

IMPLEMENTASI SENTIMENT ANALYSIS PADA OPINI MASYARAKAT INDONESIA DI *TWITTER* TERHADAP VIRUS *COVID-19* VARIAN *OMICRON* DENGAN ALGORITMA *NAÏVE BAYES*, *DECISION TREE*, DAN *SUPPORT VECTOR MACHINE*

Azriyan Arham¹⁾, Ericks Rachmat Swedia²⁾, Margi Cahyanti³⁾, dan M Ridwan Dwi Septian⁴⁾

^{1,2,4} Teknik Informatika, Universitas Gunadarma

³ Sistem Informasi, Universitas Gunadarma

^{1,2,3,4} Jalan Margonda Raya No.100, Pondok Cina, Depok

E-mail : riyandarham18@gmail.com¹⁾, ericks_rs@staff.gunadarma.ac.id²⁾, margi@staff.gunadarma.ac.id³⁾, ridwandwiseptian@staff.gunadarma.ac.id⁴⁾

ABSTRAK

Covid-19 merupakan suatu penyakit pneumonia yang menjadi penyebab pandemi tahun 2020. Selama berlangsungnya pandemi, *Covid-19* terus mengalami mutasi genetik yang menghasilkan varian-varian baru dengan fenotip, pola transmisi, dan virulensi yang beragam. WHO sendiri mengelompokkan varian *Covid-19* menjadi 3 varian, yaitu *Variant of Interest* (VoI), *Variants under Monitoring* (VUM), dan *Variants of Concern* (VoC) (Susilo et al., 2022). *Omicron* merupakan muatsi *Covid-19* yang termasuk ke dalam VoC dan diperkirakan memiliki daya transmisi lebih cepat daripada varian *Delta* sehingga diindikasikan lebih cepat menyebar (Susilo dkk., 2022). Untuk itu dilakukan penelitian tentang analisis sentimen pengguna *Twitter* di Indonesia terhadap *Covid-19* varian *Omicron*. Penelitian menerapkan tiga algoritma klasifikasi, di antaranya adalah algoritma *Naïve Bayes*, algoritma *Support Vector Machine*, dan algoritma *Decision Tree*. Penelitian dilakukan dengan menggunakan *Python* sebagai bahasa pemrograman dan *Google Colaboratory* sebagai tools. Penelitian ini menggunakan data *Tweet* sebanyak 3,931 data yang diambil dari tanggal 28 April 2022 hingga 4 Juni 2022. Tahapan pada penelitian ini terdiri dari penarikan data, data *preprocessing*, pelabelan sentimen, klasifikasi, pengujian model, visualisasi data, serta implementasi algoritma ke dalam *platform website*. Analisis yang dilakukan, algoritma akurasi tertinggi adalah algoritma *Support Vector Machine* dengan nilai akurasi sebesar 81,68%, kemudian algoritma *Naïve Bayes* menghasilkan nilai akurasi sebesar 74,25%, dan algoritma *Decision Tree* menghasilkan nilai akurasi sebesar 72,77%. Pada ketiga algoritma menggunakan data uji sebanyak 202 data. Sementara itu berdasarkan analisis yang dilakukan, didapatkan sentimen netral dengan sentimen tertinggi sebesar 63,7%, lalu sentimen negatif sebesar 21,1%, dan sentimen positif sebesar 15,2%.

Kata Kunci: *Decision Tree*, *Klasifikasi*, *Naïve Bayes*, *Omicron*, *Support Vector Machine*, *Twitter*

1. PENDAHULUAN

SARS-CoV-2, virus penyebab *Covid-19* merupakan penyakit pneumonia penyebab pandemi tahun 2020. Selama berlangsungnya pandemi, *Covid-19* terus mengalami mutasi genetik yang menghasilkan varian-varian baru dengan fenotipe, pola transmisi, dan virulensi yang beragam. WHO sendiri mengelompokkan varian *Covid-19* menjadi 3 varian, yaitu *Variant of Interest* (VoI), *Variants under Monitoring* (VUM), dan *Variants of Concern* (VoC) (Susilo dkk., 2022).

Salah satu varian dari VoC adalah *Omicron*. *Omicron* diklasifikasikan sebagai *Variants of Concern* (VoC) karena mengalami mutasi berlipat ganda, pada *omicron* terdapat fenomena pelarian kekebalan (*escape immune*), dan diperkirakan memiliki daya transmisi lebih cepat daripada varian *Delta* sehingga lebih cepat menyebar dibandingkan mutasi varian yang lainnya (Susilo dkk., 2022).

Untuk mengetahui opini masyarakat dengan lebih efektif dan efisien terkait varian *Omicron* perlu dilakukan analisis. Analisis yang dapat dilakukan adalah analisis

sentimen. Algoritma yang dapat digunakan untuk melakukan analisis sentimen adalah menggunakan algoritma klasifikasi seperti algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree*. (Sihombing & Yuliati, 2021).

Penelitian yang dilakukan oleh (Ardiani dkk., 2020) yang berjudul "*Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak*" yang bertujuan untuk melakukan analisis *Tweet* untuk analisis sentimen opini masyarakat yang tuliskan ke dalam *Twitter*. Implementasi dilakukan dengan mengklasifikasikan *Tweet* untuk mendapatkan informasi sentimen yang terkandung dalam tanggapan masyarakat, pada penelitian ini metode pengklasifikasian sentimen yang digunakan adalah *naïve bayes* dengan membagi sentimen ke dalam tiga kelas, yaitu sentimen positif, sentimen netral, dan sentimen negatif. Pada penelitian ini menggunakan

data latih yaitu sebesar 450 dan data uji sebesar 50 dengan akurasi yang dihasilkan sebesar 72%.

Algoritma *Naïve Bayes* digunakan oleh (Saputro & Sari, 2020) dengan penelitian yang berjudul “*Naïve Bayes Algorithm Performance Test for Student Study Prediction*”. Penelitian ini bertujuan untuk membuat sistem klasifikasi kelulusan pada Universitas AMIKOM Yogyakarta menggunakan teknik data mining dengan algoritma *Naïve Bayes*. Data yang digunakan sebanyak 300 data yang bersumber dari data alumni Universitas AMIKOM Yogyakarta angkatan 2012, dan 2013 dengan masing-masing data sebanyak 150 data. Penelitian ini menghasilkan rata-rata akurasi model *Naïve Bayes* sebesar 68%, nilai *precision* sebesar 61.3%, nilai *recall* sebesar 65.3%, dan nilai *f1-score* sebesar 61%.

Algoritma *Naïve Bayes* juga digunakan oleh (Pamungkas dkk., 2020) dengan penelitian yang berjudul “*Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python*”. Penelitian ini bertujuan untuk menganalisis dan membandingkan metode-metode pendekatan *supervised learning* dengan menggunakan studi kasus *data churn modelling* dari kaggle. Berdasarkan hasil perhitungan akurasi metode *supervised learning* yang dilakukan pada penelitian ini, diperoleh nilai 79,1% untuk algoritma *Naïve Bayes*.

Algoritma *Support Vector Machine* digunakan oleh (Parapat et al, 2018) dengan penelitian yang berjudul “*Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak*”. Penelitian ini bertujuan untuk melakukan klasifikasi penyimpangan tumbuh kembang anak dengan menggunakan metode SVM yang terdiri dari proses *training* sebagai pembelajaran sistem dan *testing* untuk memperoleh hasil klasifikasi. Data yang digunakan pada penelitian sebanyak 90 data yang terbagi menjadi 3 kelas. Kelas mewakili 3 jenis penyimpangan tumbuh kembang anak yaitu *Down Syndrome*, *Autisme*, dan *Attention Deficit Hyperactivity Disorder*. Hasil dari penelitian ini menghasilkan akurasi rata-rata sebesar 63,11%. Algoritma klasifikasi teks SVM ini juga dapat digunakan untuk sentimen analisis, karena pada dasarnya sentimen analisis juga mengklasifikasikan berdasarkan tiga sentimen, yaitu sentimen positif, negatif dan netral.

Berdasarkan penelitian sebelumnya yang telah dilakukan oleh (Ardiani dkk., 2020), (Pamungkas dkk., 2020), (Parapat et al, 2018), dan (Saputro & Sari, 2020), peneliti melakukan pengembangan dengan menggabungkan algoritma *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine* sebagai algoritma klasifikasi untuk melihat sentimen pengguna Twitter dalam Bahasa Indonesia terhadap Covid-19 varian *Omicron*. Penelitian yang dilakukan menggunakan 3931 data. Penelitian ini diimplementasikan ke dalam sebuah *platform website*. Penelitian ini mendapatkan nilai akurasi sebesar 81,68% untuk algoritma *Support Vector Machine*. Untuk algoritma *Naïve Bayes* menghasilkan nilai akurasi sebesar 74,25%, dan

nilai akurasi yang didapatkan oleh algoritma *Decision Tree* sebesar 72,77%.

2. RUANG LINGKUP

Adapun ruang lingkup yang ditetapkan dari penelitian ini adalah sebagai berikut.

1. *Tools* yang digunakan untuk melakukan analisis adalah *Google Colaboratory*.
2. Analisis sentimen yang dilakukan menggunakan bahasa pemrograman Python.
3. Data yang digunakan bersumber dari Twitter dengan kata kunci “*Omicron*”.
4. Data yang digunakan merupakan *Tweet* dalam Bahasa Indonesia sebanyak 3931 data yang diambil dari tanggal 28 April 2022 hingga 4 Juni 2022.
5. Data yang diperoleh diklasifikasikan menjadi positif, negatif, dan netral.
6. Algoritma yang digunakan adalah *Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree*.

3. BAHAN DAN METODE

Pada bagian ini menjelaskan bahan-bahan dan metode apa yang akan digunakan dalam penelitian ini, diantaranya.

3.1 Analisis Sentimen

Analisis sentimen merupakan suatu metode yang digunakan untuk mengekstrak dan memahami data tekstual yang berupa opini secara otomatis untuk menemukan informasi sentimen yang terkandung dalam sebuah opini. Pada analisis sentimen informasi sentimen yang terkandung dalam sebuah opini dapat diklasifikasikan menjadi tiga bagian, yaitu sentimen positif, sentimen netral, dan sentimen negatif (Sari & Wibowo, 2019).

Dalam KBBI atau Kamus Besar Bahasa Indonesia kalimat sentimen positif adalah reaksi atau sikap yang meningkatkan nilai seseorang atau sesuatu. Sebagai contoh kalimat sentimen positif adalah “*Senang sekali dengan adanya beberapa taman dan ruang baca, semoga bisa membuat kota Pontianak lebih bersinar*” (Ardiani dkk., 2020).

Dalam KBBI sentimen negatif ini adalah merupakan reaksi atau sikap yang menurunkan nilai seseorang atau sesuatu, jadi kalimat sentimen negatif akan menyebabkan penyurutan nilai pandang terhadap sesuatu, sehingga membentuk *tren down* (Ardiani dkk., 2020).

Selanjutnya pada kata sentimen netral ini sendiri di dalam KBBI yang berarti tidak berpihak, kalimat bersentimen netral ini merupakan ekspresi kalimat yang tidak bersifat positif maupun negatif. Contoh penggunaan dari kalimat sentimen netral ini adalah : “*Mengunjungi karnaval khatulistiwa di Pontianak*” (Ardiani dkk., 2020).

3.2 Twitter

Twitter adalah sebuah layanan jejaring sosial (media sosial) yang memungkinkan pengguna untuk dapat mengirim dan membaca pesan berbasis teks yang dikenal dengan sebutan kicauan (*Tweet*). Media sosial ini didirikan pada Maret tahun 2006 oleh Jack Dorsey dan situs jejaring sosialnya diluncurkan pada Juli 2006. Pada Twitter pengguna terdaftar dapat membuat *Tweet* dengan *web interface* atau *mobile application* (Zukhrufillah, 2018)

3.3 Covid-19 Varian Omicron

Omicron merupakan varian mutasi dari Covid-19. WHO sendiri mengelompokkan varian Covid-19 menjadi 3 varian, yaitu *Variant of Interest* (VoI), *Variants under Monitoring* (VUM), dan *Variants of Concern* (VoC). VoI adalah varian dengan perubahan fenotip yang berdampak pada pola transmisi, virulensi, dan antigenisitas, contohnya adalah varian Lambda. Kemudian VUM adalah jenis varian dengan perubahan fenotipe namun belum diketahui lebih lanjut mengenai peningkatan transmisi, dampak morbiditas dan mortalitas, serta virulensinya, contoh dari varian ini adalah Kappa, Iota, Eta, Epsilon, Zeta dan Theta. Kemudian VoC adalah varian dengan fenotip yang berdampak negatif dan mengalami peningkatan transmisi yang signifikan, contoh dari varian ini merupakan Alpha, Beta, Gamma, Delta, dan yang terbaru adalah varian *Omicron* (Susilo dkk., 2022).

3.4 Machine Learning

Machine learning merupakan sebuah model komputasi statistik yang berfokus pada prediksi menggunakan komputer. Algoritma *machine learning* membangun model matematika dari data sampel yang dikenal sebagai data latih atau data *training* untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk melakukan tugas. Alur kerja dari *machine learning* adalah dengan mengambil data mentah sebagai input dan menghasilkan prediksi sebagai *output* (Santoso et al., 2020).

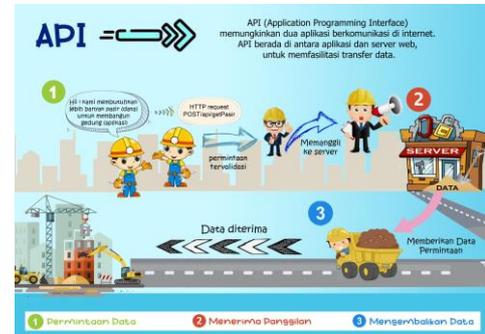
Pada *machine learning*, data yang diproses biasanya dibagi menjadi dua, yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih algoritma, sedangkan data *testing* digunakan untuk mengetahui performa dari algoritma yang telah dilatih sebelumnya ketika menemukan data baru (Fikriya dkk., 2017)

Machine learning sendiri dapat diklasifikasikan menjadi tiga jenis, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* (Santoso et al., 2020).

3.5 API dengan Bahasa Pemrograman Python

API (*Application Programming Interface*). API sendiri adalah program yang menjembatani dua aplikasi untuk saling berkomunikasi (*interface*). API memudahkan koneksi untuk beberapa sistem yang memiliki arsitektur berbeda. Aplikasi ini biasanya berjalan di *web server*, yang berfungsi untuk memberikan layanan terhadap permintaan dari *client* seperti pada gambar 1. Layanan yang diberikan dapat berupa *file html*, dapat juga berupa data. Data yang diminta

biasanya berupa JSON, *image* atau *video*. *Web Server* yang cukup terkenal adalah IIS (*Internet Information Services*) dari perusahaan Microsoft dan Apache dari The Apache Software Foundation *Windows Server* (Swedia ER dkk., 2022).



Gambar 1. API di Web Server

Python merupakan bahasa pemrograman yang menyediakan algoritma dan library yang dapat digunakan untuk membuat model *machine learning* baik *supervised learning* ataupun *unsupervised learning*. Contoh *library* pada *Python* yang dapat digunakan untuk *machine learning* adalah NumPy, Pandas, Matplotlib, dan Sklearn (*Scikit Learn*) (Retnoningsih & Pramudita, 2020).

Selain memiliki *library* yang dapat digunakan untuk *machine learning*, *Python* juga merupakan bahasa pemrograman yang sederhana jika dibandingkan dengan bahasa pemrograman lainnya seperti Java, C, C#, ataupun C++ (Retnoningsih & Pramudita, 2020).

Bahasa pemrograman *Python* ini juga dapat dijalankan di berbagai macam platform sistem operasi seperti Windows, Linux, dan Apple OS (Herwanto, 2019).

3.6 Scikit Learn

Menurut Jason dikutip oleh (Munawar & Silitonga, 2019) *Scikit Learn* adalah modul pada *Python* yang mengintegrasikan berbagai macam jenis algoritma *machine learning* untuk *supervised learning* dan *unsupervised learning*, modul ini adalah model efisien untuk data mining dan analisis data. Data mining sendiri merupakan suatu proses pencarian pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu (Wijaya et al., 2021).

3.7 Preprocessing

Preprocessing merupakan pemrosesan data sebelum dilakukannya klasifikasi. *Preprocessing* bertujuan untuk menghilangkan *noise* seperti simbol atau kata yang tidak diperlukan yang bisa memperlambat pemrosesan saat berlangsungnya klasifikasi (Musa'adah dkk., 2018). Gambar 2

merupakan alur proses *preprocessing* ((de Araujo dkk., 2022a)).



Gambar 2. Proses *Preprocessing* (de Araujo dkk., 2022b)

3.8 Klasifikasi *Naïve Bayes*

Naïve bayes adalah metode klasifikasi (*supervised learning*) dengan pendekatan probabilistik. Pendekatan ini membuat asumsi mengenai bagaimana data bisa dihasilkan dengan menempatkan model probabilistik untuk mewujudkannya (Musa'adah et al., 2018).

Menurut *Andrew McCallum* dan *Kamal Nigam* dikutip oleh (Musa'adah et al., 2018) Klasifikasi *Naïve Bayes* merupakan metode yang sederhana dengan mengasumsikan bahwa semua atribut pada data tidak bergantung satu sama lain berdasarkan konteks kelas.

Menurut (Pamungkas et al., 2020) probabilitas dalam klasifikasi *Naïve Bayes* dinyatakan dalam sebuah persamaan (1).

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

3.9 Klasifikasi *Decision Tree*

Klasifikasi *Decision Tree* merupakan klasifikasi dengan metode *unsupervised learning* yang berarti proses pembelajaran pada *Decision Tree* dilakukan dengan mengklasifikasikan data baru berdasarkan *training samples* yang sudah ada (Wijaya, 2021).

3.10 Klasifikasi *Support Vector Machine*

Support Vector Machine merupakan metode klasifikasi dengan sistem pembelajaran yang menggunakan ruang hipotesis yang berupa fungsi-fungsi linear di dalam sebuah fitur yang memiliki dimensi tinggi dan dilatih menggunakan algoritma pembelajaran berdasarkan teori optimasi (Parapat et al, 2018).

3.11 Gambaran Umum

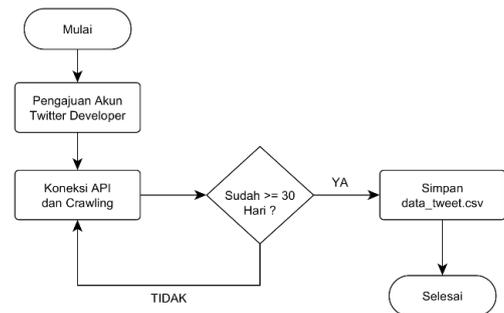
Pada implementasi '*sentiment analysis*' pada opini masyarakat Indonesia di *twitter* terhadap virus covid-19 varian *omicron* ini dilakukan proses diantaranya :

1. Pengumpulan data
2. *Preprocessing* Data
3. Pelabelan Sentimen
4. Klasifikasi
5. Visualisasi data
6. Implementasi *Website*

3.12 Pengumpulan Data

Data yang digunakan bersumber dari Twitter dan diambil (*crawling*) menggunakan Twitter API yang didapatkan melalui pengajuan menggunakan akun Twitter Developer. Data yang didapat berupa *Tweet* berbahasa Indonesia dengan total 3931 data.

Crawling dilakukan setiap sembilan hari sekali yang dilakukan dari tanggal 28 April 2022 hingga 4 Juni 2022. *Flowchart* dari alur atau tahapan dalam melakukan pengumpulan data pada gambar 3.



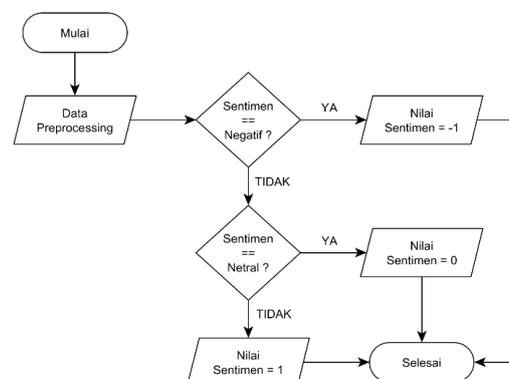
Gambar 3. Proses Pengumpulan Data

3.13 *Preprocessing* Data

Pada tahap ini dilakukan pembersihan data dengan menghilangkan *noise* seperti simbol atau kata yang tidak diperlukan dan membuat data menjadi lebih terstruktur. Tahap data *preprocessing* yang dilakukan terdiri dari normalisasi data, *filtering*, *case folding*, *stopword removal*, *stemming*, dan *tokenization*.

3.14 Pelabelan Sentimen

Pelabelan sentimen dilakukan dengan memberikan nilai terhadap masing-masing sentimen (positif, negatif, dan netral). Untuk data *Tweet* yang memiliki sentimen positif akan diberi nilai 1, untuk data *Tweet* yang memiliki sentimen netral akan diberikan nilai 0, dan untuk data *Tweet* yang memiliki sentimen negatif akan diberikan nilai -1. *Flowchart* dari proses pelabelan sentimen pada gambar 4 di bawah ini.



Gambar 4. Proses Pelabelan Sentimen

Pada proses pelabelan, didapat sebanyak 204 data *Tweet* bersentimen positif, 855 data *Tweet* bersentimen netral, dan 284 *Tweet* bersentimen negatif. *Script* untuk menampilkan jumlah masing-masing sentimen dapat dilihat pada gambar 5.

```
df['sentimen'].value_counts()
Netral      855
Negatif     284
Positif     204
Name: sentimen, dtype: int64
```

Gambar 5. Jumlah Masing-Masing Sentimen

3.15 Klasifikasi

Setelah pelabelan berhasil dilakukan, selanjutnya adalah melakukan klasifikasi. Klasifikasi dilakukan dengan menggunakan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree*. Sebelum dilakukannya klasifikasi, data *Tweet* dibagi menjadi dua bagian (*data split*) yaitu dengan rasio 85% untuk data latih (*data train*) banding 15% untuk data uji (*data test*).

3.15.1 Klasifikasi Naïve Bayes

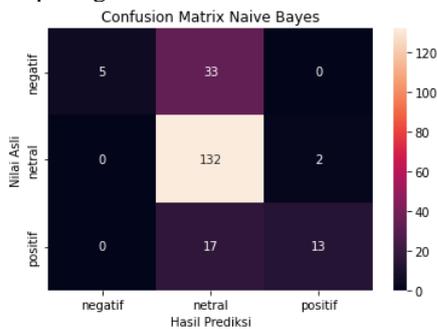
Akurasi yang didapatkan dari model yang dilatih menggunakan algoritma *Naïve Bayes* adalah sebesar 74,25 %. *Script* yang digunakan untuk menampilkan nilai akurasi dapat dilihat pada gambar 6.

```
#Pengujian Model Naïve Bayes
predictions_NB = NB.predict(test_x)
print("Nilai Akurasi Naïve Bayes : ",
accuracy_score(predictions_NB, test_y)*100,"%")
```

Nilai Akurasi Naive Bayes : 74.25742574257426 %

Gambar 6. Script Akurasi Naïve Bayes

Sedangkan *confusion matrix* dari model Naïve Bayes dapat dilihat pada gambar 7.



Gambar 7. Confusion Matrix Naïve Bayes

Kemudian untuk *classification report* dari model Naïve Bayes dapat dilihat pada gambar 8.

	precision	recall	f1-score	support
0	1.00	0.13	0.23	38
1	0.73	0.99	0.84	134
2	0.87	0.43	0.58	30
accuracy			0.74	202
macro avg	0.86	0.52	0.55	202
weighted avg	0.80	0.74	0.68	202

Gambar 8. Classification Report Naïve Bayes

3.15.2 Klasifikasi Support Vector Machine

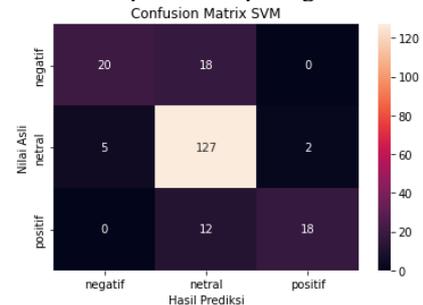
Akurasi yang didapatkan dari model yang dilatih menggunakan algoritma *Support Vector Machine* adalah sebesar 81,68 %. *Script* yang digunakan untuk menampilkan nilai akurasi dapat dilihat pada gambar 9.

```
#Pengujian Model SVM
predictions_SVM = SVM.predict(test_x)
print("Nilai Akurasi SVM : ", accuracy_score(predictions_SVM, test_y)*100,"%")
```

Nilai Akurasi SVM : 81.68316831683168 %

Gambar 9. Script Akurasi SVM

Sedangkan *confusion matrix* dari model *Support Vector Machine* dapat dilihat pada gambar 10.



Gambar 10. Confusion Matrix Support Vector Machine

Kemudian untuk *classification report* dari model *Support Vector Machine* dapat dilihat pada gambar 11.

	precision	recall	f1-score	support
0	0.80	0.53	0.63	38
1	0.81	0.95	0.87	134
2	0.90	0.60	0.72	30
accuracy			0.82	202
macro avg	0.84	0.69	0.74	202
weighted avg	0.82	0.82	0.81	202

Gambar 11. Classification Report Support Vector Machine

3.15.3 Klasifikasi Decision Tree

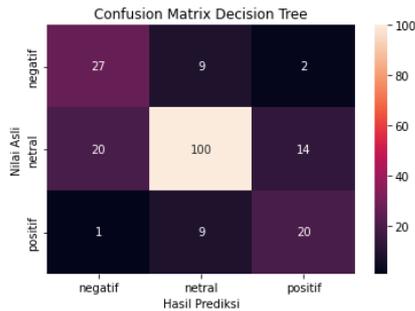
Akurasi yang didapatkan dari model yang dilatih menggunakan algoritma *Decision Tree* adalah sebesar 72,77 %. *Script* yang digunakan untuk menampilkan nilai akurasi dapat dilihat pada gambar 12.

```
#Pengujian Model Decision Tree
predictions_DT = DT.predict(test_x)
print("Nilai Akurasi Decision Tree : ", accuracy_score(predictions_DT, test_y)*100,"%")
```

Nilai Akurasi Decision Tree : 72.77227227227277 %

Gambar 12. Kode Program Akurasi Decision Tree

Sedangkan *confusion matrix* dari model *Decision Tree* pada gambar 13 yang ditunjukkan di bawah ini.



Gambar 13. Confusion Matrix Decision Tree

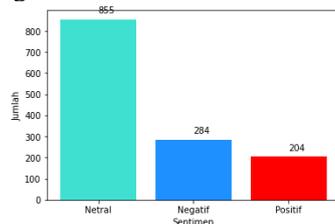
Kemudian untuk *classification report* dari model Decision Tree dapat dilihat pada gambar 14.

	precision	recall	f1-score	support
0	0.52	0.68	0.59	38
1	0.86	0.74	0.80	134
2	0.59	0.73	0.66	30
accuracy			0.73	202
macro avg	0.66	0.72	0.68	202
weighted avg	0.76	0.73	0.74	202

Gambar 14. Classification Report Decision Tree

3.16 Visualisasi Data

Pada tahap ini data *Tweet* yang sudah memiliki label sentimen ditransformasikan ke dalam bentuk visual dengan cara memetakan data *Tweet* ke dalam diagram (*chart*) dan ke dalam bentuk *word cloud*. Diagram yang digunakan adalah *bar chart* pada gambar 15.

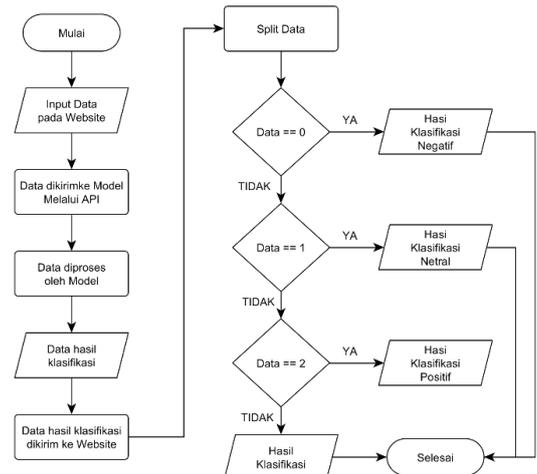


Gambar 15. Visualisasi Bar Chart

3.17 Implementasi Rancangan

Implementasi rancangan dibuat dengan menggunakan bahasa pemrograman PHP dan kerangka HTML. Tampilan pada aplikasi dibuat dengan menggunakan *form* yang terdapat pada HTML dengan *method post*. *Method post* ini berfungsi untuk melakukan pertukaran data (mengambil dan mengirim) antara halaman *website* dengan model yang sudah dibuatkan dalam bentuk API.

Kemudian proses dari aplikasi ini berawal dari *input* data yang dikirimkan ke model dan model mengembalikan data berupa angka yang mewakili sentimen. Untuk angka 0 berarti sentimen bernilai negatif, untuk angka 1 berarti sentimen bernilai netral, dan untuk angka 2 berarti sentimen bernilai positif. Alur proses dari aplikasi pada gambar 16.



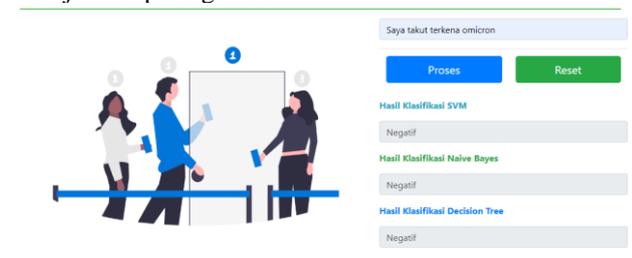
Gambar 16. Flowchart Rancangan Aplikasi

4. PEMBAHASAN

Pada bagian ini merupakan hasil dari implementasi '*sentiment analysis*' diantaranya :

4.1 Implementasi

Untuk melakukan klasifikasi sentimen, dapat memasukkan *input* berupa *text* dan menekan tombol proses. Setelah itu aplikasi akan menampilkan *output* dengan masing-masing algoritma klasifikasi (*Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree*). Pada aplikasi, *output* yang akan dihasilkan berupa keterangan mengenai sentimen, dapat berupa "Positif", "Netral", atau "Negatif". Tampilan aplikasi ditunjukkan pada gambar 17.



Gambar 17. Hasil Implementasi

4.2 Pengujian Aplikasi

Pada bagian ini akan dilakukan pengujian pada implementasi '*sentiment analysis*' pada opini masyarakat Indonesia di *twitter* terhadap virus COVID-19 varian omicron, Hasil uji coba dari aplikasi yang dibangun dapat dilihat pada tabel 1 berikut dengan keterangan positif (+), netral (=), dan negatif (-).

Tabel 1. Hasil Uji Coba

<i>Input</i>	<i>SVM</i>	<i>NB</i>	<i>DT</i>	<i>Aktual</i>
Saya berani terkena Covid-19 varian <i>Omicron</i>	=	=	+	+
Saya sedih karena Covid-19 varian <i>Omicron</i>	-	=	=	-
Saya panjatkan puji syukur karena Covid-19 varian <i>Omicron</i>	+	=	=	+
Covid-19 varian <i>Omicron</i> bahaya untuk Kesehatan	-	-	-	-
Sakit yang disebabkan oleh Covid-19 varian <i>Omicron</i> cukup parah	-	-	-	-
Saya khawatir terkena Covid-19 varian <i>Omicron</i>	-	=	-	-
Covid-19 varian <i>Omicron</i> membuat saya semangat	+	=	=	+
Covid-19 varian <i>Omicron</i> membuat kita sehat	+	=	+	+
Covid-19 varian <i>Omicron</i> membuat kita patuh pada peraturan	+	+	=	+
Covid-19 varian <i>Omicron</i> membuat kita disiplin	+	=	=	+
Covid-19 varian <i>Omicron</i> membuat saya malas bekerja	-	=	=	-

5. KESIMPULAN

Pada analisis sentimen yang dilakukan terhadap pengguna *Twitter*, didapatkan sentimen netral dengan jumlah tertinggi yaitu sebesar 63,7%, kemudian sentimen negatif sebesar 21,1%, dan sentimen positif sebesar 15,2%. Algoritma dengan akurasi tertinggi pada analisis yang dilakukan adalah algoritma *Support Vector Machine* dengan nilai akurasi sebesar 81,68%. Untuk algoritma *Naïve Bayes* menghasilkan nilai akurasi sebesar 74,25%, dan nilai akurasi yang didapatkan oleh algoritma *Decision Tree* sebesar 72,77%. Pada ketiga algoritma tersebut digunakan data uji sebanyak 202 data.

6. SARAN

Saran untuk penelitian selanjutnya dapat dilakukan *labeling* dengan metode lain seperti menggunakan kamus *Lexicon* atau pembobotan TF-IDF sehingga proses *labeling* menjadi lebih efisien. Selain itu pada penelitian selanjutnya juga dapat digunakan algoritma atau metode lain yang tidak memiliki ketergantungan pada *dataset* seperti CNN.

7. DAFTAR PUSTAKA

Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 8(2), 183–190.

de Araujo, A. L., Hardell, C., Koszek, W. A., Wu, J., & Willeminck, M. J. (2022a). Data Preparation for Artificial Intelligence. Dalam *Artificial Intelligence in Cardiothoracic Imaging* (hlm. 37–43). Springer.

de Araujo, A. L., Hardell, C., Koszek, W. A., Wu, J., & Willeminck, M. J. (2022b). Data Preparation for Artificial Intelligence. Dalam *Artificial Intelligence in Cardiothoracic Imaging* (hlm. 37–43). Springer.

Fikriya, Z. A., Irawan, M. I., & Soetrisno, S. (2017). Implementasi extreme learning machine untuk pengenalan objek citra digital. *Jurnal Sains Dan Seni ITS*, 6(1), A1–A6.

Herwanto, H. (2019). Diagnosa Statistik Pemetaan Pemahaman Bahasa Pemrograman Sebagai Acuan Untuk Mempersiapkan Penelitian Mahasiswa. *NUANSA INFORMATIKA*, 13(2), 33.

Musa’adah, Y., Wihardi, Y., & others. (2018). Klasifikasi Komentar Spam pada Youtube Menggunakan Metode Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbors. *JIKO (Jurnal Informatika Dan Komputer)*, 3(2), 54–59.

Pamungkas, F. S., Prasetya, B. D., & Kharisudin, I. (2020). Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 692–697.

Parapat, I. M. (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. Universitas Brawijaya.

Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156–165.

Santoso, R. R. (2020). Implementasi Metode Machine Learning Menggunakan Algoritma Evolving Artificial Neural Network Pada Kasus Prediksi Diagnosis Diabetes. Universitas Pendidikan Indonesia.

Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1–11.

Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(2), 681–686.

- Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 417–426.
- Susilo, A., Jasirwan, C. O. M., Wafa, S., Maria, S., Rajabto, W., Muradi, A., Fachriza, I., Putri, M. Z., & Gabriella, S. (2022). Mutasi dan Varian Coronavirus Disease 2019 (COVID-19): Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 9(1), 59–81.
- Swedia ER, Fitriani RR, Cahyanti M, Ernastuti, & Septian MRD. (2022). Feed Forward Neural Network untuk Prediksi Data mplementasi dengan Python dan Flask API pada Sistem Operasi Windows (1 ed., Vol. 978-623-351-612-9). PT. Nas Media Indonesia.
- Wijaya, Y. A. (2021). Analisa Klasifikasi menggunakan Algoritma Decision Tree pada Data Log Firewall. *JURSIMA (Jurnal Sistem Informasi Dan Manajemen)*, 9(3), 256–264.
- Zukhrufillah, I. (2018). Gejala media sosial twitter sebagai media sosial alternatif. *Al-I'lam: Jurnal Komunikasi Dan Penyiaran Islam*, 1(2), 102–109.