



PENERAPAN DATA MINING UNTUK KLASIFIKASI PANDEMI COVID-19 DI INDONESIA DENGAN ALGORITMA NAÏVE BAYES

Ermanto¹, Damas Mahardi²

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹ermanto@pelitabangsa.ac.id

Abstraksi

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). Meningkatnya akurasi ini dapat mempermudah dalam pengambilan keputusan dan upaya memberikan informasi dan pengetahuan untuk pencegahan terhadap orang yang mau melakukan aktivitas di provinsi tersebut sehingga kita dapat mencegah lebih awal dan tidak terpapar oleh covid-19 ini dan kita dapat memprediksi satu bulan kedepan kasus akan naik yang terkena covid-19. Dalam pengujian data covid-19 dengan algoritma Naïve Bayes, penelitian ini dapat dilakukan pengujian dengan memprediksi atribut yang memiliki 7 atribut dan pengujian tersebut menunjukkan bahwa algoritma naïve bayes dapat klasifikasi sebuah data covid-19 dalam menghasilkan tingkat probabilitas 0,910 yang mana hasil ini cukup baik kemudian dilakukan pengujian 700 training dan 300 data testing menghasilkan paling optimal yang memiliki nilai Accuracy 100.00 % , Precision 100.00 % dan Recall 100.00 %.

Kata Kunci: Virus Corona, Covid-19, Klasifikasi, Naive Bayes, Data Mining

Abstract

Coronavirus is a group of viruses that can cause disease in animals or humans. Several types of corona viruses are known to cause respiratory infections in humans ranging from cold coughs to more serious ones such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). This increased accuracy can facilitate decision making and efforts to provide information and knowledge for the prevention of people who want to carry out activities in the province so that we can prevent earlier and not be exposed by covid-19 and we can predict that one month the case will increase affected by covid-19. In testing the covid-19 data with the Naïve Bayes algorithm, this study can be tested by predicting attributes that have 7 attributes and the test shows that the naïve Bayes algorithm can classify a covid-19 data in producing a 0.910 probability level which is where the results are quite good then 700 training tests were performed and 300 testing data produced the most optimal which had an Accuracy value of 100.00%, Precision 100.00% and Recall 100.00%.

Keywords: Corona Virus, Covid-19, Classification,

Naive Bayes, Data Mining

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. Penularan COVID-19 juga dapat terjadi jika orang menghirup percikan yang keluar dari batuk atau napas orang yang

terjangkit COVID-19. Oleh karena itu, penting bagi kita untuk menjaga jarak lebih dari 1 meter dari orang yang sakit.

WHO terus mengkaji perkembangan penelitian tentang cara penyebaran COVID-19 dan pertama muncul di negara China yaitu di Wuhan. Pengujian data untuk klasifikasi data pasien positif dan negatif, kami dapat mengidentifikasi dampak parah akibat coronavirus dan mungkin dapat mengidentifikasi kasus dan klasifikasi data yang terkena virus.

Manusia merupakan makhluk sosial yang saling berinteraksi secara langsung sehingga tingkat penyebaran pandemi Covid-19 semakin pesat, Indonesia merupakan salah satu negara yang terinfeksi pandemi Covid-19, salah satu penyebab virus corona mudah menyebar di Indonesia adalah karena Indonesia merupakan negara dengan Sektor pariwisata. Sektor pariwisata merupakan salah satu faktor yang berperan penting dalam pertumbuhan perekonomian Indonesia dan memiliki kontribusi devisa terbesar kedua di Indonesia setelah devisa hasil ekspor Kelapa Sawit.

Sektor pariwisata memiliki dampak jangka pendek dan jangka panjang pada perekonomian Indonesia, dampak jangka pendek dapat dirasakan secara langsung, sedangkan dampak jangka Panjang dapat dilihat dengan bertambahnya pendapatan nasional, namun dengan adanya Covid-19 semuanya tak lagi sama, sektor pariwisata yang sekarang mengalami kelesuan sehingga daya beli menurun secara drastis karena berkurangnya pengunjung baik turis lokal maupun turis mancanegara, yang secara otomatis pendapatan dan devisa yang di hasilkan dari sektor pariwisata semakin menurun, bukan hanya sektor pariwisata yang mengalami kelumpuhan sementara, tetapi para karyawan dari jenis perusahaan lainnya ikut merasakan dampak dari pandemi Covid-19.

Yang dimana pekerjaan atau kegiatan yang biasanya dilakukan diluar rumah secara langsung sekaran terpaksa harus dilakukan di dalam rumah, serta ada banyak pula karyawan yang terancam pemberhentian hak kerja (PHK) karena banyak pekerjaan yang tidak memungkinkan untuk dikerjakan dirumah, seperti halnya kegiatan produksi yang bergantung pada mesin yang berada di tempat produksi.

PHK ini juga dilakukan karena kurangnya pembelian dari konsumen dan dibatasinya ekspor ke negara tertentu sehingga akan menghambat ekspor dan mengurangi pendapatan perusahaan, bahkan perusahaan bisa mengalami kerugian. Ada pun penyebab lain dari di PHK nya para karyawan yaitu karena kelangkaan bahan baku untuk diproduksi yang di impor dari negara luar seperti dari negara Tiongkok sehingga akan menghambat kegiatan industri. Perusahaan yang berhenti beroperasi dan

peningkatan jumlah angka pengangguran dapat menghambat dan mengurangi produk domestik bruto (PDB) serta menghambat pertumbuhan ekonomi Indonesia.

Data mining adalah salah satu bentuk implementasi yang diterapkan untuk mencari sebuah model dan pola yang mampu melakukan prediksi pada suatu data berdasarkan data sebelumnya di periode waktu tertentu. Data mining adalah bentuk penggalian data yang digunakan untuk menggali pengetahuan dari jumlah data yang besar. Salah satu algoritma yang digunakan dalam teknik data mining yang memakai teori klasifikasi adalah Naïve Bayes merupakan teknik prediksi, algoritma Naive Bayes Pengklasifikasi bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema bayes. Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition), penggunaan algoritma ini dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Berdasarkan uraian masalah diatas penelitian ini peneliti akan menggunakan algoritma naïve bayes, maka sehubungan dengan hal tersebut, maka dalam penelitian ini mengambil sebuah judul. "Penerapan Data Mining Untuk Klasifikasi Pandemi Covid 19 Di Indonesia Dengan Algoritma Naïve Bayes".

2. Tinjauan Studi

2.1 Covid-19 (Corona Virus Disease 2019)

Penyakit Coronavirus tahun 2019, dikenal sebagai COVID-19, adalah penyakit menyebar cepat yang disebabkan oleh Sindrom Pernafasan Akut Parah Coronavirus 2 (SARS-CoV2). COVID-19 sekarang dianggap pandemi yang telah mempengaruhi negara-negara di semua yang dihuni benua. Sejak kasus pertama COVID-19 dilaporkan di Wuhan, China, pada Desember 2019, jumlah kematian di seluruh dunia telah meningkat dengan cepat.

Karena itu infeksi tinggi dan angka kematian, pemerintah miliki mengimplementasikan berbagai kebijakan yang ditujukan untuk mengurangi penyebaran virus ini dan dampaknya. Seperti itu tindakan dimulai dengan perintah pemerintah China untuk Karantina Wuhan pada 23 Januari 2020, untuk sebagian besar Baru- baru ini, beberapa negara menyatakan keadaan darurat dan menerapkan karantina yang ketat dan menjaga jarak sosial.[7] Sebuah coronavirus baru telah diidentifikasi dengan kasus pasien pertama yang dikonfirmasi pada bulan Desember 2019 di kota Wuhan, provinsi Hubei, di Cina. Sejak itu, jumlah kasus yang dikonfirmasi telahmeningkat drastis. [8].

Coronavirus merupakan suatu kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia. Beberapa jenis corona virus 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. Penularan COVID-19 juga dapat terjadi jika orang menghirup percikan yang keluar dari batuk atau napas orang yang terjangkit COVID-19 dan COVID-19, penyakit yang disebabkan oleh virus SARS-CoV-2, telah dinyatakan pandemi oleh Organisasi Kesehatan Dunia (WHO) [9].

2.2 Algoritma Naive Bayes

Algoritma Naive Bayes Pengklasifikasi bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema bayes. Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition), penggunaan algoritma ini dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Misalkan X adalah set atribut data dan h kelas variabel dan jika kelas memiliki hubungan dengan atribut maka diperlukan X dan h sebagai variabel acak dan menangkap hubungan peluang P(h|X) ini peluang posterior untuk h dan sebaliknya perior P(h) [4]. Naive Bayes mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas label kelas label kelas y. () ∏ () = = = d i i P X Y y P X Y y 1 dengan tiap set atribut X = {X1 , X 2 ,2, X d} terdiri dari d atribut. Tahapan algoritma naive bayes:

1. Menyiapkan data training.
2. Setiap data dipresentasikan sebagai vektor berdimensi-n yaitu X=(x1,x2,x3, xn).
3. n adalah gambaran dari ukuran yang dibuat di test dari n atribut yaitu A1,A2,A3,An
4. M adalah kumpulan kategori yaitu C1,C2,C3, Cm.
5. Diberikan data test X yang tidak diketahui kategorinya, maka classifier akan memprediksi bahwa X adalah milik kategori dengan posterior probability tertinggi berdasarkan kondisi X.
6. Naive bayes classifier menandai bahwa test X yang tidak diketahui tadi ke kategori C1 jika dan hanyajika P(Ci|X)>P(Cj|X) untuk 1 ≤ j ≤ m, j ≠ i.
7. Kemudian kita perlu memaksimalkan P(Ci|X) P(Ci|X)= P(X|Ci). P(Ci) P(X).
8. Dimana x adalah nilai-nilai atribut dalam sampel X dan probabilitas P(x1|Ci), P(x2|Ci),

3. Metode Penelitian

Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat langkah – langkah dalam tahapan penelitian yang akan dilakukan berikut



Gambar 1. Tahapan Penelitian

Dalam pengujian data menggunakan teknik klasifikasi dan tahapan - tahapan pada data mining dengan algoritma *Naive Bayes*, pengolahan data dan yang akan dijadikan dataset dalam penelitian ini Dari data covid-19 tersebut akan dibagi menjadi 70% data training dan 30% data testing sebagai data yang akan diujikan dan menghasilkan nilai sebagai hasil prediksi data covid-19.

4. Hasil dan Pembahasan

4.1 Split Validation

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian *data training* dan sebagian *data testing*. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik sampling random untuk data training (70%) dan data testing (30%). Contoh perhitungan untuk pengambilan data testing adalah sebagai berikut :

Jumlah data keseluruhan (N) = 1000 Jumlah data testing = 30% x 1000 = 300 Jumlah sample (n) = 300

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(C_i|X)}$$

$$P(X|C_i)P(C_i)$$

Interval sampling (k) = N/n = 1000/300 = 3,3
 Unsur pertama yang diambil untuk data testing (D)= 300.

Tabel 1. Model Split Validation

P(xn|Ci), dapat diperkirakan dari d

K-Fold Cross Validation	Data training	Data Testing
1	700	300
2	700	300
3	700	300

Dari hasil diatas diperoleh data testing sebanyak 300 data, maka sisanya dijadikan data training sebanyak $1000 - 300 = 700$ data Perhitungan dari algoritma naïve bayes, akurasi dilakukan dengan cara jumlah TP + TN dibagi jumlah total data testing yang diuji.

Tabel 2. Accuracy

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP 27	FP 0
	FALSE	FN 0	TN 273

Nilai precision dihitung dengan cara membagi jumlah data benar yang bernilai positif (True Positive) dibagi dengan jumlah data benar yang bernilai positif (True Positive) dan data salah yang bernilai positif (False Positive).

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} * 100\% \\
 &= \frac{27 + 273}{27 + 273 + 0 + 0} * 100\% \\
 &= \frac{300}{300} * 100\% \\
 &= 1 * 100\% \\
 &= 100.00\%
 \end{aligned}$$

Tabel 3. Precision

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP 27	FP 0
	FALSE	FN 0	TN 273

Nilai recall dihitung dengan cara membagi data benar yang bernilai positif (True Positive) dengan hasil penjumlahan dari data benar yang bernilai positif (True Positive) dan data salah yang bernilai negatif (False Negative).

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} * 100\% \\
 &= \frac{27}{27 + 0} * 100\% \\
 &= \frac{27}{27} * 100\% \\
 &= 1 * 100\% \\
 &= 100.00\%
 \end{aligned}$$

Tabel 4. Recall

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP 27	FP 0
	FALSE	FN 0	TN 273

Dari data sampel yang diutarakan yaitu sebanyak dan 300 data yang menyatakan Tingkat ke accuracy, recall dan persicion naïve bayes yang lebih baik dalam tingkat accurecy 100.00 %, Recall 100.00% dan Percision 100% maka di putuskan dalam memprediksi data covid-19 menghasilakn tingkat ke akuratan data yang maksimal sehingga mendapatkan hasil yang lebih baik.

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} * 100\% \\
 &= \frac{27}{27 + 0} * 100\% \\
 &= \frac{27}{27} * 100\% \\
 &= 1 * 100\% \\
 &= 100.00\%
 \end{aligned}$$

Tabel 5. Prediksi

No	Algoritma	Accury	Percision	Recall
1	Naïve Bayes	100.00%	100.00%	100.00%

Hasil pengujian data yang telah dilakukan. Berdasarkan analisa data testing pengujian dari data training yang dihasilkan terdapat hasil yang relevan dengan komposisi data yang akurat, data Covid-19 di indonesia dengan pengujian algoritma naïve bayes menghasilkan hasil optimal, berdasarkan hasil yang sudah didapatkan dalam penelitian ini maka algoritma yang menghasilkan tingkat accuracy, recall dan precision yang baik atau semua atribut dalam meningkatkan hasil accuracy, recall dan precision. Meningkatnya akurasi ini dapat mempermudah dalam pengambilan keputusan dan

upaya memberikan informasi dan pengetahuan untuk pencegahan terhadap orang yang mau melakukan aktivitas di provinisi tersebut sehingga kita dapat mencegah lebih awal dan tidak terpapar oleh covid-19 ini dan kita dapat memprediksi satu bulan kedepan kasus akan naik yang terkena covid-19.

5. Kesimpulan

Berdasarkan hasil pengujian dengan menggunakan algoritma Naïve Bayes dapat diambil suatu kesimpulan sebagai berikut :

1. Dalam pengujian data Covid-19 hasil accuracy, precision, dan recall dengan algoritma Naïve Bayes, penelitian ini juga dilakukan pengujian dengan klasifikasi atribut yang memiliki 7 atribut. Hasil pengujian tersebut menunjukkan bahwa algoritma naïve bayes dapat klasifikasi sebuah data covid-19 dalam menghasilkan tingkat probalitas 0,910 yang mana hasil ini cukup baik .
2. Hasil pengujian data Covid-19 yang dilakukan pengujian 700 training dan 300 data testing menghasilakn paling optimal dalam pengujian data Covid-19 sehingga menghasilkan tingkat akurat yang paling tinggi dalam klasifikasi dan dapat meramalkan satu bulan kedepan atau memprediksi satu bulan kedepan akan meningkat kasus yang terkena covid-19.
3. Dari hasil pengujian dan penelitian ditemukan hasil terpapar oleh covid-19 di Indonesia cukup tinggi terkena virus corona atau Covid-19 sehingga dapat melakukan pencegahan dan membuat saran agar melakukan lock down atau pembatasan sosial bersekala besar dan rapid test .

Daftar Pustaka

- [1] De Salazar PM, Niehus R, Taylor A ,Buckee C, Lipsitch M. Using predicted imports of 2019- nCoV cases to determine locations that may not be identifying all imported cases (2020
- [2] C. E. Lopez, M. Vasu, and C. Gallemore, “Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset,” pp. 1–4, 2020.
- [3] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, “Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network,” 2020.
- [4] Suyanto, Data Mining. Yogyakarta: Informatika, 2017.
- [5] Retno Tri vulandari, Data Mining. Yogyakarta: Gava Media, 2017.
- [6] G. widi N. Dicky Nofriansyah, Algoritma Data Mining Dan pengujian. Yogyakarta: Cv Budi Utama, 2015.

- [7] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the Landscape of Artificial Intelligence Applications against COVID-19," pp. 1–14, 2020.
- [8] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight covid-19," *Physiol. Genomics*, vol. 52, no. 4, pp. 200–202, 2020, doi:10.1152/physiolgenomics.00029.2020.
- [9] Y. Li et al., "COVID-19 Evolves in Human Hosts," *SSRN Electron. J.*, pp. 1–11, 2020, doi:10.2139/ssrn.3562070.
- [10] Y. Silalahi, Kristiani Silalahi., Murfi, Hendri., Satria, "Studi Perbandingan Pemilihan Fitur untuk Support Vector Machine pada Klasifikasi Penilaian Risiko Kredit," vol. 1, no. 2, pp. 119–136, 2017.
- [11] O. Villacampa, "(Weka - Thesis) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis," *ProQuest Diss. Theses*, no. 63, p. 188, 2015.
- [12] M. Rahmadi, F. Kaurie, and T. Susanti, "Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan Weka Tools," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 134, 2020, doi: 10.30865/jurikom.v7i1.1761.