



ANALISIS SENTIMEN TERHADAP MASYARAKAT INDONESIA DI MASA PPKM MENGUNAKAN ALGORITMA NAÏVE BAYES

Ahmad Turmudi Zy¹, Aswan Sunge², Riani³, Edy Widodo⁴

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹turmudi@pelitabangsa.ac.id, ²aswan.sunge@pelitabangsa.ac.id, ³riani.fauqiyah@gmail.com

⁴ewidodo@pelitabangsa.ac.id

Abstrak

Coronavirus merupakan suatu kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia. Beberapa jenis coronavirus diketahui menyebabkan infeksi saluran nafas pada manusia mulai dari batuk pilek hingga yang lebih serius seperti Middle East Respiratory Syndrome (MERS) dan Severe Acute Respiratory Syndrome (SARS). Salah satu topik yang menjadi perbincangan oleh masyarakat saat ini diantaranya di media sosial twitter adalah kebijakan pemerintah mengenai Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). PPKM merupakan kebijakan Pemerintah Indonesia untuk menangani COVID-19 yang dibuat sejak awal tahun 2021. Pemberlakuan PPKM menimbulkan prokontra dari masyarakat. Berdasarkan hasil survei SRMC yang diberitakan melalui laman saifulmunjani.com disebutkan bahwa secara nasional terdapat 44% memilih menjalankan PPKM secara ketat walaupun disisi lain penghasilan menurun, dan terdapat 40% memilih untuk menghentikan PPKM dengan peningkatan risiko penularan COVID-19. Berdasarkan permasalahan yang terjadi, maka menjadi dasar dalam penelitian ini yang bertujuan untuk mengetahui bagaimana sentimen masyarakat terhadap penerapan kebijakan PPKM di Indonesia melalui tweet dan komentar di platform media sosial Twitter menggunakan analisis sentiment

Kata Kunci: PPKM, Naïve Bayes, Covid19

Abstract

Coronavirus is a group of viruses that can cause disease in animals or humans. Several types of coronavirus are known to cause respiratory tract infections in humans ranging from coughs and colds to more serious ones such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). One of the topics currently being discussed by the public, including on social media Twitter, is the government's policy regarding the Enforcement of Restrictions on Community Activities (PPKM). PPKM is a policy of the Government of Indonesia to deal with COVID-19 that has been made since early 2021. The implementation of PPKM raises pros and cons from the community. Based on the results of the SRMC survey reported through the saifulmunjani.com page, it was stated that nationally, 44% chose to strictly implement PPKM even though on

the other hand income decreased, and 40% chose to stop PPKM with an increased risk of COVID-19 transmission. Based on the problems that occurred, it became the basis of this research which aims to find out how the public sentiment towards the implementation of PPKM policies in Indonesia through tweets and comments on the Twitter social media platform using sentiment analysis

Keywords: PPKM, Naïve bayes, Covid19

1. Pendahuluan

Coronavirus merupakan suatu kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia. Beberapa jenis coronavirus diketahui menyebabkan infeksi saluran nafas pada manusia mulai dari batuk pilek hingga yang lebih serius seperti Middle East Respiratory Syndrome (MERS) dan Severe Acute Respiratory Syndrome (SARS). Coronavirus jenis baru

yang ditemukan menyebabkan penyakit COVID-19, Orang dapat tertular COVID-19 dari orang lain yang terjangkit virus ini. COVID-19 dapat menyebar dari orang ke orang melalui percikan-percikan dari hidung atau mulut yang keluar saat orang yang terjangkit COVID-19 batuk atau mengeluarkan napas. Percikan-percikan ini kemudian jatuh ke benda-benda dan permukaan-permukaan disekitar. Orang yang menyentuh benda atau permukaan tersebut lalu menyentuh mata, hidung atau mulutnya, dapat terjangkit COVID-19.

Salah satu topik yang menjadi perbincangan oleh masyarakat saat ini diantaranya di media sosial twitter adalah kebijakan pemerintah mengenai Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). PPKM merupakan kebijakan Pemerintah Indonesia untuk menangani COVID-19 yang dibuat sejak awal tahun 2021. Sebelumnya pemerintah telah melakukan pembatas sosial berskala besar (PSBB) sebelum adanya PPKM yang dilakukan di sejumlah wilayah Indonesia. PPKM berlangsung di beberapa wilayah yang menjadi titik penyebaran infeksi COVID-19, yakni di Pulau Jawa dan Bali.

Media sosial yang populer khususnya di Indonesia yakni Twitter. Twitter merupakan salah satu media sosial yang didirikan oleh Jack Dorsey yang umumnya berfungsi untuk mengirmkan pesan yang disebut dengan kicauan atau (tweet). Penggunaan twitter umumnya diakses para pengguna di Indonesia mencapai 59% dan menduduki peringkat ke-5 media sosial yang sering digunakan pada tahun 2020. Dari data yang ada menjadikan platform twitter menjadi media sosial yang cukup memiliki pengaruh bagi para pengguna di Indonesia. Pemberlakuan PPKM menimbulkan prokontra dari masyarakat. Berdasarkan hasil survei SRMC yang diberitakan melalui laman saifulmunjani.com disebutkan bahwa secara nasional terdapat 44% memilih menjalankan PPKM secara ketat walaupun disisi lain penghasilan menurun, dan terdapat 40% memilih untuk menghentikan PPKM dengan peningkatan risiko penularan COVID-19. Media sosial merupakan media yang memiliki berbagai informasi karena adanya suatu perkembangan teknologi yang memudahkan pengguna untuk mengakses dan digunakan untuk mencurahkan hatinya termasuk permasalahan penerapan PPKM melalui platform tersebut.

2. Landasan Pemikiran

Menurut penelitian ini dapat diperoleh kesimpulan yang digunakan berasal dari media sosial twitter. Data tersebut diklarifikasikan kedalam dua kelas yaitu 98 data opini positif dan 36 data opini negatif. Dari hasil pengujian dengan menggunakan algoritma naive bayes tersebut didapatkan rata-rata akurasi sebesar 77,18%. [1].

Menurut penelitian ini mengklasifikasikan sentimen terhadap komentar film di media sosial twitter untuk mendapatkan kecenderungan opini terhadap film apakah cenderung beropini positif atau negatif. Algoritma yang digunakan pada penelitian ini adalah

Algoritma Naive Bayes. Berdasarkan hasil eksperimen, analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90% dengan rincian nilai precision 92%, recall 90% dan f-measure 90%. [2]

Menurut penelitian ini untuk memisahkan tweets berbahasa Indonesia pada media sosial twitter kedalam kategori positif, negatif dan netral. Metode yang digunakan yaitu Naive Bayes Classifier (NBC) dengan feature selection Particle Swarm Optimization (PSO) yang diterapkan pada dataset untuk mengurangi atribut yang kurang relevan pada saat proses klasifikasi. Hasil pengujian menunjukkan bahwa algoritma Naive Bayes Classifier dengan feature selection Particle Swarm Optimization (PSO) menggunakan parameter term frequency (TF) diperoleh hasil akurasi sebesar 97,48%.

Menurut penelitian ini bertujuan untuk mengklafikasi hasil data opini mahasiswa yang dituliskan pada form keputusan angket mahasiswa terhadap dosen dengan menggunakan metode

Naive Bayes Classifier (NBC) sehingga menghasilkan klasifikasi secara otomatis. Hasil penelitian ini mempunyai tingkah precision 75% , recall 75% dan akurasi 80%. [3]

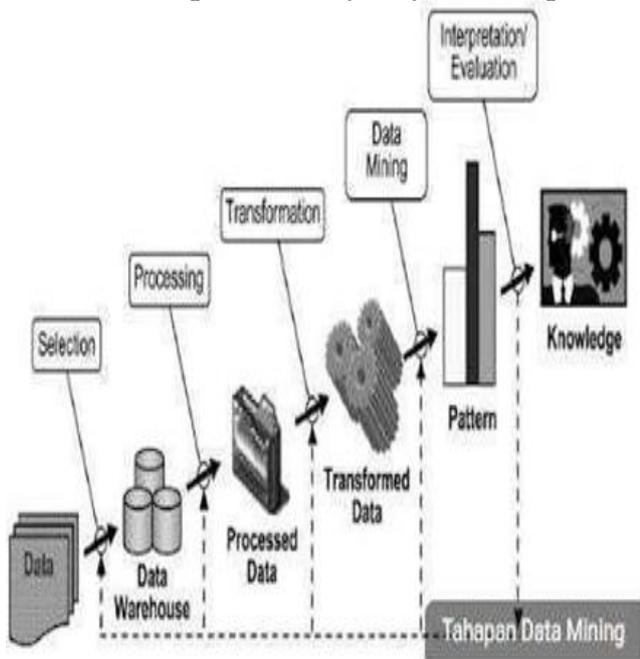
Analisis yang dilakukan yaitu mengklasifikasi tweet yang berisi komentar negatif dan positif menggunakan metode Naive Bayes Classifier.

Sementara metode Naive Bayes Classifier dikombinasikan dengan fitur yang dihitung bobotnya menggunakan TF-IDF. Klasifikasi tweet pada penelitian ini terdiri dari kelas sentimen negatif dan positif. Hasil pengujian pada penelitian didapatkan dengan menggunakan tools Rapid Miner yang memperlihatkan bahwa akurasi klasifikasi dengan ditambahkan fitur TF-IDF menghasilkan tingkat akurasi sebesar 55.80%. [4]

Analisis Sentimen atau opinion mining adalah suatu bidang studi yang menganalisis pendapat atau pandangan seseorang, sikap orang, dan emosi seseorang terhadap entitas seperti produk, layanan, individu, masalah, peristiwa, topik dan atributnya. Analisis sentimen di definisikan sebagai pendapat penulis untuk entitas tertentu. Dalam proses pengambilan keputusan seseorang dipengaruhi oleh pendapat yang di bentuk oleh pemikiran berbagai orang. Sebagai contoh kecil ketika seseorang ingin membeli produk secara online, orang tersebut biasanya akan mulai dengan mencari ulasan dan berbagai pendapat yang ditulis oleh orang lain pada berbagai penawaran.

Data mining merupakan teknik pengekraksi pengetahuan secara implisit yang berguna dari data memiliki jumlah yang besar [3]. Data ini juga telah banyak digunakan untuk menyelesaikan masalah bisnis diantaranya segementasi pelanggan, retensi pelanggan, penilaian kredit, rekomendasi produk, pemasaran langsung, penjualan silang, deteksi penipuan, dan lain sebagainya. Data mining adalah suatu tahap eksplorasi dari “deteksi informasi dalam file” prosedur untuk menentukan desain dalam koleksi data yang cukup besar termasuk pendekatan terhadap hubungan simulasi kecerdasan buatan, Machine Learning dan juga perekaman sistem.

Berikut gambaran tahap tahap data mining:



Gambar 1 Tahapan Data Mining

Text Mining atau penambang data dapat didefinisikan secara luas sebagai proses intensif pengetahuan dimana pengguna berinteraksi dengan berbagai kumpulan dokumen dari awal ke waktu dengan menggunakan seperangkat alat analisis. Penambahan text juga mendapatkan aspirasi dan arahan dari penelitian tentang adanya penambangan data (Data Mining). Maka dari itu, tidak mengherankan untuk menemukan bahwa penambangan teks dan penambangan data menunjukkan kesamaan arsitektur tingkat tinggi. Sebagai contoh, dari kedua sistem ini bergantung pada rutinitas preprocessing[6]. Text mining atau sering disebut penambangan data adalah penambangan yang dilakukan oleh computer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui. Informasi tersebut berasal dari informasi yang diekstraksi secara otomatis yang di ambil dari berbagai sumber data teks.

Naïve Bayes merupakan metode klasifikasi berdasarkan probabilitas dalam statistik yang digunakan oleh Thomas Bayes yang memprediksi peluang dimasa akan datang berdasarkan peluang sebelumnya (teorema bayes). Kemudian metode ini dikombinasikan dengan “naïve” dimana kondisi antar atribut saling bebas tidak berhubungan satu sama lain[10].

Prediksi Bayes didasarkan pada teorema Bayes dengan formula umum. Sebagai berikut:

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

Keterangan:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik

- P(H|X) : Probabilitas hipotesis berdasarkan kondisi (posteriori probability)
- P(H) : Probabilitas hipotesis (priori probability)
- P(X|H) : Probabilitas berdasarkan kondisi pada data sampel
- P(X) : Probabilitas X

K-fold cross validation adalah teknik yang dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah instance tidak banyak). K-fold cross validation merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. K-fold cross validation diawali dengan membagi data sejumlah n-fold yang diinginkan. Dalam proses cross validation data akan dibagi dalam n buah partisi dengan ukuran yang sama D1, D2, D3...Dn selanjutnya proses testing dan training dilakukan sebanyak n kali. Dalam iterasi ke-i partisi Di akan menjadi data testing dan sisanya akan menjadi data training. Untuk penggunaan jumlah fold terbaik untuk uji validitas, dianjurkan menggunakan 10-fold cross validation dalam model [11].

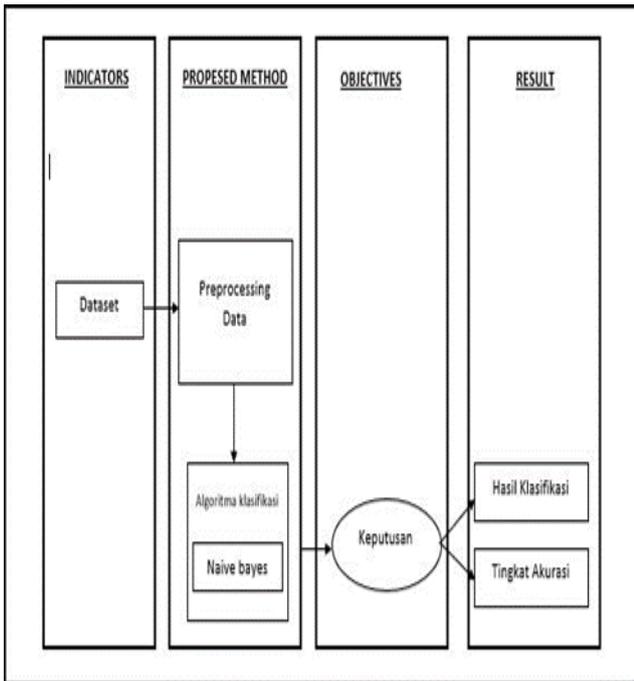
Confusion matrix atau sering juga dikenal dengan error matrix. Confusion matrix memberikan sebuah informasi perbandingan dan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi sebenarnya. Confusion matrix berbentuk table matrix yang menggambarkan sebuah kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya di ketahui.

1. Presisi
Presisi adalah pengukuran seberapa ketepatan informasi yang diberikan oleh sistem.
2. Recall
Recall adalah pengukur tingkat keberhasilan pengambilan data yang relevan.

RStudio merupakan Bahasa pemrograman komputer yang memungkinkan pengguna untuk memprogramkan algoritma dan menggunakan alat yang telah dikembangkan melalui R oleh pengguna lainnya. R merupakan Bahasa pemrograman tingkat tinggi dan juga merupakan lingkungan untuk analisis data dan grafik. Desain R dipengaruhi sangat kuat oleh dua Bahasa pemrograman computer yang telah ada sebelumnya, yaitu Bahasa S yang dikembangkan oleh Becker, Chamber, dan Wilk, serta Bahasa pemrograman Scheme yang dikembangkan oleh Susman. Dengan demikian Bahasa ini sangat mirip dengan Bahasa S, tetapi implementasi dan semantiknya ditopang oleh Scheme [12].

Penulis membuat kerangka pemikiran ini untuk mempermudah dalam memahami arah penelitian dari proses-proses hingga hasil yang akan diperoleh,

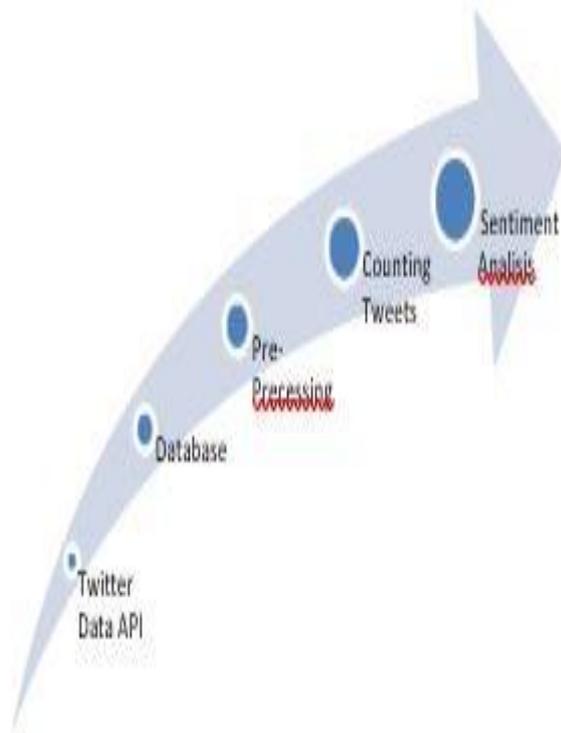
kerangka berpikir ini dapat di lihat pada gambar di bawah.



Gambar 2 Kerangka Pemikiran

3. Metode Penelitian

Penelitian ini melalui beberapa proses, alur urutan proses pada penelitian ini akan diuraikan pada metode penelitian ini, berikut ini merupakan flowchart yang menunjukkan tahapan penelitian dimulai dari pengumpulan data hingga diperoleh kesimpulan:



Gambar 1 Flow Tahapan Penelitian

Proses pengumpulan data dilakukan dengan mengambil data tweet menggunakan twitter API sebagai

sumber data informasi. Dengan cara memasukan Api Key Secret yang diperoleh dengan cara mendaftar melalui twitter development. Data dikumpulkan menggunakan pencarian keyword “PPKM”. Objek yang digunakan dalam penelitian ini yaitu data tweet berbahasa Indonesia yang terdapat pada media sosial Twitter. Tweet yang digunakan ialah tweet-tweet yang mengandung sentiment positif dan sentiment negatif. Dari masing-masing sentimen tersebut, diambil jumlah data 1000 data persentimen sehingga total tweet yang digunakan sebagai data berjumlah 2000 data.

3.1 Data Pre-Processing

Pre-processing dilakukan pada data training untuk membantu proses pembelajaran algoritma. Proses dilakukan untuk merubah data dari yang tidak beraturan menjadi beraturan atau terstruktur sehingga dapat diproses dengan mudah oleh algoritma. Text pre-processing pada penelitian ini dilakukan dengan 6 tahap yaitu:

1. **Cleansing:** Cleansing adalah proses penghapusan karakter yang tidak dibutuhkan untuk mengurangi noise. Karakter yang dihapus berupa tanda baca, simbol-simbol, nama pengguna, hastag, emoticon, URL dan simbol yang tidak diperlukan akan dihapus.
2. **Case Folding:** Proses ini mengubah semua huruf kedalam huruf kecil yang tadinya tidak selaras menjadi selaras yaitu huruf kecil semua atau lower case.
3. **Tokenizing:** Proses pemisahan kalimat menjadi kata dasar yang nantinya disebut token atau term.
4. **Normalization:** Tahap ini dilakukan untuk menormalkan kata-kata yang disingkat diperpanjang menjadi kata yang normal sesuai KBBI. Pada tahap ini juga menterjemahkan bahasa daerah menjadi bahasa indonesia.
5. **Filtering:** Kata-kata yang sering muncul secara umum kurang relevan dengan teks akan dihapus. Tahap ini disebut juga stopword removing, maksudnya menghapus kata yang tidak memiliki arti dan tidak berpengaruh terhadap analisis sentimen.
6. **Stemming:** Tahap ini dilakukan untuk mengubah kata berimbuhan menjadi kata dasar nya.

3.2 Pembagian Data

Informational index dalam penelitian ini adalah tanggapan mengenai aturan ppkm dimasa pandemi Coronavirus dengan catchphrase "ppkm" pada media sosial twitter dari tanggal 6 september 2021 sampai 10 september 2021. Information tanggapan yang diperoleh sebanyak 1689 tanggapan. Penentuan jumlah information preparing dan testing dengan perbandingan rasio untuk masing-masing information positif dan negatif sebanyak 735 information tanggapan positif dan 559 information tanggapan negatif, untuk information netral tidak digunakan karena tidak memberikan informasi yang penting. Pada penelitian ini untuk seluruh information dibagi ke dalam 5 kombinasi

perbandingan rasio information preparing dan information testing seperti pada tabel dibawah ini.

Tabel 1 Rasio Data Training dan Data Testing

No.	Jumlah Data	Rasio Data Training: Testing (%)	Data Training	Data Testing
1		20:80	800	3200
2		40:60	1600	2400
3	4000	60:40	2400	1600
4		80:20	3200	800
5		90:10	3600	400

4. Pembahasan

4.1 Simulasi Perhitungan Manual Algoritma

Misalnya terdapat empat buah dokumen yang telah melalui tahapan preprocessing, dua dokumen diambil dari kelas positif dan dua dokumen diambil dari kelas negatif. Dari 2 dokumen dengan jumlah kata sebanyak 24 kata dari 39 kosakata yang ada, sedangkan 2 dokumen kelas negatif terdiri dari 23 kata dari yang ada. Berdasarkan jumlah kata tersebut, dapat dihitung nilai probabilitas untuk setiap kelasnya dengan menggunakan persamaan berikut :

1. Probabilitas kata kelas positif
Contoh perhitungan probabilitas untuk kata "PPKM" yang terdapat dalam kelas positif.

$$P(w_i|c) = \frac{\text{count}(w_i, C) + 1}{((\sum w \in V \text{count}(w, c) + |V|)}$$

$$P \text{ppkm positif} = \frac{(2 + 1)}{(24 + 39)} = 0,047619048$$

2. Probabilitas kata kelas negative
Contoh perhitungan probabilitas untuk kata "PPKM" yang terdapat dalam kelas negatif.

$$P(w_i|c) = \frac{\text{count}(w_i, C) + 1}{((\sum w \in V \text{count}(w, c) + |V|)}$$

$$P \text{ppkm negatif} = \frac{(2 + 1)}{(23 + 39)} = 0,048387097$$

Nilai probabilitas kata pada masing-masing kelas tersebut, kemudian disimpan pada database yang nantinya akan digunakan untuk menguji data baru. Misal ingin diketahui kelas data dari tanggapan baru "vaksin covid aman". Langkah pertama yang dilakukan untuk melakukan klasifikasi adalah memecah kalimat dalam tanggapan tersebut menjadi kata per kata, selanjutnya menghitung nilai probabilitas dari kata pada masing-masing kelas dengan menggunakan tabel probabilitas kata yang telah diperoleh sebelumnya, sedangkan probabilitas masing-masing kelas ditentukan dengan menggunakan persamaan berikut.

Tabel 1 Nilai Probabilitas Tanggapan Baru

Kelas	Vaksin	Covid	Aman	Nilai Probabilitas
Positif (P=0,5)	0,03174 6032	9048	0,03174 6032	0,00002 3995
Negatif (P=0)	0,01612 9032	0,04838 7097	0,01612 9032	0,00000 6294

Untuk menggambarkan seberapa baik kinerja machine learning dalam mengklasifikasikan data, salah satunya metode yang digunakan sebagai pengukuran kinerja klasifikasi yaitu confusion matrix. Confusion matrix mengandung nilai true positive, true negative, false positive, false negative. Nilai dari true positive dan true negative memberikan informasi bahwa ketika classifier dalam melakukan klarifikasi data yang bernilai benar, dan sedangkan false negative dan false positive memberikan informasi bahwa Ketika classifier salah dalam melakukan pengklasifikasian data.

Pengukuran efektif dapat dilakukan dengan perhitungan nilai akurasi, nilai precision, dan nilai recall. Misalkan diperoleh confusion matrix hasil perhitungan klasifikasi menggunakan algoritma naïve bayes sebagai berikut:

Tabel 2 Confusion Matrix Perhitungan Naïve Bayes

Aktual	Prediksi	
	Positif	Negatif
Positif	382 (TP)	18 (FN)
Negatif	120 (FP)	280 (TN)

Berdasarkan tabel diatas diperoleh nilai TP sebesar 382, FN sebesar 18, FP sebesar 120 dan TN sebesar 280. Sehingga untuk menghitung nilai akurasi, precision dan recall tersebut adalah sebagai berikut

$$\text{Akurasi} = \frac{((TP + TN))}{((TP + FP + TN + FN))}$$

$$\text{Akurasi} = \frac{((382 + 280))}{((381 + 120 + 18 + 280))}$$

$$\text{Akurasi} = \frac{((662))}{((800))}$$

$$\text{Akurasi} = 0.8275$$

$$\text{Akurasi} = 82,75\%$$

$$\text{Precision} = \frac{TP}{((TP + FP))}$$

$$\text{Precision} = \frac{382}{((382 + 120))}$$

$$\text{Precision} = \frac{382}{((502))}$$

$$\text{Precision} = 0.7609562$$

$$\text{Precision} = 76,09\%$$

$$\text{Recall} = \frac{TP}{((TP + FN))}$$

$$\text{Recall} = \frac{x_{y,2}}{y}$$

$$\text{Recall} = 0,955$$

$$\text{Recall} = 95,5\%$$

Ser., vol. 893, no. 1, 2017, doi: 10.1088/1742-6596/893/1/012051.

- [10] L. Singh, L. Bode, C. Budak, K. Kawintiranon, C. Padden, and E. Vraga, "Understanding high- and low-quality URL Sharing on COVID-19 Twitter streams," *J. Comput. Soc. Sci.*, vol. 3, no. 2, pp. 343–366, 2020, doi: 10.1007/s42001-020-00093-6.

- [11] T. Olofinlua, "Twitter: social communication in the twitter age," *Information, Commun. Soc.*, vol. 22, no. 13, pp. 2037–2038, 2019, doi: 10.1080/1369118x.2019.1620824.

- [12] S. Haryati, A. Sudarsono, and E. (2015) Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.