



## *Sentiment Analysis of Public Opinion on Twitter Toward BSI Bank Using Machine Learning Algorithms*

### **Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma *Machine Learning***

Ratna Andini Husen<sup>1\*</sup>, Rizki Astuti<sup>2</sup>, Lili Marlia<sup>3</sup>, Rahmaddeni<sup>4</sup>, Lusiana Efrizoni<sup>5</sup>

<sup>1,2,3,4,5</sup>Teknik Informatika, STMIK Amik Riau, Indonesia

E-Mail: <sup>1</sup>201003182011@sar.ac.id, <sup>2</sup>2010031802007@sar.ac.id, <sup>3</sup>2010031802008@sar.ac.id,  
<sup>4</sup>rahmaddeni@sar.ac.id, <sup>5</sup>lusiana.efrizoni@sar.ac.id

Diterima 30 Agustus 2021; September direvisi 05 2021; Diterima 15 September 2022 2  
Penulis Korespondensi: Ratna Andini Husen

#### **Abstract**

Public opinion expressed through social media, especially Twitter, has become an important source of information for companies and financial institutions, including BSI Bank. Sentiment analysis of public opinion can help BSI Bank in understanding the views and perceptions of the public towards their services. This study aims to develop and apply machine learning algorithms, namely SVM, naïve bayes and logistic regression algorithms to analyze public opinion sentiment towards BSI Bank contained in tweets on Twitter. The tweet Data used in this study was taken from kaggle's website with a total of 24,401 data, containing reviews of user comments related to ransomware at BSI Bank. The results of the experiments that have been done obtained that SVM produces 0.88% accuracy, naïve bayes produces 0.76% accuracy, and logistic regression produces 0.86% accuracy. Based on the experimental results that SVM get a performance performance that is superior to the naïve bayes algorithm and logistic regression . In this context, SVM can be a good choice for sentiment analysis in general. This study revealed that the percentage of negative sentiment towards BSI Bank is higher than positive sentiment. These findings indicate a significant concern and dissatisfaction among the public towards the company's services. While there are some positive sentiments identified, it is important for BSI Bank to take seriously the concerns underlying these negative sentiments.

*Keyword: Bank of BSI, Logistic Regression, Naïve Bayes, SVM, Twitter*

#### **Abstrak**

Opini publik yang terekspresikan melalui media sosial, khususnya Twitter, telah menjadi sumber informasi yang penting bagi perusahaan dan lembaga keuangan, termasuk Bank BSI. Analisis sentimen opini publik dapat membantu Bank BSI dalam memahami pandangan dan persepsi masyarakat terhadap layanan mereka. Penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan algoritma *machine learning* yaitu algoritma SVM, *naïve bayes* dan *logistic regression* untuk menganalisis sentimen opini publik terhadap Bank BSI yang terdapat dalam *tweet* di Twitter. Data *tweet* yang digunakan dalam penelitian ini diambil situs dari kaggle dengan jumlah data 24.401, berisi tentang ulasan komentar pengguna terkait *ransomware* pada Bank BSI. Hasil dari percobaan yang telah dilakukan diperoleh bahwa SVM menghasilkan akurasi 0,88%, *naïve bayes* menghasilkan akurasi 0,76%, dan *logistic regression* menghasilkan akurasi 0,86%. Berdasarkan dari hasil percobaan bahwa SVM mendapatkan performa kinerja yang lebih unggul dari pada algoritma *naïve bayes* dan *logistic regression* . Dalam konteks ini, SVM dapat menjadi pilihan yang baik untuk analisis sentimen secara umum. Penelitian ini mengungkapkan bahwa persentase sentimen negatif terhadap Bank BSI lebih tinggi daripada sentimen positif. Temuan ini menunjukkan adanya keprihatinan dan ketidakpuasan yang signifikan di antara masyarakat terhadap layanan perusahaan. Meskipun ada beberapa sentimen positif yang teridentifikasi, penting bagi Bank BSI untuk secara serius memperhatikan dan menangani masalah-masalah yang mendasari sentimen negatif ini.

Kata Kunci: Bank BSI, Logistic Regression, Naïve Bayes, SVM, Twitter

#### **1. PENDAHULUAN**

Opini publik merupakan aspek penting dalam dunia bisnis, terutama dalam sektor perbankan. Di zaman yang sangat maju ini, hiburan virtual seperti Twitter, telah menjadi platform dasar bagi orang-orang untuk berbagi perspektif, kesimpulan, dan pengalaman mengenai berbagai produk dan layanan, termasuk layanan

perbankan. Bank sebagai institusi keuangan juga tidak terkecuali dari eksposur terhadap opini publik yang tergambar dalam media sosial.

Medsos atau *social media* adalah sebuah media yang biasa digunakan untuk memperkenalkan diri, berbagi dalam bentuk informasi, berinteraksi atau berkolaborasi satu sama lain. Media sosial Twitter telah menjadi platform penting bagi penggunaannya untuk menyampaikan pendapat, termasuk opini publik tentang berbagai topik, termasuk lembaga keuangan. Di Indonesia, Twitter bukan hanya dipergunakan sebagai sarana untuk mengekspresikan emosi, tetapi juga sebagai sarana untuk memberikan masukan atau kritik terhadap suatu objek, pelayanan atau kebijakan pemerintah [1].

Pada penelitian ini peneliti memanfaatkan dataset publik dari sosial media *Twitter* yang diunduh dari Kaggle untuk melakukan analisis sentimen terhadap ulasan yang berkaitan dengan Bank BSI (Bank Syariah Indonesia). BSI termasuk bank syariah terbesar yang ada di Indonesia bentuk penggabungan dari tiga bank Syariah dari himpunan bank milik negara (Himbara), yaitu PT Bank BRI Syariah (BRIS), PT Bank Syariah Mandiri (BSM), dan PT Bank BNI Syariah (BNIS) yang dimana di dalam pelaksanaan kegiatannya dilakukan dengan sistem anti riba atau sistem ekonomi syariah [2]. Topik yang lagi hangat diperbincangkan dalam kurun waktu terakhir ini yaitu dampak serangan *ransomware* pada Bank BSI.

Penelitian yang dilakukan mengulas beberapa penelitian sebelumnya yang dilakukan di bidang analisis sentimen. Berdasarkan penelitian sebelumnya yang berjudul analisis perbandingan sentimen *corona virus diseases-2019* (covid 19) pada *Twitter* menggunakan algoritma *logistic regresion* dan *Support Vector Machine* (SVM) melakukan suatu analisis sentimen terhadap bahasan yang saat ini sering menjadi *trending topic* di twitter yaitu "*CoronaVirus Disease 2019* (covid19)". Penyebaran virus corona juga ramai diperbincangkan di Twitter, saat ini virus corona sedang menarik pandangan dunia internasional. Pada penelitian ini menghasilkan nilai akurasi tertinggi sebesar 92,13% dalam data pelatihan dan 91,15% dalam data pengujian sedangkan algoritma *logistic regresion* memperoleh nilai dari hasil akurasi sebesar 87,79% pada data pelatihan dan 87,68% pada data pengujian [3]. Sedangkan pada penelitian berjudul analisis sentimen Gofood berdasarkan *twitter* menggunakan metode *Naïve Bayes* dan SVM penelitian ini dilakukan dengan tujuan menganalisa opini masyarakat akan kinerja Gojek (Gofood) di Indonesia. itu mendapatkan nilai netral sebesar 92,8% , bernilai positif sebesar 5,2% dan bernilai negatif sebesar 2,0%. Perbandingan dari hasil akurasi yang didapat, algoritma SVM memperoleh nilai akurasi yang jauh lebih besar dari algoritma *naïve bayes*. Adapun hasil akurasi algoritma SVM sebesar 83% dan 98,5% sedangkan algoritma *naïve bayes* mendapatkan nilai akurasi 74,6% dan 91,5% [4].

Berdasarkan penelitian sebelumnya, pada penelitian ini akan digunakan beberapa algoritma *machine learning* seperti SVM, *naïve bayes*, dan *logistic regression*. SVM adalah teknik *machine learning* yang didasarkan dari teori struktural pembelajaran statistik [5]. Keunggulan dari *naïve bayes* yaitu seperti kesederhanaan, kecepatan, dan akurasi yang ada pada algoritma ini cenderung lebih tinggi. Telah banyak penelitian yang dilakukan untuk klasifikasi sentimen menggunakan *naïve bayes* [6]. *Logstic regression* adalah metode yang sering digunakan dalam menganalisis data dengan satu atau lebih variabel prediktor yang menggambarkan variabel respon. Variabel respon *logistic regression* sendiri mempunyai nilai 1 (ya) dan 0 (tidak), sehingga akan menciptakan variabel respon *Bernoulli* [7]. Tujuan penelitian yang akan dilakukan yaitu untuk melakukan klasifikasi atau pengelompokan ulasan positif dan negatif pengguna *Twitter* terhadap Bank BSI. Penelitian yang dilakukan dapat digunakan oleh Bank BSI sebagai tolak ukur untuk meningkatkan layanan kepada pengguna berdasarkan ulasan positif dan negatif.

## 2. METODE PENELITIAN

Penelitian ini memiliki beberapa proses. Data dan informasi yang obyektif akan dijadikan acuan dalam penelitian, dengan data tersebut diharapkan penelitian yang dihasilkan akan menjadi penelitian yang berkualitas. Pada Gambar 1 ini berisi tahapan-tahapan pengerjaan seperti *data collection*, *pre-processing data*, *word embedding*, *modelling*, *evaluation model*, dan *analysis sentiment*.

### 2.1 Data Collection (Pengumpulan Data)

Pengumpulan data kini menjadi komponen penting dari semua operasi bisnis, baik itu data klien untuk melakukan hal yang sederhana layanan atau data perusahaan untuk memastikan pengoperasian infrastruktur penting. Dalam lingkungan operasional saat ini dan dengan terus berkembangnya ekonomi digital, data adalah aset perusahaan yang penting. Terlepas dari fungsi dan pentingnya data, hal tersebut memang benar adanya sulit untuk mendorong bisnis untuk melindungi data mereka sendiri [8].

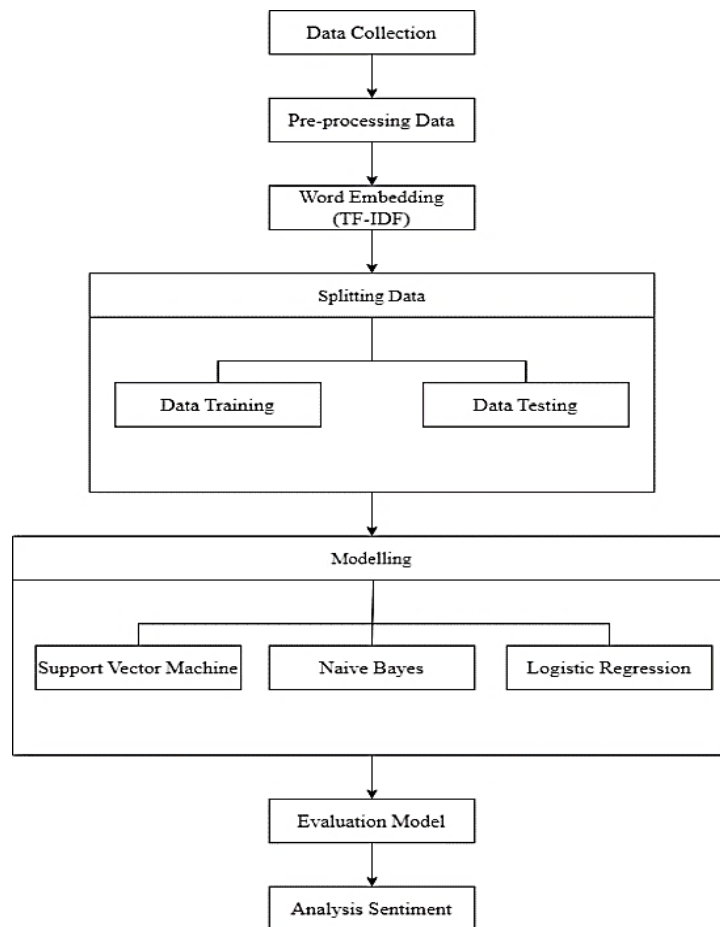
Data yang digunakan dalam proses perbandingan algoritma ini tentang komentar nasabah bank BSI yang bersumber dari Kaggle.jumlah data yang digunakan 24.401 komentar yang tersusun dalam file *csv*.

**Tabel 1.** Isi dataset

1,URL,Date,Tweet,ID,Username,Replies,Reetweets,Likes,Quotes,conversationId,Language,Links,Media,Retweeted Tweet,Bookmarks
2,https://twitter.com/SaidRberdaya/status/1658980771284279297,2023-05-17 23:40:12+00:00,"@bankbsi_id Layanan BAYAR ke @TakafulKeluarga sudah bisa blm ya ,kok jadi repot begini ðŸ« ",1658980771284279297,SaidRberdaya,0,0,0,1658830871812775937,in,,,0
3,https://twitter.com/ariemayw/status/1658980516857802753,2023-05-17 23:39:12+00:00,@NurmansyahAff @berlianidris @Paltiwest @bankbsi_id Apa mungkin dah ganti vendor?,1658980516857802753,ariemayw,1,0,0,0,1658792156524425218,in,,,0
4,https://twitter.com/mputhreeeee/status/1658977499853000705,2023-05-17 23:27:12+00:00,@bankbsi_id Uдах bisa tarik tunai via ATM atau ke teller gak min?,1658977499853000705,mputhreeeee,5,0,0,0,1658830871812775937,in,,,0
5,https://twitter.com/rahayusheey/status/1658975344765054976,2023-05-17 23:18:39+00:00,@bankbsi_id Jangan lupa perbaikan mbanking untuk topup!!!!,1658975344765054976,rahayusheey,4,0,0,0,1658830871812775937,in,,,0
6,https://twitter.com/ndaalfatih_/status/1658975122894774272,2023-05-17 23:17:46+00:00,"@bankbsi_id Min, BSI udah amann dipake belum minn??",1658975122894774272,ndaalfatih_,0,0,0,0,1658634424966344704,in,,,0
7,https://twitter.com/bukaniqbaale/status/1658974757671542784,2023-05-17 23:16:19+00:00,"@abcdxyumi Mending di telfon bank BSI, hallo @bankbsi_id",1658974757671542784,bukaniqbaale,0,0,0,0,1658796272202190848,in,,,0
8,https://twitter.com/sir_usil/status/1658974702591942658,2023-05-17 23:16:06+00:00,"Rekrutmen berdasarkan identitas bukan kualitas bisa merusak tatanan sistem operasi perbankan di Indonesia.

**2.2 Pre-processing Data**

*Pre-processing* yaitu langkah untuk menghapus kesalahan atau faktor lain yang dideteksi tidak cocok dan dapat membuat nilai dari hasil proses data yang diolah menurun [9]. Pemeriksaan kembali data dapat dilakukan pada tahap pra-pemrosesan/pembersihan, dalam tahap ini dilakukan proses penghapusan redudansi, *outlier* dan data kosong agar dapat memastikan data masukan yang diproses sudah benar atau sudah bersih, untuk dapat menetapkan maka hasil perhitungan algoritma juga akan memmberikan hasil yang tepat [10].



**Gambar 1.** Metode Penelitian

### 2.3 Word Embedding (Pembobotan Kata)

*Word Embedding* atau pembobotan kata adalah teknik yang mengubah sebuah kata menjadi vektor atau larik yang berada dalam sekumpulan angka. Saat membuat sebuah model *machine learning*, teks sebagai penerima inputnya, tentunya *machine learning* tidak bisa langsung menerima teks yang kita miliki, kata harus diubah terlebih dahulu menjadi angka dengan mengacu pada kamus kata. Biasanya, jika tidak menggunakan penyematan kata, setiap kata akan dikonversi menjadi bilangan bulat berdasarkan posisi bilangan tersebut di kamus [11]. Adapun pembobotan kata yang digunakan yaitu TF-IDF. TF-IDF yaitu perhitungan yang digunakan untuk menguji pentingnya kata (istilah) dalam *document* dan *corpus* [12]. Tahapan ini dilakukan agar dapat menilai *term* atau bobot kata yang ada pada *document* pada semua *document* yang ada di *corpus*. Rumus untuk menghitung tf-idf yaitu :

$$\text{TF-IDF} = \text{TF} * \text{IDF} = \text{TF} * \log(n/\text{df}) \quad (1)$$

Keterangan:

TF = Text frequency  
IDF = Teks frekuensi dalam *document*  
n = Jumlah *document*

### 2.4 Splitting Data

Setelah tahap *word embedding* dengan menggunakan tf-idf, tahapan selanjutnya yaitu melakukan *splitting data* agar dapat menjadi data pelatihan dan data pengujian. Pembagian ini berarti memberdayakan kerangka kerja untuk meninjau kumpulan data terlebih dahulu, kemudian, pada saat itu, pengujian informasi akan diselesaikan untuk memastikan tingkat akurasi. Pada sistem ini *splitting data* dapat ditetapkan sesuai dengan kemauan pengunanya [13]. Dengan menggunakan *split data* 90:10, 80:20, 70:30, dan 60:40. Data pengujian yaitu data yang sebelumnya tidak pernah digunakan dalam penelitian namun berguna untuk menilai berhasil atau tidaknya penelitian, sedangkan data pelatihan yaitu data yang akan digunakan dalam melakukan penelitian.

### 2.5 Modelling

Pada penelitian yang dilakukan adalah melakukan komparasi akurasi dari beberapa algoritma. Adapun algoritma yang digunakan yaitu sebagai berikut:

#### 2.5.1 Support Vector Machine (SVM)

SVM atau *Support Vector Machine* disebut sebagai algoritma yang cenderung umum untuk mengklasifikasikan teks dan memberikan kinerja yang baik di banyak bidang. Keunggulan SVM untuk mengidentifikasi *hyperplane* yang berbeda antara dua kelas yang berbeda dimaksimalkan, dan SVM memastikan jarak antara data yang paling dekat dengan *hyperplane* yang dimaksimalkan [14]. SVM mengklasifikasikan *instance* pelatihan yang termasuk dalam salah satu dari dua kelas dengan memasang batas pemisah (*hyperplane*) antar kelas sehingga margin antara batas dan salah satu kelas dimaksimalkan. Kelas dari sebuah *instance* baru ditentukan tergantung pada sisi *hyperplane* mana ia berada [15]. Berikut rumus perhitungan svm:

Titik data :  $x_i = \{x_1, x_2, \dots, x_n\} \in R^n$  Kelas data :  $y_i \in \{-1, +1\}$

Pasangan data dan kelas :

$$\{(x_i, y_i)\}_{i=1}^N \quad (2)$$

Maksimalkan fungsi:

$$Ld = \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad \text{syarat : } 0 \leq \alpha_i \leq C \quad \text{dan} \quad \sum_{i=1}^N \alpha_i y_i = 0 \quad (3)$$

Menghitung nilai w dan b:

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad b = -\frac{1}{2} (w \cdot x^+ + w \cdot x^-) \quad (4)$$

Fungsi keputusan klasifikasi  $\text{sign}(f(x))$  :

$$f(x) = w \cdot x + b \quad \text{atau} \quad f(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b \quad (5)$$

Keterangan :

N	= Jumlah data
n	= Banyak fitur
C	= Nilai konstanta
m	= Jumlah support vector/titik data yang memiliki $\alpha_i > 0$
$K(x, x_i)$	= Fungsi kernel
Ld	= Dualitas Lagrange Multiplier $\alpha_i$ : nilai mutu setiap titik data

### 2.5.2 Naïve Bayes

*Naïve bayes* merupakan algoritma *machine learning* yang masuk pada kategori *supervised classification* [16]. *Naïve bayes* adalah bentuk paling sederhana dari pengklasifikasi jaringan Bayes [17]. Pengklasifikasi Naïve Bayes berakar dari teorema Bayes, yang megasumsikan bahwa data tidak berhubungan secara statistik. Berikut merupakan rumus dari algoritma Naïve Bayes.

$$P(c | x) = P(x | c) P(c) P(x) \quad (6)$$

Keterangan :

$P(c   x)$	= Probabilitas <i>posterior</i> pada kelas c saat diberikan atribut x
$P(c)$	= Probabilitas prior pada class c
$P(x   c)$	= Probabilitas pada atribut x saat diberikan kelas c
$P(x)$	= Probabilitas prior pada atribut x

### 2.5.3 Logistic Regression

*Logistic regression* atau regresi logistik termasuk salah satu tipe analisis regresi yang biasa digunakan untuk mendeskripsikan hubungan antara suatu variabel terikat dan suatu variabel bebas yang melibatkan satu atau beberapa variabel bebas dengan variabel terikatnya. Algoritma *logistic regression* terutama digunakan untuk klasifikasi biner [18]. Tipe kategori dapat berupa 0 dan 1, benar atau salah, besar atau kecil. Jenis variabel bebasnya adalah variabel kategori. Hal inilah yang memisahkan *logistic regression* dengan regresi berganda atau *liniar regression* lainnya [19].

### 2.6 Model Evaluation

*Model evaluation* dilakukan dengan cara menghitung nilai akurasi yang diperoleh pada sebuah sistem untuk suatu data pengujian tertentu, dimana setiap data pengujian mempunyai dua peluang hasil yaitu positif dan negatif. Jika keluaran data uji tepat dengan target yang diinputkan maka data akan diberi label dengan benar, sedangkan data uji yang keliru pada target akan mendapatkan label salah. Oleh karena itu, semua nilai yang dihasilkan dapat direpresentasikan sebagai *confusion matrix*.

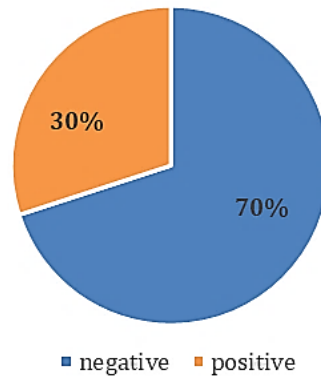
### 2.7 Sentiment Analysis

*Sentiment analysis* atau analisis sentimen, juga dikenal sebagai Penambahan Opini, adalah bidang penelitian yang bertujuan untuk menganalisis opini, emosi, opini, sikap, dan sentimen publik terhadap suatu entitas berdasarkan produk, layanan, isu, organisasi, acara, atau fitur tertentu. Analisis sentimen dapat dilakukan dengan berbagai macam algoritma, seperti *Naive Bayes* (NB), *Decision Trees*, *K-Nearest Neighbor* (KNN), *Neural Network*, dan *Support Vector Machine* (SVM) [20].

## 3. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dari analisis sentimen dengan tiga algoritma, yaitu SVM, *naïve bayes*, dan *logistic regression*, terhadap komentar-komentar terkait Bank BSI telah berhasil diamati. Pada penelitian ini data yang digunakan yaitu sejumlah komentar yang berkaitan dengan Bank BSI yang diambil dari platform Twitter.

Dataset bersumber dari kaggle berjumlah 24.401 dataset dengan 15 *feature* sebelum dilakukannya *preprocessing* data. Dalam *dataset* terdapat 30% sentimen berlabel positif yaitu 7.197 sentimen dan 70% berlabel negatif yaitu 16.796 sentimen, visualisasi data diperlihatkan dalam *pie chart* dibawah.



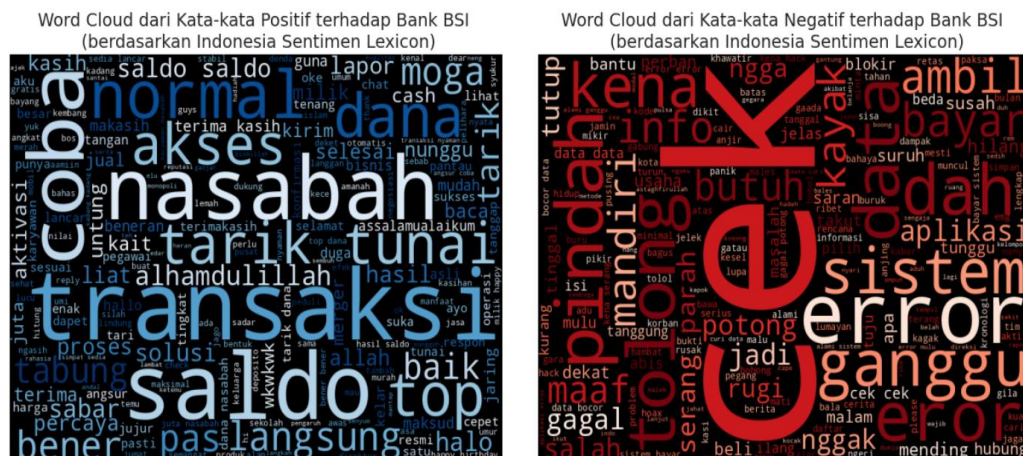
**Gambar 2.** Visualisasi *Pie Chart* Sentimen Bank BSI

Dapat disimpulkan dari Gambar 2 bahwa pengguna twitter cenderung memberikan ulasan negatif terhadap Bank BSI. Melalui proses pra-pemrosesan, seperti penghilangan tanda baca, tokenisasi, dan pembersihan data. Hasil setiap proses ini dapat dilihat dalam bentuk Tabel 2.

**Tabel 2.** Hasil *preprocessing data*

text_clean	text_preprocessed
id layanan bayar ke sudah bisa blm ya kok jadi repot begini	['id', 'layan', 'bayar', 'blm', 'ya', 'repot', '']
id apa mungkin dah ganti vendor	['id', 'dah', 'ganti', 'vendor']
id udah bisa tarik tunai via atm atau ke teller gak min	['id', 'udah', 'tarik', 'tunai', 'via', 'atm', 'teller', 'gak', 'min']
id jangan lupa perbaiki mbanking untuk topup	['id', 'lupa', 'baik', 'mbanking', 'topup']
id min bsi udah amann dipake belum minn	['id', 'min', 'bsi', 'udah', 'amann', 'dipake', 'minn']
mending di telfon bank bsi hallo id	['mending', 'telfon', 'bank', 'bsi', 'hallo', 'id']
rekrutmen berdasarkan identitas bukan kualitas bisa merusak tatanan sistem operasi perbankan di indonesia	['rekrutmen', 'dasar', 'identitas', 'kualitas', 'rusak', 'tatanan', 'sistem', 'operasi', 'perban', 'indonesia', 'cc', 'ri', 'indonesia', 'marufamin', 'id']
cc ri indonesia marufamin id	['marufamin', 'id']

Pada tabel 2 telah dilakukan proses pembersihan data, dimana pada proses ini data yang sebelumnya masih terdapat huruf kapital, simbol, angka, dan kata imbuhan telah menjadi kata dasar dan tidak terdapat lagi huruf kapital, simbol dan angka. Karena pada proses ini akan menentukan tingkat akurasi yang diperoleh nantinya dengan algoritma *machine learning*.



**Gambar 3.** *Word cloud* Positif dan Negatif

Pada Gambar 3, terdapat hasil visualisasi menggunakan *word cloud* sehingga terdapat 2 *word cloud* yaitu *word cloud* positif dan *word cloud* negatif. Kedua *word cloud* tersebut akan menampilkan kata-kata yang paling sering disebutkan oleh pengguna *Twitter*, membuat analisis lebih mudah daripada harus melihat *tweet* satu per satu.

Pengujian menggunakan algoritma SVM, *naive bayes*, dan *logistic regression* menghasilkan akurasi yang berbeda-beda. Dalam percobaan ini, 24.401 data dengan 15 fitur digunakan sebelum melakukan *pre-processing data*. Pada tahap *pre-processing* dilakukan pelabelan data untuk mengetahui sentimen positif atau



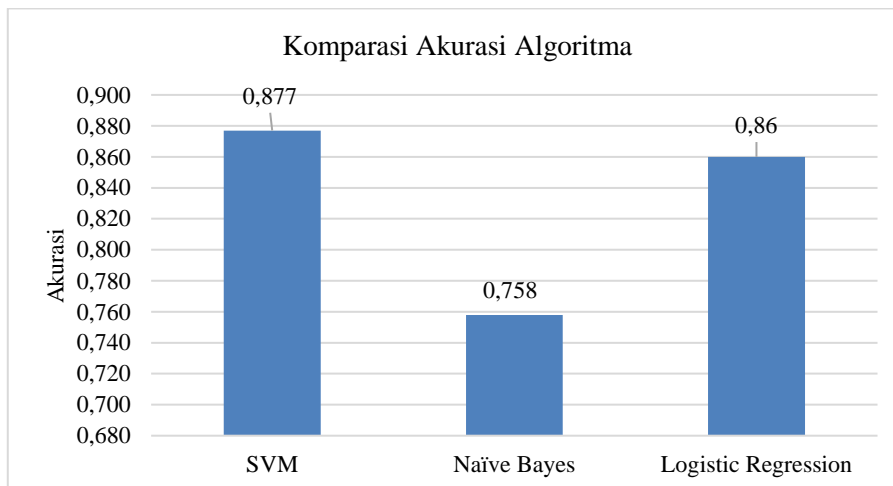
negatif. Kemudian dilakukan *pre-processing* data. Langkah selanjutnya adalah memodelkan algoritma SVM, *naive bayes* dan *logistic regression* dengan *split data* 90:10, 80:20, 70:30, dan 60:40. Dengan hasil sebagai berikut.

**Tabel 3.** Hasil komparasi splitting data

<i>Splitting Data</i>	<i>Support Vector Machine</i>	<i>Naïve Bayes</i>	<i>Logistic Regression</i>
90:10	88%	76%	86%
80:20	88%	75%	86%
70:30	88%	74%	86%
60:40	88%	73%	86%

Berdasarkan tabel 3, dengan beberapa splitting data yang digunakan, akurasi tertinggi terdapat pada splitting data 90:10 yaitu untuk algoritma SVM sebesar 88%, *Naïve Bayes* sebesar 76%, dan *Logistic Regression* sebesar 86%. Algoritma pertama SVM mendapatkan akurasi 88%. Algoritma kedua *Naïve Bayes* mendapatkan akurasi 76%. Algoritma ketiga *Logistic Regression* mendapatkan akurasi 86%. Dapat disimpulkan bahwa nilai akurasi model algoritma SVM memperoleh nilai akurasi yang lebih baik yaitu 88% dengan selisih 2% dibandingkan dengan algoritma *Logistic Regression* sedangkan dengan algoritma *Naïve Bayes* mendapatkan selisih 12% dan kinerja akurasi model yang mendapatkan nilai sangat rendah adalah algoritma *Naïve Bayes* mendapatkan akurasi model yaitu 76%.

Hasil komparasi akurasi algoritma SVM, *naive bayes*, dan *logistic regression* divisualisasikan dengan diagram batang yang disajikan pada gambar 4.



**Gambar 4.** Visualisasi komparasi akurasi dengan beberapa algoritma

Berdasarkan pada gambar 4, hasil komparasi akurasi dari algoritma SVM mendapatkan nilai akurasi sebesar 88%, *naive bayes* sebesar 76%, dan *logistic regression* sebesar 86%. Hasil komparasi algoritma pada gambar 4 menunjukkan bahwa algoritma SVM jauh lebih baik dalam melakukan analisis sentimen pada penelitian ini.

#### 4. KESIMPULAN

Hasil penelitian yang dilakukan dapat menunjukkan persentase opini masyarakat terhadap bank BSI yang tergolong dalam kategori positif dan negatif serta dapat dijadikan acuan untuk meningkatkan pelayanan perusahaan dan memberikan visibilitas kepada masyarakat. Persentase sentimen positif cenderung lebih sedikit dibandingkan dengan sentimen negatif yaitu sebesar 30%, sedangkan sentimen negatif mendapatkan persentase sebesar 70%. Hasil komparasi dengan menggunakan split data 90:10, 80:20, 70:30, dan 60:40 dengan menggunakan algoritma SVM, *Naïve Bayes*, dan *Logistic Regression* maka didapatkan hasil akurasi tertinggi yaitu dengan menggunakan algoritma SVM pada split data 90:10 dengan nilai akurasi 88%. Perusahaan dapat mempertimbangkan untuk menguji dan membandingkan kinerja metode lain, seperti algoritma pembelajaran mendalam (misalnya, jaringan saraf) atau metode berbasis aturan (misalnya, analisis teks berbasis aturan). Dengan mencoba berbagai metode, perusahaan dapat memastikan bahwa mereka menggunakan pendekatan yang paling efektif dalam menganalisis sentimen publik.

## REFERENSI

- [1] Rosdiana, E. Tungadi, Z. Saharuna, and M. Utomo, "Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar," *Proceeding SNTEI*, Sep. 2019, [Online]. Available: <https://dev.twitter.com>
- [2] K. Utari, Y. Septiyana, Asnaini, and K. Elwardah, "Efisiensi Keputusan Merger Tiga Bank Syariah di Indonesia pada Masa Pandemi Covid-19 (Studi di BSI Kc Bengkulu Panorama)," *Journal Ekombis Review*, vol. 10, pp. 311–324, 2022, doi: 10.37676/ekombis.v10iS1.
- [3] Kelvin, J. Banjarnahor, E. Indra, and S. Sinurat, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," *JUSIKOM PRIMA*, vol. 5, no. 2, Feb. 2022.
- [4] M. Petiwi, A. Triayudi, and I. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 542, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [5] D. Oktavia, Y. Ramadahan, and Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 407–417, Aug. 2023, doi: 10.30865/klik.v4i1.1040.
- [6] R. Amelia, Darmansah, N. Prastiwi, and M. Purbaya, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 338, Apr. 2022, doi: 10.30865/jurikom.v9i2.3895.
- [7] D. Devito, R. Wihandika, and A. Widodo, "Ekstraksi Ciri Untuk Klasifikasi Gender Berbasis Citra Wajah Menggunakan Metode Histogram of Oriented Gradients," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 8, pp. 8002–8011, Aug. 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] K. E. Todt, "Data Privacy and Protection : What Businesses Should Do," *The Cyber Defense Review*, vol. 4, no. 2, pp. 39–46, 2019, doi: 10.2307/26843891.
- [9] T. Hidayat, M. Priyatna, A. Sutanto, A. Khudri, and R. Khomarudin, "Informasi Sebaran Titik Panas Berbasis WebGIS untuk Pemantauan Kebakaran Hutan dan Lahan di Indonesia," *Jurnal Teknologi Lingkungan*, vol. 20, no. 1, 2019, [Online]. Available: <http://pycsw.org>
- [10] S. Widaningsih and S. Yusuf, "Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, Sep. 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [11] A. Rapsanjani and E. Junianto, "Implementasi Probabilistic Neural Network dan Word Embedding untuk Analisis Sentimen Vaksin Sinovac," *JURNAL RESPONSIF*, vol. 3, no. 2, pp. 233–242, Aug. 2021, [Online]. Available: <http://ejurnal.ars.ac.id/index.php/jti>
- [12] F. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, Nov. 2019.
- [13] M. Muafa and L. Iswari, "Pengembangan Aplikasi Berbasis Web dengan Rshiny untuk Data Klasifikasi Menggunakan Metode Naïve Bayes," *Prosiding Automata*, vol. 3, no. 1, Jan. 2022.
- [14] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar, and S. Alfarizi, "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023," *Jurnal Teknik Komputer AMIK BSI*, vol. 6, no. 1, Jan. 2020, doi: 10.31294/jtk.v4i2.
- [15] C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson, and E. Ifeachor, "Early Detection of Alzheimer's Disease with Blood Plasma Proteins Using Support Vector Machines," *IEEE J Biomed Health Inform*, vol. 25, no. 1, pp. 218–226, Jan. 2021, doi: 10.1109/JBHI.2020.2984355.
- [16] M. Alsharif, A. Kelechi, K. Yahya, and S. Chaudhry, "Machine learning algorithms for smart data analysis in internet of things environment: Taxonomies and research trends," *Symmetry (Basel)*, vol. 12, no. 1, Jan. 2020, doi: 10.3390/SYM12010088.
- [17] L. Yu, S. Gan, Y. Chen, and M. He, "Correlation-Based Weight Adjusted Naive Bayes," *IEEE Access*, vol. 8, pp. 51377–51387, 2020, doi: 10.1109/ACCESS.2020.2973331.
- [18] A. B. Amjoud and M. Amrouch, "Transfer Learning for Automatic Image Orientation Detection Using Deep Learning and Logistic Regression," *IEEE Access*, vol. 10, pp. 128543–128553, 2022, doi: 10.1109/ACCESS.2022.3225455.
- [19] F. Pramakrisna, F. Adhinata, and N. Tanjung, "Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 11, no. 2, pp. 90–97, Jun. 2022, doi: 10.34148/teknika.v11i2.466.
- [20] Imam and I. Santoso, "Analisis Sentimen Pada Twitter Terhadap Gagalnya Pelaksanaan Piala Dunia Di Indonesia Menggunakan Metode Naïve Bayes," *Ikraith-Informatika*, vol. 7, no. 2, Jul. 2023, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/issue/archive>