

# Sentiment Analysis Pemutusan Hubungan Kerja Akibat Pandemi Covid-19 Menggunakan Algoritma NaïveBayes Dan PSO

Wiyanto<sup>[1]\*</sup>, Zulia Setyaningsih<sup>[2]</sup>

Program Studi Teknik Informatika, Fakultas Teknik<sup>[1], [2]</sup>

Universitas Pelita Bangsa

Bekasi, Jawa Barat, Indonesia

[wiyanto@pelitabangsa.ac.id](mailto:wiyanto@pelitabangsa.ac.id)<sup>[1]</sup>, [juliasetyaningsih@gmail.com](mailto:juliasetyaningsih@gmail.com)<sup>[2]</sup>

**Abstract** - The Pandemic Covid-19 in Indonesia in 2020 had an impact on Termination of Employment (PHK), this has received various public opinions on social media. At a time when the poverty rate is high and unemployment increases every year, it becomes a factor of public disapproval of Termination of Employment (PHK). It is necessary to classify public opinion into a negative opinion or a positive opinion on this issue. The purpose of this study is to analyze the sentiment towards layoffs to determine negative or positive opinions using the Naïve Bayes algorithm by adding feature selection. The research stages consist of data collection, text preprocessing, feature selection, and application of algorithms. The testing process in this study uses the Rapid Miner application. The test results in this study using the Naïve Bayes Algorithm, the accuracy value is 93.57% and for addition to the Naïve Bayes + PSO feature selection the accuracy value is 93.71%. The best accuracy value in sentiment analysis of layoffs in the covid-19 pandemic is the addition of the PSO feature selection in the Naïve Bayes Algorithm, which is 0.14% better.

**Keywords** - PHK, Sentiment Analysis, Naïve Bayes, PSO, Rapidminer

**Abstrak** - Pandemi Covid-19 di Indonesia tahun 2020 berimbas pada Pemutusan Hubungan Kerja (PHK), hal ini mendapatkan opini masyarakat yang beragam di dalam media sosial. Pada saat tingkat kemiskinan tinggi dan pengangguran meningkat setiap tahun menjadi faktor ketidaksetujuan masyarakat terhadap Pemutusan Hubungan Kerja (PHK). Perlu Pengklasifikasian opini masyarakat ke dalam opini negatif atau opini positif pada persoalan ini. Tujuan dari penelitian ini ialah menganalisa sentimen terhadap PHK untuk menentukan opini negatif atau positif menggunakan algoritma Naïve Bayes dengan penambahan seleksi fitur. Tahapan penelitian terdiri dari pengumpulan data, *preprocessing text*, seleksi fitur dan penerapan Algoritma. Proses pengujian pada penelitian ini menggunakan aplikasi Rapidminer. Hasil pengujian pada penelitian ini dengan Algoritma Naïve Bayes nilai *accuracy* sebesar 93,57% dan untuk penambahan dengan seleksi fitur Naïve Bayes + PSO nilai *accuracy* sebesar 93,71%. Nilai *accuracy* terbaik dalam *sentiment analysis* PHK pada *pandemic covid-19* ialah dengan penambahan seleksi fitur PSO pada Algoritma Naïve Bayes yaitu lebih baik 0,14%.

**Kata Kunci** - PHK, Analisis Sentimen, Naïve Bayes, PSO, Rapidminer

## I. PENDAHULUAN

Coronavirus Disease-2019 (COVID-19) merupakan salah satu virus berbahaya yang menyebabkan penyakit infeksi saluran pernapasan berkepanjangan dan sampai mengakibatkan meninggal dunia. Virus ini pertama kali di temukan di kota Wuhan negara China pada bulan Desember 2019. Indonesia merupakan negara ke-24 yang terkena dampak pandemi COVID-19 yaitu sebesar 97.286 jiwa positif COVID-19 dengan 55.354 jiwa sembuh dan sebanyak 4.714 meninggal dunia[1]. Pandemi COVID-19 menjadi krisis kesehatan dunia yaitu dengan meningkatnya angka kematian akibat COVID-19 di seluruh dunia. Virus ini juga menyebabkan menurunnya ekonomi dunia, yang mana hal ini terjadi karena kebijakan *social distancing*, *lockdown*, atau Pembatasan Sosial Berskala Besar (PSBB) guna meminimalisir angka terkena COVID-19. Faktor-faktor tersebut membuat roda perekonomian melambat sehingga efeknya terhadap pasar tenaga kerja perusahaan. Dampak dari perusahaan terhadap produksi secara tidak langsung akan menurun dengan adanya *social distancing* yaitu dengan memberikan kebijakan terhadap karyawannya misalnya dirumahkan, bekerja bergantian, pengurangan gaji, atau Pemutusan Hubungan Kerja (PHK). Dalam sebuah riset data bulan April 2020 pada Data Kementerian Ketenagakerjaan (Kemaker) sebanyak 2.084.593 pekerja yang dirumahkan dan terkena PHK dari dampak *pandemic COVID-19*, sebanyak dua juta pekerja berasal dari 116.370 perusahaan [2]. Akibat dari PHK tersebut berdampak pada perekonomian Indonesia semakin meningkat dengan angka kemiskinan tertinggi.

Adanya pro dan kontra terhadap Pemutusan Hubungan Kerja (PHK) yang terjadi di berbagai perusahaan Indonesia. Pengambilan keputusan PHK tidaklah mudah dengan melihat kondisi di tengah pandemi COVID-19. Asosiasi Serikat Pekerja (Aspek) Indonesia meminta agar tidak memutuskan PHK massal terhadap pekerja saat tengah pandemi COVID-19 [3]. Dalam sidang Paripurna di Istana Negara yang diadakan pada 18 Juni 2020, Presiden Jokowi mengungkapkan bahwa semua yang berkaitan dengan ekonomi, manufaktur, industri, terutama padat karya untuk memberikan prioritas agar tidak melakukan PHK secara besar-besaran [4]. Hasil sidang Paripurna tersebut menjadi sorotan Presiden dengan adanya PHK di berbagai bidang usaha maupun perusahaan yang ada di Indonesia. Banyak masyarakat yang menolak adanya PHK, karena bernilai

tidak baik dalam perekonomian mereka. Pada hal lain, keputusan yang dilakukan perusahaan untuk melakukan PHK terhadap karyawannya merupakan jalan terakhir selama pandemi COVID-19.

Pemutusan Hubungan Kerja (PHK) menjadi pembicaraan yang sangat menarik karena telah menimbulkan opini masyarakat dalam media sosial. Media sosial telah menjadi kumpulan data-data atau teks yang mengandung sentimen, salah satunya adalah *Twitter*. *Twitter* adalah media sosial yang ideal untuk ekstraksi opini masyarakat pada isu-isu secara spesifik. *Twitter* memungkinkan mengandung 145 juta *tweet* dikirim dalam sehari [5]. *Tweet* mengandung teks informal berisi berbagai opini, emosi, kata tidak baku, dan bahasa yang dipersingkat [6]. Data *tweet* tersebut digunakan untuk analisis sentimen (*Text Mining*). Analisis Sentimen adalah proses mengolah data tekstual dengan mengekstraknya secara otomatis untuk mendapatkan suatu informasi sentimen yang mengandung kalimat opini, hal ini dilakukan pada sebuah masalah atau objek seseorang untuk mendapatkan opini yang mengandung negatif atau positif [7].

Analisis Sentimen dapat mengklasifikasikan opini yang terkandung dalam *Twitter* bisa menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* adalah salah satu algoritma klasifikasi yang mudah untuk analisis sentimen. Kekurangan menggunakan algoritma ini adalah adanya masalah pada proses pengolahan data yaitu pada pembobotan proses [8].

Pada penelitian ini akan dilakukan analisis sentimen terhadap Pemutusan Hubungan Kerja (PHK) dengan menggunakan algoritma yaitu *Naïve Bayes* ditambah *feature selection PSO*. Peneliti berharap dapat meningkatkan nilai *accuracy*, sehingga akan ditemukan hasil terbaik untuk analisis sentimen terhadap Pemutusan Hubungan Kerja (PHK) perusahaan di Indonesia.

Merujuk kepada penelitian sebelumnya berkaitan dengan analisa sentimen, dalam tema yang dirujuk oleh peneliti ialah Analisa Sentiment Pemindahan Ibu Kota Negara dengan *Feature Selection* Algoritma *Naïve Bayes* dan *Support Vector Machine* oleh Faried Zamachsari dkk [9], metode yang digunakan adalah *Naïve Bayes* dan *Support Vector Machine* dengan tambahan *Feature Selection PSO* menghasilkan *Accuracy NB* = 62,09 %, *NB+PSO* = 88,24%, *SVM* = 63,19%, *SVM+PSO* = 77,77%. Jika dibandingkan penelitian yang dilakukan penulis dan dengan penelitian sebelumnya maka dalam hal analisis sentimen penerapan Algoritma *Naïve Bayes* dengan tambahan *Feature Selection PSO* dalam mendapatkan selisih *accuracy* pada Algoritma *NB* = 93,57% - 62,09 % = 31,48%, jadi *accuracy* dalam dalam penggunaan *NB* lebih baik 31,48% dibandingkan penelitian sebelumnya, dan pada *accuracy* dalam penggunaan *NB+PSO* = 93,71% - 88,24% = 5,47%, jadi *accuracy* dalam dalam penggunaan *NB+PSO* lebih baik 5,47% dibandingkan penelitian sebelumnya.

## II. LANDASAN TEORI

### A. Data Mining

Data mining ialah gabungan dari berbagai disiplin ilmu pada bidang komputer yang didefinisikan sebagai proses

penemuan berbagai pola baru dari kumpulan suatu data yang besar meliputi berbagai metode tertentu. *Data mining* atau biasa disebut penambahan data memiliki dua kegunaan dalam proses mining yaitu deskriptif dan prediktif. Deskriptif ialah *data mining* yang digunakan untuk mencari berbagai pola yang dapat dipahami manusia untuk menjelaskan tentang karakteristik suatu data. Sedangkan prediktif ialah *data mining* yang digunakan untuk membentuk sebuah *model knowledge* yang akan digunakan untuk melakukan suatu prediksi tertentu [10].

### B. Text Mining

*Text Mining* merupakan cabang ilmu *data mining* untuk menganalisis data berupa teks atau dokumen untuk memproses suatu teks dari peramalan suatu pola tertentu dan kecenderungan melalui suatu pola statistik agar menjadi sebuah informasi. Tujuannya ialah menganalisis dari sebuah pendapat, sikap, penilaian, sentimen, evaluasi, emosi seseorang terhadap topik, layanan, organisasi, individu, atau kegiatan tertentu yang digambarkan pada sebuah media sosial. *Text Mining* digunakan untuk suatu klasterisasi, klasifikasi, pencairan informasi dan mengestrak suatu informasi [11]. *Text mining* membantu untuk mengolah data atau teks yang tidak struktur (*unstructured*) agar dapat diproses dan digali untuk mendapatkan informasi dari data atau suatu teks yang tidak struktur tersebut. Terdapat beberapa teknik melakukan *Text mining* yaitu dengan algoritma *Naïve Bayes*, *single pass clustering*, *support vector machine*, dan lain-lain [12].

### C. Naïve Bayes

*Naïve Bayes* merupakan metode klasifikasi paling sederhana yang digunakan untuk analisis sentimen, dengan nilai *accuracy* yang tinggi dalam klasifikasi teks. Metode ini ditemukan oleh Thomas Bayes pada abad ke-18 dengan menerapkan teorema bayes. *Teorema Bayes* adalah teorema yang digunakan dalam statistik untuk menghitung probabilitas suatu hipotesis [5]. Dalam *Teorema Bayes*, suatu probabilitas bersyarat dinyatakan sebagai:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan:

P : probabilitas

X : objek data (x1, x2, x3, ..., xn)

H : hipotesis

### D. Particle Swarm Optimazation (PSO)

*PSO (Particle Swarm Optimazation)* merupakan metode optimasi yang paling sederhana untuk mengubah beberapa parameter dari atribut. Dalam penerapan *PSO* dapat dengan menyeleksi atribut (*attribute selection*) dan *feature selection*, serta meningkatkan bobot dari *attribute (attribute weight)* pada semua *attribute* yang digunakan dalam suatu optimasi tersebut [13]. Kelebihan dari penggunaan *PSO* ialah sederhana, mudah diterapkan dan mempunyai kecepatan dalam konvergensi. Suatu kalimat opini ataupun data mentah dapat dikonversikan ke dalam numerik dan TF-IDF digunakan untuk mengubah data opini menjadi numerik.

E. Feature Selection

Pemilihan fitur (*feature selection*) adalah proses mendapatkan subset dari fitur asli yang diatur menurut pemilihan fitur tertentu dengan kriteria yang memilih fitur relevan dari kumpulan data. Perannya adalah untuk mengompresi skala pemrosesan data, di mana fitur yang berlebihan dan tidak relevan akan dihapus [7].

F. Sentiment Analysis

Pada umumnya analisis sentimen bagian dari *Text mining* dapat disebut juga dengan penambahan opini (*opinion mining*), dengan istilah tersebut bisa diartikan deteksi sikap-sikap terhadap orang atau objek dengan tujuan untuk menemukan sentimen positif dan sentimen negatif terhadap seseorang, perusahaan, kelompok ataupun sebuah situasi tertentu, yang mengandung nilai-nilai sikap yang lebih kompleks seperti menyukai (*liking*), mencintai (*loving*), membenci (*hating*), menghargai (*valuing*), dan menginginkan (*desiring*) [10]. Analisis sentimen memiliki tugas yaitu untuk mengelompokkan suatu teks yang didapat kedalam sebuah kalimat kemudian menentukan dari pendapat yang dikemukakan bersifat positif maupun negatif [14].

Adapun sentimen positif menurut KBBI (Kamus Besar Bahasa Indonesia) adalah sikap seseorang yang bisa meningkatkan nilai dari perasaan seseorang. Sedangkan sentimen negatif adalah sikap seseorang yang bisa menurunkan nilai seseorang sehingga bisa menyebabkan penyurutan nilai pandang seseorang (*down*) [15].

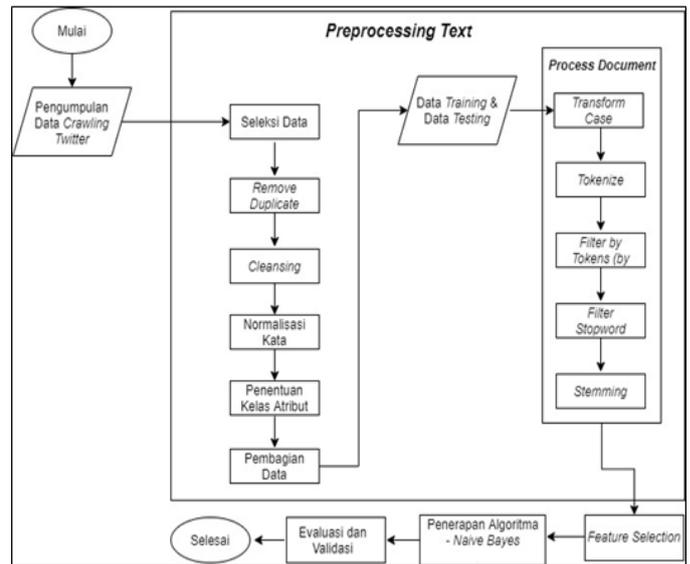
G. Twitter

*Twitter* ialah sebuah situs yang dimiliki dan dioperasikan langsung oleh *Twitter, Inc.*, yang menawarkan salah satu jejaring sosial berupa *microblog* yang memungkinkan penggunanya mengirim dan membaca tweet yang telah dipublish. *Microblog* ialah suatu jenis alat komunikasi secara online dimana pengguna dapat memperbarui status tentang mereka sendiri baik sedang memikirkan ataupun melakukan sesuatu terkait pandangan seseorang terhadap suatu objek yang sedang terjadi. Dalam *microblog* umumnya hanya dibatasi sebanyak 140 karakter yang ditampilkan di halaman pengguna. *Twitter* memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan-aturan khusus. Pesan dalam *Twitter* disebut dengan *tweet*. *Tweet* adalah teks atau tulisan yang ditulis oleh pengguna untuk mengungkapkan opini atau pandangan seseorang terhadap objek yang sedang dibicarakan [16].

III. METODE PENELITIAN

A. Tahapan Penelitian

Berikut ialah tahapan yang dilakukan oleh peneliti dalam penelitian ini dapat digambarkan sebagai berikut :



Gambar 1 Tahapan Penelitian

Pada gambar tahapan penelitian diatas data yang berasal dari *twitter* akan diproses dalam *preprocessing text* yang melalui tahapan-tahapan pada *process document*. Pada proses yang harus diperhatikan yaitu:

1. Cleansing

*Cleansing* merupakan suatu proses untuk penghapusan karakter non-alfabetis dalam mengurangi noise suatu data. Karakter yang dihapus dari data tersebut berupa tanda baca, simbol-simbol yang tidak diperlukan.

2. Normalisasi Kata

Normalisasi kata yaitu langkah menormalkan suatu kata yang disingkat atau diperpanjang menjadi kata yang normal berdasarkan Kamus Besar Bahasa Indonesia (KBBI).

3. Data Training dan Data Testing

Pada penelitian ini dataset yang dikumpulkan sebanyak 1200 data yang berasal dari *twitter*, setelah melalui tahapan *remove duplicate* data yang diperoleh sebanyak 1000 dataset. Penelitian ini menggunakan proposisi 7:3 dari 10 yaitu membagi data training sebanyak 70% sedangkan data testing sebanyak 30%. Data training dilakukan analisis sentiment secara manual yaitu memberikan sentiment negatif dan sentiment positif. Sedangkan untuk data testing akan dilakukan uji terhadap data training sebanyak 300 data.

4. Process Document

Pada proses ini data yang sudah dipecah menjadi data training dan data testing kemudian akan diproses untuk untuk tahap selanjutnya yaitu:

a. Case Folding

Proses *Case Folding* ini dilakukan untuk mengubah semua karakter yang mengandung *alfabet* dan sudah dibersihkan sebelumnya menjadi huruf kecil (*lowercase*).

b. Tokenizing

Pada proses ini, semua kata hasil dari tweet dipisahkan dan dihilangkan tanda bacanya, serta dihilangkan jika ada simbol atau apapun yang bukan merupakan huruf.

c. *Filter by Tokens (by Length)*

Tahap ini dilakukan untuk menghilangkan atau menghapus kata berdasarkan Panjang karakter yang ditentukan. proses filtering hasil dari tokenisasi data diambil kata-kata yang memiliki minimal 3 karakter dan maksimal 25 karakter. Untuk kata yang terdiri kurang dari 4 karakter atau lebih dari 25 karakter akan dihapus.

d. *Filter Stopword*

Kata-kata yang sering muncul dan kurang relevan dengan teks yang ada sehingga akan dihapus. Tahapan *filter stopwords* ini disebut juga dengan *stopword removing*, yaitu menghapus kata-kata yang tidak mempunyai makna dan tidak memiliki suatu pengaruh terhadap analisis sentimen.

e. *Stemming*

Proses *stemming* ini dilakukan untuk mengubah suatu kata yang berimbuhan dari setiap kata yang sudah difilter menjadi kata dasar yang akan diolah.

Penggunaan *Feature Selection* PSO dengan parameter *inertia weight* yang ada pada operator *optimize weight (PSO)*. *Optimize weight* tersebut berfungsi untuk mengoptimalkan bobot berdasarkan korelasi antara fitur satu dengan fitur yang ada dalam *dataset* dengan mempertimbangkan nilai label. Setelah fitur dilakukan penghitungan pada bobot maka dilakukan *inertia weight* dan *population size* pada parameter yang ada. Tujuannya ialah untuk mengoptimalkan tingkat bobot pada nilai atribut tersebut. Dalam parameter lain akan diatur untuk posisi terbaik (*Pbest*) pada partikel dan posisi terbaik secara global untuk mendapatkan nilai bobot yang tinggi. Dan dan perpindahan partikel dilakukan pada parameter *population size*. Dalam penggunaan penambahan *Feature Selection* PSO ini mendapatkan hasil *accuracy* yang lebih baik.

B. *Proses Pengujian*

Pengujian yang digunakan dalam penelitian ini ialah menggunakan *tools Software Rapid Miner versi 9.5.1.0* menggunakan algoritma *Naive Bayes* dengan penambahan seleksi fitur untuk mendapatkan nilai *Accuracy*, *Precision* dan *Recall* sesuai tujuan yang diharapkan. Pada pengujian ini dilakukan sebanyak 4 kali pengujian.

Pada proses pengujian pada tahap performa menggunakan Teknik evaluasi dan validasi dimana ditampilkan pada tabel *confusion matrix*. Adapun tabel *confusion matrix* adalah:

TABEL 1. CONFUSION MATRIX

Class	Actual Class		
		+	-
Predicate Class	+-	TP	FP
	--	FN	TN

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

Proses pengujian pada penelitian ini akan dilakukan terhadap *data training* dan *data testing* yang sudah ditentukan. Adapun beberapa tahapan pengujian yang akan dilakukan adalah:

- a. Pengujian *dataset* dengan menggunakan algoritma *Naive Bayes (NB)*
- b. Pengujian *dataset* dengan menggunakan algoritma *Naive Bayes (NB)* dengan *Feature Selection (PSO)*

IV. HASIL DAN PEMBAHASAN

Pengumpulan data dilakukan dengan melakukan proses *crawling* data dari *API Search Twitter*, berupa data Pemutusan Hubungan Kerja (PHK) dalam Pandemi *Covid-19* di Indonesia, *crawling* dilakukan di tool *Rapid Miner* sebanyak 2000 data *tweet* yang berasal dari halaman *Twitter* dengan pencarian menggunakan *hashtag (#)* PHK kemudian diklasifikasikan menggunakan algoritma *Naive Bayes*. Penelitian ini juga menggunakan algoritma optimasi yaitu PSO untuk mengetahui optimasi mana yang paling tepat digunakan pada algoritma *Naive bayes* untuk klasifikasi data Pemutusan Hubungan Kerja (PHK) dalam Pandemi *Covid-19*. Model penelitian yang digunakan pada penelitian ini diuji menggunakan *tools Rapid Miner 9.5* yang kemudian diperoleh nilai *accuracy*, *presisi*, *recall*.

A. *Hasil Penelitian*

Pada pembagian data dalam penelitian ini dilakukan dengan cara menentukan *data training* dan *data testing*. Dalam proses ini terdapat 1000 data dengan pembagian data sebesar 70% untuk *data training* dan 30% untuk *data testing*.

Dataset adalah kumpulan data yang terdiri dari data positif dan data negatif yang sudah ditentukan dari kelas atribut oleh peneliti. Adapun hasil dari data training positif dan data negatif adalah sebagai berikut :

TABEL 2. DATA TRAINNING POSITIF DAN NEGATIF

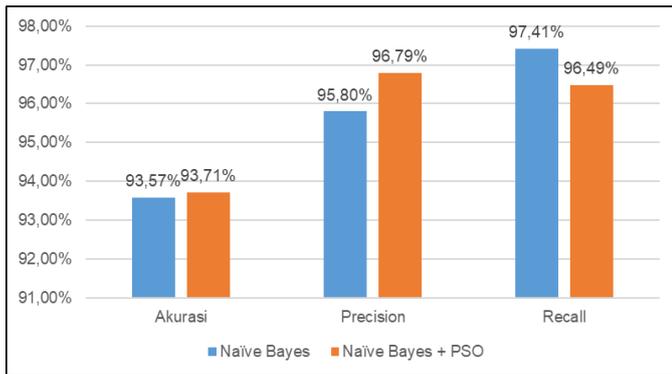
Akun	Data Crawling	Data Positif	Data Negatif	Data Akhir
PHK	1000	44	656	700

Dari hasil pengujian yang dilakukan dari algoritma *Naive Bayes* dengan menggunakan seleksi *fitur attribute weight (PSO)* dapat dianalisis dengan hasil *accuracy*, *precision* dan *recall* dari masing-masing pengujian tersebut. Adapun hasil dari pengujian dapat dilihat sebagai berikut:

TABEL 2. HASIL ACCURACY, PRECISION DAN RECALL DARI PENGUJIAN MODEL

Algoritma	Accuracy	Precision	Recall
Naïve Bayes	93,57%	95,80%	97,41%
Naïve Bayes + PSO	93,71%	96,79%	96,49%

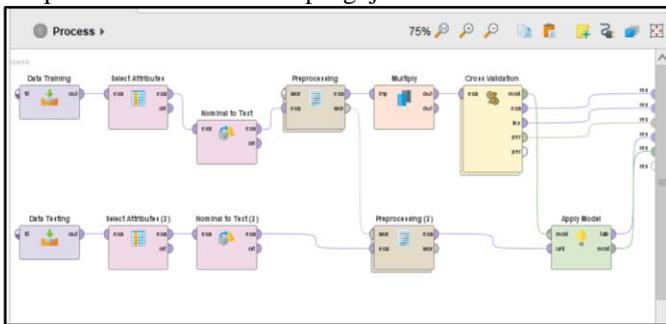
Tabel 2 menunjukkan bahwa model pengujian yang memiliki *accuracy*, *precision*, *recall* paling baik adalah klasifikasi data Pemutusan Hubungan Kerja (PHK) akibat pandemic Covid-19 dengan algoritma *Naïve Bayes* yang di optimasi menggunakan *PSO* dimana nilai *accuracy*-nya sebesar 93,57%, nilai *precision* sebesar 95,80%, dan nilai *recall* sebesar 97,41%. Hasil klasifikasi data Pemutusan Hubungan Kerja (PHK) akibat pandemic Covid-19 dengan algoritma *Naive Bayes + PSO* memperoleh *accuracy* sebesar 93,71%, presisi sebesar 96,79%, dan *recall* sebesar 96,49%.



Gambar 2. Perbandingan Accuracy, Precision, dan Recall dari Pengujian Model Penelitian

B. Pengujian Menggunakan Rapid Miner

Pada penelitian pengujian yang dilakukan 2 kali yaitu algoritma *naïve bayes* dengan penambahan seleksi fitur. Adapun model utama untuk pengujian adalah:



Gambar 3 Model Pengujian Utama

C. Pengujian Naïve Bayes

Pengujian dilakukan dengan menggunakan *software* uji berupa *Rapid Miner 9.5.1* bertujuan untuk mengetahui apakah perhitungan yang telah dilakukan sudah sesuai dengan metode perhitungan *naïve bayes* yang diterapkan pada penelitian ini.

Gambar 4 Hasil Prediksi Naïve Bayes Rapid Miner 9.5.1

Langkah selanjutnya ialah melakukan pengecekan pada setiap atribut dan model yang digunakan pada penelitian ini. Kemudian melakukan evaluasi dari hasil yang didapat dengan melakukan analisis pada setiap keluaran *variable* dan karakteristik informasi yang dihasilkan oleh model *text mining* yang digunakan. Dibawah ini merupakan hasil *confusion matrix* dari *Rapid Miner*.

accuracy: 93.57% +/- 2.71% (micro average: 93.57%)			
	true positif	true negatif	class precision
pred. positif	16	17	48.48%
pred. negatif	28	639	95.80%
class recall	36.36%	97.41%	

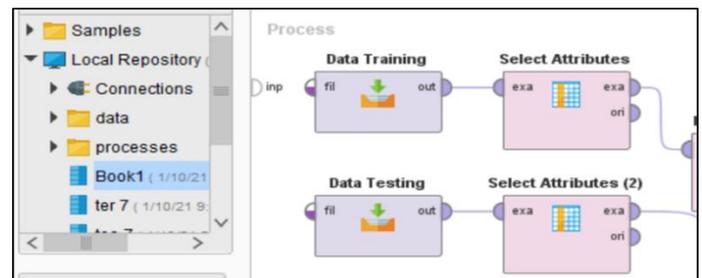
Gambar 5 Hasil Confusion Matrix Naïve Bayes

Adapun perhitungan untuk menentukan nilai *accuracy* sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{16+639}{16+17+28+639} = 0,9357$$

D. Pengujian Naïve Bayes + PSO

Pengujian dilakukan dengan mengeksekusi proses yang dengan menggunakan seleksi fitur *PSO* lalu *Rapidminer* akan mengeksekusi proses tersebut seperti gambar dibawah ini:



Gambar 6 Proses Eksekusi Pengujian Fitur NB + PSO

Hasil *accuracy* pengujian *Naive Bayes* dengan *PSO* dapat dilihat pada gambar berikut:

accuracy: 93.71% +/- 2.54% (micro average: 93.71%)			
	true positif	true negatif	class precision
pred. positif	23	23	50.00%
pred. negatif	21	633	96.79%
class recall	52.27%	96.49%	

Gambar 7 Hasil Pengujian *Confusion Matrix NB+PSO*

Adapun perhitungan untuk menentukan nilai *accuracy* sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

$$= \frac{23+633}{23+23+21+633} = \frac{656}{700} = 0,9371$$

Berdasarkan gambar diatas didapatkan hasil *Confusion Matrix* dari tabel *Performance* pada proses pengujian *Naive Bayes* dengan nilai *accuracy* 93.57%, nilai *precicion* 95.80% dan nilai *recall* 97.41%, sedangkan dengan penambahan seleksi fitur *naive bayes* menghasilkan nilai *accuracy* 93.71%, nilai *precicion* 96.79% dan nilai *recall* 96.49%.

## V. PENUTUP

Dari hasil penelitian yang dilakukan untuk analisis sentimen terhadap Pemutusan Hubungan Kerja (PHK) di Indonesia menggunakan algoritma *naive bayes* dengan penambahan seleksi fitur PSO dapat disimpulkan bahwa pada penelitian ini memberikan hasil opini masyarakat terhadap PHK di Indonesia yaitu sentimen negatif dan sentimen positif, dengan perolehan nilai *accuracy* yang didapat menggunakan algoritma *naive bayes* tanpa seleksi fitur yaitu 93,57%, dan untuk penggunaan algoritma *naive bayes + PSO* menghasilkan nilai *accuracy* sebesar 93,71%. Sehingga pada penelitian ini algoritma yang terbaik dengan penambahan seleksi fitur PSO pada algoritma *naive bayes* dengan selisih 0,14%. Dari penitian yang terdahulu berkaitan dengan analisis sentimen bahwa terjadi gap dengan penelitian sebelumnya maka dalam hal analisis sentimen penerapan Algoritma *Naive Bayes* dengan tambahan *Feature Selection PSO* dalam mendapatkan selisih *accuracy* pada Algoritma NB = 93,57% - 62,09 % = 31,48%, jadi *accuracy* dalam dalam peggungan NB lebih baik 31,48% dibandingkan penelitian sebelumnya, dan pada *accuracy* dalam peggungan NB+PSO = 93,71% - 88,24% = 5,47%, jadi *accuracy* dalam dalam peggungan NB+PSO lebih baik 5,47% dibandingkan penelitian sebelumnya.

Dari penelitian ini diharapkan kedepanya dapat dilanjutkan dengan menggunakan teknik ataupun metode yang lebih relevan untuk meningkatkan *accuracy* yang lebih baik.

## DAFTAR PUSTAKA

- [1] WHO Covid-19, "WHO Coronavirus Disease (COVID-19) Dashboard," *World Health Organization*, 2020. .
- [2] P. P. Agus, "Tidak Hanya di Indonesia, PHK Massal Terjadi di Seluruh Dunia," *CNBC Indonesia*, 2020. .
- [3] "Undang-undang Nomor 13 Tahun 2003 tentang Ketenagakerjaan," no. 13, Bandung: Citra Umbara, 2003.
- [4] C. A. Gian, "Korban PHK Akibat Corona Jadi Sorotan Jokowi," *CNBC Indonesia*, 2020. .
- [5] S. Liao, J. Wang, R. Yu, K. Sato, dan Z. Cheng, "ScienceDirect CNN for situations understanding based on sentiment analysis of twitter data," *Procedia Comput. Sci.*, vol. 111, no. 2015, hal. 376–381, 2017, doi: 10.1016/j.procs.2017.06.037.
- [6] A. Sharma dan U. Ghose, "Sentimental Analysis of Twitter Data with respect to General Elections in India," *Procedia Comput. Sci.*, vol. 173, no. 2019, hal. 325–334, 2020, doi: 10.1016/j.procs.2020.06.038.
- [7] T. A. Lorosae dan B. D. Prakoso, "ANALISIS SENTIMEN BERDASARKAN OPINI MASYARAKAT PADA," hal. 25–30, 2018.
- [8] O. Somantri dan D. Apriliani, "Opinion mining on culinary food customer satisfaction using naive bayes based-on hybrid feature selection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 15, no. 1, hal. 468–475, 2019, doi: 10.11591/ijeeecs.v15.i1.pp468-475.
- [9] F. Zamachsari, G. V. Saragih, Susafa'ati, dan W. Gata, "Analisis Sentimen Pemindahan Ibu Kota Negara dengan Feature Selection," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 3, hal. 504–512, 2020.
- [10] Dr. Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*, Edisi Revi. Bandung: Informatika Bandung, 2019.
- [11] W. A. Luqyana, I. Cholissodin, dan R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, hal. 4704–4713, 2018.
- [12] S. Al Faraby, F. Informatika, U. Telkom, dan D. Frekuensi, "Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String Kernel for Classification indonesian news," *e-Proceeding Eng.*, vol. 5, no. 1, hal. 1701–1710, 2018.
- [13] V. Kevin, S. Que, A. Iriani, dan H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization ( Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization )," vol. 9, no. 2, hal. 162–170, 2020.
- [14] M. W. Fuad Nur Hasan, "ANALISIS SENTIMEN ARTIKEL BERITA TOKOH SEPAK BOLA DUNIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION," *J. AKRAB JUARA*, vol. 3, no. November, hal. 42–55, 2018.
- [15] L. Ardiani dan H. Sujaini, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak Implementation of Sentiment Analysis of Community Responses to Development in Pontianak City," vol. 8, no. 2, hal. 44–51, 2020, doi: 10.26418/justin.v8i2.36776.
- [16] I. T. S. A. Pamungkas, "ANALISIS SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE ( SVM )," no. 1, hal. 69–79, 2018.