuhan pemodelan algoritma.

a. *Splitting* data

Data yang telah melalui tahap *preprocessing* dibagi menjadi dua bagian, yaitu data latih dan data uji. proses ini dilakukan agar model dapat dilatih menggunakan data *training* (latih) dan kemudian dievaluasi dengan data *testing* (uji). Pembagian data ini dilakukan dengan 3 skenario yaitu, data latih : data uji, yaitu 80%:20%, 70%:30%, 90% :10% untuk melihat perbandingan mana yang menghasilkan model paling optimal.

b. *TF-IDF*

Setelah proses pembagian data, tahap berikutnya adalah transformasi data teks menjadi format numerik menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* yang memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam sebuah dokumen dibandingkan dengan seluruh kumpulan dokumen.

```
  (0, 1244)     0.5904818384279484
  (0, 1225)     0.8070509268235495
  (1, 1501)     0.5582180109035844
  (1, 524)      0.37621433911870195
  (1, 118)      0.3433837616644216
  (1, 1371)     0.27900133145516204
  (1, 1180)     0.5925379925504882
  (2, 989)      0.35322213306080963
  (2, 1291)     0.27398970923845445
```

Sumber: Hasil Penelitian (2025)

Gambar 13. Hasil Penerapan Pembobotan *TF-IDF* Pada Teks

3.4. *Data Mining*

Pada tahap *data mining*, pembangunan model klasifikasi dilakukan untuk menganalisis sentimen dari data teks yang telah ditransformasikan. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naïve Bayes* dan *Support Vector Machine (SVM).*

a. *Naive Bayes*

Setelah data dibagi menjadi data latih dan data uji, model *Naïve Bayes* dibangun menggunakan *library Multinomial Naïve Bayes*. Gambar 14 menunjukkan penerapan *Naïve Bayes*.

```python
from sklearn.naive_bayes import MultinomialNB

# 1. Inisialisasi dan training model
nb_model = MultinomialNB(
    alpha=1.0,
    fit_prior=True
)
nb_model.fit(X_train_tfidf, y_train)

# 2. Prediksi dan evaluasi
y_pred_nb = nb_model.predict(X_test_tfidf)
y_pred_proba_nb = nb_model.predict_proba(X_test_tfidf)[:, 1]
```

Sumber: Hasil Penelitian (2025)

Gambar 14. Penerapan *NB* Dengan *Library Multinomial Naïve Bayes* Pada Teks

b. *Support Vector Machine (SVM)*

Setelah data dibagi menjadi data latih dan data uji, model *Support Vector Machine (SVM)* dibangun menggunakan *kernel linear*. Gambar 15 menunjukkan penerapan *SVM.*

```python
# 1. Inisialisasi training model SVM
svm_model = LinearSVC(
    C=1.0,
    class_weight='balanced',
    random_state=42,
    max_iter=1000
)
# 2. training model SVM
svm_model.fit(X_train_tfidf, y_train)

# Prediksi dan evaluasi
y_pred_svm = svm_model.predict(X_test_tfidf)
```

Sumber: Hasil Penelitian (2025)

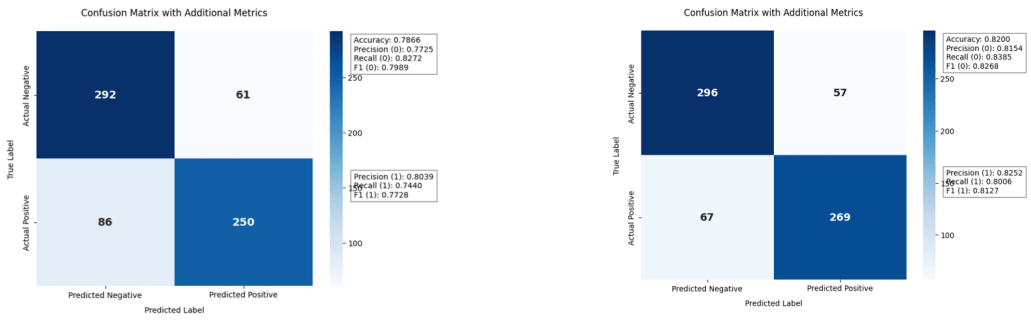
Gambar 15. Penerapan *SVM* Dengan *Kernel Linear* Pada Teks

3.5. *Evaluation*

Tahap evaluasi dilakukan untuk mengukur dan membandingkan performa model klasifikasi sentimen yang dibangun menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*. Evaluasi dilakukan berdasarkan hasil prediksi terhadap data uji pada tiga skenario pembagian data, yaitu 80:20, 70:30, dan 90:10.

Penilaian dilakukan menggunakan Confusion Matrix, yang menghasilkan empat nilai utama: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan nilai-nilai tersebut, dihitung metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi ini menjadi dasar dalam menentukan algoritma yang paling optimal untuk analisis sentimen tweet terkait PHK di Indonesia.

a. *Confusion Matrix* pada skenario 80:20

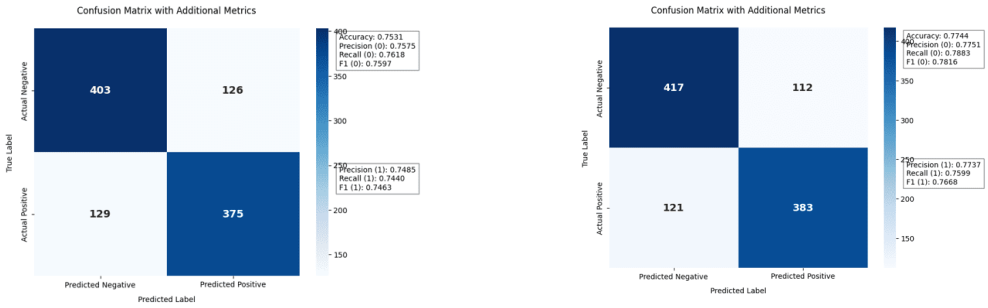


(a) (b)

Sumber: Hasil Penelitian (2025)

Gambar 16. (a) *Confusion Matrix Naive Bayes* Dengan Skenario 80:20, (B) *Confusion Matrix SVM* Dengan Skenario 80:20

b. *Confusion Matrix* pada skenario 70:30

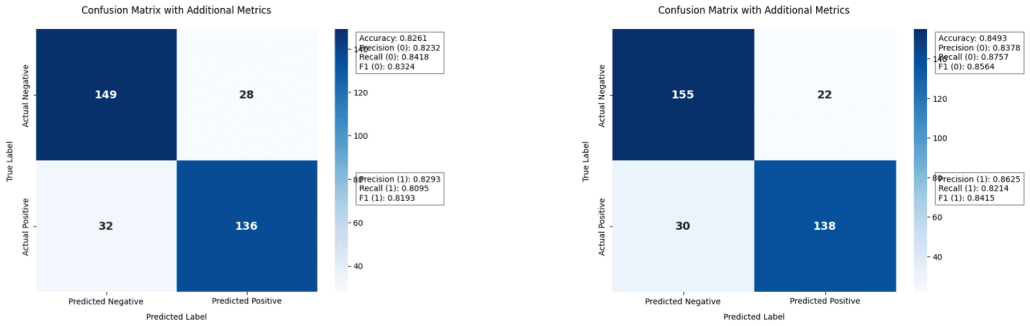


(a) (b)

Sumber: Hasil Penelitian (2025)

Gambar 17. (a) *Confusion Matrix Naive Bayes* Dengan Skenario 70:30, (B) *Confusion Matrix SVM* Dengan Skenario 70:30

c. *Confusion Matrix* pada skenario 90:10



(a) (b)

Sumber: Hasil Penelitian (2025)

Gambar 18. (a) *Confusion Matrix Naive Bayes* Dengan Skenario 90:10, (B) *Confusion Matrix SVM* Dengan Skenario 90:10

3.6. *Knowledge*

Hasil dari evaluasi model klasifikasi menunjukkan bahwa algoritma *Support Vector Machine (SVM)* memberikan akurasi yang lebih tinggi dibandingkan *naïve bayes* pada semua skenario pembagian data. Tabel 1 menunjukan perbandingan akurasi kedua algoritma pada skenario pembagian data 80:20, 70:30, dan 90:10.

Tabel 1. Tabel Perbandingan Akurasi Algoritma Pada Tiga Skenario Pembagian Data

Algoritma	Skenario		
	80:20	70:30	90:10
Naïve Bayes	78,66%	75,31%	82,61%
Support Vector Machine (SVM)	82,00%	77,44%	84,93%

Sumber: Hasil Penelitian (2025)

Berdasarkan hasil tersebut, *SVM* pada skenario 90:10 dipilih sebagai model paling optimal karena menghasilkan akurasi tertinggi sebesar 84,93%. Hasil ini menunjukkan bahwa

pembagian data dengan proporsi lebih besar untuk pelatihan dapat meningkatkan kinerja model secara signifikan. Visualisasi tambahan ditampilkan pada dalam bentuk *wordcloud* dari algoritma *naïve bayes* terlihat pada Gambar 19, dan *wordcloud* dari algoritma *SVM* ditunjukan pada Gambar 20.

Sumber: Hasil Penelitian (2025)

Gambar 19. *Wordcloud* sentimen pada *naïve bayes*

Sumber: Hasil Penelitian (2025)

Gambar 20. *Wordcloud* sentimen pada *SVM*

4. Kesimpulan

Penelitian ini menerapkan tahapan *Knowledge Discovery in Databases (KDD)* untuk menganalisis sentimen masyarakat terhadap isu Pemutusan Hubungan Kerja (PHK) di Indonesia menggunakan data *Twitter*. Hasil evaluasi menunjukkan bahwa algoritma *Support Vector Machine (SVM)* menghasilkan performa klasifikasi terbaik dibandingkan *Naïve Bayes*, dengan akurasi tertinggi sebesar 84,93% pada skenario pembagian data 90:10. Skenario ini dinilai paling optimal dalam konteks dataset yang digunakan karena memberikan porsi data pelatihan yang lebih besar, sehingga memungkinkan model mempelajari pola sentimen secara lebih efektif. Visualisasi *wordcloud* turut memberikan gambaran distribusi kata-kata dominan dalam opini publik. Menunjukkan potensi penggunaan SVM dalam analisis opini publik berbasis media sosial, khususnya pada isu ketenagakerjaan. Pengembangan di kemudian hari dapat dilakukan dengan memperluas data, mempertimbangkan sentimen netral, dan mengeksplorasi metode lain seperti deep learning untuk meningkatkan akurasi dan generalisasi model.

Daftar Pustaka

Charisma, M. O., Hamzah, M. F., Erwin, M., Nurbaiti, I., & Kurniawan, F. (2024). Klasifikasi Sentimen Terhadap Kebijakan PHK 55 Ribu Karyawan oleh BT Group menggunakan Algoritma Klasifikasi Naive Bayes. *Journal of Computer and Information Systems Ampera*, *5*(2), 2775–2496. https://journal-computing.org/index.php/journal-cisa/index

Duarte, F. (2025). *X (Formerly Twitter) User Age, Gender, & Demographic Stats (2025)*. https://explodingtopics.com/blog/x-user-stats

Herlawati, Srisulistiowati, D. B., Agustin, S. C., Syafina, P. H., Rachmatin, N., & Setiawati, S. (2024). Metode Naïve Bayes dan Support Vector Machine untuk Mengolah Sentimen Ulasan dan Komentar di Platform Digital. *Journa of Students' Research in Computer Science*, *5*(2), 197–212. https://doi.org/https://doi.org/10.31599/dby15h32

Jansevskis, M., & Osis, K. (2023). Knowledge Discovery Frameworks and Characteristics. *Baltic Journal of Modern Computing*, *11*(4), 686–702. https://doi.org/10.22364/bjmc.2023.11.4.08

Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label Classifier Performance Evaluation with Confusion Matrix. *Computer Science & Information Technology (CS & IT)*, 01–14. https://doi.org/10.5121/csit.2020.100801

Miftahusalam, A., Nuraini, A. F., Khoirunisa, A. A., & Pratiwi, H. (2022). Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer. *Seminar Nasional Official Statistics*, *2022*(1), 563–572.

Ovirianti, N. H., Zarlis, M., & Mawengkang, H. (2022). Support Vector Machine Using A Classification Algorithm. *SinkrOn*, *7*(3), 2103–2107. https://doi.org/10.33395/sinkron.v7i3.11597

Retnaningsih, H. (2025). *Info Singkat: Pemutusan Hubungan Kerja (PHK), No. XVII/1/I/P3DI/Januari/2025*. https://berkas.dpr.go.id/pusaka/files/info_singkat/Info Singkat-XVII-1-I-P3DI-Januari-2025-175.pdf

Saddam, M. A., Kurniawan D, E., & Indra, I. (2023). Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, *8*(3), 226–233. https://doi.org/10.30591/jpit.v8i3.4884

Widianto, S. (2024, Desember 27). Badai PHK 2025, Ketika Kebijakan Tak Ramah Industri Mengancam Nasib Pekerja. *Pikiran Rakyat*. https://www.pikiran-rakyat.com/ekonomi/pr-018913867/badai-phk-2025-ketika-kebijakan-tak-ramah-industri-mengancam-nasib-pekerja

Wijaya, H., & Hakim, L. (2023). Analisis Sentimen Kebijakan Jaminan Hari Tua (Jht) Pada Twitter Menggunakan Naïve Bayes. *Jurnal Algoritma, Logika dan Komputasi*, *6*(1), 509–518. https://doi.org/10.30813/j-alu.v6i1.3552