ISSN: 2477-5126
e-ISSN: 2548-9356

Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan *Support Vector Machine*

Mohd. Amiruddin Saddam[1*)], Erno Kurniawan D[2], Indra[3]
[1,2,3] Fakultas Teknologi Informasi, Universitas Budi Luhur , Jakarta
[1,2] Jl. Ciledug Raya, Jakarta Selatan, 12260, Indonesia
email: [1]2111600827@student.budiluhur.ac.id, [2]2111600413@student.budiluhur.ac.id, [3]indra@budiluhur.ac.id

***Abstract* – Termination of employment (PHK) on a large scale has a very significant impact on society and the economy. Mass layoffs have led to an increase in the number of unemployed people. Many people who have lost their jobs without a stable source of income struggle to find new jobs. This exacerbated the situation on the labor market and increased the number of unemployed people. Mass layoffs can also reduce economic activity and consumption. The sentiment analysis carried out aims to determine public sentiment regarding the phenomenon of mass layoffs that are currently happening in Indonesia based on positive and negative categories. In this study, the classification method used is the SVM method, which is one of the supervised learning methods in machine learning and also uses Nave Bayes as a comparison method. After classification, the next stage is the testing process using the K-fold cross-validation method. From the various sentiments obtained from Twitter data, it can be concluded that there are around 108 positive sentiments and 333 negative sentiments related to mass layoffs, while the results obtained from the test results using the SVM method show an accuracy of up to 84% while using the Nave Bayes method shows an accuracy of up to 74.1 percent.**

***Abstrak* – Pemutusan Hubungan Kerja (PHK) secara besar-besaran berdampak sangat signifikan terhadap masyarakat dan perekonomian. PHK massal telah menyebabkan peningkatan jumlah pengangguran. Banyak orang yang kehilangan pekerjaan tanpa sumber pendapatan yang stabil berjuang untuk mendapatkan pekerjaan baru. Ini memperburuk situasi pasar tenaga kerja dan meningkatkan jumlah pengangguran. PHK massal juga dapat mengurangi kegiatan ekonomi dan konsumsi. Analisis sentimen yang dilakukan, bertujuan untuk mengetahui sentimen masyarakat mengenai fenomena PHK massal yang banyak terjadi saat ini di Indonesia berdasarkan kategori positif dan negatif. Pada penelitian ini, metode klasifikasi yang digunakan adalah metode SVM yang merupakan salah satu metode pembelajaran terawasi dalam pembelajaran mesin dan juga menggunakan Naïve Bayes sebagai metode pembanding. Setelah klasifikasi, tahap selanjutnya adalah proses pengujian dengan menggunakan metode *K-Fold Cross Validation*. Dari berbagai sentimen yang diperoleh dari data Twitter, dapat disimpulkan terdapat sekitar 108 sentimen positif dan 333 sentimen negatif terkait PHK Massal, sedangkan hasil yang diperoleh dari hasil pengujian menggunakan metode SVM menunjukkan akurasi mencapai 84% sedangkan jika menggunakan metode Naïve Bayes menunjukkan akurasi hingga 74,1%.**

***Kata Kunci* – Klasifikasi, PHK Massal, Analisis Sentimen, SVM, Naive Bayes, Text Mining, Twitter.**

*) **penulis korespondensi**: Mohd. Amiruddin Saddam
Email: 2111600827@student.budiluhur.ac.id

I. PENDAHULUAN

Pemutusan Hubungan Kerja (PHK) secara besar-besaran berdampak sangat signifikan terhadap masyarakat dan perekonomian. PHK massal telah menyebabkan peningkatan jumlah pengangguran. Banyak orang yang kehilangan pekerjaan tanpa sumber pendapatan yang stabil berjuang untuk mendapatkan pekerjaan baru. Ini memperburuk situasi pasar tenaga kerja dan meningkatkan jumlah pengangguran. PHK massal juga dapat mengurangi kegiatan ekonomi dan konsumsi. Banyak orang yang kehilangan pekerjaan mungkin tidak mampu membeli barang dan jasa, mengurangi permintaan dan mengurangi aktivitas ekonomi. Selain itu, PHK massal dapat berdampak negatif bagi perusahaan yang membuatnya. Reputasi perusahaan menjadi jelek dan mereka dapat kehilangan talenta dan sumber daya yang berharga. Secara keseluruhan, PHK massal memiliki efek negatif yang meluas dan memengaruhi banyak aspek masyarakat dan ekonomi. Oleh karena itu, perusahaan harus berhati-hati dalam menerapkan PHK dan mempertimbangkan alternatif lain seperti pengurangan jam kerja atau pengurangan upah untuk mengurangi dampak negatifnya.

Berkat perkembangan teknologi, banyak orang yang mengungkapkan pendapat dan ekspresinya melalui teknologi seperti jejaring sosial, salah satu media sosial yang banyak digunakan oleh masyarakat untuk berkomunikasi Pendapat dan ekspresi mereka adalah jejaring sosial twitter, dimana di twitter orang dapat menuliskan pendapat mereka. komentar dalam bentuk pesan. banyak orang bisa melihatnya [1].

Analisis sentimen juga dapat dimasukkan dalam bidang NLP (*Natural Language Programming*), yaitu proses pengolahan data dengan tujuan untuk mengidentifikasi dan memahami masalah yang dihadapi bisnis dan perasaan mereka sendiri atas topik mereka. Analisis sentimen, juga dikenal sebagai "*polling*", adalah bidang ilmu yang menganalisis pendapat, perasaan, peringkat, moral, dan sentimen orang terhadap produk, layanan, organisasi, atau produk, lembaga, dan individu. Untuk melakukan analisis sentimen, membutuhkan data yang berisi opini terhadap suatu isu. Ada banyak cara untuk mendapatkan data tersebut, salah satunya melalui jejaring sosial. Berbeda dengan sebelumnya, opini publik hanya dapat diketahui melalui jajak pendapat, organisasi opini, surat kabar, stasiun radio dan televisi [2].

Analisis sentimen dilakukan dengan tujuan untuk mengetahui sentimen publik terhadap PHK massal yang terjadi di Indonesia berdasarkan kategori positif dan negatif. Setelah konsultasi masyarakat, akan dilakukan klasifikasi

untuk mendapatkan hasil sebagai dokumen penilaian bagi pemangku kepentingan.

Peneliti mengambil beberapa langkah, termasuk *pre-processing*, untuk mengumpulkan komentar publik. Setelah mengumpulkan komentar masyarakat atas PHK massal pada tahap *pre-processing*, dilakukan klasifikasi berdasarkan opini publik agar pemangku kepentingan dapat menggunakan sebagai dokumen penilaian bagi pihak terkait. Pada tahap klasifikasi, penulis memilih untuk menggunakan metode SVM karena merupakan metode pembelajaran yang diawasi pada *machine learning* dan juga SVM memiliki nilai akurasi yang cukup tinggi. Sedangkan penulis memilih untuk menggunakan Naïve Bayes sebagai metode pembanding karena NBC memiliki perhitungan yang sederhana dan tidak kompleks. Selain itu NBC juga banyak digunakan oleh beberapa ilmuwan data dalam melakukan klasifikasi. *K-Fold Cross Validation* akan digunakan untuk mengevaluasi model.

II. PENELITIAN YANG TERKAIT

Analisis sentimen pernah dibahas oleh Tuhuteru, dkk pada penelitiannya yang berjudul "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode *Support Vector Machine* dan *Naive Bayes Classifier*" [3]. Pada penelitian ini dilakukan analisa sentiment masyarakat pulau ambon mengenai pelayanan PT. PLN menggunakan *Naïve Bayes Classifier*.

Pada penelitian yang dilakukan Satrya, dkk [4], mengusulkan model ansambel tingkat kedua yang disebut model ansambel multi-level untuk analisis sentimen. Ini menggabungkan *Multinomial Naive Bayes*, pembelajaran ansambel berbasis *Boosting*, dan pembelajaran ansambel berbasis *Bagging*. Data diproses menggunakan *CountVectorizer* untuk mengubah teks menjadi angka dan *over sampling* untuk menangani data yang tidak seimbang. Hasil percobaan menunjukkan bahwa model yang diusulkan memiliki hasil terbaik dalam hal akurasi, presisi, dan skor F1 untuk rangkaian pengujian dibandingkan dengan model individual.

Analisis sentimen yang diperoleh dalam penelitian yang dilakukan oleh Fitri, dkk [5] menunjukkan bahwa pengguna Twitter di Indonesia lebih banyak memberikan komentar netral. Pada penelitian ini diperoleh akurasi sebesar 86,43% dari data pengujian menggunakan Algoritma Naïve Bayes pada tools RapidMiner, dimana akurasi tersebut lebih tinggi dibandingkan dengan algoritma lainnya yaitu *Decision Tree* dan *Random Forest* yaitu sebesar 82,91%.

Hasil dari penelitian yang dilakukan oleh Zulfikar, dkk [6] difokuskan pada model atau pola yang dihasilkan oleh Algoritma *Multinomial Naive Bayes*. Hasil klasifikasi kicauan pengguna media sosial terhadap kebijakan *new normal* diperoleh hasil yang baik dengan nilai akurasi sebesar 90,25%. Setelah mengklasifikasikan kicauan pengguna media sosial terkait kebijakan normal baru, hasilnya menunjukkan lebih dari 70% setuju dan mendukung kebijakan normal baru.

Dalam sebuah studi yang dilakukan oleh Fridom et al.[7], sentimen positif mendominasi analisis sentimen pada tweet terkait obesitas dengan menambang teks sentimen negatif dan sentimen netral. Nilai akurasi dengan algoritma *Naïve Bayes Classifier* berada pada kategori "*Excellent Classifier*". Artinya algoritma *Naive Bayes Classifier* berhasil memprediksi kategori sentimen pada pencarian ini.

Syahputra, dkk [8] mengatakan dalam penelitiannya bahwa, Setelah membandingkan algoritma SVM dan Naive Bayes menggunakan model evaluasi *cross-validation*, peneliti juga menemukan bahwa algoritma ini memiliki nilai akurasi yang lebih tinggi dibandingkan Naïve Bayes. Tingkat akurasi algoritma Naïve Bayes adalah 85%, sedangkan algoritma SVM mencapai akurasi yang lebih tinggi dengan tingkat akurasi 86%. Walaupun akurasi ditentukan dengan pengujian teknik dekomposisi 8020, algoritma Naïve Bayes memiliki tingkat akurasi 80% lebih tinggi dibandingkan dengan algoritma SVM yang memiliki tingkat akurasi 79%.

Pada penelitian yang dilakukan Gata & Bayhaqy [9] mengenai analisis sentimen terkait Islamofobia. Pendapat dan opini publik yang diungkapkan dalam jumlah besar di Twitter setidaknya dapat memberikan analisis keseluruhan tentang sentimen anti-Muslim terhadap Muslim di seluruh dunia pada hari berita kerusuhan Serangan terhadap Islamisasi Gereja Kristus di Selandia Baru diumumkan. laporan. Para peneliti menggunakan algoritme Naïve Bayes dan SVM yang dikombinasikan dengan SMOTE untuk menyeimbangkan data berlabel VADER guna meningkatkan hasil pengujian algoritme.

Saddam, dkk [10] pada penelitiannya, menggunakan analisis sentimen untuk mengetahui pendapat masyarakat tentang penanggulangan banjir di ibu kota Jakarta berdasarkan kategori positif, dan negatif. Setelah terkumpul sentimen masyarakat terkait penanganan banjir, dilakukan klasifikasi berdasarkan sentimen tersebut agar pihak terkait dapat menggunakannya untuk bahan evaluasi penanganan banjir.

Pada penelitiannya, Alita, dkk [11] menganalisis penggunaan emoji pada postingan forum bahasa Belanda berdasarkan kosakata dan memberikan peningkatan akurasi hasil klasifikasi. Sedangkan pada penelitian ini digunakan 3 karakteristik untuk mendeteksi ironi yaitu unigram, negatifitas dan jumlah seruan. Hasil riset menunjukkan bahwa fitur negatif dan jumlah panggilan review dapat meningkatkan akurasi sebesar 6% dibandingkan dengan hanya menggunakan fitur unigram. Pengujian dan analisis dilakukan untuk menentukan pengaruh emoji dan sarkasme pada analisis sentimen, sebagaimana diklasifikasi oleh Naïve Bayes, SVM, dan *random forest*.

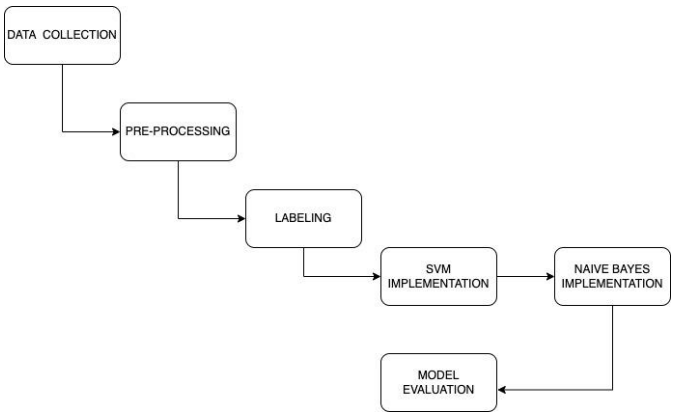
Fatihah, dkk [12] menganalisis sentimen publik terkait kebijakan *work from the home* yang diterapkan pemerintah selama pandemi COVID-19 menggunakan *Support Vector Machine* dengan *Randomized Search* untuk mendapatkan akurasi yang maksimal. Selain itu, penelitian ini juga membandingkan metode *pre-processing* yang tepat digunakan untuk model yang dibangun.

Analisis sentimen dilakukan oleh Kartika et al [13] . dapat memprediksi peristiwa bencana dalam kelas yang telah ditentukan. Algoritma *Support Vector Machine* (SVM) dijalankan untuk melakukan klasifikasi data menjadi tiga (3) kategori: saksi, bukan saksi dan tidak diketahui. Total data yang akan diproses untuk ketiga kelas label tersebut adalah 3000. Algoritma SVM memiliki dua metode yaitu *one-to-one* (OVO) untuk label dua lapis dan *one-to-one* (OVA) untuk lebih dari dua label kelas. *Multilayer* SVM dengan metode OVA digunakan dalam penelitian ini. Dibandingkan dengan penelitian dan metode sebelumnya, metode OVA dengan RBF, memiliki nilai akurasi klasifikasi sebesar 87,03%.

Penelitian sebelumnya telah membuktikan bahwa metode SVM memiliki akurasi yang tinggi dalam melakukan klasifikasi dan berhasil diterapkan dalam proses analisa sentiment dalam berbagai topik. Berdasarkan hal tersebut, penulis menggabungkan metode SVM dan *Naïve Bayes Classifier* sebagai metode pembanding yang kemudian akan diuji menggunakan metode *K-Fold Cross Validation*.

III. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode kuantitatif. Sampel yang digunakan adalah data Twitter dengan kata kunci "PHK massal" atau "phkmassal" atau "phk massal". Langkah-langkah yang dilakukan dalam penelitian ini adalah seperti yang digambarkan pada Gbr 1.



Gbr 1 Tahapan Penelitian

A. *Data Collection*

Pengumpulan data Twitter adalah proses melakukan ekstraksi data pengguna dan tweet dari Twitter berdasarkan kata kunci tertentu. Penambangan data dilakukan melalui penggunaan *Application Programming Integration* (API)[14]. Pengumpulan dilakukan menggunakan pustaka SNScrape yang disediakan dalam bahasa pemrograman Python. Hanya tweet dalam bahasa Indonesia yang dikumpulkan dalam 5 hari, sejak 17 November 2022 sampai dengan 23 November 2022, sebanyak 441 tweet dengan kata kunci "PHK massal" atau "PHK" atau "Phk Massal".

B. *Pre-processing*

Pre-processing adalah serangkaian langkah yang dilakukan untuk membersihkan data, melakukan integrasi data, mengubah data, dan melakukan reduksi data. Langkah ini sangat penting karena jika Anda menganalisis data yang belum ditangani dengan hati-hati, dapat menghasilkan informasi yang salah. Jika terdapat banyak informasi yang tidak relevan atau banyak *noise* data, maka kualitas data akan menurun. Untuk menghasilkan data berkualitas tinggi, penting untuk memahami dan menerapkan dengan benar metode *pre-processing* yang digunakan untuk setiap studi kasus [12]. Tahapan *pre-processing* yang dilakukan adalah *data cleansing*, *case folding*, *tokenizing*, *stopword removal*, *normalization*, dan *stemming*. Tahap *pre-processing* dilakukan dengan Python menggunakan library NLTK yang disediakan.

C. *Data Cleansing*

Langkah pertama dalam *pre-processing* adalah pembersihan data. Data yang diperoleh dari Twitter dipilih selama pembersihan data berdasarkan relevansinya dengan topik yang dipilih yaitu Sentimen Masyarakat terhadap fenomena PHK Massal.

D. *Case Folding*

Langkah ini merupakan proses modifikasi teks yang memiliki ketidakteraturan penggunaan huruf pada teks komentar sehingga menghasilkan teks komentar yang tidak konsisten. Prosedur *case folding* ini berfungsi untuk mengubah huruf dalam teks komentar yang dibersihkan ke bentuk standar yaitu semua huruf kecil [15].

E. *Tokenizing*

Langkah selanjutnya adalah tokenisasi. Selama proses ini, karakter, URL, emoji, tagar, dan karakter lainnya dihapus. Proses ini merupakan proses penting dalam analisis sentimen untuk meminimalisir kesalahan penulisan. Setelah menghapus karakter, token membagi kalimat menjadi beberapa kata.

F. *Stopword Removal*

Langkah keempat adalah menghapus stopword. Dalam proses ini, kata-kata enkripsi kemudian disaring. Kata-kata yang tidak memiliki arti akan dihilangkan atau dihapus. Hal ini akan sangat membantu proses analisis sentimen.

G. *Normalization*

Langkah selanjutnya adalah normalisasi kata. Kata-kata yang telah dipilih, dibuat menjadi sandi, dan dihilangkan secara tipografis dari stopwords akan dinormalisasi ulang untuk membuat sama kata-kata dengan makna yang sama tetapi kata-kata berbeda, atau memperbaiki slang menjadi kata-kata standar.

H. *Stemming*

Langkah terakhir adalah stemming. Proses perubahan bentuk kata menjadi kata dasar dengan mencari kata dasar dari setiap kata yang disaring untuk setiap kata imbuhan yang akan diubah menjadi kata dasar, tujuan dari kata dasar adalah untuk memaksimalkan dan mengoptimalkan proses pengolahan kata [16]. Pada tahap stemming, penelitian ini menggunakan pustaka sastra Indonesia dengan bahasa python

I. *Labeling*

Algoritma pembelajaran mesin (ML) adalah rangkaian dua langkah dari data proses, pelatihan, dan pengujian. Oleh karena itu, kita perlu menandai beberapa tweet dengan komentar sebenarnya untuk digunakan selama fase pelatihan [17]. Proses pembuatan label dilakukan untuk menentukan jenis sentimen yang didapat, apakah positif atau negatif.

J. *SVM Implementation*

Setelah proses pembuatan label, langkah selanjutnya adalah melakukan implementasi model *support vector machine*. *Support Vector Machine* (SVM) terbukti bermanfaat dalam pembelajaran mesin, khususnya untuk melakukan klasifikasi data multikelas. Ada dua pendekatan untuk SVM multilayer. Pertama, ini memproses semua data langsung menjadi formula yang dioptimalkan. Bagian kedua

menjelaskan multilayer dalam rantai SVM biner. Ide di balik *binary* SVM adalah membangun *multiclass classifier* dari *binary* menggunakan teknik *one-to-one* [11].

SVM adalah algoritma yang hanya dapat diterapkan untuk mengklasifikasikan data linier. Untuk mengklasifikasikan data non-linear, penyesuaian harus dilakukan dengan menambahkan fungsi kernel. Kernel adalah fungsi yang mendefinisikan transformasi pemetaan dari beberapa data ruang *input* menjadi ruang vektor baru dengan dimensi yang lebih tinggi sehingga kedua kelas pada akhirnya dapat dipisahkan oleh garis *hyperplane* [18]. Dengan menggunakan fungsi kernel ini, fungsi keluaran dari algoritma SVM dapat ditulis pada persamaan (3)

$$f(x_d) = \sum_{i=1, x_i \epsilon sv}^{ns} \alpha_i y_i K(x_i, x_d) + b \quad (1)$$

K. Naïve Bayes Implementation

Naïve Bayes Classifier adalah teknologi *pre-processing* yang dapat melakukan klasifikasi fitur yang menambahkan banyak skalabilitas, akurasi, dan efisiensi dalam klasifikasi teks. Naïve Bayes adalah pengklasifikasi probabilistik yang menambahkan kombinasi nilai frekuensi dari kumpulan data. Proses klasifikasi menghitung nilai probabilitas untuk setiap label kelas *input*. Algoritma ini didasarkan pada teorema Bayes dan mengasumsikan bahwa setiap atribut independen atau tidak ditugaskan ke variabel kelas nilai apa pun dan tidak terpengaruh oleh atribut lainnya. Nilai atribut dapat berupa data statis atau kelas. Teorema Bayes bekerja dengan mencari nilai probabilitas posterior (P(X I Y)) berdasarkan probabilitas sebelumnya (P(Y I X)). (P(X))
adalah probabilitas dari nilai X dan pembuktiannya adalah (P(Y)). X dan Y adalah kejadian. Persamaan (1) dihasilkan dari nilai yang ditentukan [19].

$$P(X|Y) = \frac{P(Y|X).P(X)}{P(Y)} \quad (2)$$

Algoritma Naïve Bayes memiliki banyak instruksi untuk menganalisis sampel, sehingga nilai perhitungan probabilitas posterior dapat disesuaikan dengan persamaan (2).

$$P(A|B_1 \dots B_N) = P(A).\frac{P(B_1 \dots B_N|A)}{P(B_1 \dots B_N)} \quad (3)$$

L. Model Evaluation

Langkah selanjutnya adalah menguji model menggunakan teknik *K-fold cross-validation*. Evaluasi kinerja model dilakukan berdasarkan *error metrics* untuk mendapatkan akurasi model. Untuk mengevaluasi kinerja model, digunakan metode *K-Fold Cross-Validation*. Jumlah putaran yang digunakan untuk evaluasi dan pelatihan adalah 10 data. Data tersebut kemudian dibentuk untuk mendapatkan nilai yang benar. Dalam evaluasi ini, *confusion matrix* digunakan untuk mendapatkan akurasi, presisi, *recall*, dan *f1-score*. *Confusion matrix* dapat digunakan untuk menganalisis kinerja model pada setiap *fold*. Terdapat empat komponen utama dalam *confusion matrix* yaitu: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Komponen tersebut digunakan dalam menghitung *confusion matrix*.

1. Akurasi memberikan gambaran umum tentang kinerja model secara keseluruhan. Akurasi mengukur sejauh mana model dapat membuat prediksi yang benar dalam kategori yang benar. Ini dihitung dengan membagi jumlah prediksi yang benar dengan jumlah total prediksi. Persamaan (4) dapat digunakan untuk menghitung akurasi pada model.

$$\frac{TP + TN}{Total} \quad (4)$$

2. Presisi dihitung dengan membagi jumlah prediksi positif yang benar dengan jumlah total prediksi positif yang dibuat. Presisi memberikan informasi tentang seberapa baik model dapat secara akurat mengidentifikasi kelas positif. Persamaan (5) dapat digunakan untuk menghitung nilai presisi.

$$\frac{TP}{FP + TP} \quad (5)$$

3. *Recall* (juga dikenal sebagai sensitivitas atau *true positive rate*) mengukur sejauh mana model dapat mengidentifikasi dengan benar semua contoh sebenarnya dari kelas positif. *Recall* dihitung dengan membagi jumlah prediksi positif yang benar dengan jumlah *instance* yang benar-benar termasuk dalam kelas positif. *Recall* memberikan informasi tentang seberapa baik model menemukan semua contoh yang benar dari kelas positif. Persamaan (6) dapat digunakan untuk mendapatkan nilai dari *recall*.

$$\frac{TP}{FN + TP} \quad (6)$$

4. *F1-Score* merupakan nilai rata-rata dari penjumlahan presisi dan *recall*. *F1-Score* digunakan untuk menyeimbangkan nilai presisi dan juga *recall*. Persamaan (7) dapat digunakan untuk menghitung *F1-Score*.

$$2 * \frac{precision * recall}{precision + recall} \quad (7)$$

IV. HASIL DAN PEMBAHASAN

Sebelum melakukan analisis, beberapa langkah perlu dilakukan mulai dari eksplorasi data dan *pre-processing* hingga pengujian. Pada pembahasan sebelumnya telah dijelaskan bahwa ada 6 langkah *pre-processing* yang dilakukan pada penelitian ini, langkah *pre-processing* yang pertama adalah pembersihan data. Pada tahap pembersihan data, peneliti meninjau data untuk relevansi subjek dari setiap *dataset*, seperti yang ditunjukkan pada Tabel I. Verifikasi didasarkan pada relevansi yang dilihat dari *dataset*, menghubungkan data dengan topik yang dibahas, yaitu redundansi kolektif.

Tabel I
Hasil *Data Cleansing* Sesuai Relevansi

Periode	Total Data	*Relevant*	*Not Relevant*
17-19 Nov 2022	258	250	8
20-21 Nov 2022	105	92	13
22-23 Nov 2022	114	99	15

Selain melakukan pembersihan data berdasarkan relevansi, penulis juga melakukan pengelompokan data berdasarkan sumber yang didapat seperti pada Tabel II.

Tabel II
Hasil Pengelompokan Data Berdasarkan Sumber

Periode	Total Data	Media	Individu
17-19 Nov 2022	258	76	182
20-21 Nov 2022	105	14	91
22-23 Nov 2022	114	30	84

Setelah dilakukan pembersihan data, tahapan selanjutnya adalah melakukan *pre-processing* data. Ada beberapa tahapan yang dilakukan saat *pre-processing*. Pada Tabel II, menampilkan hasil dari proses *case folding* yang termasuk dalam tahapan *pre-processing*.

Tabel III
Tweet Hasil *Case Folding*

Tweet Sebelum Case Folding	Tweet Setelah Case Folding
Ridwan Kamil Siapkan 2 Langkah Atasi Ancaman PHK Massal Industri di Jabar: BLT dan Belanja Produk Lokal https://t.co/RdeqwXWezE https://t.co/WP0auAdOnt	ridwan kamil siapkan 2 langkah atasi ancaman phk massal industri di jabar: blt dan belanja produk lokal https://t.co/rdeqwxweze https://t.co/wp0auadont
👍 on @YouTube: PHK Massal Dan Roller-Coaster Bisnis Online Marketplace https://t.co/R2f9yOaVyW	👍 on @youtube: phk massal dan roller-coaster bisnis online marketplace https://t.co/r2f9yoavyw
Cara Gubernur Jabar atasi ancaman krisis ekonomi dan PHK massal https://t.co/Bv2mw8kGlJ #mandiri_itu_keren	cara gubernur jabar atasi ancaman krisis ekonomi dan phk massal https://t.co/bv2mw8kglj #mandiri_itu_keren

Setelah dilakukan proses *case folding*, tahapan selanjutnya adalah tahapan *tokenizing* yang dijelaskan pada Tabel III.

Tabel IV
Tweet Hasil Tokenisasi

Tweet Sebelum Tokenisasi	
ridwan kamil siapkan 2 langkah atasi ancaman phk massal industri di jabar: blt dan belanja produk lokal https://t.co/rdeqwxweze https://t.co/wp0auadont	ridwan,kamil,siapkan, langkah,atasi,ancaman,phk, massal,industri,di,jabar,blt,dan, belanja,produk,lokal
👍 on @youtube: phk massal dan roller-coaster bisnis online marketplace https://t.co/r2f9yoavyw	on,phk,massal,dan,r ollercoaster,bisnis,online, marketplace
cara gubernur jabar atasi ancaman krisis ekonomi dan phk massal https://t.co/bv2mw8kglj #mandiri_itu_keren	cara,gubernur,jabar,atasi, ancaman, krisis,ekonomi,dan,phk,massal, itukeren

Tahapan selanjutnya adalah melakukan penghapusan terhadap *stopword*. Pada Tabel IV menunjukkan hasil yang didapatkan dari proses *stopword*.

Tabel V
Tweet Hasil Stopword Removal

Tweet Sebelum Stopword Removal	Tweet Setelah Stopword Removal
ridwan,kamil,siapkan, langkah,atasi,ancaman,phk, massal,industri,di,jabar, blt,dan, belanja,produk,lokal	ridwan,kamil,siapkan,l angkah,atasi,ancaman, phk,massal,industri,jabar, blt,belanja,produk,lokal
on,phk,massal,dan,r ollercoaster,bisnis,online, marketplace	on,phk,massal, rollercoaster,bisnis, online,marketplace
cara,gubernur,jabar,atasi, ancaman, krisis,ekonomi,dan, phk,massal, itukeren	gubernur,jabar,atasi, ancaman,krisis,ekonomi, phk,massal,itukeren

Setelah dilakukan proses *case folding*, *tokenizing*, dan *removal stopword*, Langkah selanjutnya adalah melakukan normalisasi tweet seperti yang ditunjukkan pada Tabel V.

Tabel VI
Tweet Hasil Normalisasi

Tweet Sebelum Stopword Removal	Tweet Setelah Stopword Removal
ridwan,kamil,siapkan,l angkah,atasi,ancaman, phk,massal,industri,jabar, blt,belanja,produk,lokal	ridwan,kamil,siapkan, langkah,atasi,ancaman, phk,massal,industri,jabar, blt,belanja,produk,lokal
on,phk,massal, rollercoaster,bisnis, online,marketplace	on,phk,massal,rollercoaster, bisnis,online,marketplace
gubernur,jabar,atasi, ancaman,krisis,ekonomi, phk,massal,itukeren	gubernur,jabar,atasi, ancaman,krisis,ekonomi, phk,massal,itukeren

Tahapan *pre-processing* yang terakhir adalah *stemming*. Pada Tabel VI menunjukkan hasil *stemming* pada tweet yang sudah dilakukan normalisasi.

Tabel VII
Tweet Hasil *Stemming*

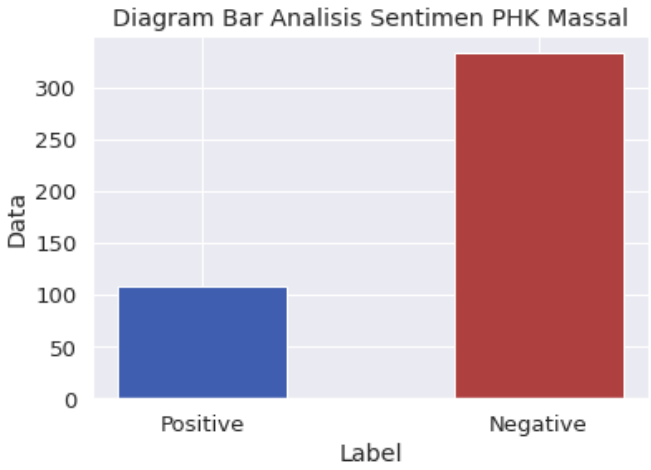
Tweet Sebelum Stemming	Tweet Setelah Stemming
ridwan,kamil,siapkan, langkah,atasi,ancaman, phk,massal,industri,jabar, blt,belanja,produk,lokal	ridwan,kamil,siap, langkah,atas,ancam, phk,massal,industri, jabar,blt,belanja,produk, lokal
on,phk,massal,rollercoaster, bisnis,online,marketplace	on,phk,massal, rollercoaster,bisnis, online,marketplace
gubernur,jabar,atasi, ancaman,krisis,ekonomi, phk,massal,itukeren	gubernur,jabar,atas, ancam,krisis,ekonomi, phk,massal,itukeren

Tahapan selanjutnya adalah tahapan pembuatan label. Dalam tahapan pembuatan label terdapat 2 cara yaitu pelabelan secara manual maupun otomatis. Untuk proses pelabelan secara manual, dapat dilakukan menggunakan kamus kata yang sudah diberikan bobot. Sedangkan jika membuat label secara otomatis, dapat menggunakan Pustaka yang sudah dibuat oleh bahasa pemrograman. Akan tetapi tidak semua bahasa pemrograman menyediakan pelabelan dalam berbagai bahasa. Pada penelitian ini, penulis membuat label secara manual menggunakan kamus seperti yang ditunjukkan pada Tabel VII.

Tabel VIII
Pelabelan Tweet Secara Manual

Tweet	Kata Positif	Kata Negatif	Label
ridwan,kamil,siap, langkah,atas,ancam, phk,massal,industri, jabar,blt,belanja, produk, lokal	Siap(1), Langkah(4), belanja(2), produk(3), kamil(3), massal(3), jabar(3)	Atas(-4),ancam(-4),phk(-4)	Positif
on,phk,massal, rollercoaster,bisnis, online,marketplace	Bisnis(3)	phk(-4)	Negatif
gubernur,jabar,atas, ancam,krisis, ekonomi, phk,massal,itukeren	ekonomi(3), massal(3)	Atas(-4), ancam(-4), krisis(-4)	Negatif

Untuk mengetahui jumlah sentimen positif dan negatif, penulis menggambarkan grafik menggunakan diagram bar seperti pada Gbr 2.



Gbr 2 Diagram Bar Sentimen Masyarakat Terhadap PHK Massal

Pada Gbr 2, dapat dilihat bahwa ada 108 tweet sentimen positif sedangkan untuk sentimen negatif mencapai 333 tweet. Setelah dilakukan proses pembuatan label, Langkah selanjutnya adalah melakukan klasifikasi menggunakan metode *Support Vector Machine* dan juga Naïve Bayes.

Support Vector Machine (SVM) adalah metode klasifikasi yang menggunakan *machine learning* (pembelajaran terbimbing) untuk memprediksi kelas dari suatu sampel atau model berdasarkan hasil proses pelatihan. Pembuatan peringkat dilakukan dengan mencari keputusan *hyperplane* atau batas yang memisahkan satu kelas dengan kelas lainnya, yang dalam hal ini bertindak untuk memisahkan tweet tentang sentimen positif (bertanda +1) dari tweet tentang sentimen perasaan negatif (bertanda -1). SVM menemukan nilai *hyperplane* menggunakan vektor dan nilai *margin* tambahan. Pada penelitian ini data *input* dengan representasi vektor diperoleh dari proses penimbangan. Dengan melakukan pelatihan SVM *classifier* akan dihasilkan nilai atau pola yang akan digunakan dalam proses pengujian SVM untuk memberikan skor sentimen pada tweet [18].

Sebelum menerapkan model SVM, kumpulan data yang diperoleh dibagi menjadi 308 data latih dan 133 data uji, kemudian dilakukan ekstraksi fitur. Penambangan fitur digunakan untuk menemukan informasi potensial dan merepresentasikan kata-kata sebagai vektor fitur. Vektor ini akan digunakan sebagai masukan untuk metode klasifikasi pada tahap selanjutnya. Salah satu teknik ekstraksi fitur adalah dengan menggunakan IDF TF. *Term frequency* atau TF dihitung berdasarkan jumlah kemunculan setiap kata dalam setiap dokumen, sedangkan *inverse document frequency* atau IDF dihitung berdasarkan jumlah kemunculan setiap kata dalam semua dokumen [19]. Proses dalam SVM ini menggunakan kernel linier yang kemudian akan melatih model sentimen untuk klasifikasi data. Laporan klasifikasi dapat dilihat pada Gbr 3 di bawah ini.

```
Here is the classification report:
              precision    recall  f1-score   support

     Negatif       0.90      0.95      0.93       110
     Positif       0.69      0.48      0.56        23

    accuracy                           0.87       133
   macro avg       0.79      0.72      0.74       133
weighted avg       0.86      0.87      0.86       133
```

Gbr 3 Laporan Klasifikasi Menggunakan Metode SVM

Pada Gbr 3 terlihat bahwa nilai presisi data yang mengandung sentimen negatif sebesar 90%, sedangkan nilai presisi data yang mengandung sentimen positif sebesar 69%. Nilai *recall* data yang mengandung sentimen negatif sebesar 95%. Sedangkan nilai *recall* untuk data dengan sentimen positif sebesar 48%, nilai *f1-score* untuk data dengan sentimen negatif sebesar 93%, sedangkan nilai *f-score* untuk data dengan sentimen positif sebesar 56%, dan nilai *support* untuk data. sentimen negatif sebesar 110, sedangkan nilai *support* untuk data dengan sentimen positif sebesar 23.

Algoritma *Naïve Bayes Classifier* adalah algoritma yang digunakan untuk mencari nilai dengan probabilitas tertinggi kemudian mengklasifikasikan data uji ke dalam kategori yang paling sesuai. *Naïve Bayes Classifier* adalah teknologi *pre-processing* yang dapat melakukan klasifikasi fitur yang menambahkan banyak skalabilitas, akurasi, dan efisiensi pada klasifikasi teks. Sebagai *classifier*, *Naïve Bayes Classifier* dianggap efisien, sederhana dan sensitif untuk pemilihan fitur [20] .

```
The classification report is:
              precision    recall  f1-score   support

     Negatif       0.74      1.00      0.85        65
     Positif       1.00      0.04      0.08        24

    accuracy                           0.74        89
   macro avg       0.87      0.52      0.46        89
weighted avg       0.81      0.74      0.64        89
```

Gbr 4 Laporan Klasifikasi Menggunakan Naive Bayes

Pada Gbr 4 terlihat bahwa nilai presisi data yang mengandung sentimen negatif sebesar 74%, sedangkan nilai presisi data yang mengandung sentimen positif sebesar 100%. Nilai *recall* data yang mengandung sentimen negatif sebesar 100%. Sedangkan nilai *recall* untuk data dengan sentimen positif sebesar 4%, nilai *f1-score* untuk data dengan sentimen negatif sebesar 85%, sedangkan nilai *f1-score* untuk data dengan sentimen positif sebesar 8%, dan nilai *support* untuk data. sentimen negatif sebesar 65, sedangkan nilai *support* untuk data dengan sentimen positif sebesar 24.

Langkah terakhir yang perlu dilakukan dalam analisis sentimen adalah mengevaluasi pola yang digunakan. Selama evaluasi, teknik yang digunakan adalah *K-fold cross-validation. Cross-validation* yang merupakan metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi data menjadi dua segmen, satu segmen digunakan untuk mempelajari atau melatih data dan segmen lainnya digunakan untuk mempelajari atau melatih data, digunakan untuk melakukan validasi model. Dalam *k-fold cross validation*, set pelatihan dan validasi harus berpotongan secara berurutan sehingga setiap data memiliki kesempatan untuk dilakukan validasi [21]. Pada tahap evaluasi, data akan dibagi menjadi 10 *fold* yang berarti akan dilakukan 10 kali proses pengujian seperti pada Tabel 8.

Tabel IX
Hasil Pengujian Menggunakan *K-Fold Cross Validation*

K-Fold	SVM Accuracy	NB Accuracy
K-2	0.827922	0.741573
K-3	0.821277	0.741573
K-4	0.818182	0.741573
K-5	0.808302	0.741573
K-6	0.817999	0.741573
K-7	0.818182	0.741573
K-8	0.814356	0.741573
K-9	0.811578	0.741573
K-10	0.827742	0.741573
K-11	0.840909	0.741573

Pada tabel IX, dapat dilihat bahwa hasil pengujian menggunakan *K-Fold Cross Validation* sebanyak 10 iterasi menunjukkan algoritma SVM lebih baik untuk digunakan dalam melakukan analisis terhadap sentimen masyarakat terhadap fenomena PHK Massal yang terjadi dibandingkan dengan Naïve Bayes dapat. Nilai Akurasi SVM mencapai 84% sedangkan nilai akurasi Naïve Bayes mencapai 74.1%.

V. KESIMPULAN

Dari penelitian mengenai analisis sentimen terkait PHK Massal yang dilakukan dapat disimpulkan bahwa:

1. Dari 441 data tweet mengenai PHK Massal, terdapat 333 sentimen negatif dan 108 sentimen positif.
2. Akurasi yang dihasilkan antara analisis sentimen menggunakan metode *Support Vector Machine* (SVM) dan Naïve Bayes memiliki perbedaan yang cukup jauh. Dengan menggunakan metode SVM tingkat akurasi yang dihasilkan mencapai 84% sedangkan dengan menggunakan Naïve Bayes mencapai 74,1%.
3. Penelitian yang dilakukan dapat menganalisis sentimen masyarakat terkait fenomena PHK Massal, sehingga diharapkan hasil analisis sentimen ini dapat dijadikan sebagai bahan evaluasi bagi pihak terkait

Untuk pengembangan penelitian selanjutnya, peneliti memberikan rekomendasi untuk menambah metode yang digunakan untuk melakukan analisa sentimen masyarakat. Selain itu, pada penelitian ini, peneliti hanya mengambil beberapa data sentimen masyarakat sehingga nilai akurasi yang didapat pun menjadi tidak maksimal sehingga peneliti merekomendasikan untuk menggunakan data sentimen dengan periode yang panjang dan parameter lainnya sebagai pendukung, sehingga bisa mendapatkan nilai akurasi yang lebih maksimal.

DAFTAR PUSTAKA

[1] H. Setiawan, E. Utami, and S. Sudarmawan, 'Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes', *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 43–51, Jul. 2021, doi: 10.31603/komtika.v5i1.5189.

[2] I. Romli, S. Prameswari R, and A. Z. Kamalia, 'Sentiment Analysis about Large-Scale Social Restrictions in Social Media Twitter Using Algoritm K-Nearest Neighbor', *Jurnal Online Informatika*, vol. 6, no. 1, p. 96, Jun. 2021, doi: 10.15575/join.v6i1.670.

[3] H. Tuhuteru and A. Iriani, 'Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier', *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 3, pp. 394–401, Oct. 2018, doi: 10.30591/jpit.v3i3.977.

[4] W. F. Satrya, R. Aprilliyani, and E. H. Yossy, 'Sentiment analysis of Indonesian police chief using multi-level ensemble model', in *Procedia Computer Science*, Elsevier BV, 2023, pp. 620–629. doi: 10.1016/j.procs.2022.12.177.

[5] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, 'Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm', in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.

[6] W. Budiawan Zulfikar, A. Rialdy Atmadja, and S. F. Pratama, 'Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes', *Scientific Journal of Informatics*, vol. 10, no. 1, 2023, doi: 10.15294/sji.v10i1.39952.

[7] F. Fridom Mailo and L. Lazuardi, 'Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia', *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health*, vol. 4, no. 1, 2019.

[8] R. Syahputra, G. J. Yanris, and D. Irmayani, 'SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter', *Sinkron*, vol. 7, no. 2, pp. 671–678, May 2022, doi: 10.33395/sinkron.v7i2.11430.

[9] W. Gata and A. Bayhaqy, 'Analysis sentiment about islamophobia when Christchurch attack on social media', *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 4, pp. 1819–1827, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14179.

[10] M. A. Saddam, E. K. Dewantara, and A. Solichin, 'Sentiment Analysis of Flood Disaster Management in Jakarta on Twitter Using Support Vector Machines', *Sinkron*, vol. 8, no. 1, pp. 470–479, Jan. 2023, doi: 10.33395/sinkron.v8i1.12063.

[11] D. Alita, S. Priyanta, and N. Rokhman, 'Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter', *Journal of Information Systems Engineering and Business Intelligence*, vol. 5, no. 2, p. 100, Oct. 2019, doi: 10.20473/jisebi.5.2.100-109.

[12] Fatihah Rahmadayana and Yuliant Sibaroni, 'Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.

[13] M. Kartika Delimayanti, R. Sari, M. Laya, M. Reza Faisal, and dan Pahrul, 'Pemanfaatan Metode Multiclass-SVM pada Model Klasifikasi Pesan Bencana Banjir di Twitter', *Edu Komputika*, vol. 8, no. 1, 2021, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/edukom

[14] H. R. Alhakiem and E. B. Setiawan, 'Aspect-Based Sentiment Analysis on Twitter Using Logistic Regression with FastText Feature Expansion', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 840–846, Nov. 2022, doi: 10.29207/resti.v6i5.4429.

[15] S. Kurniawan, W. Gata, D. Ayu Puspitawati, Nurmalasari, M. Tabrani, and K. Novel, 'Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online', *Jurnal RESTI*, vol. 3, no. 2, pp. 176–183, 2019.

[16] E. Putri Nirwandani and R. Cahya Wihandika, 'Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naïve Bayes', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputere*, vol. 5, no. 3, pp. 1039–1047, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[17] N. N. Alabid and Z. D. Katheeth, 'Sentiment analysis of twitter posts related to the covid-19 vaccines', *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, pp. 1727–1734, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1727-1734.

[18] R. K. Putri and M. Athoillah, 'Support Vector Machine Untuk Identifikasi Berita Hoax Terkait Virus Corona (Covid-19)', *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 3, 2021.

[19] S. Passura Backar, H. Darwis, and W. Astuti, 'Hybrid Fourier Descriptor Naïve Bayes dan CNN pada Klasifikasi Daun Herbal', *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, vol. 8, no. 2, pp. 126–133, 2023.

[20] I. Esa Tiffani, 'Optimization of Naïve Bayes Classifier By Implemented Unigram, Bigram, Trigram for Sentiment Analysis of Hotel Review', *Journal of Soft Computing*, vol. 1, no. 1, pp. 1–7, 2020.

[21] S. Ilahiyah and A. Nilogiri, 'Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network', *JUSTINDO (Jurnal Sistem & Teknologi Informasi Indonesia)*, vol. 3, no. 2, pp. 49–56, 2018.