



# ***JITE (Journal of Informatics and Telecommunication Engineering)***

Available online <http://ojs.uma.ac.id/index.php/jite> DOI : 10.31289/jite.v4i1.3798

Received: 21 Mei 2020

Accepted: 30 Juni 2020

Published: 20 Juli 2020

## ***Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm***

**Ricky Risnantoyo<sup>1)\*</sup>, Arifin Nugroho<sup>1)</sup> & Kresna Mandara<sup>1)</sup>**

1) Program Studi Magister Ilmu Komputer, Fakultas Ilmu Komputer, STMIK Nusa Mandiri, Indonesia

\*Corresponding Email: 14002301@nusamandiri.ac.id

### **Abstrak**

Abstrak – Wabah virus corona yang terjadi di hampir seluruh negara di dunia menimbulkan dampak tidak hanya di sektor kesehatan, namun juga di sektor lainnya seperti pariwisata, keuangan, transportasi, dll. Hal ini memunculkan berbagai macam sentimen dari publik dengan munculnya corona virus sebagai trending topik di media sosial Twitter. Twitter dipilih oleh publik karena dapat menyebarkan informasi secara real time dan dapat melihat reaksi pasar secara cepat. Pada penelitian ini digunakan data “tweet” atau kicauan publik terkait “Corona Virus” untuk melihat bagaimana polaritas sentimen yang muncul. Digunakan teknik text mining dan tiga algoritma klasifikasi *machine learning* antara lain *Naive Bayes*, *Support Vector Machine* (SVM), *K-Nearest Neighbour* (K-NN) untuk membangun model klasifikasi tweet terhadap sentimen apakah mempunyai polaritas positif, negatif, atau netral. Hasil pengujian tertinggi dihasilkan oleh algoritma *Support Vector Machine* (SVM) dengan nilai akurasi 76,21%, nilai precision 78,04%, dan nilai recall 71,42%.

**Kata Kunci:** *Machine Learning, Corona Virus, Twitter, Analisis Sentimen.*

### **Abstract**

*Corona virus outbreaks that occur in almost all countries in the world have an impact not only in the health sector, but also in other sectors such as tourism, finance, transportation, etc. This raises a variety of sentiments from the public with the emergence of corona virus as a trending topic on Twitter social media. Twitter was chosen by the public because it can disseminate information in real time and can see market reactions quickly. This research uses "tweet" data or public tweet related to "Corona Virus" to see how the sentiment polarity arises. Text mining techniques and three machine learning classification algorithms are used, including Naive Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) to build a tweet classification model of sentiments whether they have positive, negative, or neutral polarity. The highest test results are generated by the Support Vector Machine (SVM) algorithm with an accuracy value of 76.21%, a precision value of 78.04%, and a recall value of 71.42%.*

**Keywords:** *Machine Learning, Corona Virus, Twitter, Sentiment Analysis.*

**How to Cite:** Risnantoyo, R., Nugroho, A. & Mandara, K. (2020). Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm. *JITE (Journal Of Informatics And Telecommunication Engineering)*. 4 (1): 86-96

## I. PENDAHULUAN

Pada 30 Januari 2020, Organisasi Kesehatan Dunia (WHO) sudah mendeklarasikan Covid-19 sebagai Darurat Kesehatan Masyarakat secara Global (*Global Public Health Emergency*). Corona virus adalah jenis virus yang menyebabkan penyakit mulai dari gejala ringan sampai berat. Corona virus disease 2019 (Covid-19) adalah penyakit jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia.

Dampak dari pandemi Covid-19 ini tidak hanya menyerang sektor kesehatan, namun juga menyerang sektor-sektor lainnya seperti pariwisata, keuangan, bisnis, hingga transportasi. Hal ini memunculkan berbagai macam sentimen dari berbagai pihak dari seluruh penjuru dunia karena pandemi ini berlangsung hampir di seluruh negara di dunia ini.

Twitter merupakan media sosial yang dibuat oleh Jack Dorsey pada tahun 2006. Pada tahun 2019 berdasarkan press-release twitter, ada 500 juta "tweet" atau kicauan oleh pengguna twitter per harinya (Ying, 2019). Sebanyak 500 juta tweet digunakan untuk mem-post hal tentang diri pengguna dan berbagi informasi, isi tweet juga dapat mengekspresikan perasaan (Bolen, Huina, & Xiajung 2011). Opini melalui tweet inilah yang dapat dimanfaatkan untuk melihat bagaimana

sentimen yang dimunculkan terhadap kejadian-kejadian yang menarik bagi publik salah satunya terkait Virus Corona yang menjadi trending di lini sosial media twitter.

Studi komputasional dari opini pengguna internet, sentimen dan emosi melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks disebut dengan analisis sentimen atau opinion mining (Parikh & Movassate, 2009). Analisis sentimen akan mengelompokkan (klasifikasi) polaritas dari teks ke dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam bentuk kalimat atau dokumen tersebut bersifat positif, negatif atau netral (Maning, Raghavan, & Schutze, 2008). Dalam penelitian ini analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek dalam hal ini adalah cuitan atau tweet "Corona Virus" apakah mengandung polaritas sentimen negatif, positif, atau netral.

Penelitian tentang analisis sentimen yang menggunakan dataset dari jejaring sosial Twitter berfokus pada masalah klasifikasi (Parikh & Movassate, 2009). Penelitian ini menggunakan metode *text mining* dan *machine learning* untuk mengambil dan mengklasifikasi polaritas

opini dari sumber data. Beberapa algoritma *machine learning* yang digunakan dalam analisis sentiment antara lain algoritma *Naïve Bayes* (G. Gautam, Y. Divakar, 2014), (Romelta, 2012), algoritma *Support Vector Machine* (SVM) (Nurajijah & Riana, 2019), dan *K-Nearest Neighbour* (Prasanti, Fauzi, & Furqon, 2018) (Bayhaqy, Sfenrianto, & Nainggolan, 2018) yang menghasilkan tingkat akurasi yang cukup baik. Pada penelitian ini penulis bermaksud menerapkan tiga algoritma *machine learning* sekaligus antara lain *Naive Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbour* (K-NN) untuk mencari performansi terbaik yang dihasilkan oleh algoritma tersebut apabila digunakan dalam analisis sentiment data twitter terkait “Corona Virus”.

## II. STUDI PUSTAKA

### A. Analisis Sentimen

Sentiment analyst atau analisis sentimen dalam bahasa Indonesia adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen positif maupun sentimen negatif. Hasil sistem prototipe mencapai tinggi presisi (75-95% tergantung pada data) dalam

mencari sentimen pada halaman web dan artikel berita. (Nasukawa.T.& Yi, 2003)

Menurut (Liu, 2012) sentiment analysis atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistic dan teks mining yang memiliki tujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Tugas analisis sentimen yaitu mengelompokkan teks ke dalam kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen yang dianalisis apakah bersifat positif, negatif, atau netral.

### B. Naïve Bayes

Naïve Bayes merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan pada Naïve Bayes classifier adalah Teorema Bayes yang dinyatakan pertama kali oleh Thomas Bayes [11]. Nilai probabilitas yang digunakan dinyatakan secara sederhana sebagai berikut :

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

### C. Support Vector Machine (SVM)

SVM merupakan metode yang dapat menyelesaikan permasalahan secara linier maupun permasalahan non-linier. Dalam menyelesaikan permasalahan non-linier digunakan konsep kernel pada ruang kerja berdimensi tinggi, dengan mencari hyperplane yang dapat memaksimalkan margin antar kelas data. Hyperplane berguna dalam memisahkan 2 kelompok class +1 dan class -1 dimana setiap class memiliki pattern masing-masing. Dalam mengambil keputusan dengan metode SVM digunakan fungsi kernel  $K(X_i, X_d)$ . Kernel yang digunakan pada penelitian ditunjukkan dengan persamaan :

$$K(x_i, x_d) = (X_i^T X_d + C)^{\gamma}, \gamma > 0 \quad (2)$$

### D. K-Nearest Neighbour (K-NN)

Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang

paling banyak ditemui pada k buah tetangga terdekat titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean dengan rumus seperti pada persamaan :

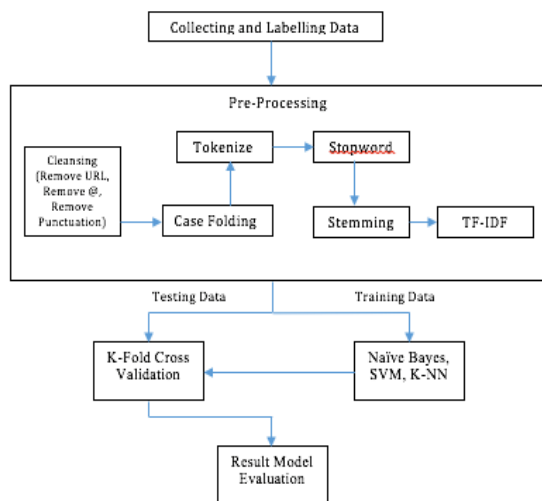
$$distance = \sqrt{\sum_{i=1}^n (X_{training}^i - X_{testing})^2} \quad (3)$$

Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data tes (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah K buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut. Nilai K yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai K yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai K yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan cross-validation.

## III. METODE PENELITIAN

Dalam analisis sentimen, terdapat beberapa tahapan yang dapat dilakukan untuk mendapatkan hasil yang terbaik.

Tahapannya terdiri atas tahap pengumpulan dan pelabelan data, lalu dilanjutkan dengan tahap *pre-processing* data (*Remove URL (Universal Resource Locator), @annotation Removal, Punctuation Removal, Case folding, Tokenize, Filter Stopword, dan Stemming*), melakukan *training data* dengan algoritma Naïve Bayes, SVM, dan K-NN, lalu melakukan *testing data* dengan metode *K-Fold Cross Validation*. Berikut tahapan metode yang dilakukan dalam penelitian ini:

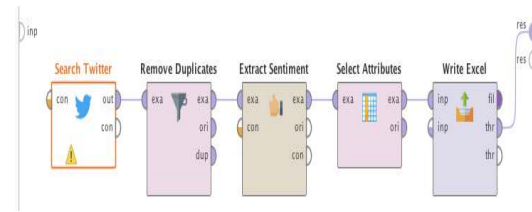


Gambar 1. Tahapan Penelitian

### A. Pengumpulan dan Pelabelan Data

Proses pengumpulan data dilakukan dengan API Twitter aplikasi Rapidminer dengan melakukan *query* “Corona Virus sebanyak kurang lebih 833 records. Lalu selanjutnya data akan diberikan label/scoring sentimen (Positif, Negatif, atau Netral) dengan menggunakan *dictionary sentiment* VADER (*Valence*

*Aware Dictionary And Sentiment Reasoner*) pada aplikasi Rapidminer.



Gambar 2. Pengumpulan dan Labeling Data

Gambar 2 menunjukkan proses mendapatkan data dari twitter dengan menggunakan operator “search” twitter dengan *query* “Corona Virus” dan menyimpan data hasilnya ke dalam file excel dengan menggunakan operator “Write Excel” yang sebelumnya sudah dilakukan filter data yang duplikasi dengan operator “Remove Duplicates”.

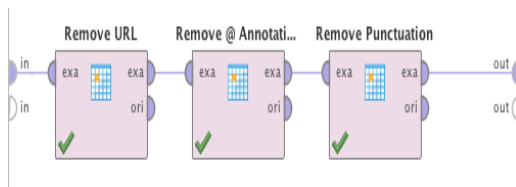
Tabel 1 Contoh Pelabelan Tweet dengan VADER

Text	Sentiment
@HonaCostello Google is the same engine that told me that corona virus from WHO website can only be contracted from mouth and nose, I get other viruses have different ways, but specifically there's a lot more secure sources than a vimeo video.	Positive
@yesnicksearcy @BeverlyMcCord2 @djrothkopf For a lock down I sure have seen a lot of Corona Virus commercials. Evidently the advertising agencies are deemed “essential”	Positive
RT @Israel_MOD: In the past two days, the Israel Institute for Biological Research (IIBR), has completed a groundbreaking scientific develo...	Neutral
The show will continue now with new restrictions due to the corona virus changing the way they will now run production. We look forward to being involved in our 11th Melb series in a row	Neutral
RT @nicolerafiee: i was not aware miss roni was over, u guys are all out there hangin with people, why didn't anyone tell me the macaroni v...	Negative
RT @Kingkeraun: For the first time in my life i feel like a terrible parent all because of the corona virus https://t.co/s6sZRBvIry	Negative

### B. Pre-Processing

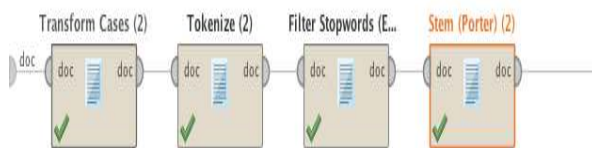
Tahapan selanjutnya setelah pelabelan data adalah proses *pre-processing*. Pada tahap ini data disiapkan menjadi data yang siap untuk dianalisa. Terdapat beberapa tahapan dalam proses

ini antara lain melakukan pembersihan terhadap data antara lain menghilangkan URL, menghilangkan *username*, menghilangkan tanda baca, lalu melakukan proses tokenisasi, menghilangkan *stopword*, dan proses *stemming*. Gambar 3 menunjukkan operator “Sub-proses” yang didalamnya terdapat operator “Remove URL”, “Remove @annotation”, dan “Remove Punctuation”.



Gambar 3. Isi dari operator “Sub-process”

Gambar 4 menunjukkan isi dari operator “Process Document” yang didalamnya terdapat operator “Transform Cases”, “Tokenize”, “Filter Stopword”, dan “Stem” pada kamus bahasa Inggris.



Gambar 4. Isi dari Operator “Process Document”

Berikut penjelasan detail dari prosedur pre-processing diatas :

### 1. Remove URL

Pada sebuah tweet biasanya terdapat beberapa tautan URL (Uniform Resource Locator) yang dimasukkan oleh pengguna. URL ini biasanya dicantumkan untuk

memberikan info lebih detil karena keterbatasan dari sebuah tweet yang hanya 280 kata. Pada proses ini dilakukan proses menghilangkan URL tersebut. Contohnya adalah “Corona Virus Awareness??https://t.co?LbMPHDGAOm” diubah menjadi “Corona Virus Awareness”.

### 2. @ Annotation Removal

Pada sebuah tweet biasanya pengguna juga memasukkan atau melakukan tagging pengguna lainnya dengan menggunakan notasi @xxxxxx. Pada proses ini dilakukan tahapan menghilangkan username pada setiap tweet. Contohnya adalah “RT @MasroorShykh1: The problems of the government and the arbitrary traders take into account the public difficulties” menjadi “RT : The problems of the government and the arbitrary traders take into account the public difficulties”.

### 3. Punctuation Removal

Pada proses ini dilakukan prosedur menghilangkan tanda baca yang ada pada sebuah tweet antara lain . , : “ “ ‘ ‘ , dsb. Contohnya adalah “luxembourg officially stopped league this season due to corona virus.. no champion and no relegation..” menjadi “luxembourg officially stopped league this season due to corona virus no champion and no relegation”.

#### **4. Case Folding**

Pada penulisan sebuah tweet terdapat beberapa bentuk huruf yang digunakan oleh pengguna, baik huruf besar maupun huruf kecil. Pada tahap ini seluruh huruf yang ada diubah kedalam huruf kecil. Contohnya adalah tweet "Some of you might hear Sweden is supposedly handling the Corona Virus well" diubah menjadi "some of you might hear sweden is supposedly handling the corona virus well".

#### **5. Tokenization**

Pada proses tokenization ini dilakukan proses pembuatan token pada kata atau pemotongan sebuah kalimat menjadi kata per kata. Contohnya adalah kalimat "some of you might hear sweden is supposedly handling the corona virus well" dipisahkan kata perkata menjadi "some, of, you, might, hear, sweden, is, supposedly, handling, the, corona, virus, well".

#### **6. StopWord Removal**

Tahapan ini merupakan proses menghilangkan kata tertentu pada sebuah tweet yang dianggap tidak mengandung arti (stopword). Pada dasarnya stopwords merupakan daftar kata pada sebuah bahasa. Stopword cenderung untuk dihilangkan pada penelitian terkait text mining karena stopwords digunakan berulang-ulang pada sebuah kalimat

sehingga stopwords dihilangkan agar penelitian dapat lebih fokus pada kata-kata yang lebih penting. Contoh stopwords pada bahasa Inggris antara lain is, am, are, the, at, on, in, dll.

#### **7. Stemming**

Stemming (lemmatization) adalah proses mengubah kata berimbuhan menjadi kata dasar. Aturan-aturan bahasa Inggris seperti grammar, tenses dihilangkan agar kata tersebut menjadi kata dasar. Contoh prosedur stemming pada bahasa Inggris antara lain kata fishing, fished, fisher di stem menjadi fish.

#### **8. Weighting Word**

Weighting word atau pembobotan kata adalah mekanisme pemberian nilai sebuah kata berdasarkan frekuensi kemunculannya pada sebuah dokumen. Metode yang digunakan pada penelitian ini adalah TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF merupakan metode pembobotan yang menggabungkan konsep Term Frequency dan Document Frequency. Algoritma TF-IDF ini mengalikan antara Term Frequency dengan Inverse Document Frequency. Term Frequency yaitu jumlah kemunculan sebuah term pada sebuah dokumen. Inverse Document Frequency yaitu pengurangan dominasi term yang sering muncul di berbagai dokumen, dengan memperhitungkan kebalikan frekuensi

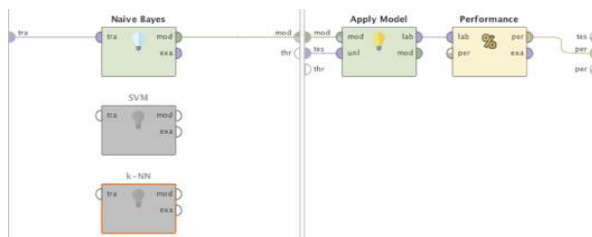


dokumen yang mengandung suatu kata. Formulasnya adalah sebagai berikut :

$$W_{d,t} = TF_{d,t} * IDF_t \quad (4)$$

### I. Klasifikasi dan Evaluasi Analisis Sentimen

Setelah prosedur pre-processing pada data, langkah selanjutnya adalah klasifikasi terhadap analisis sentiment yang ada. Pada tahap ini diterapkan sejumlah algoritma data mining. Gambar 5 menunjukkan isi dari operator "Cross Validation" pada aplikasi Rapidminer. Pada penelitian ini digunakan tiga algoritma machine learning yaitu algoritma Naive Bayes, SVM, dan K-NN



Gambar 5. Gambar Cross Validation Model

Setelah proses klasifikasi analisis sentimen telah selesai, langkah selanjutnya adalah mengevaluasi dengan mengukur keakurasian dan kualitas dari hasil tersebut. Evaluasi yang dilakukan adalah dengan pengujian performa dan akurasi sehingga menghasilkan nilai accuracy, precision, dan recall.

Accuracy (A) adalah total nilai True Positif dan True Negatif dibagi dengan jumlah keseluruhan data.

$$A = \frac{(TP+TN)}{(TP+FP+FN+TN)} \times 100\% \quad (5)$$

Precision (P) adalah prosentase nilai True Positif dari seluruh nilai Positif yang diprediksi.

$$P = \frac{TP}{(TP+FP)} \times 100\% \quad (6)$$

Recall (R) adalah persentase prediksi Positif dibandingkan dengan True Positif.

$$R = \frac{TP}{(TP+FN)} \times 100\% \quad (7)$$

Parameter TP, FP, FN, TN berdasarkan Confusion Matrix seperti pada table dibawah :

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<b>TP</b> (True Positive)	<b>FP</b> (False Positive) Type I Error
	0 (Negative)	<b>FN</b> (False Negative) Type II Error	<b>TN</b> (True Negative)

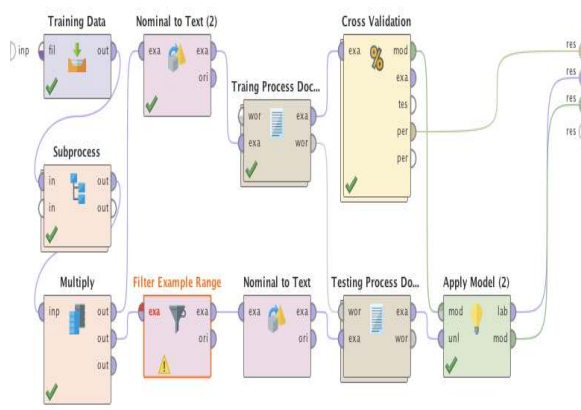
Gambar 6. Tabel Confusion Matrix

## IV. HASIL DAN PEMBAHASAN

Hasil Bab ini menjabarkan hasil penelitian dan pembahasan analisis



performansi dari setiap algoritma *machine learning*



Gambar 6 Proses Utama pada Aplikasi Rapidminer

Gambar 6 menjelaskan operator utama yang digunakan pada aplikasi Rapidminer. Prosedur diawali dengan operator “*Training Data*” yaitu melakukan pembacaan data file excel yang berisi teks dan sentimen di *Tweeter* sebagai *data training*. Operator “*Subproses*” dan “*Process Document*” digunakan dalam proses *pre-processing* data. Lalu operator “*Cross Validation*” digunakan dalam pada proses pembangunan model klasifikasi serta evaluasi model dilakukan dengan menggunakan *data testing* yang berasal dari multiply *data training* dengan melakukan sebanyak sepuluh kali validasi silang (*10-fold cross validation*).

Berikut ini hasil yang didapat setelah melakukan penelitian dengan ketiga algoritma klasifikasi antara lain algoritma *Naive Bayes*, SVM dengan *kernel linear*, dan K-NN dengan nilai terbaik yang didapat adalah nilai K=5.

Tabel 3 Confusion Matrix Algoritma Naïve Bayes

Metode	True Positif	True Netral	True Negatif
Pred. Positif	261	41	58
Pred. Netral	16	87	19
Pred. Negatif	74	33	407

Tabel 4 Confusion Matrix Algoritma Support Vector Machine (SVM)

Metode	True Positif	True Netral	True Negatif
Pred. Positif	263	39	44
Pred. Netral	10	60	4
Pred. Negatif	78	62	436

Tabel 4 Confusion Matrix Algoritma K-Nearest Neighbour (K-NN)

Metode	True Positif	True Netral	True Negatif
Pred. Positif	249	41	59
Pred. Netral	38	74	32
Pred. Negatif	64	46	393

Tabel 5 Nilai akurasi, Presisi, dan Recall Algoritma Klasifikasi

Metode	Akurasi	Precision	Recall
Naive Bayes	75,81%	75,03%	70,83%
SVM	76,21%	78,04%	71,42%
K-NN	71,89%	67,52	66,02%

Hasil yang didapat secara berturut untuk nilai akurasi algoritma Naive Bayes, SVM, dan K-NN adalah 75,81%, 76,21%, dan 71,89%. Hasil untuk nilai presisi untuk ketiga algoritma tersebut berturut-turut adalah 74,33%, 77,59%, dan 66,95%. Hasil

nilai recall dari ketiga algoritma tersebut adalah 70,83%, 71,42%, dan 66,03%. Dapat dilihat bahwa algoritma SVM memberikan hasil akurasi, presisi dan recall yang lebih baik untuk jenis dataset sosial media twitter dibandingkan algoritma Naive Bayes dan K-NN dengan perbedaan nilai yang cukup tipis (< 5%).

#### IV. SIMPULAN

Pada penelitian ini percobaan dilakukan untuk mengklasifikasi analisis sentimen yang didapat dari tweet “Corona Virus” dengan menggunakan tiga algoritma klasifikasi yaitu *Naive Bayes*, SVM, dan K-NN. Didapatkan hasil bahwa algoritma SVM memberikan hasil yang lebih baik dibandingkan dua algoritma lainnya dengan nilai akurasi, presisi, dan recall sebesar 76,21% 71,42%, dan 77,59%. Perbedaan nilai akurasi, presisi, dan recall ketiga algoritma ini tidak terlampau cukup jauh (<5%), artinya ketiga algoritma cukup baik dalam memberikan klasifikasi untuk analisis sentimen dengan data media sosial twitter. Model pada penelitian ini dapat dipertimbangkan untuk diterapkan pada kasus klasifikasi teks lainnya dengan tujuan pengambilan keputusan, kebijakan publik, dsb.

Pada penelitian selanjutnya diharapkan dapat digunakan jumlah data yang lebih besar serta dapat diuji coba

dengan *dataset* dari media sosial lainnya untuk membuktikan lebih lanjut bahwa ketiga algoritma ini cukup baik dalam memberikan klasifikasi analisis sentimen dengan *dataset* dari media sosial.

#### DAFTAR PUSTAKA

- W. K. Chen, (1993). *Linear networks and systems*. Belmont, CA: Wadsworth, pp. 123-135.
- Lin, Ying, (2019). *10 Twitter Statistics Every Marketer Should Know*.
- Bolen, Johan., Mao, Huina., Zeng, Xiaojung, (2011). Twitter mood predicts the stock market, *Journal of Computational Science*, 2, pp.1-8.
- Maning, D.C., P. Raghavan, and H. Schutze, (2008). *Introduction to Information Retrieval*, Cambridge University Press.
- Parikh, R., & Movassate, M, (2009). *Sentiment Analysis of User Generated Twitter Updates using Various Classification*, CS224 N Final Report.
- G. Gautam, Y. Divakar, (2014). *Sentiment analysis of twitter data using machine learning approaches and semantic analysis*. Seventh International Conference on Contemporary Computing (IC3), Noida, 2014, pp. 437-442.
- E. Romelta, (2012). Opinion Mining di Twitter untuk Customer Feedback Smartphone dengan Pembelajaran Mesin. *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, Vol. 1.
- N. Nurajijah and D. Riana, (2019). Algoritma naïve bayes, decision tree dan svm untuk klasifikasi persetujuan pembiayaan nasabah koperasi syariah. *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 2, pp. 77-82, 2019. doi: 10.14710/jtsiskom.7.2.2019.77-82.
- A. A. Prasanti, M. A. Fauzi, and M. T. Furqon, (2018). Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode N-Gram Dan Neighbor Weighted K-Nearest Neighbor (NW-KNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 594-601.
- Bayhaqy, Achmad., Sfenrianto, S., Nainggolan, K., (2018). *Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naive Bayes*. International Conference on Orange Technologies (ICOT), 1-6.
- Nasukawa, T. and Yi, J., (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. *Proceedings of the 2nd*

International Conference on Knowledge Capture, Florida, 23-25 October 2003, 70-77.

- B. Liu, 2012. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.