

Analisis Data Mining Untuk Klasifikasi Kasus Covid 19

Diterima:
10 Juni 2024

Revisi:
10 Juli 2024

Terbit:
1 Agustus 2024

¹Mohammad Nova Kurniawan, ²Satria Wijaya, ³Dwi Puji

Yoga Satria, Erna Daniati S, Aidina Ristyawan

Fakultas Teknik, Sistem Informasi, Universitas Nusantara PGRI Kediri

mohammadnovak@gmail.com, satriawijayasncd@gmail.com,

yogasatriamotovlog@gmail.com, ernadaniati@unpkediri.ac.id,

adinaristi@unpkediri.ac.id

Abstrak-Wabah virus COVID-19 adalah salah satu wabah terbesar yang pernah melanda umat manusia dan melemahkan ekonomi di hampir banyak negara. Tujuan penelitian ini adalah untuk menganalisis data kasus COVID-19 agar dapat mengetahui prediksi pada masalah kasus COVID-19. Pandemi COVID-19 telah memberikan dampak yang signifikan di Daerah Nganjuk. Untuk mengatasi pandemi ini secara efektif, diperlukan analisis data yang mendalam guna memahami karakteristik kasus Positif yang terdata. Penelitian ini bertujuan untuk menerapkan teknik data mining dalam klasifikasi kasus COVID-19. Data yang digunakan dalam penelitian ini mencakup kasus positif, positif baru, meninggal total, meninggal baru, sembuh total dan masih sakit mengenai pasien COVID-19 dengan menggunakan metode klasifikasi seperti Naive Bayes dan menggunakan tools Rapidminer.

Kata Kunci: COVID-19, Klasifikasi, Naive Bayes

***Abstract-**The COVID-19 virus outbreak is one of the biggest outbreaks that has ever hit humanity and weakened the economy in almost many countries. The purpose of this study is to analyze data on COVID-19 cases in order to determine predictions on the problem of COVID-19 cases. The COVID-19 pandemic has had a significant impact on the Nganjuk Region. To effectively address this pandemic, in-depth data analysis is needed to understand the characteristics of recorded Positive cases. The data used in this study includes positive cases, new positives, total deaths, new deaths, complete recovery and still sick regarding COVID-19 patients using classification methods such as Naive Bayes and using Rapidminer tools.*

Keyword: COVID-19, Clasificasion, Naive Bayes

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Nama: Erna Daniati

Departemen: Sistem Informasi Fakultas Teknik,

Institusi: Universitas Nusantara PGRI Kediri, Email:

ernadaniati@unpkediri.ac.id

ID Orcid: [<https://orcid.org/0009-0008-9471-4421>]

Handphone: 0813-3524-2202

1. Pendahuluan

Pada tahun 2020 ini, dunia dihebohkan dengan munculnya virus bernama virus corona atau Covid-19. Virus Covid-19 pertama kali terdeteksi pada akhir tahun 2019 dan diperkirakan akan semakin intensif atau menyebar pada awal tahun 2020. Virus ini pertama kali terdeteksi di Wuhan, Tiongkok, kemudian menyebar ke seluruh dunia, termasuk Indonesia [12] COVID-19 telah menyebar ke 196 negara, dengan 414.179 kasus terkonfirmasi dan 18.440 kematian dari pasien positif. Sedangkan COVID-19 pertama kali terkonfirmasi di Indonesia pada 2 Maret 2020. Hingga 25 Maret, jumlah kasus terkonfirmasi di Indonesia sebanyak 790 kasus dengan kematian 58 orang dan kesembuhan 31 orang. Jumlah pasien COVID-19 di Indonesia meningkat pesat. Kasus Covid-19 di Indonesia akan terus meningkat selama tahun 2020-2021[1][6]. Pandemi COVID-19 tidak hanya berdampak pada kesehatan tetapi juga bidang lainnya. Pandemi ini telah menyebabkan ambruknya perekonomian banyak negara, kenaikan harga barang-barang kebutuhan pokok, dan sepi pasar akibat pandemi ini. Dalam penelitian ini, peneliti berencana menganalisis penyebaran COVID-19 menggunakan data mining [14]. Pada penelitian ini penulis bertujuan untuk menerapkan data mining pada data kasus COVID-19 di Indonesia dengan menggunakan algoritma K-Means. Penambangan data roses menemukan pola dan pengetahuan yang menarik dari suatu data dalam jumlah yang besar. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan [10]. [9].

2. Tinjauan Pustaka

2.1 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[2].

Istilah data mining memiliki beberapa pandangan, seperti knowledge discovery ataupun pattern recognition. Istilah knowledge discovery atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Sedangkan istilah untuk pattern recognition atau pengenalan pola tepat untuk digunakan karena guna menemukan pola yang tersembunyi di dalam bongkahan data[3].

2.2 Covid 19

Coronavirus merupakan keluarga besar virus yang menyebabkan penyakit pada manusia dan hewan. Pada manusia biasanya menyebabkan penyakit infeksi saluran pernapasan, mulai flu biasa hingga penyakit yang serius seperti Middle East Respiratory Syndrome (MERS) dan Sindrom Pernafasan Akut Berat/Severe Acute Respiratory Syndrome (SARS).[4][14].

2.3 Algoritma Naïve Bayes

Naïve Bayes merupakan salah satu penerapan Teorema Bayes. Naïve 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[7][13].

Untuk mendapatkan nilai probabilitas pada sebuah sampel diberikan sebuah Teorema Bayes, seperti pada persamaan 1:

$$P(H_i | X) = \frac{P(H_i) \cdot P(X|H_i)}{P(X)} \quad (1)$$

Di mana $P(H)$ adalah nilai probabilitas prior dari hipotesis pada sebuah sampel biasa disebut dengan prior. $P(X)$ merupakan evidence dari probabilitas data pelatihan. $P(H|X)$ adalah nilai probabilitas H yang mempengaruhi X (posterior density), sedangkan $P(X|H)$ merupakan probabilitas X terhadap H yang disebut dengan likelihood.

Untuk menjelaskan Teorema Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis tersebut. Oleh karena itu, Teorema Bayes di atas disesuaikan seperti persamaan 2:

$$P(C|F_1, F_2, \dots, F_n) = \frac{P(C)P(F_1|C)P(F_2|C)\dots P(F_n|C)}{P(F_1, \dots, F_n)} \quad (2)$$

Di mana variabel CC merepresentasikan kelas, sementara variabel F_1, \dots, F_n merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka, rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas CC (posterior) adalah peluang munculnya kelas CC (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik sampel pada kelas CC (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga evidence). Karena itu, persamaan 2 dapat pula ditulis secara sederhana seperti persamaan 3:

$$P(C|F_1, \dots, F_n) = \frac{P(C) \cdot P(F_1|C) \cdot P(F_2|C) \cdot \dots \cdot P(F_n|C)}{Evidence} \quad (3)$$

Dari persamaan di atas, dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi persamaan 4 dan 5:

$$P(C | F_1 \dots F_n) = P(C) \cdot P(F_1 | C) \cdot P(F_2 | C) \dots P(F_n | C) \quad (4)$$

$$P(C | F_1 \dots F_n) = P(C) \cdot \prod_{i=1}^n P(F_i | C) \quad (5)$$

Persamaan 4 dan 5 merupakan model dari Teorema Naïve Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinu digunakan rumus Densitas Gauss, seperti persamaan 6:

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sigma_{ij} \sqrt{2\pi}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (6)$$

e

$$\sqrt{2\pi\sigma_{ij}}$$

Keterangan:

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 : Deviasi standar, menyatakan varian dari seluruh atribut.

2.4 Klasifikasi

Klasifikasi adalah tipe analisis data yang dapat membantu orang menentukan kelas label dari sampel yang ingin di klasifikasi. Klasifikasi merupakan Metode supervised learning, metode yang mencoba menemukan hubungan antara atribut masukan dan atribut target. Tujuan klasifikasi untuk meningkatkan kehandalan hasil yang diperoleh dari data.[5]

2.5 Rapidminer

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis , terhadap data mining, text mining dan analisis prediksi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik[6].

3. Metode Penelitian

3.1 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. Permasalahan-permasalahan yang ada

Selanjutnya dianalisa untuk mengetahui bagaimana carapenyelesaian 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[7][12].

3.2 Mengumpulan Data

Prosedur sistematis yang digunakan untuk mengumpulkan data yaitu dengan metode penelitian kuantitatif. Pelaksanaan metode penelitian kuantitatif focus pada penggunaan angka ,table, grafik dan diagram untuk menampilkan hasil data yang diperoleh dan akan digunakan untuk bahan penganalisan data terhadap metode *naive bayes*. [7].

Tahap eksperimen dilakukan dengan menerapkan Naive bayes sebagai metode yang akan diimplementasikan. Kemudian hasil model algoritma tersebut dievaluasi berdasarkan tingkat Akurasi, Presisi, dan Recall untuk dilihat hasilnya[15].

3.3 Implementasi

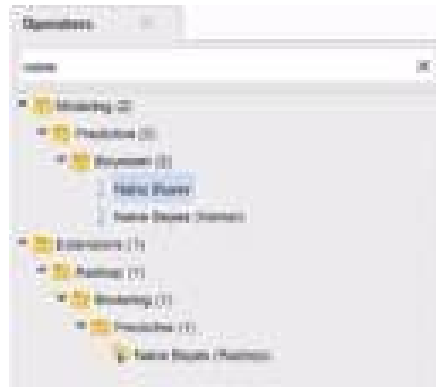
Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah tools. Tools yang akan digunakan dalam implementasi penelitian ini adalah dengan menggunakan *software* Rapidminer[7]

4. Hasil dan Pembahasan

4.1 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[7].

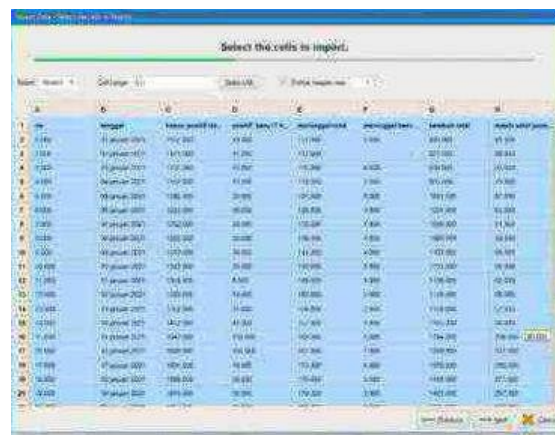
Selanjutnya pembuatan lembar proses, dimana didalam lembar proses ini selanjutnya berisi proses penerapan algoritma maupun proses pengujian menggunakan metode-metode pengujian yang ada pada operators RapidMiner.



Gambar 1. Operator Naïve Bayes

4.1.1 Import Data

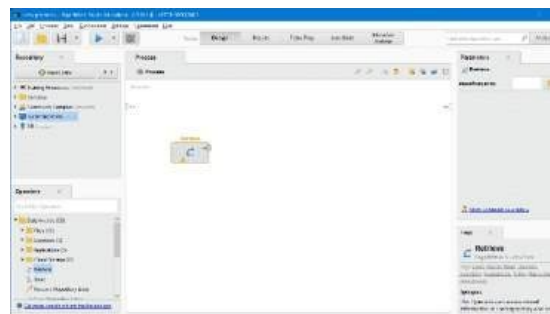
Sebelum menerapkan algoritma Naïve Bayes maka data dengan format *.xls di import terlebih dahulu ke dalam RapidMiner dengan klik “Add Data”. Selanjutnya pilih lokasi penyimpanan dataset yang akan digunakan. Karena dataset penulis berbentuk file yang memiliki format *.xls dan bukan berasal dari database seperti sql maka pilih lokasi dari “My Computer”. Selanjutnya pilih data import ke dalam RapidMiner.



Gambar 2. Import Data

4.1.2 Membuat Lembar Proses

Setelah data diimport, data dimasukkan ke halaman Proses. Dengan menyeret data yang diimport.



Gambar 3. Lembar Proses

4.1.3 Menerapkan Algoritma Naïve Bayes

Langkah selanjutnya adalah memasukkan Algoritma Naïve Bayes pada proses yang telah terbentuk iatas untuk mengetahui pengelompokkan dari proses yang telah dibuat. Cari Naïve Bayes pada Gambar 4 dibawah ini merupakan implementasi pada Rapid Miner.



Gambar 4. Menerapkan Algoritma

Setelah itu lakukan running pada proses untuk mendapatkan hasil Klasifikasi dari Algoritma Naïve Bayes dapat dilihat pada gambar 5.

A screenshot of the Rapid Miner software interface showing a table of classification results. The table has multiple columns, including 'ID', 'Actual Class', 'Predicted Class', and 'Confidence'. The rows are color-coded, with yellow highlighting the predicted class and blue highlighting the actual class. The table shows a high degree of accuracy in the predictions.

Gambar 5. Hasil Klasifikasi

5. Kesimpulan Dan Saran

5.1 Kesimpulan

Penelitian ini berhasil menerapkan algoritma Naïve Bayes Classifier untuk klasifikasi kasus COVID-19 di Kabupaten Nganjuk menggunakan RapidMiner. Berdasarkan hasil analisis, beberapa kesimpulan dapat diambil:

1. **Efektivitas Algoritma Naïve Bayes:** Algoritma Naïve Bayes terbukti efektif dalam mengklasifikasikan data kasus COVID-19. Hasil klasifikasi menunjukkan bahwa algoritma ini mampu memberikan prediksi yang akurat berdasarkan data historis yang dimiliki.
2. **Confusion Matrix:**

Kelas A (Kasus Positif Baru):

- 16 instance diprediksi benar sebagai Kelas A (True Positive).
- 0 instance diprediksi salah sebagai Kelas B atau Kelas C.
- Precision untuk Kelas A adalah 100,00%, dan Recall untuk Kelas A juga 100,00%.

Kelas B (Kasus Sembuh):

- 14 instance diprediksi benar sebagai Kelas B (True Positive).
- 2 instance diprediksi salah sebagai Kelas C (False Positive).
- Precision untuk Kelas B adalah 87,50%, dan Recall untuk Kelas B adalah 87,50%.

Kelas C (Kasus Meninggal):

- 14 instance diprediksi benar sebagai Kelas C (True Positive).
- 2 instance diprediksi salah sebagai Kelas B (False Positive).
- Precision untuk Kelas C adalah 87,50%, dan Recall untuk Kelas C adalah 87,50%.

5.2 Saran

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

6. Daftar Pustaka

- [1] Zulfa Nabila, Auliya Rahman Isnain, Pemarta, dan Zaenal Abidin, *Analisis data mining untuk clustering kasus COVID-19 di Provinsi Lampung menggunakan algoritma K-Means*. 2021
- [2] Charles Zai. *Implementasi Data Mining Sebagai Pengolahan Data*.2002
- [3] S.Susanto And D. Suryadi, *Pengantar Data Mining Menggali Pengetahuan Bongkahan Data*.2010.
- [4] K. K. R. Indonesia, "Pertanyaan dan Jawaban Terkait Coronavirus Disease 2019 (COVID-19) Update 6 Maret 2020," vol. 2019, pp. 1–9, 2020
- [5] Senna Hendrian, *Algoritma klasifikasi data mining untuk memprediksi siswa yang memperoleh bantuan dana pendidikan*.2018
- [6] D. Aprilla, D. A. Baskoro, L. Ambarwati, And I. W. S. Wicaksana, *Belajar Data Mining Dengan Rapid Miner*. 2013.
- [7] Alvina Felicia Watratan, Arwini Puspita. B, Dikwan Moeis, *Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia (JACOST)* Vol. 1 No. 1 (2020)
- [8] N. Dwitri, J. A. Tampubolon, S. Prayoga, And P.P. P. A. N. W. F. I. R. H. Zer, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," Vol. 4, No. 1, Pp. 128–132, 2020.
- [9] J. Han, M. Kamber, And J. Oei, *Data Mining Concepts and Techniques*. 2012.
- [10] Yuliana, "Corona Virus Diseases (Covid-19)(Sebuah Tinjauan Literatur," Vol. 2, No. February, Pp. 187–192, 2020.
- [11] W. F. T. F. Gorahe Vivi Lusya dkk., "Dampak Pandemi Covid 19 Terhadap Kesejahteraan Masyarakat di Desa,"2021. [Daring]. Tersedia pada: <https://www.kemkes.go.id/>
- [12] Leo Agustino, *Analisis Kebijakan Penanganan Wabah Covid-19: Pengalaman Indonesia*. 2020.
- [13] D. T. Larose and C. D. Larose, *Discovering Knowledge In Data An Introduction To Data Mining*. 2014.
- [14] Afanin Hamidah, Abdulloh Hamid, Dian C. Rini Novitasari, *Peramalan Jumlah Kasus COVID-19 Di Kabupaten Nganjuk Menggunakan Metode Arima*.2021.
- [15] Mufti Ari Bianto, Kusrini Sudarmawan, *perancangan sistem klafikasi penyakit jantung Menggunakan naive bayes*.2019