

Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine

Debby Alita^{1*}, RB Ali Shodiqin²

^{1,2}S1 Informatika, Universitas Teknokrat Indonesia, Indonesia

^{1*}debbyalita@teknokrat.ac.id, ²rbalishodiqinn@gmail.com

Abstrak: Pemberian vaksin di Indonesia saat ini sudah mencapai tahapan vaksin booster, berbagai jenis vaksin telah diberikan kepada masyarakat Indonesia dari vaksin Sinovac, AstraZeneca, Sinopharm, Moderna, Pfizer, dsb. Tidak sedikit masyarakat Indonesia yang menggunakan beberapa jenis vaksin yang ditawarkan hingga vaksin booster akan tetapi ada beberapa masyarakat yang berpendapat masih terinfeksi virus covid ini dengan gejala berat. Pendapat lain juga terdapat kemudian terdapat vaksin Pada tahun 2019 masyarakat dihebohkan dengan adanya virus baru dari Wuhan, China yaitu virus corona atau disebut COVID-19 (*Corona Virus Disease 2019*). Pemerintah mengajak masyarakat untuk melakukan vaksin Covid-19 agar terbentuk *herd immunity* atau kekebalan kelompok terdapat virus Covid-19. Analisis sentimen dapat digunakan untuk melakukan evaluasi suatu kinerja pelayanan dan sebagainya. Maka penulis akan melakukan suatu perbandingan antara metode *Naive Bayer Classifier* dan *Support Vector Machine* untuk mengetahui metode mana yang lebih efisien dalam mengetahui pandangan akurasi masyarakat terhadap vaksin Covid-19. Hasil pengujian performa dari kedua metode menunjukkan bahwa performa metode *Naive Bayes Classifier* (*Accuracy 72.88%*, *Precision 43.49%*, *Recall 54.95%*, dan rata-rata performa 57.10%) lebih tinggi dibandingkan dengan rata-rata performa metode *Support Vector Machine* (*Accuracy 77.00%*, *Precision 75.00%*, *Recall 7.70%*, dan rata-rata performa 53.52%). Berdasarkan nilai rata-rata *performance* metode *Naive Bayes Classifier* dapat dinilai lebih efisien dibandingkan metode *Support Vector Machine*.

Kata Kunci: Analisis Sentimen; Naive Bayes Classifier; Support Vector Machine; Text Mining; Vaksinasi Covid-19;

Abstract: Vaccine administration in Indonesia has now reached the booster vaccine stage, various types of vaccines have been given to the Indonesian people from the Sinovac, AstraZeneca, Sinopharm, Moderna, Pfizer vaccines, etc. Not a few Indonesian people use several types of vaccines that are offered up to booster vaccines, but there are some people who think they are still infected with this Covid virus with severe symptoms. Another opinion is that there is also a vaccine. In 2019, people were shocked by a new virus from Wuhan, China, namely the corona

virus or called COVID-19 (Corona Virus Disease 2019). The government invites the public to get the Covid-19 vaccine in order to form herd immunity or group immunity to the Covid-19 virus. Sentiment analysis can be used to evaluate a service performance and so on. So the author will conduct a comparison between the Naive Bayer Classifier method and the Support Vector Machine to find out which method is more efficient in knowing people's accurate views of the Covid-19 vaccine. The performance test results of the two methods show that the performance of the Naive Bayes Classifier method (Accuracy 72.88%, Precision 43.49%, Recall 54.95%, and average performance 57.10%) is higher than the average performance of the Support Vector Machine method (Accuracy 77.00% , Precision 75.00%, Recall 7.70%, and average performance 53.52%). Based on the average performance value of the Naive Bayes Classifier method, it can be considered more efficient than the Support Vector Machine method

Keywords: Covid-19 Vaccination; Naive Bayes Classifier; Support Vector Machine; Text Mining; Twitter Sentiment Analysis;

1. PENDAHULUAN

Akhir-akhir ini masyarakat dihebohkan dengan adanya virus baru dari Wuhan, China yaitu virus corona atau disebut COVID-19 (*Corona Virus Desease* 2019). COVID-19 telah memakan korban jiwa di Indonesia hingga mencapai 157.754 jiwa. Melihat pesatnya penyebaran Covid-19 dan bahaya yang akan muncul jika tidak segera ditangani akan sangat berdampak buruk bagi suatu Negara. Telah banyak dilakukan beberapa cara oleh pemerintah, institusi, hingga kalangan masyarakat untuk memutus rantai penyebaran virus corona. Salah satu cara yang sangat mungkin untuk mencegah penyebaran virus ini dengan menerapkan protokol kesehatan yaitu 5M (Mencuci tangan, Memakai masker, Menjaga jarak, Menjauhi kerumunan dan Mengurangi Mobilitas), tidak hanya menerapkan 5M tetapi pemerintah mengajak masyarakat untuk melakukan vaksin Covid-19 supaya adanya *herd immunity* atau kekebalan kelompok terdapat virus Covid-19, Sehingga pemerintah Indonesia saat ini sudah menggiatkan program vaksinasi kepada masyarakat Kemenkes.

Kegiatan vaksinasi haruslah mempertimbangkan segala aspek, mulai dari aspek kelayakan vaksin yang akan digunakan, resiko pasca pemakaian, sampai tahapan dan prosedur dari pemberian vaksin hingga nantinya sampai ke masyarakat. Semua itu haruslah dipertimbangkan secara terperinci agar kegiatan vaksinasi dapat berjalan dengan baik dan terhindar dari hal-hal yang justru akan merugikan, tidak hanya itu program vaksinasi juga harus mempertimbangkan berbagai masukan di antaranya adalah melihat respon dan opini masyarakat terhadap program vaksinasi yang saat ini sedang berjalan[1].

Dengan adanya hal ini membuat kebanyakan orang merasa was-was dan resah bahkan ketakutan dengan adanya virus ini. Selain itu penelitian dan produksi vaksin dilakukan dalam waktu yang dapat dibilang singkat membuat warga ragu untuk melakukan vaksin. Hal ini juga dikarenakan sempat beredar kasus efek samping vaksin yang mengganggu kesehatan bahkan meninggal. Sehingga membuat masyarakat umum ingin mengungkapkan pendapat, aspirasi dan kritikan, namun keterbatasan waktu dan ruang membuat aspirasi masyarakat tidak tersampaikan. Dengan adanya jejaring sosial masyarakat umum dapat mengungkapkan segala hal yang ada contohnya seperti *twitter*[2], [3]. Setiap *twit* dari *netizen* tidak mengandung makna yang selaras, sehingga perlu dilakukan analisis terhadap opini *netizen* di *twitter* mengenai COVID-19 dengan klasifikasi positif dan negatif. Hal tersebut menunjukkan adanya peluang sumber data yang sangat besar yang dapat dimanfaatkan untuk menghasilkan suatu *knowledge* yang

bermanfaat[4]–[6]. Pemanfaatan data yang bersumber dari media sosial merupakan suatu terobosan baru untuk dijadikan sebagai alternatif sumber data. Analisis sentimen merupakan cara mengumpulkan pendapat khalayak umum menggunakan jejaring sosial yang di dalamnya mengandung pelayanan umum, serta isu terkini. Analisis sentimen dapat digunakan untuk melakukan evaluasi suatu kinerja pelayanan dan sebagainya.

Jenis analisis sentimen yang sering digunakan dalam kalangan penelitian yaitu, analisis sentimen dalam bentuk dokumen atau kalimat. Terdapat *teks mining* yang dapat bekerja dalam komputer dengan tujuan mengolah informasi lama secara eksplisit sehingga menghasilkan temuan informasi baru[7]. Dalam penjelasan latar belakang diatas maka penulis akan melakukan suatu perbandingan antara metode *Naive Bayer Classifier* dan *Support Vector Machine* untuk mengetahui metode mana yang lebih efisien dalam mengetahui pandangan akurasi masyarakat terhadap vaksin Covid-19. Metode *Support Vector Machine* bekerja dengan membagi dua kelompok data menggunakan fungsi linear dalam sebarang ruang fitur berdimensi tinggi dengan proses menemukan garis pemisah (*hyperlane*) terbaik sehingga dapat menemukan ukuran margin yang maksimal antara ruang input dengan ruang ciri menggunakan kaidah kernel [8].

2. METODE PENELITIAN

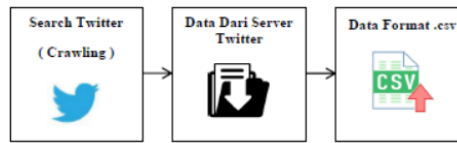
Untuk mempermudah pelaksanaan penelitian maka dibuat kerangka tahapan penelitian agar penelitian dapat dilakukan secara berurutan sehingga lebih efektif dan efisien karena penelitian dilakukan secara terstruktur[9], [10]. Kerangka tahapan penelitian ini dibuat dan disesuaikan dengan langkah-langkah proses *Text Mining*. Kerangka Tahapan Penelitian dapat dilihat pada Gambar 1.



Gambar 1 Kerangka Penelitian

Berikut ini merupakan penjelasan tahapan-tahapan dari kerangka penelitian diatas:

1. *Crawling Data*, yaitu mencari data *twitter* berisi opini masyarakat yang mengandung unsur komentar positif maupun negatif terhadap vaksin Covid-19.



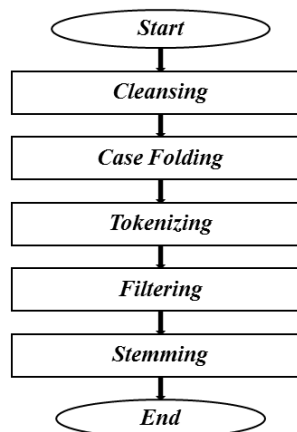
Gambar 2. Proses Crawling Data

2. *Labeling*, pemberian label terhadap *crawling data* pelabelan ini bertujuan untuk menentukan data yang akan diklasifikasikan kedalam beberapa kelas klasifikasi, label yang telah ditentukan yaitu Sentimen Positif dan Sentimen Negatif.

A	B	C
Username	Normalisasi Tweet	Label
yunussaprin	Vaksinasi diberikan kepada masyarakat terlepas dari riwayat apakah mereka pernah terinfeksi Covid-19 dengan atau tanpa gejala	Positif
TseuYee	Tidak ada batasan jumlah pendaftaran untuk tanggungan	Positif
Pitra_PAR	Vaksinasi kedua selesai. Terimakasih Yesus walaupun dosis telah ditingkatkan namun hingga saat ini tidak ada hal aneh yang dirasakan tubuh	Positif
kominfolkip	Sebanyak 163 lansia di Kelurahan Rawasen mendapatkan vaksinasi Covid-19. Sohib Jakpus, sudah dapat jadwal vaksin untuk lansia belum? Kalau sudah dapat jadwal vaksin nya jangan ditunda-tunda yah	Positif
ha3r	Momen hari ini, semoga bumi ini cepat membaik	Positif
dhyaaqs	Alhamdulillah selesai, terimakasih untuk Kementerian Kesehatan RI dan Plazzasiarnd Official untuk kesempatan vaksin Covid19. Ternyata divaksin tidak menyneramkan itu, sampai jumpa divaksin kedua	Positif
SDGsCoLab	Pakar farmasi kini mengembangkan vaksin COVID-19 generasi kedua berbentuk pil atau tablet.	Positif
newsmerahpu	Vaksinasi saat Ramadhan tidak batalkan puasa	Positif
SCRANIK08	Keputusan Menteri Putuskan Rantai Covid-19 Operasi Berentang - National Tank Force (NTP)	Positif
11serandines	Pemerintah vaksin ini merupakan salah satu aksar kita untuk melawan Covid-19	Positif

Gambar 3. Proses Labeling Crawling Data

3. *Pre-Processing*, melakukan *Cleansing*, *Case Folding*, *Tokenizing*, dan *Stemming* untuk menghasilkan data bersih.



Gambar 4 Proses Pre-processing

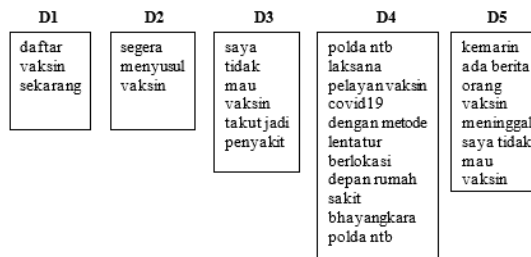
4. *Ekstrasi Fitur*, melakukan perhitungan frekuensi relatif dari suatu kata yang muncul menggunakan Algoritma TF-IDF, Tahapan dimana memberikan pada setiap kata dengan menggunakan Algoritma Trem Frekuensi Invert Document Frequency

No	Teks
1.	daftar vaksinasi sekarang
2.	segera menyusul vaksin
3.	saya tidak mau vaksin takut jadi penyakit
4.	polda ntb laksana pelayan vaksin covid 19 dengan metode lentatur berlokasi depan rumah sakit bhayangkara polda ntb
5.	kemarin ada berita orang vaksin meninggal saya tidak mau vaksin

Gambar 5. Proses Ekstrasi

Sampel data tersebut selanjutnya akan dilakukan perhitungan frekuensi relatif dari suatu kata yang muncul pada setiap kelas/dokumen.





Gambar 6 Proses Ekstraksi

Berdasarkan gambar diatas dapat diketahui bahwa jumlah dokumen ada 5 yaitu D1, D2, D3, D4, dan D5. Sebelum melakukan perhitungan *TF-IDF* kita akan mengitung frekuensi kemunculan kata dalam setiap dokumen. Berikut ini tabel kemunculan angka dalam setiap dokumen.

Tabel 1 Proses TF-IDF

No.	Kosa Kata	Frekuensi				
		D1	D2	D3	D4	D5
1	Daftar	1	0	0	0	0
2	Vaksin	1	1	1	1	2
3	Sekarang	1	0	0	0	0
4	Segera	0	1	0	0	0
5	Menyusul	0	1	0	0	0
6	Saya	0	0	1	0	1
7	Tidak	0	0	1	0	1
8	Mau	0	0	1	0	1
9	Takut	0	0	1	0	0
10	Jadi	0	0	1	0	0
11	Penyakit	0	0	1	0	0
12	Polda	0	0	0	1	0
13	Laksana	0	0	0	1	0
14	Pelayanan	0	0	0	1	0
15	covid 19	0	0	0	1	0
16	Dengan	0	0	0	1	0
17	Metode	0	0	0	1	0
18	Lentatur	0	0	0	1	0
19	Berlokasi	0	0	0	1	0
20	Depan	0	0	0	1	0
21	Rumahsakit	0	0	0	1	0
22	Bhayangkara	0	0	0	1	0
23	Kemarin	0	0	0	0	1
24	Ada	0	0	0	0	1
25	Berita	0	0	0	0	1
26	Orang	0	0	0	0	1
27	Meninggal	0	0	0	0	1

Persamaan yang digunakan untuk menghitung *TF-IDF* adalah sebagai berikut:

$$TF - IDF_{(d,t)} = TF_{(d,t)} \times IDF_{(t)}$$

Keterangan: t = kata
 d = dokumen

Dimana:

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d}$$

$$IDF_{(t)} = \log \left(\frac{\text{total dokumen } d}{\text{jumlah dokumen yang mengandung kata } d} \right)$$

a. Menghitung Nilai TF

Pertama akan menghitung nilai $TF_{(d,t)}$ kata per kata. Berikut ini merupakan perhitungan nilai $TF_{(d,t)}$:

- 1) Kata "daftar"
Pada D1

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{3} = 0.3333$$

Pada D2

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{3} = 0.00$$

Pada D3

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{7} = 0.00$$

Pada D4

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{12} = 0.00$$

Pada D5

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{10} = 0.00$$

2) Kata "vaksin"

Pada D1

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{3} = 0.3333$$

Pada D2

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{3} = 0.3333$$

Pada D3

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{7} = 0.1429$$

Pada D4

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{12} = 0.0833$$

Pada D5

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{2}{10} = 0.2000$$

3) Kata "sekarang"

Pada D1

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{3} = 0.3333$$

Pada D2

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{3} = 0.0000$$

Pada D3

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{7} = 0.0000$$

Pada D4

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{12} = 0.0000$$

Pada D5

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{10} = 0.0000$$

4) Kata "segera"

Pada D1

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{3} = 0.0000$$

Pada D2

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{3} = 0.3333$$

Pada D3

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{7} = 0.0000$$

Pada D4

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{12} = 0.0000$$

Pada D5

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{10} = 0.0000$$

5) Kata "menyusul"

Pada D1

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{3} = 0.0000$$

Pada D2

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{1}{3} = 0.3333$$

Pada D3

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{7} = 0.0000$$

Pada D4

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{12} = 0.0000$$

Pada D5

$$TF_{(d,t)} = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} = \frac{0}{10} = 0.0000$$

Begitu seterusnya perhitungan nilai $TF_{(d,t)}$ untuk masing-masing kata. Untuk hasil seluruh perhitungan nilai $TF_{(d,t)}$ dapat dilihat pada tabel dibawah ini:

Tabel 2. Hasil TF

Kosa Kata	Frekuensi					Nilai $TF_{(d,t)}$				
	D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
Daftar	1	0	0	0	0	0.3333	0.0000	0.0000	0.0000	0.0000
Vaksin	1	1	1	1	2	0.3333	0.3333	0.1429	0.0833	0.2000
Sekarang	1	0	0	0	0	0.3333	0.0000	0.0000	0.0000	0.0000
Segera	0	1	0	0	0	0.0000	0.3333	0.0000	0.0000	0.0000

Kosa Kata	Frekuensi					Nilai $TF_{(d,t)}$				
	D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
Menyusul	0	1	0	0	0	0.0000	0.3333	0.0000	0.0000	0.0000
Saya	0	0	1	0	1	0.0000	0.0000	0.1429	0.0000	0.1000
Tidak	0	0	1	0	1	0.0000	0.0000	0.1429	0.0000	0.1000
Mau	0	0	1	0	1	0.0000	0.0000	0.1429	0.0000	0.1000
Takut	0	0	1	0	0	0.0000	0.0000	0.1429	0.0000	0.0000
Jadi	0	0	1	0	0	0.0000	0.0000	0.1429	0.0000	0.0000
Penyakit	0	0	1	0	0	0.0000	0.0000	0.1429	0.0000	0.0000
Pelda	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Laksana	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Pelayan	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
covid 19	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Dengan	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Metode	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Lentatur	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Berlokasi	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Depan	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
rumahsakit	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
bhayangkara	0	0	0	1	0	0.0000	0.0000	0.0000	0.0833	0.0000
Kemarin	0	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.1000
Ada	0	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.1000
Berita	0	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.1000
Orang	0	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.1000
meninggal	0	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.1000

5. Klasifikasi data akan dilakukan ke dalam kelas **Sentimen Positif** dan **Sentimen Negatif** dengan menggunakan Algoritma *Naïve Bayes Classifier* dan *Support Vector*

Debby Alita : * Penulis Korespondensi



Copyright © 2023, Debby Alita.

Machine Klasifikasi, melakukan pengelompokan *tweet* berdasarkan kelas klasifikasi yang ditentukan.

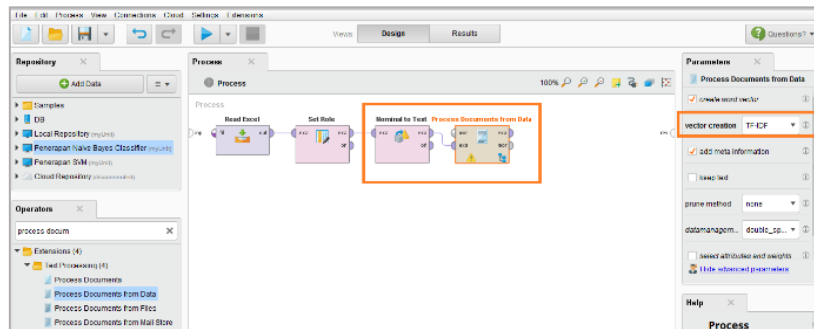
Tabel 3. Proses Klasikasi Data

No.	Kosa Kata	Positif	Negatif
1	daftar	0.6990	
2	vaksin	0.0000	0.0000
3	sekarang	0.6990	
4	segera	0.6990	
5	menyusul	0.6990	
6	saya		0.3979
7	tidak		0.3979
8	mau		0.3979
9	takut		0.6990
10	Jadi		0.6990
11	penyakit		0.6990
12	polda	0.6990	
13	laksana	0.6990	
14	pelayan	0.6990	
15	covid 19	0.6990	
16	dengan	0.6990	
17	metode	0.6990	
18	lentatur	0.6990	
19	berlokasi	0.6990	
20	depan	0.6990	
21	rumahsakit	0.6990	
22	bhayangkara	0.6990	
23	kemarin		0.6990
24	ada		0.6990
25	berita		0.6990
26	orang		0.6990
27	meninggal		0.6990

- Pengujian, melakukan pengukuran performa kedua algoritma dengan menghitung *Accuracy*, *Recall*, dan *Precision*.
- Hasil dari pengujian performa kedua metode akan dibandingkan untuk melihat kemampuan kedua metode.

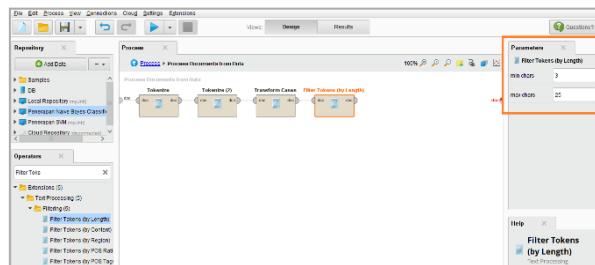
3. HASIL DAN PEMBAHASAN

Analisis sentimen dilakukan terhadap tanggapan pengguna Twitter yang diambil dengan cara *crawling* menggunakan API. Selanjutnya seluruh isi tweet diberikan label Positif dan Negatif. Pemberian label ini dilakukan bersama Dosen Universitas Teknokrat Indonesia yaitu Bapak Auliya Rahman. Penerapan kedua metode pada penelitian ini menggunakan *tools* Rapid Miner. Data yang digunakan pada penelitian ini adalah data dengan format teks sehingga sebelum membentuk model pengujian pada *tools* akan dilakukan normalisasi data menggunakan metode **TF-IDF**. Untuk melakukan proses TF-IDF pada Rapid Miner digunakan operator-operator sebagai berikut.



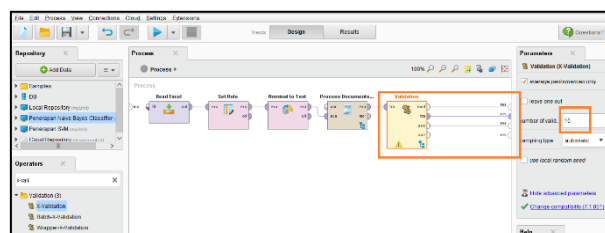
Gambar 7 TF-IDF

Setelah memilih metode TF-IDF klik dua kali pada operator *Documents from Data* untuk masuk ke dalam *preprocessing*.



Gambar 8 Pre-processing

Untuk menerapkan dan melakukan pengujian sehingga terbentuk *confusion matrix* dari pengujian data yang telah di *preprocessing* gunakan operator *X-Validation*. *X-Validation* atau sering dikenal sebagai *cross validation* disini merupakan metode pengujian klasifikasi yang dapat mengukur performa penerapan metode *Naive Bayes Classifier* dan *Support Vector Machine*. *Cross Validation* digunakan karena pengujian dilakukan kepada seluruh data sebanyak *n-fold*. Sehingga seluruh data berperan sebagai data *training* dan data *testing*. Pada pengujian ini dilakukan dalam *10 fold* atau *default* dari Rapid Minner. Langkah yang dilakukan adalah memasukkan operator *X-Validation* ke dalam *process* dan hubungkan *connector* dengan operator sebelumnya dan ke *result* seperti pada gambar 9.

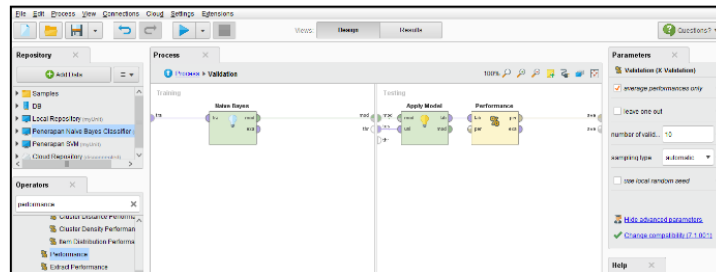


Gambar 9 Cross Validation

Implementasi *Naive Bayes Classifier*

Untuk menerapkan metode *Naive Bayes Classifier* dilakukan dengan *double click* pada operator *X-Validation* sehingga masuk ke halaman *Validation*. Pada tab *Training* masukkan operator dari metode *Naive Bayes* lalu hubungkan *connector training* dan *connector model*. Pada operator inilah data akan di *training* dengan metode *Naive Bayes Classifier*. Pada tab *Testing* masukkan operator *Apply Model* untuk membaca dan melakukan pengujian terhadap *training model* yang telah terbentuk dari *Naive Bayes Classifier* lalu hubungkan *connector model* dan *connector testing*. Selanjutnya masukkan operator *Perfarmace* untuk

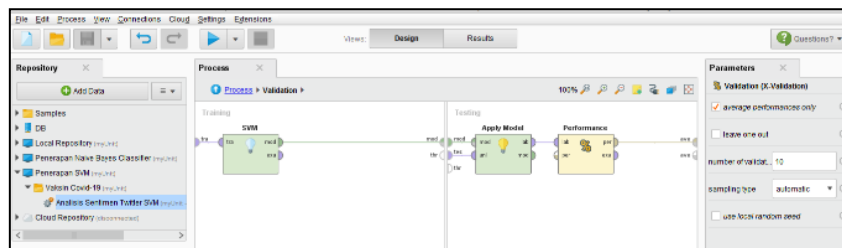
mengukur dan menampilkan hasil performa meliputi *accuracy*, *precision*, dan *recall* dari pengujian yang dilakukan.



Gambar 10 Naive Bayes Classifier

Implementasi Support Vector Machine

Untuk menerapkan metode *Support Vector Machine* dilakukan dengan *double click* pada operator *X-Validation* sehingga masuk ke halaman *Validation*. Pada tab *Training* masukkan operator dari metode *Naive Bayes* lalu hubungkan *connector training* dan *connector model*. Pada operator inilah data akan di *training* dengan metode *Support Vector Machine*. Pada tab *Testing* masukkan operator *Apply Model* untuk membaca dan melakukan pengujian terhadap *training model* yang telah terbentuk dari *Support Vector Machine* lalu hubungkan *connector model* dan *connector testing*. Selanjutnya masukkan operator *Perfarmace* untuk mengukur dan menampilkan hasil performa meliputi *accuracy*, *precision*, dan *recall* dari pengujian yang dilakukan.



Gambar 11 Support Vector Machine

Setelah membentuk model dari kedua metode didapatkan hasil *confusion matrix* yang berbeda. Hal ini menunjukkan performa dari kedua metode tersebut pun memiliki perbedaan. Berikut ini merupakan *performance vector* dari metode *Naive Bayes Classifier*.

PerformanceVector

```
PerformanceVector:
accuracy: 72.88% +/- 2.84% (mikro: 72.87%)
ConfusionMatrix:
True: Positif Negatif
Positif:      600    105
Negatif:      166    128
precision: 43.49% +/- 4.49% (mikro: 43.54%) (positive class: Negatif)
ConfusionMatrix:
True: Positif Negatif
Positif:      600    105
Negatif:      166    128
recall: 54.95% +/- 10.60% (mikro: 54.94%) (positive class: Negatif)
ConfusionMatrix:
True: Positif Negatif
Positif:      600    105
Negatif:      166    128
```

Gambar 12 Performance Vector Naive Bayes

Berikut ini merupakan *performance vector* dari metode *Support Vector Machine*.

Debby Alita : * Penulis Korespondensi



Copyright © 2023, Debby Alita.

PerformanceVector

```

PerformanceVector:
accuracy: 77.88% +/- 1.26% (mikro: 77.88%)
ConfusionMatrix:
True:  Positif Negatif
Positif:      760   215
Negatif:       6    18
precision: 75.00% (positive class: Negatif)
ConfusionMatrix:
True:  Positif Negatif
Positif:      760   215
Negatif:       6    18
recall: 7.70% +/- 5.31% (mikro: 7.73%) (positive class: Negatif)
ConfusionMatrix:
True:  Positif Negatif
Positif:      760   215
Negatif:       6    18

```

Gambar 13 Performance Vector Support Vector Machine

Hasil *performance vector* dari algoritma *Naive Bayes Classifier* menunjukkan bahwa algoritma ini mampu mengelompokkan data dengan ketepatan klasifikasi dan hasil pengujian *data training* mencapai nilai 72.88%. Nilai *precision* menunjukkan bahwa tingkat ketepatan informasi yang dihasilkan oleh algoritma *Naive Bayes Classifier* dengan klasifikasi sentimen yang diminta mencapai nilai 43.49% dan keberhasilan algoritma *Naive Bayes Classifier* dalam menemukan/mengklasifikasi data mencapai nilai 54.95%.

Sedangkan hasil *performance vector* dari algoritma *Support Vector Machine* menunjukkan bahwa algoritma ini mampu mengelompokkan data dengan ketepatan klasifikasi dan hasil pengujian *data training* mencapai nilai 77.88%. Nilai *precision* menunjukkan bahwa tingkat ketepatan informasi yang dihasilkan oleh algoritma *Support Vector Machine* dengan klasifikasi sentimen yang diminta mencapai nilai 75.00% dan keberhasilan algoritma *Support Vector Machine* dalam menemukan/mengklasifikasi data mencapai nilai 7.70%.

4. KESIMPULAN

Pembahasan yang telah dilakukan diatas dapat ditarik kesimpulan dari penelitian ini. Dimana kesimpulan dari penelitian yang telah dilakukan antara lain analisis sentimen tweet menunjukkan jumlah Sentimen Positif sebanyak 790 *tweet* dan jumlah sentimen negatif sebanyak 361 *tweet*. Maka dikatakan bahwa lebih banyak pengguna Twitter yang memberikan respon positif terhadap kegiatan Vaksin Covid-19. Penerapan metode *Naive Bayes Classifier* menunjukkan hasil *performance vector* dengan nilai *accuracy* 72.88%, *precision* 43.49% dan *recall* 54.95%. Penerapan metode *Support Vector Machine* menunjukkan hasil pengujian dengan nilai *Accuracy* 77.88%, *Precision* 75.00%, dan *Recall* 7.70%. Metode *Naive Bayes Classifier* memiliki kemampuan yang lebih baik dalam menganalisis sentimen Twitter dengan nilai rata-rata *performance* 57.10% dibandingkan dengan *Support Vector Machine* hanya mencapai rata-rata *performance* 40,77%.

5. REFERENCES

- [1] R. I. Kemenkes, "Pedoman Kesiapsiagaan Menghadapi Coronavirus Disease (COVID-19)," *Direktorat Jenderal Pencegah dan Pengendali Penyakit*, 2020.
- [2] U. Yaqub, N. Sharma, R. Pabreja, S. A. Chun, V. Atluri, and J. Vaidya, "Location-based sentiment analyses and visualization of Twitter election data," *Digit. Gov. Res. Pract.*, vol. 1, no. 2, pp. 1–19, 2020.
- [3] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Edutic-Scientific J. Informatics Educ.*, vol. 7, no. 1, 2020.

-
- [4] R. R. SURYONO and B. Indra, "P2P Lending Sentiment Analysis in Indonesian Online," 2020.
 - [5] C. Colón-Ruiz, "Semi-Supervised Generative Adversarial Network for Sentiment Analysis of drug reviews." Institute of Electrical and Electronics Engineers (IEEE), 2021, doi: 10.36227/techrxiv.17075054.
 - [6] C. Du and L. Huang, "Sentiment Analysis Method based on Piecewise Convolutional Neural Network and Generative Adversarial Network," *International Journal of Computers Communications & Control*, vol. 14, no. 1. Agora University of Oradea, pp. 7–20, 2019, doi: 10.15837/ijccc.2019.1.3374.
 - [7] I. Ahmad, H. Sulistiani, and H. Saputra, "The Application Of Fuzzy K-Nearest Neighbour Methods for A Student Graduation Rate," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, pp. 47–52, 2018.
 - [8] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, pp. 86–91, 2020.
 - [9] S. Setiawansyah, Q. J. Adrian, and R. N. Devija, "Penerapan Sistem Informasi Administrasi Perpustakaan Menggunakan Model Desain User Experience," *J. Manaj. Inform.*, vol. 11, no. 1, pp. 24–36, 2021.
 - [10] M. N. D. Satria, "Application of SAW in the Class Leader Selection Decision Support System," *Chain J. Comput. Technol. Comput. Eng. Informatics*, vol. 1, no. 1, pp. 27–31, 2023.