

KLASIFIKASI DIAGNOSIS PENYAKIT COVID-19 MENGUNAKAN METODE DECISION TREE

Rashid Ridho¹, Hendra²

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta,
Jl. Cempaka Putih Tengah 27, 10510, Jakarta
2016470050@ftumj.ac.id¹, hendra@umj.ac.id²

Abstrak

Covid-19 yaitu penyakit yang disebabkan oleh virus corona yang menyerang saluran pernafasan sehingga menyebabkan demam tinggi, batuk, flu, sesak nafas serta nyeri tenggorokan. Penyakit ini dapat menyebar melalui tetesan kecil (*droplet*) dari hidung atau mulut pada saat bersin. Kriteria orang yang terinfeksi virus corona dengan gejala-gejala demam dengan suhu tubuh lebih dari 38° C, batuk flu, dan nyeri tenggorokan, dianjurkan untuk memeriksakan kesehatan di fasilitas pelayanan kesehatan. Tetapi beberapa orang takut untuk memeriksakan diri ke pelayanan kesehatan secara langsung karena mereka masih takut terdiagnosis positif Covid-19. Kurangnya edukasi tentang Covid-19 ini, membuat masyarakat menjadi panik dan kurang kooperatif dalam menghadapi pandemi ini. Oleh sebab itu diperlukan suatu sistem yang dapat melakukan diagnosis dan edukasi tentang penyakit Covid-19 berbasis *web*. Pada penelitian ini menggunakan metode *decision tree* untuk mengklasifikasi 125 *dataset* yang di dapat dari rekam medis RSUD Koja. Hasil dari penelitian ini adalah sistem yang dapat memprediksi positif atau negatif *user* terdiagnosis Covid-19 berdasarkan jenis kelamin, usia, gejala dan kondisi *user*. Pengujian dari penelitian ini dihitung dengan menggunakan metode *confusion matrix*, dengan hasil yang di dapat yaitu nilai akurasi 90%.

Kata Kunci: Diagnosis, Covid-19, *decision tree*.

Abstract

Covid-19 is a disease caused by the corona virus which attacks the respiratory tract, causing high fever, cough, flu, shortness of breath and sore throat. This disease can spread through small droplets from the nose or mouth when sneezing. Criteria for people infected with the corona virus with symptoms of fever with a body temperature of more than 38 ° C, cough, flu, and sore throat, it is recommended that they have their health checked in a health care facility. But some people are afraid to go to health services directly because they are still afraid of being diagnosed with Covid-19. Lack of education about Covid-19 has made people panic and less cooperative in dealing with this pandemic. Therefore we need a system that can make diagnosis and education about the web-based Covid-19 disease. In this study, a decision tree method was used to classify 125 datasets obtained from the medical records of Koja Hospital. The results of this study are a system that can predict positively or negatively a user diagnosed with Covid-19 based on gender, age, symptoms and user's condition. The test of this research is calculated using the confusion matrix method, with the results obtained, namely an accuracy value of 90%.

Keyword: Diagnosis, Covid-19, Decision Tree

Pendahuluan

Saat ini dunia khususnya Indonesia sedang berjuang melawan pandemi penyakit Covid-19 yang disebabkan oleh SARS-CoV-2 (kelompok virus yang menginfeksi sistem pernapasan). Penyebaran virus ini sangatlah cepat hingga memakan banyak korban jiwa diberbagai negara. Menurut data yang dirilis Gugus Tugas Percepatan Penanganan Covid-19 Republik Indonesia, jumlah kasus terkonfirmasi positif hingga 26 Oktober 2020 adalah 389.712 orang dengan jumlah kematian 13.299 orang. Dari data tersebut dapat disimpulkan bahwa tingkat kematian yang disebabkan oleh Covid-19 di Indonesia adalah sekitar 3,4%.

Penyakit Covid-19 menyerang saluran pernafasan dan menyebabkan demam tinggi, batuk, flu, sesak nafas dan nyeri tenggorokan. Covid-19 dapat menular dari manusia ke manusia. Penyakit ini dapat menyebar melalui tetesan kecil (droplet) dari hidung atau mulut pada saat bersin. Tetesan itu kemudian mendarat di benda atau permukaan yang disentuh oleh orang yang sehat, lalu orang sehat itu menyentuh mata, hidung atau mulut mereka. Covid-19 dapat menginfeksi siapa saja, tetapi akan lebih berbahaya atau bahkan fatal bila menyerang orang lanjut usia, ibu hamil, perokok, dan penderita penyakit tertentu.

Kriteria orang yang terinfeksi virus corona dengan gejala-gejala demam dengan suhu tubuh lebih dari 38° C, batuk kering, dan sesak napas. Orang yang mempunyai gejala-gejala tersebut dianjurkan untuk memeriksakan kesehatan di fasilitas pelayanan kesehatan. Tetapi masih saja ada beberapa orang takut untuk memeriksakan diri ke pelayanan kesehatan secara langsung karena mereka masih takut terdiagnosis positif Covid-19. Kurangnya edukasi tentang Covid-19 ini, membuat masyarakat menjadi panik dan kurang kooperatif dalam menghadapi pandemi ini.

Berdasarkan latar belakang diatas, dibutuhkan sebuah aplikasi untuk mengklasifikasi diagnosis penyakit Covid-19 menggunakan metode *Decision Tree* berbasis *website*. Metode *Decision Tree* dipilih karena dapat menggambarkan pohon keputusan

terhadap masalah yang sedang dihadapi. Diharapkan aplikasi yang berbasis web ini dapat membantu masyarakat agar lebih mempunyai kesadaran dalam menghadapi pandemi ini.

Tinjauan Pustaka

A. Penyakit Covid-19

Menurut (Keputusan Menteri Kesehatan Republik Indonesia, 2020) *Coronavirus Disease 2019 (COVID-19)* adalah penyakit menular yang disebabkan oleh *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)*. Tanda dan gejala umum infeksi COVID-19 antara lain gejala gangguan pernapasan akut seperti demam, batuk dan sesak napas. Masa inkubasi rata-rata 5-6 hari dengan masa inkubasi terpanjang 14 hari. Pada kasus COVID-19 yang berat dapat menyebabkan pneumonia, sindrom pernapasan akut, gagal ginjal, dan bahkan kematian.

B. Klasifikasi

Merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data baru untuk diklasifikasi.

C. *Decision Tree* (Pohon Keputusan)

Decision Tree atau Pohon keputusan adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*) simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan (*branch node*) dan simpul daun atau *leaf node* (Hermawati, 2013).

D. Algoritma ID3

Algoritma ID3 adalah algoritma pembelajaran pohon keputusan yang sederhana dikembangkan oleh J. Ross Quinlan pada tahun 1993. Algoritma ID3 melakukan prosedur pencarian secara menyeluruh pada semua kemungkinan pohon keputusan. Ide dasar dari algoritma

ID3 adalah untuk membangun pohon keputusan dengan menggunakan sebuah *top-down*, pencairan menyeluruh melalui set yang diberikan untuk menguji masing-masing atribut pada setiap simpul pohon (Ture *et al.*, 2009). Secara ringkas, langkah kerja algoritma ID3 dapat digambarkan sebagai berikut (Defiyanti dan Pardede, 2009) :

1. Input sampel training, label training, dan atribut.
2. Hitung *Entropy* dan *Information Gain* dari setiap atribut dengan rumus

$$Entropy (S) = (-P_+ \text{Log}_2 P_+) + (-P_- \text{Log}_2 P_-)$$

Dimana:

S = ruang (data) sample yang digunakan untuk training.

P₊ = jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.

P₋ = jumlah yang bersolusi negatif (tidak mendukung) pada data sample untuk kriteria tertentu.

$$Gain (S, A) = Entropy (S) - \sum_{v=1}^v \frac{|S_v|}{|S|} * Entropy (S_v)$$

Dimana:

S = ruang (data) sample yang digunakan untuk training.

A = atribut.

V = Suatu nilai yang mungkin untuk atribut A.

Nilai(A) = Himpunan yang mungkin untuk atribut A.

|S_v| = Jumlah sample untuk nilai V.

|S| = jumlah seluruh sample data.

Entropy(S_v) = entropy untuk sample-sample yang memiliki nilai V.

3. Bentuk simpul yang berisi atribut tersebut.
4. Ulangi proses perhitungan information gain yang akan terus dilaksanakan sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam

perhitungan nilai information gain.

E. Confusion Matrix

(Rosandy, 2016) *Confusion matrix* adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Precision* atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

Tabel 1
Model *Confusion Matrix*

	Actual	
Prediksi	TP	FP
	FN	TN

Terdapat empat istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive* (TP) merupakan data positif yang diprediksi benar. *True Negative* (TN) merupakan data negatif yang diprediksi benar. *False Positive* (FP) merupakan data negatif namun diprediksi sebagai data positif. *False Negative* (FN) merupakan data positif namun diprediksi sebagai negatif. Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Metode Penelitian

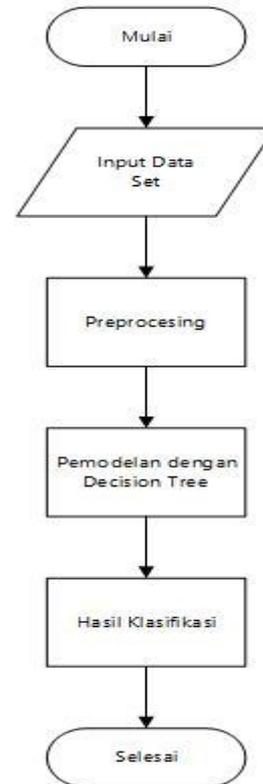
A. Pengumpulan Data

Pada penelitian ini data diperoleh dari rekam medis RSUD Koja yang pernah melakukan tes diagnosis *swab test* PCR, sebanyak 125 data dengan data positif sebanyak 65 dan negatif sebanyak 60. Pada Tabel 2 dan 3 menjelaskan kode atribut yang dipakai.

Tabel 2
Gejala-Gejala Covid-19 yang menjadi atribut

Nomor	Gejala	Kode Gejala
1	Demam	G1
2	Batuk Kering	G2
3	Pilek	G3
4	Sakit Tenggorokan	G4
5	Sesak Napas	G5
6	Menggigil	G6
7	Sakit Kepala	G7
8	Lemah	G8
9	Nyeri Otot	G9
10	Mual atau Muntah	G10
11	Nyeri Abdomen	G11
12	Diare	G12

Pengolahan data yang digunakan dalam penelitian ini menggunakan metode *Decision Tree* dengan algoritma ID3. Aplikasi yang dipakai untuk melakukan pengolahan data menggunakan RapidMiner Studio. Di bawah ini merupakan gambar 1 skema proses perancangan yang akan dilakukan sebagai berikut:



Gambar 1 Flowchart Pengolahan Data

Tabel 3
Kode Atribut Kondisi

Nomor	Kondisi	Kode
1	Hamil	K1
2	Diabetes	K2
3	Jantung	K3
4	Hipertensi	K4
5	Keganasan (Tumor)	K5
6	Ginjal	K6

B. Pengolahan Data

1) Input Data Set

Setelah data didapatkan, Dataset atau data *training* di *input* kedalam aplikasi RapidMiner Studio untuk dilakukan *preprocessing* dan pemodelan *decision tree*.

2) Preprocessing

Pada penelitian ini proses *preprocessing* dilakukan sebagai berikut :

1. Menentukan Nilai atribut

Nilai atribut yang mempunyai lebih dari dua faktor diubah menjadi *binominal* seperti atribut Jenis Kelamin, gejala-gejala, kondisi penyerta, dan hasil. Sedangkan yang mempunyai lebih dari faktor diubah

menjadi *polynomial* seperti nama dan usia.

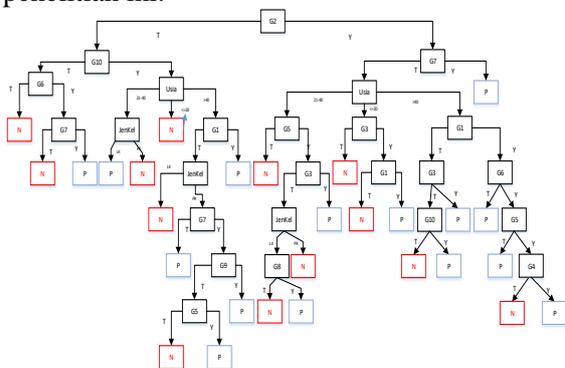
2. Menentukan Kolom Label dan Id
 Pada penelitian ini kolom hasil adalah yang menjadi label, untuk mengubah menjadi label pilih menu pengaturan kemudian pilih *Change Role* kemudian pilih label. Sedangkan nama tidak terpengaruh ke dalam penelitian maka kolom nama menjadi id, untuk mengubah kolom nama menjadi id pilih menu pengaturan kemudian pilih *Change Role* kemudian pilih id
3. Mengkategorikan Usia
 Pada penelitian ini kolom usia di kategorikan menjadi 3 yaitu diatas usia 40 Tahun (>40), antara 21 sampai dengan 40 Tahun (21...40), dan di bawah 21 Tahun (<=20).

3) Pemodelan *Decision Tree*

Pada pemodelan *decision tree* menggunakan algoritma ID3. Algoritma ini menghitung nilai *entropy* untuk semua atribut kemudian membandingkan nilai *gain* pada semua atribut. Angka *gain* yang terbesar akan menjadi node pada model *decision tree*. Atribut yang menjadi node tidak akan dihitung kembali. Hal tersebut dilakukan terus hingga mendapatkan node daun atau sebuah fakta.

4) Hasil Klasifikasi

Hasil Klasifikasi dari penelitian ini adalah model *decision tree*, nanti model ini yang akan menjadi acuan data baru. Pada gambar 2 menunjukkan model *decision tree* pada penelitian ini.

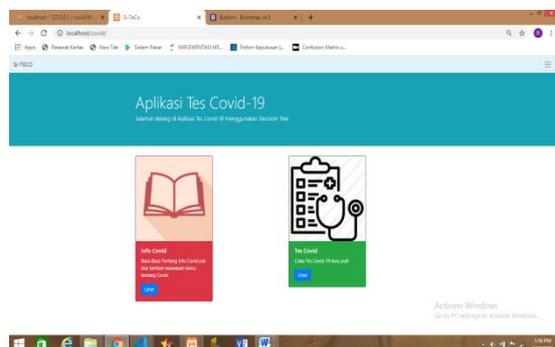


Gambar 2 Model *Decision Tree*

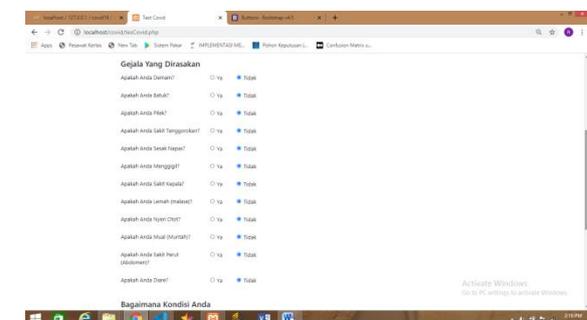
Pada penelitian ini, mengkatagorikan usia menjadi 3 kategori yaitu di bawah atau sama dengan 20 Tahun, 21 sampai dengan 40 Tahun dan diatas 40 Tahun. Dapat diketahui untuk usia diatas 40 Tahun adalah yang paling rentan terinfeksi virus Covid-19. Pada penelitian ini, terdapat 20 atribut dan hanya 12 atribut yang menjadi model *decision tree*. Hal ini dikarenakan 8 atribut tersebut mempunyai nilai *gain* yang rendah daripada atribut lainnya saat mencari node model *decision tree*, 12 atribut yang menjadi model *decision tree* adalah jenis kelamin, usia, demam, batuk, pilek, sakit tenggorokan, sesak napas, menggigil, sakit kepala, lemah, nyeri otot, dan mual, sedangkan 8 atribut yang tidak menjadi model *decision tree* adalah nyeri abdomen (nyeri perut), diare, hamil, diabetes, jantung, hipertensi, keganasan (tumor) dan gagal ginjal.

Hasil dan Pembahasan

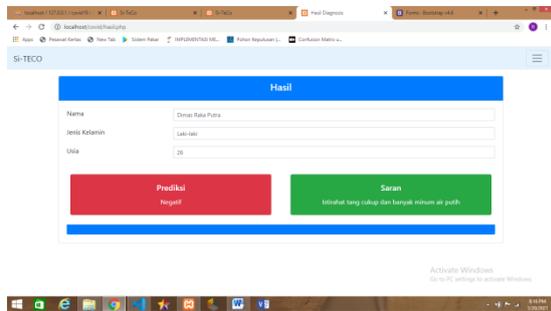
Implementasi dilakukan dengan menggunakan bahasa pemrograman PHP dengan database MySQL. Berikut adalah halaman utama pada aplikasi Tes Covid-19 menggunakan metode *decision tree*.



Gambar 3 Tampilan Halaman Utama



Gambar 4 Tampilan Halaman Tes Covid-19



Gambar 5 Tampilan Hamalan Hasil

Pengujian akurasi model dilakukan menggunakan metode *confusion matrix* dengan menghitung nilai *accuracy*. Data diperoleh dari jawaban kuesioner yang diberikan kepada orang yang pernah melakukan diagnosis Covid-19 dengan menggunakan PCR test.

Pada penelitian ini terdapat 20 data responden yang menjadi data uji. Data uji ini akan dibandingkan dengan data set untuk mendapatkan nilai akurasi model *decision tree*. Perbandingan Data uji dengan model dapat dilihat pada Tabel 4 dan Tabel 5 di bawah ini.

Tabel 4
Perbandingan Data Uji dengan Model

No	Nama	Hasil data uji	Hasil data set	Keterangan
1	Sempel 1	P	P	Sesuai
2	Sempel 2	N	N	Sesuai
3	Sempel 3	N	N	Sesuai
4	Sempel 4	N	N	Sesuai
5	Sempel 5	P	P	Sesuai
6	Sempel 6	P	P	Sesuai
7	Sempel 7	P	P	Sesuai
8	Sempel 8	P	P	Sesuai
9	Sempel 9	N	N	Sesuai
10	Sempel 10	N	N	Sesuai
11	Sempel 11	P	N	Tidak Sesuai

No	Nama	Hasil data uji	Hasil data set	Keterangan
12	Sempel 12	N	P	Tidak Sesuai
13	Sempel 13	N	N	Sesuai
14	Sempel 14	P	P	Sesuai
15	Sempel 15	P	P	Sesuai
16	Sempel 16	N	N	Sesuai
17	Sempel 17	P	P	Sesuai
18	Sempel 18	N	N	Sesuai
19	Sempel 19	N	N	Sesuai
20	Sempel 20	P	P	Sesuai

Tabel 5
Data Akurasi Model

	Aktual P	Aktual N
Prediksi P	9 (TP)	1 (FP)
Prediksi N	1 (FN)	9 (TN)

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$accuracy = \frac{9 + 9}{9 + 9 + 1 + 1} \times 100\%$$

$$accuracy = 0.9 \times 100\% = 90\%$$

Penutup

Dari hasil penelitian ini, maka dapat diambil beberapa kesimpulan yaitu:

1. Dalam penelitian ini terdapat 20 atribut yang digunakan dan 12 atribut yang memiliki pengaruh yaitu, Jenis Kelamin, Usia, Demam, Batuk, Pilek, Sakit Tenggorokan, Sesak Napas, Menggigil, Sakit Kepala, Lemah, Mual, dan Nyeri Otot dalam mendiagnosis penyakit Covid-19.
2. Aplikasi Diagnosis Covid-19 ini digunakan sebagai alat diagnosis

dini untuk mengetahui positif atau negatif Covid-19 yang hasilnya ditentukan menggunakan metode *Decision Tree* dari data yang sudah diteliti.

3. Akurasi dari aplikasi yang dibangun dalam penelitian ini memiliki persentase 90% dari 20 data uji berdasarkan perbandingan hasil perhitungan manual dan hitungan aplikasi sehingga dapat disimpulkan bahwa aplikasi yang dirancang cukup baik dan berjalan secara fungsionalitas, serta dapat menampilkan informasi yang diinginkan.

MENGANALISA KELANCARAN PEMBAYARAN. *Jurnal TIM Darmajaya*, 02, 52-62.

- Sidette, J. A., Sedyono, E., & Nurhayati, O. D. (2014). Pendekatan Metode Pohon Keputusan Menggunakan Algoritma ID3 Untuk Sistem Informasi Pengukuran Kinerja PNS. *Jurnal Sistem Informasi Bisnis*, 75-86.
- Sukamto, R., & Sahalahuddin, M. (2013). *Rekayasa Perangkat Lunak*. Bandung: Bandung Informatika.
- Yasin, V. (2012). *Rekayasa Perangkat Lunak Berorientasi Objek : Pemodelan, Arsitektur dan Perancangan (Modeling, Architecture and Design)*. Jakarta: Mitra Wacana Media.

DAFTAR PUSTAKA

- Adi, N. (2010). *Rekayasa Perangkat Lunak Menggunakan UML dan Java*. Yogyakarta: Andi Offset.
- Asy'ari, R. O. (2020). *Pengertian Covid-19 Dan Bentuk Partisipasi Dalam Memerangnya*. Surabaya: Universitas Negeri Surabaya.
- Fais, M. N., Susanto, A., & Listyorini, T. (2014, November 2). PENGEMBANGAN SISTEM PARKIR DI UNIVERSITAS MURIA KUDUS DENGAN MENGGUNAKAN ENKRIPSI DATA DAN TEKNOLOGI BARCODE. *SIMETRIS*, 173-180.
- Hermawati, F. (2013). *Data Mining*. Yogyakarta: Andi Offset.
- Keputusan Menteri Kesehatan Republik Indonesia No 341. (2020, Juli 13). *Pedoman Pencegahan Dan Pengendalian Coronavirus Disease 2019 (Covid-19)*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Ma Chung. (2018). *Studi Kasus Sistem Berbasis Pengetahuan*. Malang: CV. Seribu Bintang.
- Rosandy, T. (2016, Mei 01). PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK