



Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia

Alvina Felicia Watratan¹, Arwini Puspita. B², Dikwan Moeis³

^{1,2,3}Sistem Informasi, STMIK Profesional Makassar

¹vinawatratan@gmail.com*, ²arwinipuspita77@gmail.com, ³dikwan.moeis@gmail.com

Abstract

The COVID-19 pandemic is the first and foremost health crisis in the world. Coronavirus is a collection of viruses from the subfamily Orthocoronavirinae in the Coronaviridae family and the order of Nidovirales. This group of viruses that can cause disease in birds and mammals, including humans. In humans, coronaviruses cause generally mild respiratory infections, such as colds, although some forms of disease such as; SARS, MERS, and COVID-19 are more deadly. Anticipating and reducing the number of corona virus sufferers in Indonesia has been carried out in all regions. Among them by providing policies to limit activities out of the house, school activities laid off, work from home (work from home), even worship activities were laid off. This has become a government policy based on considerations that have been analyzed to the maximum, of course. Therefore this research was carried out as an anticipation step towards the Covid-19 pandemic by predicting the spread of Covid-19, especially in Indonesia. The research method applied in this research is problem analysis and literature study, collecting data and implementation. The application of the naive bayes method is expected to be able to predict the spread rate of COVID-19 in Indonesia. The results of the Naive Bayes method classification show that 16 data from 33 data were tested in Covid-19 cases per province with an accuracy of 48.4848%, where of the 33 data tested in the Covid-19 case per province tested there were 16 data that were successfully classified correctly.

Keywords: Covid-19, Naive Bayes, WEKA Application

Abstrak

Pandemi COVID-19 merupakan krisis kesehatan yang pertama dan terutama di dunia. Corona virus adalah sekumpulan virus dari subfamili Orthocoronavirinae dalam keluarga Coronaviridae dan ordo Nidovirales. Kelompok virus ini yang dapat menyebabkan penyakit pada burung dan mamalia, termasuk manusia. Pada manusia, corona virus menyebabkan infeksi saluran pernapasan yang umumnya ringan, seperti pilek, meskipun beberapa bentuk penyakit seperti; SARS, MERS, dan COVID-19 sifatnya lebih mematikan. Mengantisipasi dan mengurangi jumlah penderita virus corona di Indonesia sudah dilakukan di seluruh daerah. Diantaranya dengan memberikan kebijakan membatasi aktifitas keluar rumah, kegiatan sekolah dirumahkan, bekerja dari rumah (*work from home*), bahkan kegiatan beribadah pun dirumahkan. Hal ini sudah menjadi kebijakan pemerintah berdasarkan pertimbangan-pertimbangan yang sudah dianalisa dengan maksimal tentunya. Oleh karena itu penelitian ini dilakukan sebagai langkah antisipasi terhadap pandemi Covid-19 dengan memprediksi tingkat penyebaran Covid-19 terutama di Indonesia. Metode penelitian yang di terapkan pada penelitian ini ialah analisis masalah dan studi literatur, mengumpulkan data dan implementasi. Penerapan metode naive bayes diharapkan mampu memprediksi tingkat penyebaran COVID-19 di Indonesia. Hasil penelitian pada klasifikasi metode Naive Bayes menunjukkan bahwa 16 data dari 33 data yang di uji dalam kasus Covid-19 per provinsi dengan keakuratan sebesar 48,4848%, di mana dari 33 data yang di uji dalam kasus Covid-19 per provinsi yang diuji terdapat 16 data yang berhasil diklasifikasikan dengan benar.

Kata kunci: Covid-19, Naive Bayes, Aplikasi WEKA

1. Pendahuluan

Dunia saat ini tengah waspada dengan penyebaran sebuah virus yang dikenal dengan virus corona. *Corona viruses* (CoV) merupakan bagian dari keluarga virus yang menyebabkan penyakit mulai dari flu hingga penyakit yang lebih berat seperti *Middle East Respiratory Syndrome* (MERS-CoV) and *Severe Acute Respiratory Syndrome* (SARS-CoV). Penyakit yang disebabkan virus corona, atau dikenal dengan COVID-19, adalah jenis baru yang ditemukan pada tahun 2019

dan belum pernah diidentifikasi menyerang manusia sebelumnya (*World Health Organization*, 2019). Kasus virus corona muncul dan menyerang manusia pertama kali di provinsi Wuhan, China. Awal kemunculannya diduga merupakan penyakit pneumonia, dengan gejala serupa sakit flu pada umumnya. Gejala tersebut di antaranya batuk, demam, letih, sesak napas dan tidak nafsu makan. Namun berbeda dengan *influenza*, virus corona dapat berkembang dengan cepat hingga mengakibatkan infeksi lebih parah dan gagal organ.

Kondisi darurat ini terutama terjadi pada pasien dengan masalah kesehatan sebelumnya. Karena penularan virus corona yang sangat cepat inilah Organisasi Kesehatan Dunia (WHO) menetapkan virus corona sebagai pandemi pada 11 Maret 2020. Status pandemic atau epidemi global menandakan bahwa penyebaran COVID-19 berlangsung sangat cepat hingga hampir tak ada negara di dunia yang dapat memastikan diri terhindar dari virus corona (Widiyani, 2020).

COVID-19 telah menyebar ke 196 Negara, sebanyak 414,179 kasus terkonfirmasi positif dan sebanyak 18,440 meninggal dunia dari kasus terkonfirmasi positif. Sementara itu, COVID-19 pertama kali terkonfirmasi di Indonesia pada 2 Maret 2020 di Indonesia dan per 25 Maret telah terjadi secara kumulatif sebanyak 790 kasus terkonfirmasi positif (secara kumulatif), diantaranya 58 meninggal, 31 sembuh, dan 701 dalam perawatan.

Berdasarkan kondisi tersebut, maka Indonesia dalam status waspada terhadap ancaman virus corona tersebut dan sampai sekarang belum ditemukan vaksin COVID-19, maka bentuk pencegahan dari meluasnya penyebaran virus dapat dilakukan dengan cara memutuskan rantai penularannya. Salah satu metode untuk memutuskan rantai penularan tersebut adalah dengan melakukan pembatasan sosial (social distancing). Dengan adanya pembatasan sosial, harapannya setiap masyarakat tidak akan menjadi penular maupun tertular karena tidak melakukan kontak dengan siapapun sehingga laju penyebaran dapat menurun.

Pada penelitian ini kami akan melakukan penerapan algoritma Naive Bayes untuk memprediksi tingkat penyebaran COVID-19 di Indonesia dengan mengimplementasikan Algoritma Naive Bayes, yang berguna sebagai langkah antisipasi pandemic COVID-19.

Data Mining

Data mining adalah suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahamkan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting.

Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan. Data mining memungkinkan pemakai menemukan pengetahuan dalam data database yang tidak mungkin diketahui keberadaannya oleh pemakai. Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar.

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*.

Tahap-tahap *Data Mining* adalah sebagai berikut:

- a. *Pembersihan data (Data Cleaning)*
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.
- b. *Integrasi data (Data Integration)*
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
- c. *Seleksi data (Data Selection)*
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
- d. *Transformasi data (Data Transformation)*
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.
- e. *Proses Mining*
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan Data Mining.
- f. *Evaluasi pola (Pattern Evaluation)*
Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan.
- g. *Presentasi pengetahuan (Knowledge Presentation)*
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Implementasi

Menurut Nurdin Usman dalam bukunya yang berjudul *Konteks Implementasi Berbasis Kurikulum mengemukakan pendapatnya mengenai implementasi atau pelaksanaan sebagai berikut :*

“Implementasi adalah bermuara pada aktivitas, aksi, tindakan, atau adanya mekanisme suatu sistem. Implementasi bukan sekedar aktivitas, tetapi suatu kegiatan yang terencana dan untuk mencapai tujuan kegiatan” (Usman, 2002:70).

Pengertian implementasi yang dikemukakan di atas, dapat dikatakan bahwa implementasi adalah bukan sekedar aktivitas, tetapi suatu kegiatan yang terencana dan dilakukan secara sungguh-sungguh berdasarkan acuan norma tertentu untuk mencapai tujuan kegiatan. Oleh karena itu implementasi tidak berdiri sendiri tetapi dipengaruhi oleh objek berikutnya.

Metode Algoritma *Naive Bayes*

Naive Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada *teorema Bayes*. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*.

Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (*naif*) akan independensi dari masing-masing kondisi / kejadian.

Menurut *Olson Delen* (2008) menjelaskan *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan. *Naive Bayes Classifier* bekerja sangat baik dibanding dengan model classifier lainnya.

Hal ini dibuktikan oleh *Xhemali*, Hinde Stone dalam jurnalnya "*Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*" mengatakan bahwa "*Naive Bayes Classifier* memiliki tingkat akurasi yang lebih baik dibanding model classifier lainnya".

Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai *variable independent*, maka hanya varians dari suatu *variable* dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

a. Kegunaan *Naive Bayes* :

1. Mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis
2. Sebagai metode machine learning yang menggunakan probabilitas
3. Untuk membuat diagnosis medis secara otomatis
4. Mendeteksi atau menyaring spam

b. Kelebihan *Naive Bayes* :

1. Bisa dipakai untuk data kuantitatif maupun kualitatif
2. Tidak memerlukan jumlah data yang banyak
3. Tidak perlu melakukan data training yang banyak
4. Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
5. Perhitungannya cepat dan efisien
6. Mudah dipahami
7. Mudah dibuat
8. Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang
9. Jika digunakan dalam bahasa pemrograman, code-nya sederhana
10. Bisa digunakan untuk klasifikasi masalah biner ataupun multiclass

c. Kekurangan *Naive Bayes* :

1. Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol
2. Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain

3. Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya.
4. Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut. Banyak celah yang bisa mengurangi efektivitasnya dirancang untuk mendeteksi kata-kata saja, tidak bisa berupa gambar.

Persamaan Metode Naive Bayes

Persamaan dari teorema Bayes adalah :

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

Di mana :

- X* :Data dengan class yang belum diketahui.
- H* :Hipotesis data merupakan suatu class spesifik.
- P(H|X)* :Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas).
- P(H)* :Probabilitas hipotesis H (prior probabilitas).
- P(X|H)* :Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X)* :Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut.

Karena itu, metode *Naive Bayes* di atas disesuaikan:

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \quad 2$$

Di mana Variabel *C* merepresentasikan kelas, sementara variabel *F1 ... Fn* merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas *C* (*Posterior*) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas *C* (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*).

Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} (C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= (C|F_1)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= (C|F_1)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (independen) satu sama lain.

Dengan asumsi tersebut, maka berlaku suatu persamaan;

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (4)$$

Untuk sehingga

$$P(F_i|C, F_j) = P(F_i|C)$$

Persamaan di atas merupakan model dari teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi.

Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (5)$$

Di mana :

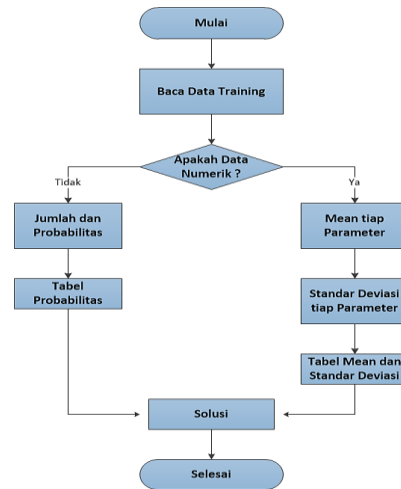
- P : Peluang
- X_i : Atribut ke i
- x_i : Nilai atribut ke i
- Y : Kelas yang dicari
- y_i : Sub kelas Y yang dicari
- μ : *mean*, menyatakan rata – rata dari seluruh atribut
- σ : *Deviiasi standar*, menyatakan varian dari seluruh atribut.

Alur dari metode Naive Bayes dapat dilihat pada gambar 1.

Adapun keterangan dari gambar 1 sebagai berikut:

1. Baca data training
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.

Adapun persamaan yang digunakan untuk menghitung nilai rata – rata hitung (mean) dapat dilihat sebagai berikut :



Gambar 1. Alur metode Naive Bayes

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (6)$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (7)$$

di mana :

- μ : rata – rata hitung (mean)
- x_i : nilai sample ke - i
- n : jumlah sampel

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (8)$$

di mana :

- σ : standar deviasi
- x_i : nilai x ke - i
- μ : rata-rata hitung
- n : jumlah sampel

- b. Cari nilai probabilitistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel mean, standard deviasi dan probabilitas.
4. Solusi kemudian dihasilkan.

2. Metode Penelitian

Metode penelitian merupakan proses atau cara ilmiah untuk mendapatkan data yang akan digunakan untuk keperluan penelitian. Penelitian dilakukan secara

bertahap dimulai dari perencanaan, menentukan fokus penelitian, waktu penelitian, pengumpulan data, analisis, dan penyajian hasil penelitian.

Adapun metode penelitian yang diterapkan pada penelitian ini adalah sebagai berikut:

1. Analisis Masalah dan Studi Literatur

Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian. Dalam hal ini mengamati permasalahan yang berhubungan dengan tingkat penyebaran COVID-19 yang terjadi khususnya di Indonesia. Permasalahan-permasalahan yang ada, selanjutnya dianalisa untuk mengetahui bagaimana cara penyelesaian terhadap masalah tersebut dan menentukan ruang lingkup permasalahan yang akan diteliti. Mempelajari dasar teori dari berbagai literatur mengenai penerapan metode *Naive Bayes*, konsep dan teori data mining dan prediksi tingkat penyebaran COVID-19 di Indonesia, melalui jurnal-jurnal dan agar mendapatkan dasar pengetahuan untuk melakukan penelitian selanjutnya.

2. Mengumpulkan Data

Prosedur sistematis yang digunakan untuk mengumpulkan data yaitu dengan metode penelitian kuantitatif. Pelaksanaan metode penelitian kuantitatif focus pada penggunaan angka, *tabel*, grafik dan diagram untuk menampilkan hasil data yang diperoleh dan akan digunakan untuk bahan penganalisisan data terhadap metode *Naive Bayes*.

Setelah data dikumpulkan maka dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada metode *Naive Bayes*.

3. Implementasi

Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah tools. Tools yang akan digunakan dalam implementasi penelitian ini adalah dengan menggunakan *Software Weka*.

Selanjutnya pengujian dilakukan untuk mengetahui apakah penelitian yang dilakukan telah sesuai dengan tujuan yang diharapkan yaitu untuk memprediksi tingkat penyebaran COVID-19 di Indonesia.

3. Hasil dan Pembahasan

3.1 Implementasi *Naive Bayes*

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam metode *Naive Bayes* data String yang bersifat konstan

dibedakan dengan data numerik yang bersifat kontinu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria nilai data numerik.

Adapun penerapan metode *Naive Bayes* menggunakan rumus excel sebagai berikut;

3.1.1 Baca data *training* (Data latih)

Untuk menentukan data yang nantinya akan dianalisis dengan metode *Naive Bayes* maka langkah pertama yang dilakukan adalah membaca data latih. Adapun data latih yang digunakan dapat dilihat pada tabel 1 berikut:

Tabel 1. Data *Training* (Data Latih)

No	Provinsi	Dalam Perawatan	Sembuh	Meninggal	Kasus Terbesar Per Provinsi
1	DKI Jakarta	4038	1276	460	Positif
2	Jawa Timur	1449	294	178	Positif
3	Jawa Barat	1237	259	100	Positif
4	Jawa Tengah	805	234	70	Positif
5	Sulawesi Selatan	527	293	51	Positif
6	Banten	404	158	60	Positif
7	Sumatera Selatan	376	73	9	Positif
8	Sumatera Barat	284	88	21	Positif
9	Kalimantan Selatan	330	24	9	Positif
10	Nusa Tenggara Barat (NTB)	163	188	7	Negatif
11	Bali	107	232	4	Negatif
12	Papua	281	48	6	Positif
13	Kalimantan Timur	187	61	3	Positif
14	Kalimantan Tengah	129	87	11	Positif
15	Sumatera Utara	125	53	24	Positif
16	Daerah Istimewa Yogyakarta	110	71	7	Positif
17	Sulawesi Tenggara	154	24	5	Positif
18	Kalimantan Utara	117	23	1	Positif
19	Kalimantan Barat	104	22	3	Positif
20	Kepulauan Riau	27	77	11	Negatif
21	Sulawesi Tengah	87	21	4	Positif
22	Sulawesi Utara	70	30	5	Positif
23	Riau	36	53	6	Negatif
24	Papua Barat	85	2	1	Positif
25	Maluku Utara	70	12	3	Positif

26	Sulawesi Barat	50	22	2	Positif
27	Jambi	66	3	0	Positif
28	Lampung	39	22	5	Positif
29	Maluku	41	17	4	Positif
30	Bengkulu	50	1	2	Positif
31	Kepulauan Bangka Belitung	21	7	1	Positif
32	Gorontalo	7	14	1	Negatif
33	Nusa Tenggara Timur (NTT)	17	1	1	Positif

Keterangan :

Positif = Kasus Pasien yang terkena Positif lebih besar
 Negatif = Kasus pasien yang terkena positif lebih rendah

3.1.2 Nilai Mean dan Standar Deviasi

Berdasarkan data analisis yang diambil pada tanggal 15 Mei 2020 di situs resmi <https://covid19.go.id/peta-sebaran> merupakan data numerik, maka terlebih dahulu mencari nilai Mean dan Standar Deviasi..

- untuk menghitung nilai rata – rata hitung (*mean*) dapat dilihat pada tabel sebagai berikut :

Tabel 2. Nilai Mean

Mean			
Kasus Terbesar Per Provinsi	Dalam Perawatan	Sembuh	Meninggal
Positif	401.8928571	115.2142857	37.35714286
Negatif	68	112.8	5.8

Mean merupakan nilai Rata-rata yang didapatkan dari hasil penjumlahan seluruh nilai dari masing-masing data, lalu dibagi dengan banyaknya data yang ada.

Berdasarkan tabel 2 diatas untuk menghitung nilai *mean* :

- Pada kategori "Positif" terdapat sekitar 28 data (per provinsi) diantaranya provinsi DKI Jakarta, Jawa Timur, Jawa Barat, Jawa Tengah, Sulawesi Selatan, Banten, Sumatera Selatan, Sumatera Barat, Kalimantan Selatan, Papua, Kalimantan Timur, Kalimantan Tengah, Sumatera Utara, Daerah Istimewa Yogyakarta, Sulawesi Tenggara, Kalimantan Utara, Kalimantan Barat, Sulawesi Tengah, Sulawesi Utara, Papua Barat, Maluku Utara, Sulawesi Barat, Jambi, Lampung, Maluku, Bengkulu, Kepulauan Bangka Belitung, Nusa Tenggara Timur (NTT) . Hasil dari nilai *mean* pada data pasien"dalam perawatan" adalah 401.8928571, data pasien "sembuh" adalah 115.2142857, dan data pasien "meninggal" adalah 37.35714286
- Pada kategori "Negatif" terdapat sekitar 5 data (per provinsi) diantaranya provinsi Nusa

Tenggara Barat (NTB), Bali, Kepulauan Riau, Riau, dan Gorontalo. Hasil nilai *mean* pada data pasien "dalam perawatan" adalah 68, data pasien "sembuh" adalah 112.8, dan data pasien "meninggal" adalah 5.8.

- Dan persamaan untuk menghitung nilai simpangan baku (*standar deviasi*) dapat dilihat sebagai berikut:

Tabel 3. Standar Deviasi

Standar Deviasi			
Kasus Terbesar Per Provinsi	Dalam Perawatan	Sembuh	Meninggal
Positif	795.2873852	244.5758774	91.65968227
Negatif	92.8477248	92.8477248	3.701351105

Standar deviasi merupakan nilai statistik yang dimanfaatkan untuk menentukan bagaimana sebaran data dalam sampel, serta seberapa dekat titik data individu ke mean atau rata-rata nilai sampel.

Berdasarkan tabel 3 untuk menghitung *standar deviasi* :

- Pada kategori "Positif" terdapat standar deviasi dengan data pasien "dalam perawatan" sekitar 795.2873852, data pasien"sembuh" sekitar 244.5758774, dan data pasien "meninggal" sekitar 91.65968227.
- Pada kategori "Negatif" terdapat standar deviasi dengan data pasien "dalam perawatan" sekitar 83.84202985, data pasien "sembuh" sekitar 92.8477248, dan data pasien "meninggal" sekitar 3.701351105.

3.1.3 Probabilitas

Probabilitas merupakan suatu nilai yang digunakan untuk mengukur tingkat terjadinya suatu kejadian yang acak. Kata probabilitas itu sendiri sering disebut dengan peluang atau kemungkinan. Probabilitas secara umum merupakan peluang bahwa sesuatu akan terjadi.

Tabel 4. Probabilitas Kasus Terbesar Per Provinsi

Probabilitas Kasus Terbesar Per Provinsi	
Kasus Terbesar Per Provinsi	Nilai
Positif	0.848484848
Negatif	0.151515152

Pada tabel 1 data kasus penyebaran COVID-19 Di Indonesia berdasarkan tiap provinsi , diketahui jumlah data training (data latih) adalah sebanyak 33 data, dimana dari 33 data tersebut terdapat 28 data dengan kategori positif, dan 5 data dengan kategori negatif.

Berdasarkan hasil pada tabel 4, terlihat bahwa nilai probabilitas tertinggi pada kategori kasus terbesar per provinsi adalah positif atau kasus pasien per provinsi yang terkena positif lebih besar sekitar 0.848484848, dan probabilitas rendah pada kategori negatif sekitar 0.151515152.

3.1.4 Menghitung Nilai Gaussian

Distribusi Gaussian adalah merupakan langkah terakhir untuk mengetahui hasil dari data latih, atau sebuah model uji data dengan mengambil nilai dari peluang dari data latih.

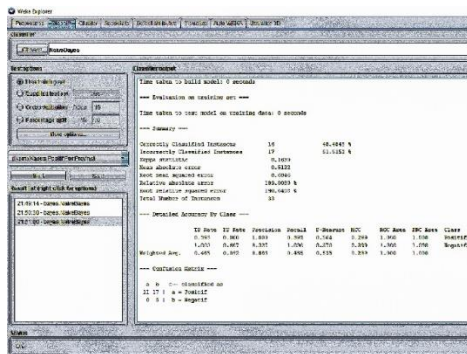
Tabel 5. Data Tes

Provinsi	Data Tes			Kasus Terbesar Per Provinsi
	Dalam Perawatan	Sembuh	Meninggal	
Aceh	3	13	1	?
Positif	0.012477561	0.0233822	0.038527102	1.12E-05
Negatif	0.032268233	0.0232408	0.089464815	6.71E-05
Max adalah				6.71E-05

Berdasarkan data uji dari provinsi Aceh dengan data pasien "dalam perawatan" yaitu 3, data pasien "sembuh" yaitu 13, dan data pasien "meninggal" yaitu 1 diprediksi algoritma *naive bayes* dengan hasil kasus terbesar per provinsi yaitu kategori "Negatif" sekitar 6.71E-05. Maka data uji dari provinsi Aceh dikatakan bahwa tingkat penyebaran COVID-19 dengan kasus pasien yang terkena positif lebih rendah .

3.2 Pengujian Metode *Naive Bayes*

Dari nilai probabilitas di atas akan diuji data sebanyak 33 data dan diselesaikan dengan menggunakan tools weka sehingga dihasilkan hasil klasifikasi dari Kasus COVID-19 Per Provinsi di Indonesia seperti pada gambar 2.



Gambar 2. Hasil Klasifikasi Metode *Naive Bayes*

Berdasarkan gambar 2 di atas, dapat dilihat persentase untuk Correctly Classified Instance adalah sebesar 48,4848% sementara persentase untuk Incorrectly Classified Instance adalah sebesar 51,5152%. Di mana dari 33 data Kasus COVID-19 Per Provinsi, ada sebanyak 16 data Kasus COVID-19 Per Provinsi berhasil diklasifikasikan dengan benar dan sebanyak 17 data Kasus COVID-19 Per Provinsi tidak berhasil diklasifikasikan dengan benar.

4. Kesimpulan

Berdasarkan penelitian tentang Implementasi Algoritma *Naive Bayes* Untuk Memprediksi Tingkat Penyebaran Covid 19 di Indonesia dapat ditarik kesimpulan sebagai berikut :

1. Metode *Naive Bayes* memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi Tingkat Penyebaran Covid 19 di Indonesia berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes* itu sendiri.
2. Berdasarkan data Kasus COVID-19 Per Provinsi yang dijadikan data training, metode *Naive Bayes* berhasil mengklasifikasikan 16 data dari 33 data yang diuji. Sehingga metode *Naive Bayes* berhasil memprediksi besarnya Kasus COVID-19 Per Provinsi dengan persentase keakuratan sebesar 48,4848%.
3. Algoritma *Naive Bayes* di dukung oleh ilmu Probabilistik dan ilmu statistika khususnya dalam penggunaan data petunjuk untuk mendukung keputusan pengklasifikasian. Pada algoritma *Naive Bayes*, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain.

5. Saran

Adapun Saran yang diberikan mengenai penelitian tentang Implementasi Algoritma *Naive Bayes* Untuk Memprediksi Tingkat Penyebaran Covid 19 di Indonesia yaitu pengujian, sebaiknya dilakukan dengan menggunakan metode lain untuk melihat metode mana yang lebih akurat dalam memprediksi Tingkat Penyebaran Covid 19 di Indonesia.

Daftar Rujukan

- [1] Mona, Nailul. 2020. Konsep Isolasi Dalam Jaringan Sosial Untuk Meminimalisasi Efek Contagious (Kasus Penyebaran Virus Corona Di Indonesia). *Jurnal Sosial Humaniora Terapan (JSHT)*, 2(2), pp. 117. doi: <https://doi.org/10.7454/jsht.v2i2.86>.
- [2] Saleh, Alfa. 2015. Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal (Citec Journal)*, 2(3), pp. 208-216. doi: <https://doi.org/10.24076/citec.2015v2i3.49>.
- [3] Bustami. 2013. Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI (Jurnal Penelitian Teknik Informatika)*, 3(2), pp. 127-146. doi: <https://doi.org/10.29103/techsi.v5i2.154>.
- [4] Ridwan, M., Suyono, H., Sarosa, M. 2013. Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, 1 (7), pp. 59-64.
- [5] Mujiasih, S. 2011. Pemanfaatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi dan Geofisika*, 12(2), pp. 189-195. doi: <http://dx.doi.org/10.31172/jmg.v12i2.100>
- [6] Yunus, N. R., & Rezki, Annisa. 2020. Kebijakan Pemberlakuan Lockdown Sebagai Antisipasi Penyebaran Corona Virus Covid-19. *SALAM; Jurnal Sosial & Budaya Syar-i*, 7(3), pp. 227-238. doi: <https://doi.org/10.15408/sjsbs.v7i3.15083>
- [7] Purwanto, A., Pramono, R., Asbari, M., Santoso, P. B., Wijayanti, L. M., Hyun, C. C., Putri, R. S. 2020. Studi Eksploratif Dampak Pandemi COVID-19 Terhadap Proses Pembelajaran Online di

- Sekolah Dasar. *EduPsyCouns: Journal of Education, Psychology and Counseling*, 2(1), pp. 1-12.
- [8] Pratiwi, R. W., & Nugroho, Y. S. 2016. Prediksi Rating Film Menggunakan Metode Naive Bayes. *Jurnal Teknik Elektro*, 8(2), pp. 60-63. doi: <https://doi.org/10.15294/jte.v8i2.7764>.
- [9] Wasiati, H., & Wijayanti, D. Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes (Studi Kasus: Di P.T. Karyatama Mitra Sejati Yogyakarta). *IJNS – Indonesian Journal on Networking and Security*, 3(2), pp. 45-51. doi: <http://dx.doi.org/10.1123/ijns.v3i2.154>.
- [10] World Health Organization. (2019). Coronavirus . Retrieved from World Health Organization: <https://www.who.int/healthtopics/coronavirus>
- [11] Fajar, Muhammad. 2020. *Estimasi Angka Reproduksi Novel Coronavirus (COVID-19) Kasus Indonesia*. Retrieved from https://www.researchgate.net/publication/340248900_ESTIMAT
- ION_OF_COVID_19_REPRODUCTIVE_NUMBER_CASE_OF_INDONESIA_Estimasi_Angka_Reproduksi_Novel_Coronavirus_COVID-19_Kasus_Indonesia. doi: 10.13140/RG.2.2.32287.92328 (diakses tanggal 15 Mei 2020)
- [12] <https://covid19.go.id/peta-sebaran> (diakses tanggal 15 Mei 2020)
- [13] DR. Derwin Suhartono, S.KOM., M.T.I. 2018. *Weka: Software untuk Memahami Konsep Data Mining*. <https://socs.binus.ac.id/2018/11/29/weka-software-untuk-memahami-konsep-data-mining/>. (diakses tanggal 29 November 2018)
- [14] Khalimy, Muiz. 2020. *Menghitung Naive Bayes dengan Excel Atribut Data Numerik*. <https://pengalaman-edukasi.blogspot.com/2020/04/menghitung-naive-bayes-dengan-excel.html> (diakses tanggal 16 April 2020)
- [15] Widiyanto, Mochammad Haldi. 2019. *Algoritma Naive Bayes*. <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>. (diakses tanggal 23 Desember 2019).