

KOMPARASI ALGORITMA *K-NEAREST NEIGHBOR* DAN *NAIVE BAYES* DALAM FAKTOR KETIDAKDISIPLINAN MASYARAKAT TERHADAP PROTOKOL KESEHATAN COVID-19

Muhammad Aliyul Amri¹, Solikhun², Rizki Alfadillah Nasution³

¹Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar

^{2,3}STIKOM Tunas Bangsa Pematangsiantar

E-mail : muhammadaliyul.amri12@gmail.com¹

Abstrak

Penelitian ini bertujuan untuk melakukan perbandingan dua metode dalam *Data Mining* klasifikasi yaitu algoritma *Naive Bayes* dan algoritma *K-Nearest Neighbor* dalam permasalahan faktor ketidaksiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19 di tempat wisata dan perbelanjaan daerah Pematangsiantar dan Simalungun. Penelitian ini menggunakan lima kriteria yang mempengaruhi ketidaksiplinan masyarakat yaitu Ketersediaan Alat, Perasaan Masyarakat, Ketegasan Hukum, Lingkungan Masyarakat dan Ekonomi. Pengumpulan data dilakukan dengan kuesioner yang disebarluaskan melalui Google Formulir kepada masyarakat sebanyak 104 orang. Hasil Penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbor* lebih baik digunakan dalam permasalahan ketidaksiplinan masyarakat terhadap protokol kesehatan COVID-19 yang memiliki tingkat akurasi sebesar 100% jika dibanding dengan algoritma *Naive Bayes* yang hanya memiliki tingkat akurasi sebesar 94,231%.

Kata kunci : Komparasi, Naive Bayes, K-Nearest Neighbor, Protokol Kesehatan COVID-19

Abstract

This study aims to compare two methods in Data Mining classification, namely the Naive Bayes algorithm and the K-Nearest Neighbor algorithm in the problem of community indiscipline factors in carrying out the COVID-19 health protocol in tourist and shopping areas in Pematangsiantar and Simalungun. This study uses five criteria that affect community indiscipline, namely Availability of Tools, Community Feelings, Legal Firmness, Community Environment and Economics. Data was collected by means of questionnaires distributed via Google Forms to 104 people. The results of this study indicate that the K-Nearest Neighbor algorithm is better used in community indiscipline problems against the COVID-19 health protocol which has an accuracy rate of 100% when compared to the Naive Bayes algorithm which only has an accuracy rate of 94.231%.

Keywords : Comparison, Naive Bayes, K-Nearest Neighbor, COVID-19 Health Protocol

1. PENDAHULUAN

Sejak kemunculan COVID-19 (*Coronavirus Disease 2019*) pada tahun 2019, masyarakat dunia harus benar-benar ekstra berhati-hati dalam menjaga kesehatannya agar tidak terinfeksi virus COVID-19. Di Indonesia untuk menjaga kesehatan masyarakat dari penyebaran virus COVID-19, Menteri Kesehatan menyelenggarakan kegiatan kesehatan yakni penyelenggaraan Protokol Kesehatan COVID-19. Kegiatan Protokol Kesehatan ini dilakukan agar dapat mengurangi penyebaran COVID-19 khususnya di tempat-tempat umum seperti tempat wisata, pasar, sekolah, tempat kerja dan sebagainya, karena dari tempat umum inilah yang menjadikan risiko penularan COVID-19 sangat besar di Indonesia.

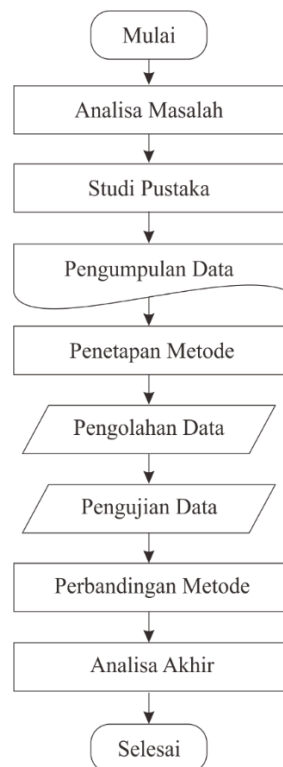
Seiring berjalannya waktu masyarakat Indonesia mulai tidak disiplin dalam menjalankan protokol kesehatan yang seharusnya dapat memutuskan rantai penyebaran penularan virus COVID-19, hal ini menyebabkan semakin meningkatnya kasus positif COVID-19 setiap bulannya di Indonesia. Masyarakat yang tidak disiplin dalam menjalankan protokol kesehatan memiliki alasan tersendiri, seperti banyak masyarakat yang mengeluhkan ketersediaan alat kesehatan, ketegasan hukum, ekonomi, lingkungan, bahkan perasaan tidak nyaman yang dirasakan masyarakat saat menjalankan protokol kesehatan merupakan faktor yang mempengaruhi ketidaksiplinan masyarakat dalam menjalankan protokol kesehatan.

Untuk itu penelitian ini menggunakan *Data Mining* klasifikasi yang membandingkan dua metode yaitu algoritma *Naive Bayes* dengan algoritma *K-Nearest Neighbor* dalam permasalahan ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19. Komparasi dua metode juga pernah dilakukan oleh penelitian sebelumnya, yang meneliti tentang menentukan tingkat keberhasilan *Immunotherapy* untuk pengobatan penyakit kanker kulit oleh F. Lia Dwi Cahyanti pada tahun 2021 [1], serta penelitian yang dilakukan oleh Pamungkas pada tahun 2021 [2], penelitian yang dilakukan oleh Gunawan pada tahun 2021 [3], penelitian yang dilakukan oleh Bianome pada tahun 2020 [4], penelitian yang dilakukan oleh Lubis pada tahun 2019 [5] dan penelitian yang dilakukan oleh Handoko pada tahun 2021 [6]. Maka penelitian ini dilakukan untuk mengetahui faktor yang mempengaruhi ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19 dengan membandingkan algoritma *Naive Bayes* dan algoritma *K-Nearest Neighbor*. Dengan tujuan agar mengetahui metode manakah yang sangat sesuai dan akurat dalam menyelesaikan permasalahan ketidakdisiplinan masyarakat terhadap protokol kesehatan COVID-19.

Penelitian ini diharapkan dapat memberikan informasi dan wawasan mengenai pengaruh ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19 dan sebagai pertimbangan algoritma yang baik dalam penelitian yang dilakukan mengenai komparasi algoritma dalam *Data Mining*.

2. METODOLOGI

Rancangan penelitian digunakan untuk menjelaskan bagaimana menguraikan dan menyelesaikan melalui tahapan-tahapan masalah yang ada dalam penelitian. Rancangan penelitian ini dapat dilihat pada gambar *flowchart* berikut :



Gambar 1. *Flowchart* Penelitian

2.1. *Data Mining*

Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer untuk menganalisis dan mengekstraksi pengetahuan secara otomatis [7]. *Data Mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang dapat membantu dalam pengambilan keputusan [8].

2.2. Naive Bayes

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (atau aturan *Bayes*) dengan asumsi ketidaktergantungan yang kuat. *Naive Bayes* merupakan salah satu metode *Data Mining* klasifikasi yang memprediksi data di masa yang akan mendatang berdasarkan data sebelumnya atau data yang sudah ada [9]. Keuntungan menggunakan metode *Naive Bayes* adalah membutuhkan data training yang kecil dalam menentukan estimasi yang diperlukan dalam proses pengklasifikasian [10].

Menurut [11] perhitungan *Naive Bayes* secara umum memiliki bentuk rumus sebagai berikut :

$$P(Y|X) = \frac{P(X|Y)P_Y}{P_X} \tag{1}$$

Keterangan :

- X : Data dari kelas yang belum diketahui
- Y : Hipotesis data yang merupakan spesifik kelas tertentu
- P(Y|X) : Probabilitas hipotesis pada kondisi tertentu X
- P(Y) : Probabilitas hopotesis Y
- P(X|Y) : Probabilitas X saat kondisi hopotesis Y
- P(X) : Probabilitas X

2.3. K-Nearest Neighbor

Prinsip sederhana yang dapat menggambarkan algoritma *K-Nearest Neighbor* yaitu “Jika hewan itu seperti ayam, bersuara *petok-petok* seperti ayam, dan berpenampilan seperti ayam, maka kemungkinan hewan itu adalah ayam”. Dengan artian jika ada suatu masukan baru yang tidak dikenali, maka algoritma KNN akan mencari objek yang terdekat dengan objek yang baru di *input* pada suatu data.

Dan secara umum rumus persamaan dalam algoritma KNN yang telah digunakan oleh penelitian terdahulu sebagai berikut [12] [13]:

$$D(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \tag{2}$$

Keterangan :

- $D(x, y)$: Jarak *Euclidean* (*Euclidean Distance*)
- x_i : Record ke-i
- x_j : Record ke-j
- a_r : Data ke-r
- i,j : 1, 2, 3, ... n

2.4. RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*) yang melakukan analisis terhadap *Data Mining* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan keputusan yang baik. *RapidMiner* sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa dan Simon Fischer di *Artificial Intelligence Unit* dari University of Dortmund [14]. *RapidMiner* yang digunakan dalam penelitian ini yaitu *RapidMiner* versi 9.8.

3. HASIL DAN PEMBAHASAN

Penelitian ini membahas tentang faktor yang mempengaruhi ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19. Data diambil melalui kuesioner google formulir, didapatlah beberapa data dimana data tersebut memiliki lima variabel utama yaitu ketersediaan alat, ketegasan hukum, lingkungan masyarakat, ekonomi dan perasaan masyarakat saat menjalankan protokol kesehatan. Dan sub faktor dari kelima faktor utama dapat dilihat pada tabel berikut :

Tabel 1. Kelas Data

Variabel (C)	Nama Field	Kelas Data
C1	Ketersediaan Alat	Minim, Mencukupi dan Berlebihan
C2	Perasaan Masyarakat	Merasa jenuh, Meremehkan dan Menikmati
C3	Ketegasan Hukum	Tidak memiliki sanksi dan Penerapan sanksi lemah
C4	Lingkungan Masyarakat	Tidak mematuhi, Mematuhi dan Sangat mematuhi
C5	Ekonomi	Kekurangan, Mencukupi dan Bertambah

Data penelitian diambil melalui kuesioner yang dibagikan kepada masyarakat ditempat-tempat keramaian di daerah Kabupaten Simalungun dan Kota Pematangsiantar. Data penelitian dapat dilihat pada tabel berikut ini :

Tabel 2. Data Penelitian

Res	C1	C2	C3	C4	Ekonomi	Label
R1	Mencukupi	Merasa Jenuh	Tidak Memiliki Sanksi	Tidak Mematuhi	Mencukupi	Iya
R2	Mencukupi	Merasa Jenuh	Penerapan Sanksi Lemah	Tidak Mematuhi	Mencukupi	Iya
R3	Minim	Merasa Jenuh	Penerapan Sanksi Lemah	Tidak Mematuhi	Mencukupi	Iya
R4	Mencukupi	Merasa Jenuh	Tidak Memiliki Sanksi	Mematuhi	Mencukupi	Tidak
R5	Minim	Menikmati	Tidak Memiliki Sanksi	Sangat Mematuhi	Kekurangan	Tidak
R6	Minim	Merasa Jenuh	Penerapan Sanksi Lemah	Mematuhi	Kekurangan	Iya
R7	Berlebihan	Merasa Jenuh	Tidak Memiliki Sanksi	Tidak Mematuhi	Mencukupi	Iya
R8	Mencukupi	Merasa Jenuh	Penerapan Sanksi Lemah	Tidak Mematuhi	Kekurangan	Iya
R9	Berlebihan	Menikmati	Penerapan Sanksi Lemah	Sangat Mematuhi	Mencukupi	Tidak
R10	Minim	Merasa Jenuh	Penerapan Sanksi Lemah	Tidak Mematuhi	Mencukupi	Iya
R11	Mencukupi	Merasa Jenuh	Tidak Memiliki Sanksi	Tidak Mematuhi	Kekurangan	Iya
R12	Mencukupi	Menikmati	Penerapan Sanksi Lemah	Tidak Mematuhi	Mencukupi	Tidak
....
R98	Berlebihan	Merasa Jenuh	Tidak Memiliki Sanksi	Sangat Mematuhi	Bertambah	Tidak
R99	Minim	Merasa Jenuh	Tidak Memiliki Sanksi	Sangat Mematuhi	Kekurangan	Iya
R100	Mencukupi	Menikmati	Penerapan Sanksi Lemah	Mematuhi	Kekurangan	Tidak
R101	Minim	Merasa Jenuh	Tidak Memiliki Sanksi	Tidak Mematuhi	Mencukupi	Iya
R102	Mencukupi	Meremehkan	Penerapan Sanksi Lemah	Tidak Mematuhi	Mencukupi	Iya
R103	Mencukupi	Merasa Jenuh	Tidak Memiliki Sanksi	Tidak Mematuhi	Mencukupi	Iya
R104	Mencukupi	Meremehkan	Penerapan Sanksi Lemah	Tidak Mematuhi	Kekurangan	Iya

3.1 Proses Naive Bayes

Pada proses *Naive Bayes* dilakukan dengan mencari jumlah nilai probabilitas setiap label menggunakan rumus probabilitas sebelumnya seperti sebagai berikut :

$$P(\text{Iya}) = \frac{\text{Jumlah Iya}}{\text{Jumlah Data Label}} = \frac{47}{104} = 0,45192$$

$$P(\text{Tidak}) = \frac{\text{Jumlah Tidak}}{\text{Jumlah Data Label}} = \frac{57}{104} = 0,54808$$

Selanjutnya mencari nilai probabilitas setiap kriteria yaitu kriteria C1, C2, C3, C4 dan C5 sebagai berikut :

Kriteria C1

$$P(\text{Minim} | \text{Iya}) = \frac{\text{Jumlah Minim Pada Iya}}{\text{Jumlah Iya}} = \frac{27}{47} = 0,57447$$

$$P(\text{Minim} | \text{Tidak}) = \frac{\text{Jumlah Minim Pada Tidak}}{\text{Jumlah Tidak}} = \frac{8}{57} = 0,14035$$

$$P(\text{Mencukupi} | \text{Iya}) = \frac{\text{Jumlah Mencukupi Pada Iya}}{\text{Jumlah Iya}} = \frac{19}{47} = 0,40426$$

$$P(\text{Mencukupi} | \text{Tidak}) = \frac{\text{Jumlah Mencukupi Pada Tidak}}{\text{Jumlah Tidak}} = \frac{25}{57} = 0,43860$$

$$P(\text{Berlebih} | \text{Iya}) = \frac{\text{Jumlah Berlebih Pada Iya}}{\text{Jumlah Iya}} = \frac{1}{47} = 0,02128$$

$$P(\text{Berlebih} | \text{Tidak}) = \frac{\text{Jumlah Berlebih Pada Tidak}}{\text{Jumlah Tidak}} = \frac{24}{57} = 0,42105$$

Kriteria C2

$$P(\text{Meremehkan} | \text{Iya}) = \frac{\text{Jumlah Meremehkan Pada Iya}}{\text{Jumlah Iya}} = \frac{13}{47} = 0,27660$$

$$P(\text{Meremehkan} | \text{Tidak}) = \frac{\text{Jumlah Meremehkan Pada Tidak}}{\text{Jumlah Tidak}} = \frac{7}{57} = 0,12281$$

$$P(\text{Merasa Jenuh} | \text{Iya}) = \frac{\text{Jumlah Merasa Jenuh Pada Iya}}{\text{Jumlah Iya}} = \frac{30}{47} = 0,63830$$

$$P(\text{Merasa Jenuh} | \text{Tidak}) = \frac{\text{Jumlah Merasa Jenuh Pada Tidak}}{\text{Jumlah Tidak}} = \frac{22}{57} = 0,38596$$

$$P(\text{Menikmati} | \text{Iya}) = \frac{\text{Jumlah Menikmati Pada Iya}}{\text{Jumlah Iya}} = \frac{4}{47} = 0,08511$$

$$P(\text{Menikmati} | \text{Tidak}) = \frac{\text{Jumlah Menikmati Pada Tidak}}{\text{Jumlah Tidak}} = \frac{28}{57} = 0,49123$$

Kriteria C3

$$P(\text{Tidak Memiliki Sanksi} | \text{Iya}) = \frac{\text{Jumlah Tidak Memiliki Sanksi Pada Iya}}{\text{Jumlah Iya}} = \frac{29}{47} = 0,61702$$

$$P(\text{Tidak Memiliki Sanksi} | \text{Tidak}) = \frac{\text{Jumlah Tidak Memiliki Sanksi Pada Tidak}}{\text{Jumlah Tidak}} = \frac{23}{57} = 0,40351$$

$$P(\text{Penerapan Sanksi Lemah} | \text{Iya}) = \frac{\text{Jumlah Penerapan Sanksi Lemah Pada Iya}}{\text{Jumlah Iya}} = \frac{18}{47} = 0,38298$$

$$P(\text{Penerapan Sanksi Lemah} | \text{Tidak}) = \frac{\text{Jumlah Penerapan Sanksi Lemah Pada Tidak}}{\text{Jumlah Tidak}} = \frac{34}{57} = 0,59649$$

Kriteria C4

$$P(\text{Tidak Mematuhi} | \text{Iya}) = \frac{\text{Jumlah Tidak Mematuhi Pada Iya}}{\text{Jumlah Iya}} = \frac{34}{47} = 0,72340$$

$$P(\text{Tidak Mematuhi} | \text{Tidak}) = \frac{\text{Jumlah Tidak Mematuhi Pada Tidak}}{\text{Jumlah Tidak}} = \frac{4}{57} = 0,07018$$

$$P(\text{Mematuhi} | \text{Iya}) = \frac{\text{Jumlah Mematuhi Pada Iya}}{\text{Jumlah Iya}} = \frac{11}{47} = 0,23404$$

$$P(\text{Mematuhi} | \text{Tidak}) = \frac{\text{Jumlah Mematuhi Pada Tidak}}{\text{Jumlah Tidak}} = \frac{31}{57} = 0,54386$$

$$P(\text{Sangat Mematuhi} | \text{Iya}) = \frac{\text{Jumlah Sangat Mematuhi Pada Iya}}{\text{Jumlah Iya}} = \frac{2}{47} = 0,04255$$

$$P(\text{Sangat Mematuhi} | \text{Tidak}) = \frac{\text{Jumlah Sangat Mematuhi Pada Tidak}}{\text{Jumlah Tidak}} = \frac{22}{57} = 0,38596$$

Kriteria C5

$$P(\text{Kekurangan} | \text{Iya}) = \frac{\text{Jumlah Kekurangan Pada Iya}}{\text{Jumlah Iya}} = \frac{26}{47} = 0,55319$$

$$P(\text{Kekurangan} | \text{Tidak}) = \frac{\text{Jumlah Kekurangan Pada Tidak}}{\text{Jumlah Tidak}} = \frac{7}{57} = 0,12281$$

$$\begin{aligned}
 P(\text{Mencukupi} \mid \text{Iya}) &= \frac{\text{Jumlah Mencukupi Pada Iya}}{\text{Jumlah Iya}} = \frac{20}{47} = 0,42553 \\
 P(\text{Mencukupi} \mid \text{Tidak}) &= \frac{\text{Jumlah Mencukupi Pada Tidak}}{\text{Jumlah Tidak}} = \frac{38}{57} = 0,66667 \\
 P(\text{Bertambah} \mid \text{Iya}) &= \frac{\text{Jumlah Bertambah Pada Iya}}{\text{Jumlah Iya}} = \frac{1}{47} = 0,02128 \\
 P(\text{Bertambah} \mid \text{Tidak}) &= \frac{\text{Jumlah Bertambah Pada Tidak}}{\text{Jumlah Tidak}} = \frac{12}{57} = 0,21053
 \end{aligned}$$

Selanjutnya perhitungan data uji, data uji yang digunakan dalam penelitian ini berjumlah 52 dari 104 data, yaitu data yang diambil dari data penelitian mulai dari responden 53 sampai responden 104. Proses dilakukan menggunakan rumus sebagai berikut :

$$\text{Prediksi}(R_n) = P(\text{Label}) * P(\text{Kriteria}_1) * P(\text{Kriteria}_n)$$

Dan proses perhitungan dilakukan sebagai berikut :

Prediksi(R53 | Iya)

$$\begin{aligned}
 &= P(\text{Iya}) * P(\text{Berlebih} \mid \text{Iya}) * P(\text{Merasa Jenuh} \mid \text{Iya}) * \\
 &\quad P(\text{Tidak Memiliki Sanksi} \mid \text{Iya}) * P(\text{Sangat Mematuhi} \mid \text{Iya}) * \\
 &\quad P(\text{Mencukupi} \mid \text{Iya}) \\
 &= 0,45192 * 0,02128 * 0,63830 * 0,61702 * 0,04255 * 0,42553 \\
 &= 0,00007
 \end{aligned}$$

Prediksi(R53 | Tidak)

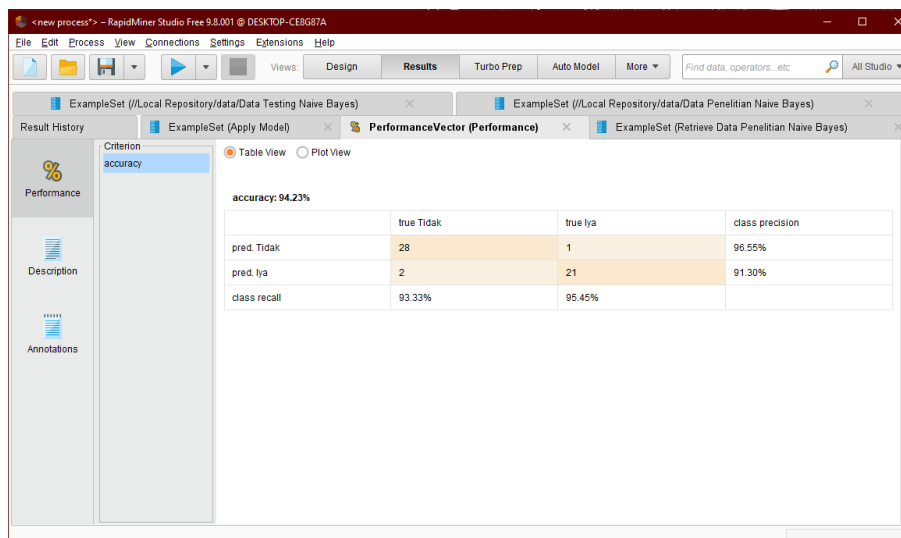
$$\begin{aligned}
 &= P(\text{Tidak}) * P(\text{Berlebih} \mid \text{Tidak}) * P(\text{Merasa Jenuh} \mid \text{Tidak}) * \\
 &\quad P(\text{Tidak Memiliki Sanksi} \mid \text{Tidak}) * P(\text{Sangat Mematuhi} \mid \text{Tidak}) * \\
 &\quad P(\text{Mencukupi} \mid \text{Tidak}) \\
 &= 0,54808 * 0,42105 * 0,38596 * 0,40351 * 0,38596 * 0,66667 \\
 &= 0,00925
 \end{aligned}$$

Proses dilakukan sampai dengan data uji ke 104 atau R104. Selanjutnya dilakukan perbandingan untuk setiap data dengan cara membandingkan nilai tertinggi dari label "Iya" dan "Tidak". Dan didapatkanlah hasil perbandingan dari label data penelitian dengan label hasil proses perhitungan *Naive Bayes* sebagai berikut :

Tabel 3. Hasil Perbandingan Label *Naive Bayes*

Label Prediksi	Label Data Uji	
	Iya	Tidak
Iya	21	2
Tidak	1	28

Proses selanjutnya dilakukan dengan menguji coba proses perhitungan dengan implementasi menggunakan aplikasi *RapidMiner*. Disapatkanlah hasil yang sama dari proses manual algoritma *Naive Bayes* seperti pada gambar berikut :



Gambar 2. Hasil Akurasi Naive Bayes

3.2 Proses K-Nearest Neighbor

Pada proses *K-Nearest Neighbor* dilakukan menggunakan rumus *euclidean distance* yang merupakan rumus numerik, maka data penelitian akan dilakukan pembobotan yaitu semakin tinggi nilai bobot maka akan semakin mempengaruhi tidak disiplin masyarakat terhadap protokol kesehatan COVID-19, seperti sebagai berikut :

Tabel 4. Pembobotan Sub Kriteria

Kriteria	Sub Kriteria	Hasil Pembobotan
Ketersediaan Alat	Berlebih	2
	Mencukupi	3
	Minim	4
Perasaan Masyarakat	Menikmati	2
	Merasa Jenuh	3
	Meremehkan	4
Ketegasan Hukum	Tidak Memiliki Sanksi	3
	Penerapan Sanksi Lemah	2
Lingkungan Masyarakat	Sangat Mematuhi	2
	Mematuhi	3
	Tidak Mematuhi	4
Ekonomi	Bertambah	2
	Mencukupi	3
	Kekurangan	4

Tabel pembobotan tersebut kemudian di implementasikan kedalam 104 data penelitian dan 52 data uji yang sama dengan proses algoritma *Naive Bayes*. Proses selanjutnya dilakukan dengan menghitung nilai jarak terdekat dari setiap data uji menggunakan rumus *euclidean distance*: tahap awal yaitu mencari nilai jarak terdekat untuk data uji ke 1 yaitu R53 sebagai berikut :

$$\begin{aligned} \text{Distance}(R1,R53) &= \sqrt{(3 - 2)^2 + (3 - 3)^2 + (3 - 3)^2 + (4 - 2)^2 + (3 - 3)^2} = 2,23607 \\ \text{Distance}(R2,R53) &= \sqrt{(3 - 2)^2 + (3 - 3)^2 + (2 - 3)^2 + (4 - 2)^2 + (3 - 3)^2} = 2,44949 \\ \text{Distance}(R3,R53) &= \sqrt{(4 - 2)^2 + (3 - 3)^2 + (2 - 3)^2 + (4 - 2)^2 + (3 - 3)^2} = 3 \end{aligned}$$

Proses perhitungan data uji R53 dilanjutkan sampai dengan data penelitian ke 104 (R104). Dan begitu juga seterusnya sampai dengan data uji ke 52 (R104). Hasil menghitung jarak terdekat dapat dilihat pada tabel berikut :

Tabel 5. Hasil Perhitungan Jarak Terdekat

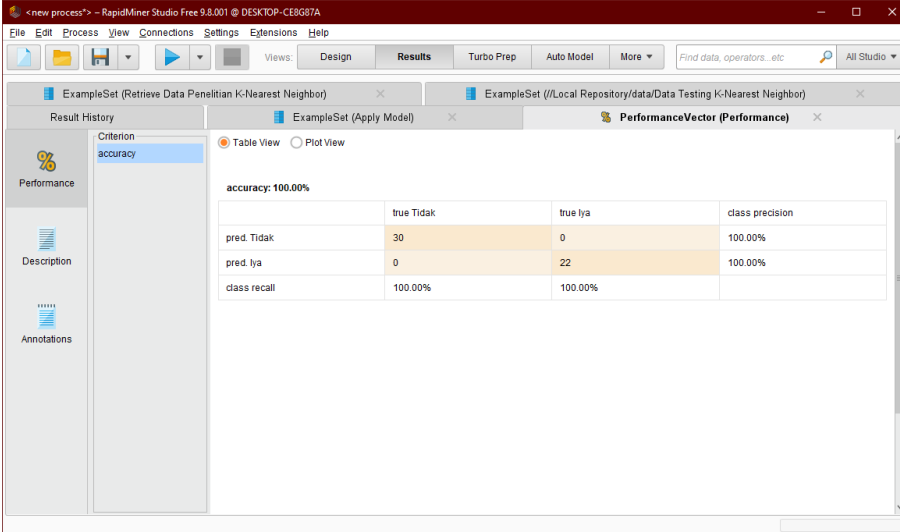
Responden	D(R53)	D(R54)	D(R55)	D(R56)	D(R104)
R1	2.23607	2.00000	1.73205	2.00000	1.73205
R2	2.44949	1.73205	2.00000	1.73205	1.41421
R3	3.00000	2.44949	2.23607	1.41421	1.73205
R4	1.41421	1.73205	1.41421	1.73205	2.00000
R5	2.44949	3.31662	1.41421	2.64575	3.16228
R6	2.64575	2.44949	1.73205	2.00000	1.73205
R7	2.00000	1.73205	2.00000	2.64575	2.00000
R8	2.64575	2.00000	1.73205	2.44949	1.00000
R9	1.41421	2.23607	2.00000	2.64575	3.16228
R10	3.00000	2.44949	2.23607	1.41421	1.73205
R11	2.44949	2.23607	1.41421	2.64575	1.41421
R12	2.64575	2.44949	1.73205	2.00000	2.23607
R13	3.00000	2.44949	2.23607	1.41421	1.73205
R14	2.23607	2.00000	1.73205	2.00000	1.73205
R15	2.23607	2.44949	2.23607	1.41421	2.64575
R16	3.00000	2.82843	1.73205	2.44949	1.73205
R17	1.00000	2.00000	1.73205	2.00000	2.64575
R18	3.16228	2.64575	2.00000	2.23607	1.41421
R19	3.16228	2.64575	2.44949	2.64575	1.41421
R20	2.44949	2.23607	1.41421	2.64575	1.41421
R21	2.00000	1.73205	2.00000	2.64575	1.41421
R22	1.73205	2.44949	1.73205	2.00000	3.00000
R23	1.41421	1.73205	1.41421	1.73205	2.00000
R24	3.00000	3.16228	1.73205	2.00000	2.64575
R25	2.64575	2.82843	1.73205	1.41421	2.64575
R26	1.73205	2.44949	1.73205	2.00000	3.00000
....
R98	1.00000	2.00000	2.64575	2.44949	3.31662
R99	2.23607	2.82843	1.73205	2.44949	2.64575
R100	2.23607	2.44949	1.00000	2.44949	2.23607
R101	2.82843	2.64575	2.00000	1.73205	2.00000
R102	2.64575	1.41421	2.64575	2.00000	1.00000
R103	2.23607	2.00000	1.73205	2.00000	1.73205
R104	2.82843	1.73205	2.44949	2.64575	0.00000

Selanjutnya menentukan nilai K atau jarak terdekat seperti jika nilai K=3 maka akan mencari nilai jarak terdekat 3 dari data. Penelitian ini menggunakan K=5, penentuan nilai K dilakukan secara *random* sesuai dengan kebutuhan penelitian. Proses mencari jarak terdekat 5 atau K5 dilakukan dengan mencari nilai minimal 5 dari proses jarak terdekat. dan kemudian dilakukan perbandingan terhadap label, baik itu membandingkan label “iya” dan “Tidak” sampai perbandingan label antara data penelitian dengan proses algoritma yang dilakukan. Didapatkanlah hasil sebagai berikut :

Tabel 6. Hasil Perbandingan Label *K-Nearest Neighbor*

Label Prediksi	Label Data Uji	
	Iya	Tidak
Iya	22	0
Tidak	0	30

Tahap selanjutnya dilakukan dengan menguji coba proses perhitungan dengan implementasi menggunakan aplikasi *RapidMiner*. Disampaikanlah hasil yang sama dari proses manual algoritma *K-Nearest Neighbor* seperti pada gambar berikut :



The screenshot shows the RapidMiner Studio interface with the following data:

	true Tidak	true Iya	class precision
pred. Tidak	30	0	100.00%
pred. Iya	0	22	100.00%
class recall	100.00%	100.00%	

Gambar 3. Hasil Akurasi *K-Nearest Neighbor*

3.3 Perbandingan Algoritma

Dari proses perbandingan kedua algoritma yang dilakukan dan telah di implementasikan menggunakan aplikasi *RapidMiner* dalam permasalahan faktor ketidakdisiplinan masyarakat terhadap protokol kesehatan didapatkanlah hasil akurasi akhir dari setiap metode yaitu algoritma *Naive Bayes* memiliki akurasi sebesar 94,23% dan algoritma *K-Nearest Neighbor* memiliki akurasi sebesar 100%. Perbandingan metode menggunakan tingkat akurasi karena semakin tinggi nilai akurasi maka semakin baik pula proses dari suatu algoritma. Maka dapat diketahui dalam perbandingan metode yang dilakukan algoritma *K-Nearest Neighbor* lebih baik dan sesuai dalam menyelesaikan proses permasalahan ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan karena *K-Nearest Neighbor* lebih kuat dalam melakukan proses terhadap 104 data penelitian dan 52 data uji yang digunakan dalam penelitian ini.

4. PENUTUP

Berdasarkan hasil dan pembahasan yang telah dilakukan dengan membandingkan algoritma *Naive Bayes* dengan algoritma *K-Nearest Neighbor* dalam permasalahan faktor yang mempengaruhi ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19 dapat disimpulkan bahwa hasil perhitungan maupun pengujian menggunakan *RapidMiner* algoritma *K-Nearest Neighbor* memiliki tingkat akurasi sebesar 100% dan lebih baik jika dibandingkan dengan algoritma *Naive Bayes* yang hanya memiliki tingkat akurasi sebesar 94,23%. Maka algoritma *K-Nearest Neighbor* lebih diunggulkan dalam menyelesaikan permasalahan faktor yang mempengaruhi ketidakdisiplinan masyarakat dalam menjalankan protokol kesehatan COVID-19. Dan dapat diketahui dari lima kriteria yaitu Ketersediaan Alat, Perasaan Masyarakat, Ketegasan Hukum, Lingkungan Masyarakat dan Ekonomi didapat hasil akurasi setiap metode bahwa nilai label “Tidak” masih lebih tinggi di banding nilai label “Iya”, untuk itu dari kasus permasalahan ketidakdisiplinan masyarakat terhadap protokol kesehatan COVID-19 di tempat-tempat keramaian Siantar dan Simalungun dapat disimpulkan bahwa sebagian besar masyarakat tidak melanggar protokol kesehatan COVID-19.

DAFTAR PUSTAKA

- [1] F. L. D. Cahyanti, W. Gata and F. Sarasati, “Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit,” *Jurnal Ilmiah Universitas Batanghari Jambi (JIUBJ)*, pp. 259-262, 2021.

- [2] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 628-634, 2021.
- [3] M. Gunawan, M. Zarlis and Roslina, "Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, pp. 513-523, 2021.
- [4] R. M. Bianome, D. R. Sina and Y. Y. Nabuasa, "DIAGNOSA HAMA DAN PENYAKIT PADA TANAMAN PADI MENGGUNAKAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR," *J-ICON*, vol. 8, pp. 156-162, 2020.
- [5] I. H. Lubis, Poningsih and I. S. Saragih, "Klasifikasi Masyarakat Wajib Pajak Di Kabupaten Simalungun Menggunakan Metode Naive Bayes," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 15-23, 2019.
- [6] D. Handoko, H. S. Tambunan and J. T. Hardinata, "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. 6, pp. 111-119, 2021.
- [7] F. A. Hermawati, Data Mining, Putri Christian ed., Yogyakarta: CV. ANDI OFFSET, 2013.
- [8] E. Prasetyo, Data Mining - Konsep dan Aplikasi Menggunakan MATLAB, Nikodemus WK ed., Yogyakarta: CV. ANDI OFFSET, 2012.
- [9] A. L. Fairuz, R. D. Ramadhani and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *Jurnal DINDA (Indonesian Journal of Data Science, IOT, Machine Learning and Artificial Intelligence)*, vol. 10, pp. 1-10, 2021.
- [10] M. Fatchan, I. N. T. K. MT, Alfian and Kurniawan, "Perbandingan Dalam Memprediksi Penyakit Liver Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *Jurnal Pelita Teknologi*, vol. 16, no. 1, pp. 15-21, 2021.
- [11] D. Sepri, "Penerapan Algoritma Naïve Bayes Untuk Analisis Kepuasan Penggunaan Aplikasi Bank," *Journal of Computer System and Informatics (JoSYC)*, vol. 2, no. 1, pp. 135-139, 2020.
- [12] S. T. Rizaldi and Mustakim, "Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 12*, pp. 130-137, 2020.
- [13] F. Gultom and T. Simanjuntak, "PREDIKSI TINGKAT KELANCARAN PEMBAYARAN KREDIT BANK DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR," *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 4, pp. 98-102, 2020.
- [14] D. A. C. D. A. Baskoro, L. Ambarwati and I. W. S. Wicaksana, Belajar Data Mining dengan RapidMiner, Remi Sanjaya ed., Para Penulis, 2013.