

SISTEMATISASI DIAGNOSIS PENYAKIT COVID-19 MENGUNAKAN *DATA MINING* DENGAN *DECISION TREE*

Hendra¹, Rashid Ridho²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta

hendra@umj.ac.id*

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 terdiagnosa 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 didapat yaitu nilai akurasi 90%.

Kata Kunci: *diagnosis, covid-19, data mining, 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 healthcare 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%.

Keywords: *diagnosis, covid-19, data mining, decision tree*

1. Pendahuluan

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. Virus ini juga bisa menyebar ketika tetesan kecil (*droplet*) itu dihirup oleh orang yang sehat ketika berdekatan dengan penderita Covid-19. 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 terdiagnosa positif Covid-19. Kurangnya edukasi tentang Covid-19 ini, membuat masyarakat menjadi panik dan kurang kooperatif dalam menghadapi pandemi ini.

1.1. Tujuan Penelitian

Tujuan penelitian ini adalah membantu masyarakat dalam mengetahui gejala awal dari penyakit Covid-19 sehingga dapat mengambil tindakan dini.

1.2. Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah membantu masyarakat untuk melakukan pengecekan terhadap gejala-gejala awal Covid-19, sehingga masyarakat yang telah menderita penyakit Covid-19 dapat segera ditangani oleh para pakar yang ahli dibidangnya.

Berdasarkan latar belakang diatas, penulis ingin membuat rancangan sistematis yang dapat melakukan diagnosis dan edukasi tentang Covid-19. Rancangan ini

menghasilkan suatu keputusan diagnosis yang sesuai dengan keputusan dari seorang dokter. Mengklasifikasi diagnosis penyakit Covid-19 menggunakan metode *decision tree* berbasis web. Metode *decision tree* dipilih karena dapat menggambarkan pohon keputusan terhadap masalah yang sedang dihadapi. Diharapkan sistematisasi yang berbasis web ini dapat membantu masyarakat agar lebih mempunyai kesadaran dalam menghadapi pandemi.

2. Tinjauan Pustaka

2.1. 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). SARS-CoV-2 merupakan coronavirus jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Ada setidaknya dua jenis coronavirus yang diketahui menyebabkan penyakit yang dapat menimbulkan gejala berat seperti *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS). 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.

2.2. Data Mining

Data mining merupakan sebuah teknik untuk menggali informasi tersembunyi untuk memperoleh manfaat lebih dari data yang ada. Mining bisa diartikan serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data. Diantara tujuan data mining adalah untuk melakukan klasifikasi, klusterisasi, menemukan pola asosiasi hingga melakukan peramalan (*predicting*). Istilah data mining lebih populer dengan sebutan KDD (Knowledge Discovery from Database). (Witten, 2011)

2.3. Decision Tree

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). Menurut (Sidette, Sedyono, & Nurhayati, 2014) pada sebuah pohon keputusan setiap node merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level node teratas dari sebuah pohon keputusan adalah node akar (*root*) yang biasanya berupa atribut yang memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya pohon keputusan melakukan strategi pencarian secara top-down untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar (*root*) sampai node akhir (*daun*) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon tersebut menjadi aturan (*rule*).

3. Metode Penelitian

3.1. Pengumpulan Data

Dalam penelitian ini pengumpulan data konkret yang berhubungan dengan perancangan dan pembuatan klasifikasi diagnosis penyakit Covid-19 dengan metode *decision tree*.

Tahapan pengumpulan informasi yang berkaitan dengan penelitian serta teori-teori pendukung lainnya. Teori tersebut meliputi *Unified Modeling Language* (UML) seperti usecase diagram, dan *activity diagram*. Selain itu, PHP, MySQL, sebagai metode yang digunakan untuk proses perancangan sistem.

Wawancara dilakukan dengan mengajukan beberapa pertanyaan mengenai kebutuhan perancangan sistem terkait penelitian.

Metode observasi dilakukan dengan cara mengamati secara langsung di RSUD Koja selama satu bulan terhitung dari pada tanggal 15 Desember 2020 sampai dengan 15

Januari 2021. Observasi dimaksudkan untuk mendapat data rekam medis penyakit Covid-19 yang ada di RSUD Koja. Pada penelitian ini, terdapat data rekam medis sebanyak 125 data yang pernah melakukan diagnosis Covid-19 di RSUD Koja. Data yang diambil adalah data rekam medis RSUD Koja dari bulan Maret 2020 sampai dengan bulan November 2020. Berdasarkan data yang ada orang yang terdiagnosa positif Covid-19 sebanyak 65 orang dan yang negatif 60 orang.

3.2. Analisis Data

Pada tahap ini, dilakukan analisis kebutuhan fungsional dan kebutuhan non-fungsional pada sistem klasifikasi diagnosis penyakit Covid-19 yaitu:

- 1) Kebutuhan Fungsional merupakan kebutuhan yang mencakup reaksi sistem pada input dan perilaku sistem pada situasi tertentu. Berikut adalah analisis kebutuhan fungsional pada sistem klasifikasi diagnosis.

Kebutuhan non-fungsional merupakan kebutuhan yang menitikberatkan pada properti perilaku yang dimiliki oleh sistem. Berikut adalah analisis kebutuhan non-fungsional pada sistem klasifikasi diagnosis penyakit Covid-19.

3.3. Perancangan Sistem

Setelah data-data dikumpulkan dan dianalisis, lalu dilakukan perancangan sistem mencakup prosedur pemeriksaan, aturan-aturan metode *decision tree* untuk mengklasifikasi diagnosis penyakit Covid-19, perancangan UML, database dan aplikasi.

4. Hasil dan Pembahasan

4.1. Data Gejala Penyakit Covid-19

Data gejala penyakit Covid-19, jenis kelamin, usia, hamil, dan riwayat penyakit digunakan untuk menjadi atribut untuk mengklasifikasi diagnosis penyakit Covid-19. Data-data tersebut diambil dari rekam medis yang ada pada RSUD Koja. Pada Tabel 1 menjelaskan kode atribut gejala-gejala Covid-19 yang digunakan pada penelitian ini.

TABEL 1

GEJALA-GEJALA COVID-19		
No.	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/ Muntah	G10
11.	Nyeri Abdomen	G11
12.	Diare	G12

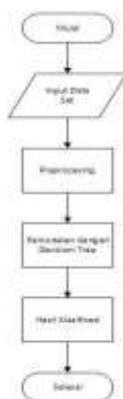
Pada Tabel 2 menjelaskan kode atribut yang digunakan.

TABEL 2
GEJALA-GEJALA COVID-19

No.	Gejala	Kode Gejala
1.	Demam	G1
2.	Batuk Kering	G2
3.	Pilek	G3
4.	Sakit Tenggorokan	G4
5.	Sesak Napas	G5
6.	Menggigil	G6

4.2. Metode Pengolahan Data

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.

1. Input *dataset*

Dataset atau data training di input

kedalam aplikasi RapidMiner Studio. Proses peng-input-an data dengan cara memilih menu import data, kemudian melakukan pemilihan file pada file yang menjadi sumber data.

2. *Preprocessing*

Proses preprocessing dilakukan sebagai berikut :

a. Menentukan nilai atribut

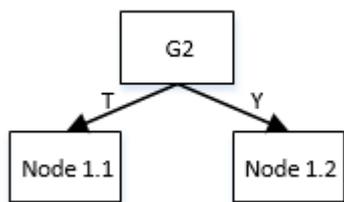
Nilai atribut yang mempunyai lebih dari dua faktor diubah menjadi binominal seperti atribut Jenis Kelamin, gejala-gejala, kondisi penyerta, dan hasil. Sedangkan yang mempunyai lebih dari faktor diubah menjadi polynominal seperti nama.

b. 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. Karena nama tidak terpengaruh ke dalam penelitian maka kolom nama menjadi id.

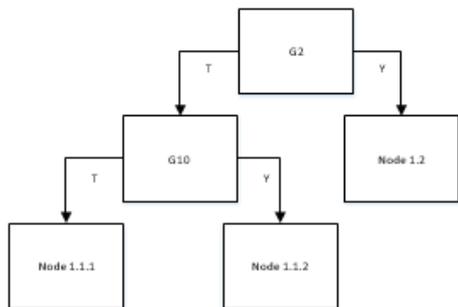
3. Mengkategorikan usia

Pada kolom usia dikategorikan menjadi 3 yaitu diatas usia 40 Tahun (>40), antara 21 sampai dengan 40 Tahun (21...40), dan di bawah 21 Tahun (<=20). Pada gambar 3.6 menunjukkan bahwa kolom usia masih bernilai integer, untuk dapat mengkategorikan usia dataset harus drag ke bagian proses, kemudian cari operator Discretize by User Specification kemudian drag ke proses kemudian hubungkan dataset dengan operator tersebut. Setelah itu akan muncul halaman edit parameters list, masukan nama kelas dan batas atas dari kelas tersebut. Setelah dibuatkan batas maka kolom usia sudah dikategorikan. Nilai entropy dan gain menunjukkan mencari node akar (root). Atribut G2 menjadi node akar (root) karena mempunyai nilai gain yang paling tinggi dengan nilai 0.2181095. Atribut G2 mempunyai 2 faktor Y dan T sehingga selanjutnya mencari node G2 T dan node G2 Y.



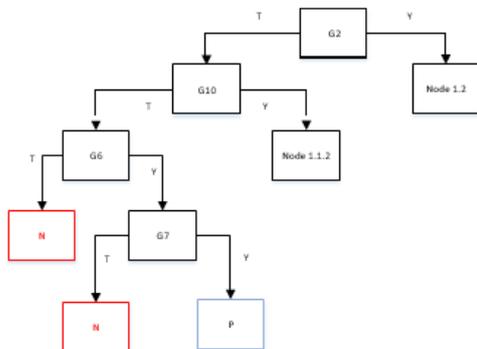
Gambar. 2. Flowchart Pengolahan Data.

Pada Gambar 2 menunjukkan *node* akar (*root*) terdapat dua cabang (T dan Y) yang menjadi *node* 1.1 dan *node* 1.2. Untuk mencari *node* 1.1 dan *node* 1.2 dapat melakukan cara yang ada di atas, sehingga dapat diketahui nilai entropy dan gain.



Gambar. 3. Model Decision Tree Node 1.1.

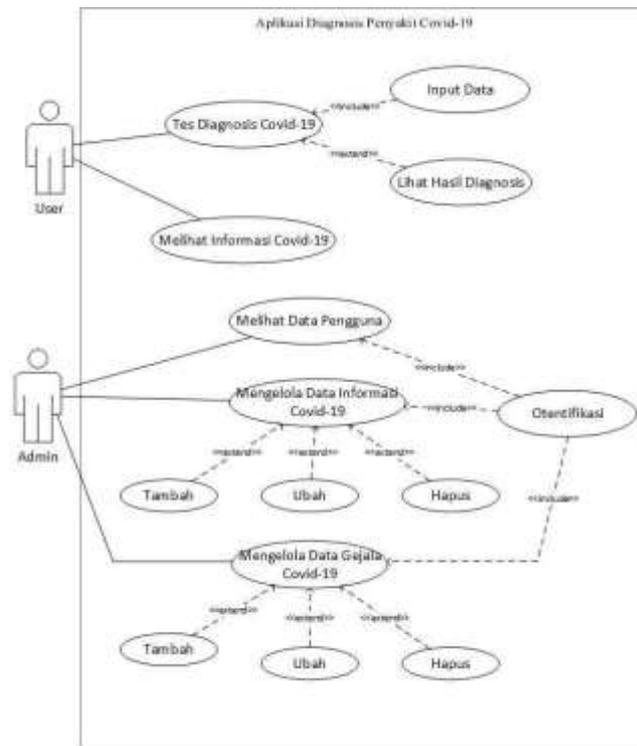
Nilai entropy dan gain pada *node* 1.1.1, diketahui bahwa atribut G6 sebagai *node* 1.1.1 karena mempunyai nilai gain paling tinggi. Data pada *node* 1.1.1, dengan data tersebut dapat dibuatkan fakta bahwa jika G6 bernilai T maka hasilnya akan N. Sedangkan G6 bernilai Y harus dicari *node* 1.1.1.2 seperti pada Gambar 4.



Gambar. 4. Model Decision Tree Node 1.1.1.2

Nilai entropy dan gain untuk mencari *node* 1.1.1.2, diketahui bahwa G7 menjadi

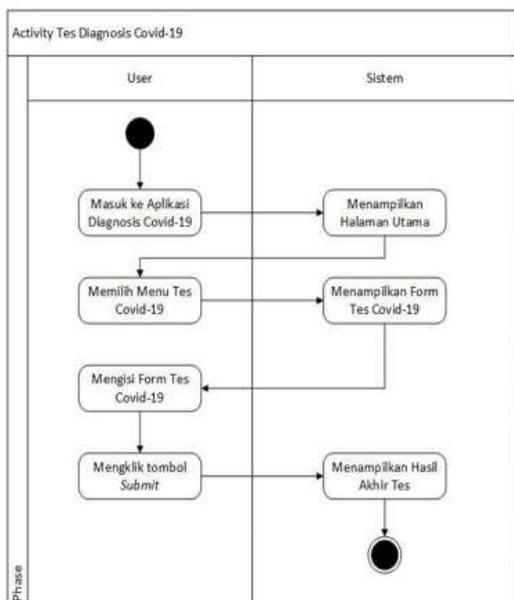
node 1.1.1.2 dengan fakta jika G7 bernilai T maka hasilnya N, sedangkan bernilai Y maka hasilnya P yang menunjukkan data dari *node* 1.1.1.2.



Gambar. 5. Use Case Diagram Aplikasi Diagnosis Penyakit

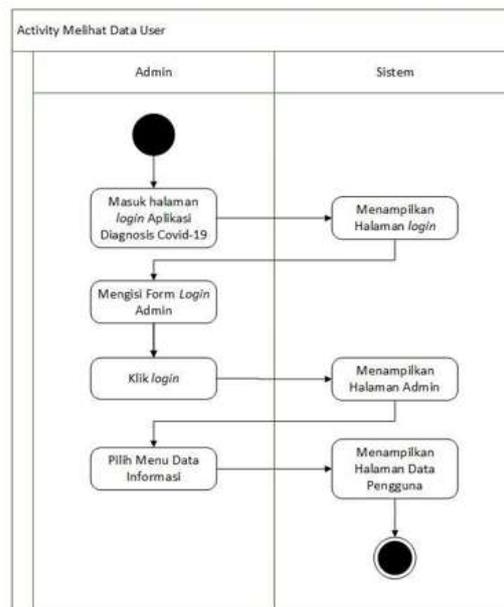
Pada Gambar 5 menjelaskan *use case* diagram pada aplikasi diagnosis penyakit Covid-19. *User* dapat melakukan tes diagnosis dan melihat informasi tentang Covid-19. Sedangkan admin setelah melakukan otentikasi, dia dapat melihat data user, mengelola informasi Covid-19, mengelola data gejala Covid19. Berikut ini adalah skenario *use case* pada aplikasi diagnosis penyakit Covid-19.

Activity diagram atau diagram aktivitas adalah tipe khusus dari diagram status yang memperlihatkan aliran dari suatu aktivitas ke aktivitas lainnya dalam suatu sistem. *Activity diagram* dalam penulisan ini menjelaskan aktivitas-aktivitas yang dapat dilakukan oleh user dan admin. Aktivitas-aktivitas tersebut di antaranya seperti pada gambar 6 sebagai berikut:



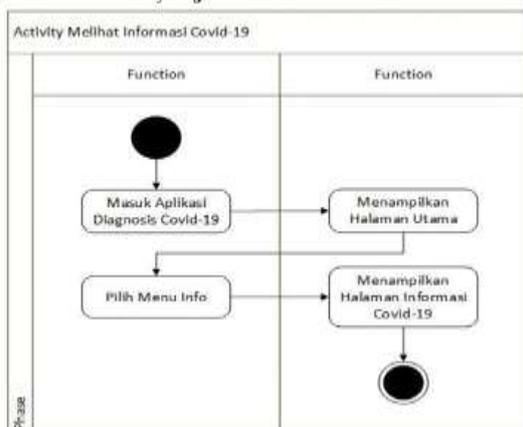
Gambar. 6. Activity Diagram Tes Covid-19

Gambar 6. pada *activity diagram* tes diagnosis Covid-19. *User* dapat melakukan tes diagnosis Covid-19 dengan menginput gejala-gejala yang dirasakan. Sistem akan mengolah hasil input dan menyimpulkan berdasarkan model *decision tree* yang sudah ada. Nantinya output yang dihasilkan adalah positif atau negatif.



Gambar. 8. Activity Diagram Melihat Data User

Pada Gambar 8 menjelaskan *activity diagram* melihat *data user* yang telah melakukan tes diagnosis dengan sistem ini. Halaman ini hanya dapat diakses oleh admin dengan melakukan login.



Gambar. 7. Activity Diagram Lihat Informasi Covid-19

Gambar 7 adalah *activity diagram* melihat informasi tentang Covid-19. *User* nantinya dapat melihat informasi tentang Covid-19 pada menu Info.

Pada penelitian ini, mengkategorikan 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.

Gejala yang banyak positif Covid-19 adalah batuk dan sakit kepala. Pada tahap pengujian sistem menggunakan 20 data uji mendapatkan hasil akurasi 90%. Berdasarkan hasil

pengujian akurasi maka sistem ini sudah berjalan fungsional dan mengeluarkan informasi yang diinginkan.

5. Kesimpulan

1. 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. Sistematisasi 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 rancangan sistematisasi yang dibangun memiliki persentase 90% dari 20 data uji berdasarkan perbandingan hasil perhitungan manual dan hitungan aplikasi sehingga dapat disimpulkan bahwa sistematisasi yang dirancang cukup baik dan berjalan secara fungsionalitas, serta dapat menampilkan informasi yang diinginkan.

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 Negri 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 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.
- Sukanto, 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.