

# ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP WABAH COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE

Harun Sujadi <sup>1</sup>, Sandi Fajar <sup>2</sup>, Cecep Roni <sup>3</sup>

<sup>1,2</sup>Program Studi Informatika Fakultas Teknik Universitas Majalengka

<sup>3</sup>Program Studi Informatika Fakultas Teknik Universitas Majalengka

Email: <sup>1</sup>harunsujadi@unma.ac.id, <sup>3</sup>ceceproni.cr@gmail.com

## ABSTRACT

Twitter is often used to express opinions about a topic or issue that is trending. In the early 2020 period in Indonesia, Twitter was enlivened by the issue of the COVID-19 virus caused by SARS-CoV-2. Many Twitter users have expressed their views on the COVID-19 issue, which has attracted the attention of several parties to be used as a reference in making new decisions or policies. Therefore, it is necessary to do a sentiment analysis to determine the polarity of the sentiments that are in the contents of the tweets. This study uses the Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) methods. with a total dataset of 1652 tweets. From the results of classification using the NBC method, the classification accuracy value is 78.3%. While the accuracy value obtained by the SVM method is 81.6%. While the results of the accuracy test using the Cross Validation method with 10 K-Fold CV results in an average accuracy value of the NBC method of 69.8% and an average accuracy value of the SVM method of 74.4%. It can be concluded that the SVM method is proven to have a higher accuracy value than the NBC method.

**Keywords:** sentiment analysis, covid-19, naïve bayes classifier, support vector machine

## 1. PENDAHULUAN

### 1.1. Latar Belakang

Saat ini dunia sedang dihadapkan dengan sebuah wabah yang sangat meresahkan, wabah tersebut telah menyebar luas ke berbagai belahan dunia sehingga pada 11 Maret 2020, WHO menetapkan wabah ini sebagai pandemi global. Wabah yang dimaksud tidak lain adalah COVID-19 (*Corona Virus Disease 2019*) yang merupakan sebuah penyakit menular yang disebabkan oleh jenis *coronavirus* yang baru ditemukan. Gejala COVID19 umumnya berupa demam 38°C, batuk kering, dan sesak nafas serta dampak paling buruk untuk manusia ialah kematian (Setiawan, 2020). Pandemi global yang terjadi pula di Indonesia ini memunculkan berbagai persoalan yang berpengaruh kepada hampir seluruh aspek kehidupan manusia, seperti sektor pendidikan, politik, transportasi, dan sektor perekonomian (Buana, 2020).

Twitter merupakan salah satu media sosial yang memungkinkan penggunaanya untuk mengirim dan membaca pesan yang berbasis teks. Pengguna Twitter banyak membicarakan isu-isu hangat dari berbagai dimensi, mulai dari *entertainment*, politik, sosial, hingga pemerintahan (Nooraeni, dkk, 2019). Pada periode awal tahun 2020, di Indonesia, Twitter diramaikan dengan isu tentang wabah Covid-19. Banyak yang menuliskan pendapatnya mengenai isu tersebut baik berupa opini positif, negatif dan ada juga yang netral. Hal ini menjadi menarik dan penting bagi pihak-pihak tertentu yang ingin mengetahui baik buruknya sentimen atau opini yang diberikan masyarakat Indonesia terhadap wabah Covid-19 tersebut.

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif (Simorangkir & Lhaksmana, 2018). Salah satu metode klasifikasi yang sering digunakan dalam proses analisis sentimen adalah *Support Vector Machine* dan *Naive Bayes Classifier*. Penggunaan metode ini dinilai sesuai karena kemudahan penggunaannya, waktu pemrosesannya yang cepat, dan memiliki tingkat akurasi yang tinggi (Rodiyansyah & Winarko, 2013).

### 1.2. Tinjauan Pustaka

#### a. Sentimen Analisis

Sentimen analisis atau bisa disebut juga opinion mining, adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya.

#### b. Text Preprocessing

*Preprocessing* merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah yang diproses oleh sistem.

c. TFIDF

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen (Nurjannah, Hamdani, & Astuti, 2013).

Berikaut adalah persamaan yang digunakan untuk mencari bobot kata dengan metode *term frequency inverse document frequency* (tfid) :

Perhitungan *term frequency* (tf)

$$tf = tf_{i,j} \tag{1}$$

Perhitungan *inverse document frequency* (idf)

$$idf_i = \log \frac{N}{df_i} \tag{2}$$

Perhitungan *term frequency inverse document frequency* (tfidf)

$$W_{i,j} = tf_{i,j} \cdot \log \left( \frac{N}{df_i} \right) \tag{3}$$

d. Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (Prasetyo, 2012) :

- a. Pembangunan model sebagai prototipe untuk disimpan sebagai memori.
- b. Penggunaan model tersebut untuk melakukan pengenalan / kladifikasi / prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya.

e. *Naïve Bayes Classifier*

*Naïve Bayes Classifier* mengasumsikan bahwa keberadaan maupun ketidakberadaan sebuah fitur dalam sebuah kelas tidak memiliki keterkaitan dengan keberadaan maupun ketidakberadaan fitur lainnya. Walaupun fitur ini bergantung antara satu fitur dengan fitur yang lainnya. *Naïve Bayes Classifier* akan tetap menganggap bahwa fitur – fitur tersebut independen dan tidak memiliki pengaruh satu sama lainnya (Rocha, 2006).

Secara umum *Naïve Bayes* dapat dinotasikan pada persmaan 2.3 berikut:

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)} \tag{4}$$

f. *Support vector machine*

*Support vector machine* (SVM) adalah suatu metode untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM memiliki prinsip dasar *linier classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem *non-linier* dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Fungsi kernel yang digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) himpunan data ke dimensi baru (dimensi yang relatif lebih tinggi) (Santosa, 2007).

g. *Cross-Validation*

*Cross-validation* (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma. *Cross-validation* membagi data menjadi dua subset yaitu data training dan data testing. Model dilatih oleh subset training dan diverifikasi oleh subset testing. Selain itu, pemilihan tipe CV dapat didasarkan pada ukuran kumpulan dataset. CV K-fold biasanya digunakan karena dapat mengurangi waktu kalkulasi dengan tetap menjaga keakuratan estimasi (Widjaya, Hiryanto, & Handhayani, 2017)

A. Penelitian Terkait

Guna mendapatkan hasil penelitian yang optimal harus melakukan kajian dari penelitian-penelitian yang terdahulu yang *linier* dengan penelitian ini sehingga bisa dijadikan referensi dalam penelitian yang sekarang dilakukan. Berikut penelitian tinjauan *linier* dengan penelitian ini:

Pada penelitian yang berjudul “Analisis Sentimen Publik Terhadap Sistem Zonasi Sekolah Menggunakan Data *Twitter* Dengan Metode *Naïve Bayes Classification*” yang ditulis oleh Rani Nooraeni, Amirah Balqis Safiruddin, Aulia Fatin Afifah, Krisna Dwi Agung Wijaya dan Nada Nabila Rosyad yang ditulis tahun 2019. Membahas tentang analisis sentimen *public* terhadap penerapan sistem zonasi sekolah menggunakan dataset dari *Twitter*. Persamaan variabel yang terkait adalah tentang penggunaan algoritma *Naïve Bayes*.

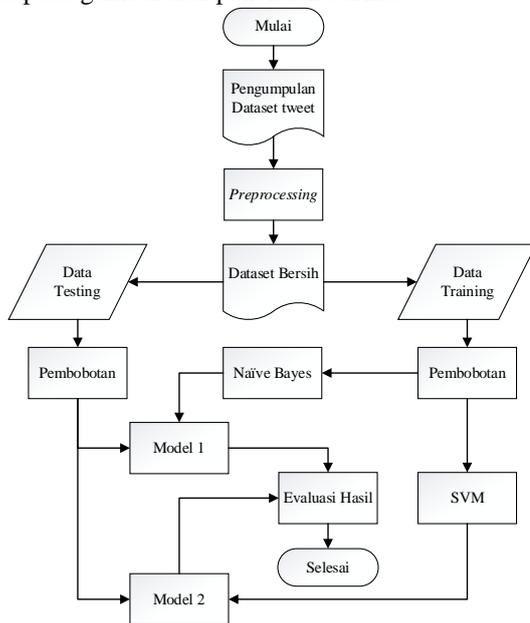
Adapun penelitian yang digagas oleh Ghulam Asrofi Buntoro yang berjudul “Analisis Sentimen Hatespeech Pada *Twitter* Dengan Metode *Naïve Bayes Classifier* Dan *Support Vector Machine*” pada tahun 2016. Menganalisis tren tagar Hatespeech (*#HateSpeech*) pada *Twitter*, benarkan semua yang diberi tagar tersebut sentimennya kebencian. Proses klasifikasi pada penelitian ini menggunakan metode klasifikasi *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) dengan preprocessing data menggunakan tokenisasi, cleansing dan filtering. Persamaan variabel yang terkait dengan penelitian ini adalah penggunaan algoritma yang digunakan.

Pada penelitian dari Wanda Athira Luqyana, Imam Cholissodin dan Rizal Setya Perdana yang berjudul “Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi *Support Vector Machine*” tahun 2018 dilakukan suatu

analisis sentimen pada kolom komentar Instagram yang berupaya untuk mengetahui sentimen dari setiap komentar yang diberikan. Untuk mengetahui setiap sentimen pada komentar digunakan fitur Term Frequency-Inverse Document Frequency (TF-IDF) dan metode klasifikasi Support Vector Machine (SVM). Variabel yang terkait disini adalah analisis sentiment pada media sosial.

**1.3. Metodologi Penelitian**

Dalam metode penelitian tugas akhir ini akan diuraikan kedalam beberapa bagian seperti yang tertera pada gambar alur penelitian berikut:



Gambar 1. Alur Penelitian

Dari gambar 1 diatas dapat dijelaskan bahwa untuk melakukan penelitian analisis sentimen ini akan dilakukan dalam beberapa langkah yang terdiri dari pengumpulan data, *preprocessing* data, *stopword removal*, *stemming*, pembobotan dengan TFIDF, lalu proses klasifikasi dengan metode *Naïve Bayes Classifier* dan *Support Vector Machine*, selanjutnya proses evaluasi hasil dengan menentukan nilai akurasi, *precision*, *recall* dan *f-score* dari confusion matics yang diperoleh dari hasil klasifikasi dan terakhir akan dilakukan proses pengujian model dengan metode *cross validation* untuk mengetahui metode mana yang memiliki akurasi yang lebih tinggi.

**2. PEMBAHASAN**

**A. Pengumpulan data**

Penelitian ini menggunakan data sekunder yaitu data *Twitter*. Data yang digunakan pada penelitian ini adalah Dataset *Sentiment Analysis* pada Covid-19 yang didapat dari *bisa.ai*. Data tersebut merupakan data kotor sebanyak 1652 *tweets*.

**B. Preprocessing Data**

Proses *preprocessing* terdiri dari beberapa tahapan yang terdiri dari *cleansing*, *case folding*, *tokenizing*, *stopwords removal* dan *stemming*. Pada proses *stemming* digunakan kamus *slangword* "slangword.txt" yang dikumpulkan oleh Nikmatun Aliyah Salsabila, dkk dalam publikasinya yang berjudul "*Colloquial Indonesian Lexicon*" dan pada proses *stopword* digunakan kamus *stoplist* "Stopword List Bahasa Indonesia.txt" yang dikumpulkan sendiri. Proses tersebut didapatkan hasil yang akan digambarkan dengan *wordcloud* sebagai berikut:



Gambar 2. Wordcloud Tahap Preprocessing

**C. Pembagian Dataset**

Pada tahap ini dataset hasil *preprocessing* akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*, dengan perbandingan dataset yaitu 10% data *testing* dan 90% data *training*.

**D. TF-IDF**

Pada tahap ini digunakan function *TfidfVectorizer* dari *library sklearn* untuk melakukan proses pembobotan kata. Setelah dilakukan proses TFIDF dan telah berhasil didapatkan nilai bobot dari setiap kata pada masing-masing kalimat *tweet*, maka akan dilakukan proses pengklasifikasian data sentimen

**E. Klasifikasi**

Pada proses klasifikasi digunakan dua buah metode klasifikasi yaitu *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM). Berikut merupakan implementasi dari masing-masing metode klasifikasi yang digunakan:

**a. Bayes Classifier (NBC)**

Pengklasifikasian sentimen berdasarkan NBC ini dilakukan dengan cara menghitung probabilitas dokumen data testing dengan merujuk pada probabilitas kata pada data training. Proses NBC ini dilakukan dengan menggunakan *function* *GaussianNB* dari *library sklearn*. berikut merupakan kode program pengaplikasian metode NBC:

```
# model NBC
model_nbc = GaussianNB()
model_nbc.fit(X_train.todense(), y_train)
model_nbc.score(X_test.todense(), y_test)

0.7833333333333333

score = model_nbc.score(X_test.todense(), y_test)
print("Score klasifikasi NBC adalah : ",score)

Score klasifikasi NBC adalah : 0.7833333333333333
```

Gambar 3. Klasifikasi Model NBC

Dari gambar program diatas dapat disimpulkan bahwa proses pengklasifikasian metode *Naïve Bayes Classifier* telah berhasil dilakukan dengan diperoleh hasil akurasi sebesar 78,3%.

b. *Support Vector Machine (SVM)*

Pada proses ini digunakan metode klasifikasi *Support Vector Machine (SVM)*. Pengklasifikasian sentimen berdasarkan SVM ini dilakukan dengan menggunakan *function svm.SVC* dari *library sklearn*. berikut merupakan kode program pengaplikasian metode SVM :

```
# model svm
model_svm = svm.SVC(kernel='linear')

model_svm.fit(X_train, y_train)
model_svm.score(X_test, y_test)

0.8166666666666667

score_svm = model_svm.score(X_test, y_test)
print("Score klasifikasi SVM adalah : ",score_svm)

Score klasifikasi SVM adalah : 0.8166666666666667
```

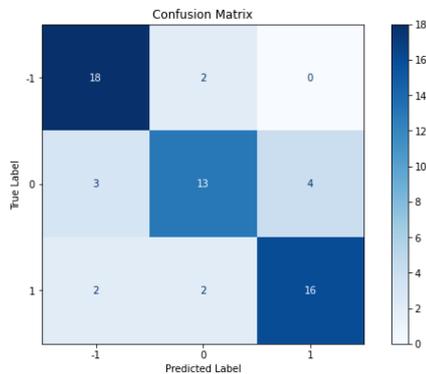
Gambar 4. Klasifikasi Model SVM

Dari gambar program diatas dapat disimpulkan bahwa proses pengklasifikasian metode *Support Vector Machine* telah berhasil dilakukan dengan diperoleh hasil akurasi sebesar **81,6%**.

F. Evaluasi Hasil

Berkut merupakan hasil *confusion matrices* dari masing-masing model yang digunakan pada penelitian ini yaitu *Naïve Bayes Classifier* dan *Support Vector Machine*.

a. Berikut merupakan *confusion matrices* yang diperoleh dari hasil klasifikasi model *Naïve Bayes Classifier* :



Gambar 5. *Confusion Matrics* Model NBC

Pada hasil prediksi positif terdapat 18 *tweet* terklasifikasi secara benar positif dan 2 terklasifikasi netral. Pada hasil klasifikasi netral terdapat 13 *tweet* terklasifikasi secara benar, 3 terklasifikasi salah pada positif dan 4 terklasifikasi salah pada negatif. Sementara pada hasil prediksi negatif terdapat 16 *tweet* terklasifikasi benar negatif dan terdapat 2 *tweet*

terklasifikasi salah pada netral dan 2 *tweet* juga terklasifikasi salah pada netral.

Dari hasil *confusion matrices* diatas diperoleh hasil akurasi, *precision*, *recall*, dan *f1-score* seperti pada gambar berikut :

	precision	recall	f1-score	support
-1	0.82	0.90	0.86	20
0	0.78	0.70	0.74	20
1	0.85	0.85	0.85	20
accuracy			0.82	60
macro avg	0.82	0.82	0.81	60
weighted avg	0.82	0.82	0.81	60

	precision	recall	f1-score	support
-1	0.78	0.90	0.84	20
0	0.76	0.65	0.70	20
1	0.80	0.80	0.80	20
accuracy			0.78	60
macro avg	0.78	0.78	0.78	60
weighted avg	0.78	0.78	0.78	60

Gambar 6. Hasil Akurasi Model NBC

Nilai rata-rata *precision* yang diperoleh adalah :

$$Precision = \frac{20}{60} \times 0.78 + \frac{20}{60} \times 0.76 + \frac{20}{60} \times 0.80 = 0.779 = 77.9\%$$

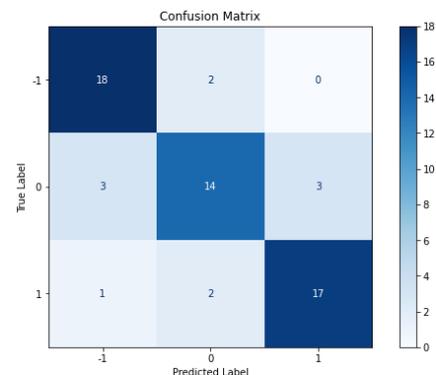
Nilai rata-rata *recall* yang diperoleh adalah :

$$Recall = \frac{20}{60} \times 0.90 + \frac{20}{60} \times 0.65 + \frac{20}{60} \times 0.80 = 0.782 = 78.2\%$$

Nilai rata-rata *f1-score* yang diperoleh adalah :

$$F1 - Score = 0.84 + 0.70 + 0.80 = 234 / 3 = 0.78 = 78\%$$

b. Berikut merupakan *Confusion Matrics* yang diperoleh dari hasil klasifikasi model *Support Vector Machine* :



Gambar 7. *Confusion Matrics* Model SVM

Pada hasil prediksi positif terdapat 18 *tweet* terklasifikasi secara benar positif dan 2 terklasifikasi netral. Pada hasil klasifikasi netral terdapat 14 *tweet*

terklasifikasi secara benar, 3 terklasifikasi salah pada positif dan 3 terklasifikasi salah pada negatif. Sementara pada hasil prediksi negatif terdapat 17 tweet terklasifikasi benar negatif dan terdapat 2 tweet terklasifikasi salah pada netral dan 1 tweet terklasifikasi salah pada netral.

Dari hasil confusion matrices diatas diperoleh hasil akurasi, precision, recall, dan f1-score seperti pada gambar berikut :

Gambar 8. Hasil Akurasi Model SVM

Nilai rata-rata *precision* yang diperoleh adalah :

$$Precision = \frac{20}{60} \times 0.82 + \frac{20}{60} \times 0.78 + \frac{20}{60} \times 0.85 = 0.816 = 81.6\%$$

Nilai rata-rata *recall* yang diperoleh adalah :

$$Recall = \frac{20}{60} \times 0.90 + \frac{20}{60} \times 0.70 + \frac{20}{60} \times 0.85 = 0.816 = 81.6\%$$

Nilai rata-rata *f1-score* yang diperoleh adalah :

$$F1 - Score = 0.86 + 0.74 + 0.85 = 2.46$$

$$F1 - Score = \frac{2.46}{3} = 0.82 = 82\%$$

### 3. ALGORITMA ATAU PROGRAM

#### G. Pengujian Model

Proses pengujian dilakukan menggunakan metode *cross-validation*. K-fold yang digunakan adalah 10 K-Fold CV, berikut merupakan hasil dari proses pengujian cross-validation.

Menguraikan hasil analisis kualitatif dan/atau kuantitatif dengan penekanan pada jawaban atas permasalahan. Isi dari pembahasan ini memuat segala sesuatu tentang kegiatan yang dilakukan dalam makalah. Mulai dari konsep, perancangan, hipotesis (bila ada), percobaan, data pengamatan, dan hasil dari data pengamatan yang ada.

Isi didukung dengan gambar, tabel, dan persamaan yang dirujuk dalam naskah.

#### a. Pengujian *Cross Validation* pada model *Naif Bayes Classifier*

```
# Cross Validation NBC
cv_result_nbc = cross_val_score(model_nbc, X_train.todense(), y_train, cv=10)
for i in range(10):
    print("Nilai akurasi prediksi sentimen dengan model NBC adalah : {0:2%}".format(cv_result_nbc[i]))
print()
print("Nilai rata-rata akurasi model NBC adalah : ", cv_result_nbc.mean())

Nilai akurasi prediksi sentimen dengan model NBC adalah : 72.222222%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 66.666667%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 81.481481%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 59.259259%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 81.481481%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 61.111111%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 62.962963%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 70.370370%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 75.925926%
Nilai akurasi prediksi sentimen dengan model NBC adalah : 66.666667%

Nilai rata-rata akurasi model NBC adalah : 0.6981481481481482
```

Gambar 9. *Cross Validation* Model NBC

Dari pengujian *cross validation* pada gambar 4.24 diperoleh hasil akurasi rata-rata yaitu **69,8%**.

#### b. Pengujian *Cross Validation* pada model *Support vector machine*

```
# Cross Validation SVM
cv_result_svm = cross_val_score(model_svm, X_train.todense(), y_train, cv=10)
for i in range(10):
    print("Nilai akurasi prediksi sentimen dengan model SVM adalah : {0:2%}".format(cv_result_svm[i]))
print()
print("Nilai rata-rata akurasi model SVM adalah : ", cv_result_svm.mean())

Nilai akurasi prediksi sentimen dengan model SVM adalah : 72.222222%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 64.814815%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 81.481481%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 64.814815%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 92.592593%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 72.222222%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 66.666667%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 74.074074%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 83.333333%
Nilai akurasi prediksi sentimen dengan model SVM adalah : 72.222222%

Nilai rata-rata akurasi model SVM adalah : 0.7444444444444444
```

Gambar 10. *Cross Validation* Model SVM

Dari pengujian *cross validation* pada gambar 4.25 diperoleh hasil akurasi rata-rata yaitu **74,4%**.

## 4. KESIMPULAN

### A. Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, penulis dapat mengambil kesimpulan bahwa :

- a. Proses pengklasifikasian sentimen covid 19 dengan metode *Naïve Bayes Classifier* dan *Support Vector Machine* dapat dilakukan dengan tahapan-tahapan sebagai berikut :
  - 1) Preprocessing data
  - 2) Pembobotan kata dengan metode TFIDF
  - 3) Pengklasifikasian sentimen dengan metode *Naïve Bayes Classifier* dan *Support Vector Machine*
  - 4) Terakhir mengevaluasi hasil dengan melakukan perhitungan akurasi, *precision*, *recall* dan *f-score* dari *confusion matrix* yang diperoleh dari proses klasifikasi.
- b. Metode *Naïve Bayes Classifier* menghasilkan nilai akurasi pengklasifikasian sebesar **78,3%**. Sedangkan nilai akurasi yang diperoleh dari hasil klasifikasi dengan metode *Support Vector Machine* adalah sebesar **81,6%**. Sementara hasil pengujian akurasi menggunakan metode *Cross Validation* dengan 10 K-Fold CV menghasilkan nilai rata-rata akurasi pada metode *Naïve Bayes Classifier* sebesar **69,8%** dan nilai rata-rata akurasi pada metode *Support Vector Machine* sebesar **74,4%**. Dari hasil perbandingan kedua nilai akurasi tersebut dapat disimpulkan bahwa metode *Support Vector Machine* terbukti memiliki nilai akurasi yang lebih tinggi dari pada metode *Naïve Bayes Classifier*

### B. Saran

Agar diperoleh hasil yang lebih maksimal, penulis memiliki beberapa saran yang dapat digunakan untuk pengembang lebih lanjut, yaitu :

- a. Jumlah dataset tweets yang digunakan lebih diperbanyak lagi, bisa ditambahkan juga dataset dari berbagai media sosial lain selain twitter sehingga sistem pengklasifikasian bisa lebih akurat lagi
- b. Jumlah pembendeharaan kata untuk proses stemming dan stopword lebih diperbanyak lagi agar proses preprocessing data bisa lebih baik.
- c. Hasil penelitian analisis sentimen terhadap wabah covid-19 ini diharapkan menjadi acuan dan suatu kajian studi untuk menghasilkan kebijakan dan pengambilan keputusan baru oleh pemerintah terkait.

### PUSTAKA

- BNPB. (2020, 6). *Data Sebaran*. Retrieved from covid19.go.id: <https://covid19.go.id/>
- Buana, A. R. (2020). Problematika Regulasi Ojek Online Dalam Masa Pembatasan Sosial Berskala Besar Covid-19 . *Adalah: Buletin Hukum dan Keadilan*.
- Clinten, B. (2019, 10). *Tekno : Pengguna Aktif Harian Twitter Indonesia*. Retrieved from tekno.kompas.com: <https://tekno.kompas.com/read/2019/10/30/16062477/pengguna-aktif-harian-twitter-indonesia-diklaim-terbanyak>
- Hadiwardoyo, W. (2020). Kerugian Ekonomi Nasional Akibat Pandemi CoviD-19 . *Baskara : Journal of Business and Entrepreneurship* .
- Kemkes. (2020, April Selasa). *Menkes Tetapkan PSBB untuk DKI Jakarta*. Retrieved from www.kemkes.go.id: <https://www.kemkes.go.id/article/view/20040700003/-menkes-tetapkan-psbb-untuk-dki-jakarta.html>
- Lee, E. (2018). Cyber physical systems: Design challenges. In Object Oriented Real-Time Distributed Computing (ISORC. *11th IEEE International Symposium*, 363-369.
- Nooraeni, R., & dkk. (2019). Analisis Sentimen Publik Terhadap Sistem Zonasi Sekolah Menggunakan Data Twitter Dengan Metode Naïve Bayes Classification. *Faktor Exacta*.
- Nurjannah, M., Hamdani, & Astuti, I. F. (2013). Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Text Mining. *Jurnal Informatika Mulawarman*, 110.
- Prasetyo, E. (2012). *Data mining konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi. Jakarta: Pusat Data dan Informasi.
- Rocha, A. D. (2006). *Naive Bayes Classifier Teaching Material*. Retrieved from www.ic.unicamp.br: <http://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906/aulas/naive-bayes-classifier.pdf>.
- Rodiyansyah, S. F., & Winarko, E. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *IJCCS*.
- Santosa, B. (2007). *Tutorial Support Vector Machine*. Surabaya: ITS.
- Setiawan, A. R. (2020). Lembar Kegiatan Literasi Saintifik untuk Pembelajaran Jarak Jauh Topik Penyakit Coronavirus 2019 (COVID-19) . *EDUKATIF: JURNAL ILMU PENDIDIKAN* , 2.
- Simorangkir, H., & Lhaksmana, K. M. (2018). Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier. *e-Proceeding of Engineering*.
- Syadid, F. (2019). Analisis Sentimen Komentar Nitizen Terhadap Calon Presiden Indonesia 2019 dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (TFIDF) dan Metode Multi Layer Perceptron (MLP) Neural Network. *Respositori UIN Syarif Hidayatullah Jakarta*.
- WHO. (2020). *Pertanyaan dan jawaban terkait Coronavirus*. Retrieved from who.int: <https://www.who.int/indonesia/news/novel-coronavirus/qa-for-public>
- Widjaya, A., Hiryanto, L., & Handhayani, T. (2017). Prediksi Masa Studi Mahasiswa Dengan Voting Feature Interval 5 Pada Aplikasi Konsultasi Akademik Online. *Journal of Computer Science and Information Systems*.