



PERBANDINGAN ALGORITMA UNTUK KLASIFIKASI ANALISIS SENTIMEN TERHADAP GENOSE PADA MEDIA SOSIAL TWITTER

Achyar Jhonathan Syahid^{*1}, Deni Mahdiana²

^{1,2}Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta
e-mail: ^{*1}1712501459@student.budiluhur.ac.id , ²deni.mahdiana@budiluhur.ac.id

Abstrak

GeNose merupakan alat pendeteksi virus *COVID-19* hanya dengan hembusan nafas yang dibuat oleh Universitas Gadjah Mada dan merupakan inovasi pertama buatan Indonesia. Sejak hadirnya *GeNose* di tengah masyarakat tentu terdapat pro dan kontra, sehingga penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap *GeNose* dengan data *tweet* dari media sosial *Twitter*. *Dataset* yang didapatkan sebanyak 3590 *tweet* dengan label sentimen netral, tidak relevan, positif dan negatif. Kemudian dari 3590 *tweet* yang digunakan hanya 637 *tweet*, diantaranya 287 *tweet* bersentimen positif dan 350 *tweet* bersentimen negatif. Penelitian ini menggunakan model CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dan algoritma yang digunakan pada penelitian ini adalah *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* untuk membandingkan performa dari ketiga algoritma tersebut dengan operator *Cross Validation* menggunakan *RapidMiner*. Hasil dari penelitian yang telah dilakukan menunjukkan performa dari algoritma *Naïve Bayes* cukup baik dengan nilai *accuracy* sebesar 72.36%, *precision* 66.91%, *recall* 77.97%. Sedangkan algoritma *K-Nearest Neighbor* mendapatkan nilai *accuracy* sebesar 64.22%, *precision* 61.79%, *recall* 55.15%. Selanjutnya algoritma *Decision Tree* mendapatkan nilai *accuracy* sebesar 65.15%, *precision* 86.89%, *recall* 27.13%. Penelitian ini bisa menjadi referensi untuk meningkatkan sosialisasi tentang *GeNose* untuk meminimalisir sentimen negatif masyarakat.

Kata kunci; *K-Nearest Neighbor*, *Decision Tree*, *Naïve Bayes*, *GeNose*

Abstract

GeNose is a tool for detecting the *COVID-19* virus with just a breath made by Gadjah Mada University and the first innovation made in Indonesia. Since the presence of *GeNose* in the community, there are many pros and cons, so this study aims to analyze public sentiment towards *GeNose* with *tweet* data from social media *Twitter*. The dataset obtained is 3590 tweets labeled neutral, irrelevant, positive and negative sentiments. Then of the 3590 tweets used only 637 tweets, which 287 tweets positive and 350 negatives. This study uses the CRISP-DM (*Cross-Industry Standard Process for Data Mining*) model and the algorithms used in this study are *K-Nearest Neighbor*, *Decision Tree* and *Naïve Bayes* to compare the performance three algorithms with *Cross Validation* operators using *RapidMiner*. The results of the research that have been conducted show that the performance of the *Naïve Bayes* algorithm is quite good with an *accuracy* value of 72.36%, *precision* 66.91%, *recall* 77.97%. While the *K-Nearest Neighbor* algorithm gets *accuracy* value 64.22%, *precision* 61.79%, *recall* 55.15%. Furthermore, the *Decision Tree* algorithm gets an *accuracy* value of 65.15%, *precision* 86.89%, *recall* 27.13%. This research can be a reference to increase socialization about *GeNose* to minimize negative public sentiment.

Keywords; *K-Nearest Neighbor*, *Decision Tree*, *Naïve Bayes*, *GeNose*



1. PENDAHULUAN

Pada tanggal 2 Maret 2020 virus *COVID-19* atau *Corona Virus Disease* pertama kali masuk ke Indonesia dan pada 11 Maret 2020 *World Health Organization* menetapkan virus *COVID-19* sebagai Pandemi Global [1]. Virus *COVID-19* adalah virus yang menginfeksi saluran pernafasan dan penyebarannya sangat cepat baik dari manusia ke manusia maupun dari manusia ke hewan [2]. Hingga saat ini penyebaran virus *COVID-19* masih berlangsung dan kasus terkonfirmasi positif di Indonesia terbanyak pada tanggal 30 Januari 2021 sebanyak 14.518 orang, namun saat ini kasus positif di Indonesia cenderung menurun sejak tanggal 31 Januari 2021 [3]. Ada beberapa alat yang ada di Indonesia saat ini untuk mendeteksi virus *COVID-19* seperti *Rapidtest Antibody*, *Rapidtest Antigen*, *Immunoassay* atau *Serologi*, *Polymerase Chain Reaction (PCR)* dan *GeNose*.

GeNose adalah alat buatan Universitas Gadjah Mada dan merupakan inovasi pertama buatan Indonesia yang bisa mendeteksi dan mendiagnosa seseorang yang tertular virus *COVID-19* hanya dengan bernafas. *Volatile Organic Compound (VOC)* yang terbentuk karena infeksi *COVID-19* yang dikeluarkan saat bernafas dapat dideteksi secara cepat dan akurat oleh *GeNose* [4].

Sejak hadirnya *GeNose* banyak argumen atau opini masyarakat mengenai *GeNose* yang bersifat positif atau negatif. Salah satu media masyarakat untuk beropini adalah melalui media sosial *Twitter*, mereka membuat *tweet* untuk menyampaikan opini mereka dan *tweet* yang mereka buat akan tersebar di *timeline Twitter*. *Tweet* yang mereka buat mempunyai peranan yang penting, karena bisa dijadikan *dataset* untuk melakukan analisis sentimen terhadap *GeNose* dan memprediksi keakuratan algoritma *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan *tweet* tersebut bersifat positif atau negatif [5].

Sentimen analisis atau bisa juga disebut dengan *opinion mining* merupakan bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi orang pada suatu produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya [6]. Sentimen analisis bertujuan untuk mengetahui apakah opini masyarakat dalam

bentuk *tweet* yang dikirimkan oleh pengguna memiliki makna positif atau negatif. Dengan melakukan sentimen analisis ini juga bisa dijadikan referensi untuk meningkatkan pengembangan suatu layanan atau produk.

Pada penelitian ini, peneliti melakukan analisis sentimen masyarakat terhadap *GeNose* dengan *crawling* data *Twitter* dan melakukan klasifikasi dengan algoritma *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* serta perhitungan TF-IDF (*Term Frequency — Inverse Document Frequency*) dengan menggunakan tools *RapidMiner*. Peneliti menggunakan ketiga algoritma tersebut karena algoritma yang digunakan pada penelitian ini merupakan 3 dari 10 algoritma teratas yang paling berpengaruh dalam *Data Mining* [7]. Berdasarkan penelitian sebelumnya yang dilakukan oleh Puspita dan Widodo [8] dengan judul penelitian “Perbandingan Metode KNN, *Decision Tree*, dan *Naïve Bayes* Terhadap Analisis Sentimen Pengguna Layanan BPJS”. *Dataset* yang diperoleh pada penelitian ini dari *tweet* media sosial *Twitter* terkait opini terhadap layanan BPJS dan didapatkan sebanyak 903 data. Tujuan dari penelitian tersebut adalah untuk mengetahui tingkat akurasi dari 3 algoritma diantaranya *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* dengan evaluasi *Cross Validation*. Akurasi tertinggi sebesar 96.13% dengan algoritma *Decision Tree*.

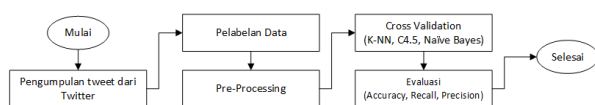
Berdasarkan penelitian yang dilakukan oleh Lubis [9] dengan judul “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data *Twitter*”. Data *tweet* yang diperoleh di penelitian ini sebanyak 170 data. Penelitian tersebut bertujuan untuk mengklasifikasi apakah opini yang disampaikan masyarakat termasuk dalam kategori riba atau bukan. Hasil menunjukkan bahwa *Accuracy* K-NN sebesar 70,59%, *Precision* 69,87% dan *Recall* 62,32% dengan evaluasi *Cross Validation*.

Berdasarkan penelitian yang dilakukan oleh Dharmawan, Arwani dan Ratnawati [10] dengan judul “Analisis Sentimen pada Sosial Media *Twitter* Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode *KNearest Neighbor*”. *Dataset* yang digunakan pada penelitian tersebut berdasarkan layanan SIAM UB yang sempat menjadi *trending topic Twitter*. Data yang diperoleh sebanyak 200

data latih dan 50 data uji. Akurasi terbaik dari penelitian tersebut dengan menggunakan nilai $K=3$, akurasi yang didapatkan sebesar 86%.

2. METODE PENELITIAN

Dataset yang digunakan pada penelitian ini bersumber dari *tweet* pada media sosial *Twitter* tentang *GeNose*. Metode yang digunakan pada penelitian ini mengadopsi model CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Fase yang ada pada CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah sebagai berikut [11] :

2.1 Business Understanding

Fase ini adalah proses untuk memahami topik permasalahan yang akan dijadikan sebagai bahan penelitian yaitu analisis sentimen masyarakat terhadap *GeNose*. Pemahaman terhadap topik permasalahan dapat dilakukan dengan mencari informasi melalui media sosial salah satunya *Twitter*, kemudian mencari *tweet* tentang *GeNose* pada media sosial *Twitter*.

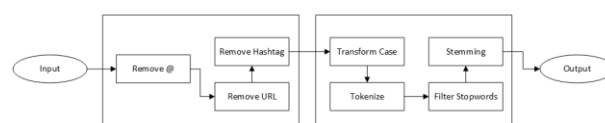
2.2 Data Understanding

Fase ini merupakan proses yang dilakukan untuk memperoleh data sesuai topik permasalahan dengan atribut yang diperlukan. Data pada penelitian ini diperoleh dari media sosial *Twitter* dari tanggal 22 Maret 2021 sampai dengan 29 Maret 2021 dengan *query* atau kata kunci “*GeNose*”. Kemudian dilakukan *data cleansing* (pembersihan data) seperti *remove duplicate* dan menghapus kata “RT” yang ada pada *tweet*.

2.3 Data Preparation (Pre-processing)

Pre-processing merupakan proses dimana data yang sudah diperoleh kemudian diolah agar menjadi data yang berkualitas. (1) *Remove @* adalah proses menghilangkan mention yang terdapat pada *tweet*. (2) *Remove URL* adalah proses menghilangkan link atau tautan yang terdapat pada *tweet*. (3) *Remove*

hashtag adalah proses untuk menghilangkan tagar atau *hashtag* yang ada pada *tweet*. (4) *Transform case* adalah proses mengubah semua huruf pada *tweet* menjadi huruf kecil. (5) *Tokenize* adalah proses memisahkan kalimat menjadi per-kata dan menghilangkan karakter yang tidak diperlukan seperti tanda baca atau angka. (6) *Filter stopwords* adalah proses menghilangkan kata yang tidak penting seperti “aku”, “di”, “yang”, “yakni” dan lain lain. (7) *Stemming* adalah proses mengubah menjadi akar kata dari proses *tokenize*. Uraian proses dari *Pre-Processing* dapat dilihat pada Gambar 2.



Gambar 2. Alur Proses Pre-Processing

2.4 Modelling

Pada fase ini, *modelling* dilakukan dengan metode *Cross Validation* lalu isi *number of folds* dengan 10. Dengan menggunakan *Cross Validation* data dibagi menjadi 10 bagian, dimana 9 data menjadi *data training* dan 1 data menjadi *data testing*, proses iterasi dilakukan sebanyak 10 kali. Skema *Cross Validation* dapat dilihat pada Gambar 3.

Iterasi ke-	Cross Validation (number of folds = 10)									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10

■ Data Uji (Data Testing)
■ Data Latih (Data Training)

Gambar 3. Skema Cross Validation

2.5 Evaluation

Pada fase ini, klasifikasi yang dilakukan dengan algoritma *K-Nearest Neighbor* (*K-NN*), *Decision Tree* dan *Naïve Bayes* yang sebelumnya menghitung *TF-IDF* pada *tools RapidMiner* dan melakukan evaluasi menggunakan *Confusion Matrix* untuk menunjukkan hasil performa dari ketiga algoritma yang digunakan. Gambaran dari *Confusion Matrix* dapat dilihat pada Gambar 4.

		Actual Values	
		Positive	Negative
Prediction	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

Gambar 4. Confusion Matrix

2. 5.1 TF-IDF

TF-IDF adalah kependekan dari *Term Frequency-Inverse Document Frequency* yang merupakan metode untuk menghitung bobot setiap kata pada suatu dokumen. Metode ini digunakan untuk mengetahui jumlah suatu kata yang sering muncul di dalam dokumen. Rumus untuk menghitung IDF ada pada persamaan (1) dan untuk menghitung bobot dokumen ada pada persamaan (2).

$$IDF_j = \log\left(\frac{n}{DF_j}\right) \quad (1)$$

$$W_{ij} = TF_{ij} * IDF_j \quad (2)$$

Keterangan:

W = Bobot dokumen (D_i) dari total nilai term (T_j)

TF_{ij} = Jumlah kemunculan term (T_j) yang ada di dalam dokumen (D_i)

n = Jumlah semua dokumen

DF_j = Jumlah dokumen yang mengandung term (T_j)

IDF = Inversed Document Frequency

2. 5.2 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* adalah metode klasifikasi data berdasarkan pembelajaran terhadap data yang sudah terklasifikasikan sebelumnya (*supervised learning*). Hasil klasifikasi data yang baru berdasarkan banyaknya data yang berdekatan dari kategori yang ada dalam K-NN. Rumus K-NN ada pada persamaan (4) dan penjabaran dari rumus K-NN ada pada persamaan (3).

$$d_i = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (3)$$

$$d_i = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (4)$$

Keterangan:

d = Jarak data ke i

n = Dimensi data

i = Variabel data

q_i = Data latih

p_i = Data uji

2. 5.3 Decision Tree

Algoritma C4.5 adalah salah satu algoritma klasifikasi yang digunakan untuk membuat *Decision Tree* atau pohon keputusan dengan menghitung nilai *Entropy* dan *Gain* pada suatu atribut. Rumus untuk menghitung nilai *Entropy* tertera dalam persamaan (5).

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (5)$$

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi dari S_i terhadap S

Setelah menghitung nilai *Entropy*, selanjutnya menghitung nilai *Gain* dengan melihat persamaan (6).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (6)$$

Keterangan:

S = Himpunan kasus

A = Atribut

n = Jumlah partisi atribut (A)

$|S_i|$ = Jumlah kasus pada partisi ke- i

$|S|$ = Jumlah kasus dalam S

2. 5.4 Naïve Bayes

Naïve Bayes adalah salah satu algoritma klasifikasi yang menggunakan metode probabilitas dan statistik untuk memprediksi peluang bahwa prediksi benar dengan asumsi yang sangat kuat (naif) pada masing masing kejadian atau kondisi. Untuk rumus klasifikasinya seperti pada persamaan (7).

$$P(C|X) = \frac{P(x|C) * P(c)}{P(x)} \quad (7)$$

Keterangan:

X = Data sampel dengan class yang belum diketahui.

C = Hipotesis bahwa X adalah data dengan klas (label) C

$P(C|X)$ = Probabilitas hipotesis benar untuk data sampel X yang diamati berdasarkan kondisi

$P(c)$ = Probabilitas hipotesis c

$P(x|C)$ = Probabilitas sampel x berdasarkan kondisi bahwa hipotesis benar

$P(x)$ = Peluang x yang merupakan data sample yang diamati

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan mengenai hasil dan pembahasan performa dari ketiga algoritma yang digunakan yaitu *K-Nearest Neighbor* (K-NN), *Decision Tree* dan *Naïve Bayes* sesuai dengan metode CRISP-DM yang digunakan pada topik penelitian ini.

3.1 Business Understanding

Pada penelitian ini, *dataset* yang digunakan bersumber dari cuitan masyarakat di *Twitter* tentang *GeNose*. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap *GeNose* berdasarkan opini masyarakat yang disajikan berupa *tweet* di *Twitter*.

3.2 Data Understanding

Pengumpulan data atau bisa juga disebut *crawling* data dilakukan dengan menggunakan aplikasi *RapidMiner*, *query* yang digunakan adalah “*GeNose*” dengan limit 10000 berdasarkan *tweet* “terbaru dan terpopuler” dengan parameter sesuai dengan yang disebutkan di metode penelitian, hasil dari proses *crawling* di-export ke dalam file *Microsoft Excel*. Hasil *crawling* memperoleh data *tweet* sebanyak 3590. Pemberian label pada *tweet* peneliti dibantu oleh seorang pakar dan selanjutnya data tersebut melalui proses *cleansing* untuk menghilangkan *tweet* dengan label netral dan tidak relevan. Data yang diperoleh setelah melalui proses *cleansing* sebanyak 637 *tweet* yang terdiri dari 287 *tweet*

bersentimen positif dan 350 *tweet* bersentimen negatif.

3.3 Data Preparation (Pre-processing)

Pada metode penelitian *pre-processing* dilakukan dalam 7 (tujuh) tahapan terhadap 637 *dataset* menggunakan *RapidMiner*. Contoh hasil dari tahapan *pre-processing* dapat dilihat pada Tabel 1 sampai dengan Tabel 7 berikut ini.

Tabel 1. *Remove @*

Sebelum	Sesudah
@BanghDeni: Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif #CTBot Adrop Ram 10 tatm 10 https://t.co/vcReWKYC Hi	Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif #CTBot Adrop Ram 10 tatm 10 https://t.co/vcReWKYC Hi

Tabel 2. *Remove URL*

Sebelum	Sesudah
Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif #CTBot Adrop Ram 10 tatm 10 https://t.co/vcReWKYCHi	Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif #CTBot Adrop Ram 10 tatm 10

Tabel 3. *Remove Hashtag*

Sebelum	Sesudah
Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif #CTBot Adrop Ram 10 tatm 10	Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif Adrop Ram 10 tatm 10

Tabel 4. *Transform Case*

Sebelum	Sesudah
Penerapan GeNose untuk Penumpang Pesawat Akan Kian Masif Adrop Ram 10 tatm 10	penerapan genose untuk penumpang pesawat akan kian masif adrop ram 10 tatm 10

Tabel 5. *Tokenize*

Sebelum	Sesudah
penerapan genose untuk penumpang pesawat akan kian masif adrop ram 10 tatm 10	penerapan genose untuk penumpang pesawat

	akan kian masif adrop ram tاتم
--	---

Tabel 6. Filter Stopwords

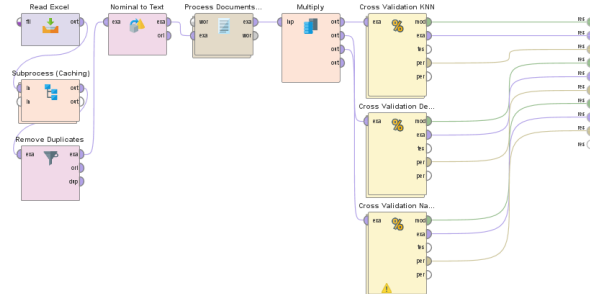
Sebelum	Sesudah
penerapan genose untuk penumpang pesawat akan kian masif adrop ram tاتم	penerapan genose penumpang pesawat kian masif

Tabel 7. Stemming

Sebelum	Sesudah
penerapan genose penumpang pesawat kian masif	terap genose tumpang pesawat kian masif

3.4 Modelling

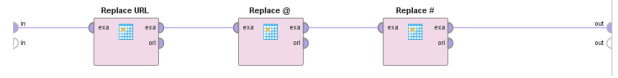
Pada tahap *modelling* penelitian ini menggunakan *Cross Validation* untuk melakukan pengujian terhadap 3 algoritma yang digunakan yaitu *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan *tweet* yang bersifat positif dan *tweet* yang bersifat negatif. Proses utama dalam menganalisis terhadap 3 algoritma dengan *tools RapidMiner* dapat dilihat pada Gambar 5.



Gambar 5. Proses Utama di dalam RapidMiner

Sebelum memasuki proses *Cross Validation*, data *tweet* yang sudah diberi label oleh pakar dalam bentuk *Microsoft Excel* akan dibaca oleh operator *Read Excel*. Selanjutnya

data akan diproses oleh operator *Subprocess* yang terdiri dari *remove @*, *remove url* dan *remove hashtag*. Proses ini merupakan tahapan dari *data preparation (pre-processing)* yang dapat dilihat pada Gambar 6.



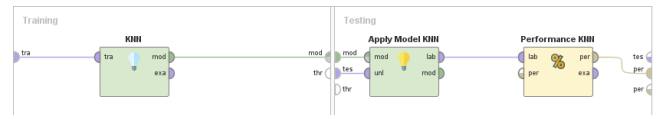
Gambar 6. Proses di dalam Operator Subprocess

Selanjutnya terdapat operator *Remove Duplicates* untuk menghapus data yang serupa. Operator *Nominal to Text* berfungsi untuk mengubah atribut yang bernilai *nominal* menjadi *string*. Selanjutnya operator *Process Documents* yang juga merupakan tahapan dari *data preparation (pre-processing)* di dalamnya terdiri dari *transform case*, *tokenize*, *filter stopwords* dan *stemming* yang dapat dilihat pada Gambar 7.

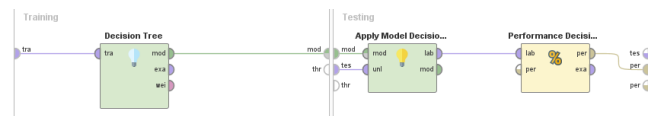


Gambar 7. Proses di dalam Operator Process Documents

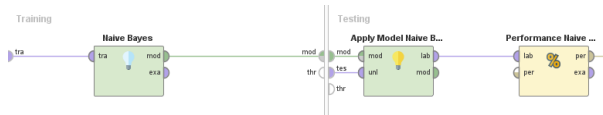
Kemudian hasil dari proses dokumentasi tersebut dilipatgandakan menggunakan operator *Multiply*. Operator ini berfungsi agar dapat menggunakan 3 operator *Cross Validation* (*number of folds* = 10) dalam waktu bersamaan, masing-masing di dalamnya terdapat algoritma *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* yang dapat dilihat pada Gambar 8 (KNN), Gambar 9 (*Decision Tree*) dan Gambar 10 (*Naïve Bayes*).



Gambar 8. Proses Cross Validation KNN



Gambar 9. Proses Cross Validation Decision Tree



Gambar 10. Proses *Cross Validation Naïve Bayes*

3.5 Evaluation

Setelah proses klasifikasi analisis sentimen terhadap 3 algoritma klasifikasi *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes* dengan *Cross Validation* (*number of folds* = 10) hasil dari pengujian performa dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pengujian Algoritma

Algoritma	Accuracy	Precision	Recall
<i>K-Nearest Neighbor</i>	64.22%	61.79%	55.15%
<i>Decision Tree</i>	65.15%	86.89%	27.13%
<i>Naïve Bayes</i>	72.36%	66.91%	77.97%

Berdasarkan hasil pengujian dari 3 algoritma, dapat dilihat bahwa algoritma *Naïve Bayes* memiliki performa yang cukup baik jika dibandingkan dengan algoritma *K-Nearest Neighbor* dan *Decision Tree*.

4. KESIMPULAN

Berdasarkan hasil dan proses analisis yang telah dilakukan pada penelitian ini, menunjukkan bahwa performa algoritma *Naïve Bayes* cukup baik karena prediksi yang diberikan cukup akurat dalam menganalisis sentimen masyarakat terhadap *GeNose* di *Twitter* jika dibandingkan algoritma *K-Nearest Neighbor* dan *Decision Tree*. Algoritma *Naïve Bayes* memperoleh nilai *accuracy* sebesar 72.36%, *precision* 66.91%, *recall* 77.97%. Sedangkan algoritma *K-Nearest Neighbor* mendapatkan nilai *accuracy* sebesar 64.22%, *precision* 61.79%, *recall* 55.15%. Selanjutnya algoritma *Decision Tree* mendapatkan nilai *accuracy* sebesar 65.15%, *precision* 86.89%, *recall* 27.13%.

5. SARAN

Dalam penelitian ini terdapat beberapa keterbatasan, adapun saran-saran untuk penelitian selanjutnya yaitu:

- Dataset* yang digunakan diharapkan lebih banyak dari penelitian sebelumnya agar keakuratan dari analisis yang dilakukan dapat meningkat dan memberikan informasi mengenai kehadiran *GeNose*.
- Pada penelitian selanjutnya disarankan menggunakan algoritma yang berbeda seperti *Support Vector Machine*, *Neural Network*, dan lain-lain sehingga diperoleh hasil analisis yang menunjukkan performa pada setiap algoritma klasifikasi.

DAFTAR PUSTAKA

- [1] Gugus Tugas Percepatan Penanganan COVID-19, "Protokol Percepatan Penanganan Pandemi Covid-19 (Corona Virus Disease 2019)," *Gugus Tugas Percepatan Penanganan COVID-19*, vol. 19, hal. 3, 2020, [Daring]. Tersedia pada: [https://covid19.go.id/storage/app/media/Protokol/Protokol Percepatan Penanganan Pandemi Corona Virus Disease 2019.pdf](https://covid19.go.id/storage/app/media/Protokol/Protokol%20Percepatan%20Penanganan%20Pandemi%20Corona%20Virus%20Disease%202019.pdf).
- [2] D. Kumar, "Corona Virus: A Review of COVID-19," *Eurasian J. Med. Oncol.*, vol. 4, no. 2, hal. 8, 2020, doi: 10.14744/ejmo.2020.51418.
- [3] Satuan Tugas Penanganan COVID-19, "Perkembangan Kasus Terkonfirmasi Positif Covid-19 Per-Hari," *covid19.go.id*, 2021. <https://covid19.go.id/peta-sebaran> (diakses Apr 04, 2021).
- [4] Satria, "Hembusan Nafas Sekelas Swab Test Untuk Deteksi Covid-19: GeNose Karya UGM," <https://ugm.ac.id/>, 2020. https://ugm.ac.id/id/berita/20051-genose-deteksi-covid-19-cepat-mudah-dan-akurat-hanya-dengan-hembusan-nafas?utm_source=dlvr.it&utm_medium=twitter (diakses Apr 04, 2021).
- [5] W. E. Nurjanah, R. S. Perdana, dan M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini

- Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, hal. 1751, 2017.
- [6] B. Liu, *Sentiment Analysis and Opinion Mining Opinion Mining*. 2012.
- [7] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.
- [8] R. Puspita dan A. Widodo, “Perbandingan Metode KNN , Decision Tree , dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” vol. 5, no. 4, hal. 646–654, 2021.
- [9] H. Lubis, “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter,” vol. 4, no. April, hal. 369–376, 2020, doi: 10.30865/mib.v4i2.2051.
- [10] L. R. Dharmawan, I. Arwani, dan D. E. Ratnawati, “Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K- Nearest Neighbor,” vol. 4, no. 3, hal. 959–965, 2020.
- [11] A. Azevedo dan M. F. Santos, “KDD, semma and CRISP-DM: A parallel overview,” *MCCSIS'08 - IADIS Multi Conf. Comput. Sci. Inf. Syst. Proc. Informatics 2008 Data Min. 2008*, no. June, hal. 184, 2008.
-