

PENERAPAN METODE DECISION TREE PADA SISTEM DIAGNOSIS PREDIKSI PENYAKIT UMUM BERDASARKAN KELUHAN (STUDI KASUS RUMAH SAKIT ISLAM JEMURSARI SURABAYA)

Irfan Hidayat^{*1)}, Rokhmatul Insani²⁾, Berlian Rahmy Lidiawaty³⁾

1. Sistem Informasi, Fakultas Rekayasa Industri, Telkom University, Indonesia

2. Sistem Informasi, Fakultas Rekayasa Industri, Telkom University, Indonesia

3. Sistem Informasi, Fakultas Rekayasa Industri, Telkom University, Indonesia

Article Info

Kata Kunci: Data Mining; Decision Tree; Sistem Diagnosis Penyakit; Cross Validation; Confusion Matrix

Keywords: *Data Mining, Decision Tree, Disease Diagnosis System; Cross Validation; Confusion Matrix*

Article history:

Received 29 July 2024

Revised 10 September 2024

Accepted 3 October 2024

Available online 1 September 2025

DOI :

<https://doi.org/10.29100/jipi.v10i3.6266>

* Corresponding author.

Corresponding Author

E-mail address:

ihawesome99@gmail.com

ABSTRAK

Sebagian masyarakat masih menganggap remeh dan enggan untuk memeriksa kondisi kesehatan ke dokter karena berbagai macam kondisi. Kurangnya penyebaran dokter di beberapa wilayah di Indonesia juga berpengaruh terhadap pelayanan kesehatan terhadap masyarakat. Penelitian ini bertujuan untuk mengembangkan sistem diagnosis penyakit umum berdasarkan gejala berbasis web menggunakan metode klasifikasi *Decision Tree*. Penelitian ini mengambil obyek penelitian di Rumah Sakit Islam Jemursari Surabaya dengan menggunakan dataset berupa data rekam medik pasien sebanyak 9934 data. Atribut dataset yang digunakan dalam sistem ini termasuk keluhan, rps, dan diagnosa. Hasil yang didapatkan dari sistem diagnosis penyakit umum berdasarkan keluhan berbasis web dengan menggunakan algoritma CART ini mampu menghasilkan akurasi sebesar 96.61 %. Kemudian kinerja model CART dievaluasi menggunakan *cross-validation*, yang menunjukkan bahwa model tidak *overfitting*. Kesalahan per *fold* stabil, dengan rata-rata nilai error 5.69 %, menunjukkan bahwa model tidak terlalu sensitif terhadap perubahan data dan dapat bekerja dengan baik pada data yang baru. Nilai error training 2.42 % lebih rendah dari nilai error testing 0.0339, tetapi selisih yang tidak signifikan yaitu sebesar 0.97 %, menunjukkan kemampuan model untuk menggeneralisasi dengan baik ke data baru.

ABSTRACT

Some people still take it for granted and are reluctant to have their health checked by a doctor because of various conditions. The lack of distribution of doctors in several regions in Indonesia also affects health services to the community. This research aims to develop a web-based system for diagnosing common diseases based on symptoms using the Decision Tree classification method. This research took the research object at the Jemursari Islamic Hospital, Surabaya, using a dataset in the form of 9934 patient medical record data. Dataset attributes used in this system include complaints, rps, and ratings. The results obtained from the general disease diagnosis system based on web-based complaints using the CART algorithm were able to produce an accuracy of 96.61%. Then the performance of the CART model was evaluated using cross-validation, which showed that the model was not overfitting. The error per fold is stable, with an average error value of 5.69 %, indicating that the model is not too sensitive to changes in data and can work well on new data. The training error value of 2.42 % is lower than the testing error value of 3.39 %, but the difference is not significant, namely 0.97 %, indicating the model's ability to generalize well to new data.

I. PENDAHULUAN

SEBUAH gejala penyakit merupakan indikasi awal sebelum terjadinya sebuah penyakit yang dapat mengancam kondisi kesehatan [1]. Virus, bakteri, dan juga pola hidup yang kurang sehat dapat menjadi penyebab terjadinya berbagai macam penyakit ringan maupun berat [1]. Ironisnya sebagian besar masyarakat terkadang menganggap remeh tentang gejala penyakit yang dialami [2]. Masyarakat baru akan melakukan

pemeriksaan ke dokter jika penyakit tersebut sudah berada di tahap kronis atau stadium lanjut [3]. Masalah lain yang sering ditemukan adalah sebagian masyarakat masih enggan untuk memeriksakan kondisi kesehatan dan berobat ke dokter karena berbagai macam kondisi seperti biaya periksa yang mahal dan kurangnya kesadaran masyarakat terhadap pentingnya kesehatan [4]. Menurut data dari Kementerian Kesehatan Republik Indonesia tahun 2021, masih terdapat pendistribusian dokter yang tidak merata pada puskesmas di masing-masing provinsi. Terdapat 6 provinsi yang memiliki persentase puskesmas dengan ketersediaan dokter kurang dari 80% yaitu Papua, Maluku, Papua Barat, Nusa Tenggara Timur, Kalimantan Tengah, dan Sulawesi Barat [5].

Penelitian ini bertujuan untuk menyelesaikan permasalahan dalam masyarakat untuk memberikan informasi kepada masyarakat agar dapat mengetahui kemungkinan penyakit yang ada pada tubuh berdasarkan keluhan yang dirasakan. Sehingga masyarakat bisa mendapatkan perawatan lebih tepat dengan mengembangkan sistem diagnosis penyakit umum berdasarkan keluhan berbasis web. *Data mining* merupakan suatu cara yang bertujuan dalam penemuan pola dari data yang dimanfaatkan untuk menyelesaikan suatu masalah melalui berbagai aturan proses [6]. Penelitian ini menggunakan metode klasifikasi *decision tree* karena dapat mengklasifikasikan data baik kategoris ataupun numerik [7]. Metode Decision Tree dipilih karena mudah dimengerti, fleksibel, dan visualnya menarik, meskipun rentan terhadap overfitting dan sulit menentukan desain pohon optimal [8]. Neural Networks unggul dalam menangani data kompleks dan perhitungan paralel, tetapi memerlukan banyak sumber daya dan waktu pelatihan [8]. Naive Bayes mudah diterapkan, efisien, dan akurat, tetapi perkiraan kelasnya bisa kurang tepat dan membutuhkan penentuan threshold manual [8]. K-NN cepat, sederhana, dan efektif untuk data besar, namun komputasinya kompleks, memori terbatas, dan sensitif terhadap atribut tidak relevan [8]. Dengan mempertimbangkan faktor-faktor ini, Decision Tree dipilih karena keseimbangan interpretabilitas, fleksibilitas, dan efisiensinya, meskipun perlu dilakukan pemangkasan dan validasi silang untuk meningkatkan stabilitas model.

Data yang di inputkan akan diolah dengan menggunakan metode *decision tree* dengan bahasa pemrograman python dan akan memunculkan hasil diagnosis penyakit berdasarkan data keluhan, sehingga dapat menjadi sebuah wawasan terhadap masyarakat dan juga dokter terhadap pasien yang sedang diperiksa, sehingga masyarakat bisa mengetahui kemungkinan penyakit yang ada pada tubuh dengan mudah. Dengan sistem ini, diharapkan dapat membantu masyarakat agar dapat mengetahui kemungkinan penyakit berdasarkan keluhan yang dialami dan dapat mempermudah dokter dalam mendiagnosis penyakit pasien.

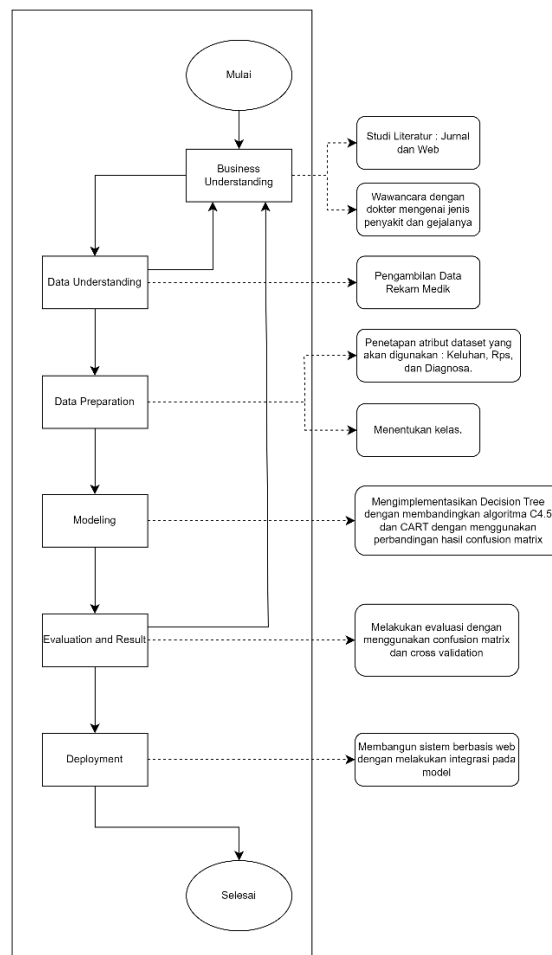
Sistem diagnosis ini dikembangkan berdasarkan data yang diperoleh dari obyek penelitian dari Rumah Sakit Islam Jemursari Surabaya. Dataset yang diperoleh berupa data rekam medik pasien pada tahun 2020 yang telah disetujui dan mendapatkan izin untuk dilakukan pengambilan data oleh Rumah Sakit Islam Jemursari. Data rekam medik merupakan data sensitif yang mengandung informasi pribadi pasien, oleh karena itu, perlindungan terhadap privasi dan kerahasiaan data menjadi prioritas utama. Data pasien yang digunakan dalam penelitian ini telah melalui proses anonimisasi, dimana semua informasi yang dapat mengidentifikasi pasien secara langsung dihapus atau dienkripsi. Hal ini dilakukan untuk mencegah identifikasi pasien dari dataset yang digunakan. Selain itu, penelitian ini mengikuti regulasi dan standar etika yang berlaku, seperti Peraturan Menteri Kesehatan Republik Indonesia mengenai perlindungan data pasien dan etika penelitian kesehatan [9]. Atribut dataset yang dapat diolah dalam sistem diagnosis ini meliputi keluhan dan diagnosis. Proses awal pengolahan data akan dilakukan dengan menghilangkan bagian yang tidak sesuai, dan pembagian antara data training dan data testing. Penelitian yang sama yang telah dilakukan dengan judul *Application of Machine Learning With The Binary Decision Tree Model In Determining The Classification of Dental Disease* [10], yang membahas tentang penggunaan metode *decision tree* untuk klasifikasi penyakit berdasarkan gejala-gejalanya. Penelitian ini menunjukkan bahwa penggunaan metode *decision tree* dapat memberikan hasil yang cukup akurat dengan nilai akurasi 92 % dalam pengklasifikasian penyakit berdasarkan gejala yang diderita oleh pasien. Berdasarkan penelitian yang berjudul Implementasi Data Mining Dalam Klasifikasi Hasil Diagnosa Pasien BPJS Menggunakan Algoritma CART [11], tentang implementasi data mining dalam klasifikasi hasil diagnosa pasien BPJS menggunakan algoritma CART menunjukkan hasil bahwa algoritma CART dapat mengklasifikasikan hasil diagnosa pasien BPJS berdasarkan gejala dengan cukup baik dengan nilai akurasi 69 %. Pada penelitian yang berjudul Comparative Analysis of C4.5 and CART Algorithms for Classification of Stroke [12], telah dilakukan penerapan jenis dari metode *decision tree* yaitu algoritma C4.5 dan CART untuk klasifikasi penyakit berdasarkan keluhan pasien. Penerapan algoritma CART mampu menyelesaikan dataset dengan berbagai tipe atribut dan menghasilkan model yang dapat digunakan untuk klasifikasi dan pengambilan keputusan secara efektif.

Hasil yang dapat diharapkan dari sistem diagnosis penyakit berdasarkan keluhan berbasis web dengan metode *decision tree* ini mampu mendapatkan tingkat akurasi yang optimal dengan menggunakan algoritma CART, sehingga dapat dijadikan sebagai sebuah sistem pendukung keputusan untuk mempermudah dokter dalam mendiagnosis penyakit pasien dan juga membantu masyarakat agar bisa mengetahui kemungkinan penyakit berdasarkan keluhan, sehingga masyarakat bisa mendapatkan perawatan lebih tepat.

II. METODE PENELITIAN

A. Alur Penelitian

Penelitian ini menggunakan metode kuantitatif data sekunder dengan data yang diperoleh dari Rumah Sakit Islam Jemursari Surabaya. Data tersebut akan dianalisis, dan untuk memperkuat informasi data yang diperoleh, dilakukan juga observasi dan wawancara. Dalam proses pengelolaan data, penelitian ini menggunakan metode *decision tree* yang dapat mengklasifikasikan data dan identifikasi data atau prediksi pada suatu kelas. Pada penelitian ini menggunakan alur pedoman metode model pada *data mining* yaitu *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai alur proses penelitian. Dapat dilihat pada gambar 1 CRISP-DM:



Gambar 1. CRISP-DM

Gambar 1 menjelaskan tentang alur penelitian dengan menggunakan model CRISP-DM. Berikut adalah penjelasan mengenai langkah-langkah prosedur penelitian dengan menggunakan model CRISP-DM [13].

1) *Business Understanding*

Tahap *business understanding* merupakan langkah awal untuk memahami tujuan dan kebutuhan dari masalah bisnis secara singkat dan akurat [14]. Pada tahap ini, pemahaman permasalahan dilakukan dengan mendalami berbagai studi literatur atau penelitian terdahulu serta melakukan wawancara dengan praktisi kesehatan atau dokter mengenai jenis-jenis penyakit umum yang sering dijumpai di masyarakat berdasarkan keluhan. Dalam penelitian ini, pendalaman masalah terkait kurangnya kesadaran masyarakat tentang kondisi kesehatan diperkuat dengan sumber referensi seperti jurnal, situs web, buku, dan lain-lain.

2) *Data Understanding*

Tahap *data understanding* melibatkan proses penggabungan data yang didapatkan dengan data yang diperlukan dalam penelitian. Penelitian ini melakukan proses pengumpulan data berdasarkan rekam medik pasien pada tahun 2020 di Rumah Sakit Islam Jemursari Surabaya. Pada proses ini memfokuskan pada pemahaman atribut dari data yang akan digunakan dalam analisis data.

3) *Data Preparation*

Tahap *data preparation* sudah mulai masuk ke dalam pengolahan data agar model yang didapatkan bisa berhasil dan sesuai dengan penelitian. Pada proses pengolahan data, penelitian ini menggunakan cara drop data untuk menghilangkan data yang tidak diperlukan atau tidak cocok dengan kebutuhan model *data mining*. Kemudian dilakukan normalisasi atau standarisasi untuk menyeragamkan bahasa yang digunakan menjadi bahasa baku atau bahasa yang digunakan dalam dunia medis. Kemudian melakukan penyeimbangan jumlah data pada setiap kelas yang terpilih sehingga mendapatkan hasil yang seimbang. Tahap selanjutnya adalah melakukan transformasi data dengan menguraikan atribut yang dimana pada awalnya atribut tersebut menjadi satu atribut kemudian diuraikan menjadi beberapa atribut.

Kriteria yang digunakan untuk menghapus data yang dianggap tidak relevan atau tidak cocok dengan model adalah sebagai berikut:

1. Data yang tidak memiliki atribut yang relevan dengan keluhan dan diagnosis pasien akan dihapus. Data yang tidak memiliki informasi ini dianggap tidak relevan dan tidak akan digunakan dalam analisis lebih lanjut.
2. Data yang menunjukkan inkonsistensi, seperti nilai yang tidak valid atau tidak masuk akal dalam atribut keluhan dan diagnosis, akan dihapus. Hal ini dilakukan untuk memastikan bahwa data yang digunakan dalam model adalah data yang valid dan dapat diandalkan.
3. Data yang memiliki banyak missing values atau data yang tidak lengkap akan dihapus.

Keputusan penghapusan data dilakukan melalui analisis data awal untuk mengidentifikasi data yang tidak relevan, diskusi dengan tim peneliti untuk validasi, dan implementasi dalam model setelah penghapusan data yang tidak relevan.

4) *Modeling*

Tahap *modeling* merupakan tahap eksekusi pembuatan model *data mining* berdasarkan hasil dari proses *data preparation*. Dalam penelitian ini, tahap pemodelan difokuskan terhadap pengelolaan data rekam medik dengan beberapa atribut yang dapat dipertimbangkan berdasarkan hasil analisis data yaitu data keluhan dan diagnosis penyakit. Penelitian ini memilih kelas berdasarkan atribut diagnosis dari berbagai jenis penyakit. Penelitian ini menerapkan metode *decision tree* dengan algoritma *Classification and Regression Trees* (CART) [10]. Algoritma CART adalah salah satu metode *decision tree* yang digunakan untuk membuat model klasifikasi dan regresi [10]. Cara kerja dasar algoritma ini melibatkan pembagian data menjadi subset-subset berdasarkan nilai atribut yang memaksimalkan pemisahan kelas target. Pada setiap node dalam tree, CART memilih atribut yang memberikan informasi terbesar untuk memisahkan data ke dalam kategori yang berbeda. Algoritma CART memiliki beberapa kelebihan dibandingkan dengan algoritma lain pada *decision tree* seperti algoritma ini mampu menangani data numerik dan kategoris dengan baik serta menghasilkan model yang mudah diinterpretasikan dan algoritma CART memiliki mekanisme pruning yang efektif untuk menghindari overfitting yang merupakan masalah umum pada model *decision tree*. Beberapa pertimbangan khusus penggunaan algoritma CART dalam penelitian ini yaitu data rekam medis yang digunakan dalam penelitian ini memiliki atribut yang jelas dan terstruktur yang cocok untuk analisis menggunakan algoritma CART dan hasil dari algoritma CART menghasilkan pohon keputusan yang mudah dipahami dan diinterpretasikan oleh praktisi medis, sehingga memudahkan dalam pengambilan keputusan klinis berdasarkan hasil model.

5) *Evaluation*

Tahap *evaluation and result* merupakan tahap dimana memastikan kesesuaian hasil model dengan tujuan yang ingin dicapai pada penelitian. Setelah memperoleh hasil dari perbandingan model, perlu dilakukan evaluasi agar dapat mengetahui kualitas dan efektivitas dari model tersebut. Pada penelitian ini, untuk memvalidasi hasil dari sebuah model yang telah diolah dilakukan dengan menggunakan *confusion matrix*. Nilai yang dihasilkan melalui *confusion matrix* adalah berupa evaluasi nilai akurasi, *precision*, *recall*, dan *f1-score* [14]. *Confusion matrix* dapat dilihat pada tabel I.

TABEL I
CONFUSION MATRIX

		Nilai Sebenarnya	
		True	False
Nilai Prediksi	True	TP (True Positive)	FN (False Negative)
	False	FP (False Positive)	TN (True Negative)

Keterangan :

- a) *TP (True Positive)* = Data Aktual dan data prediksi bersifat positif
- b) *TN (True Negative)* = Data Aktual dan data prediksi bersifat negatif
- c) *FP (False Positive)* = Data Aktual bersifat positif, namun data prediksi diklasifikasikan positif
- d) *FN (False Negative)* = Data Aktual bersifat positif, namun data prediksi diklasifikasikan negative

Nilai yang dihasilkan melalui confusion matrix adalah berupa evaluasi nilai akurasi, *precision*, *recall*, dan *f1-score*. Nilai tersebut dapat diperoleh melalui perhitungan:

$$accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (1)$$

$$precision = \frac{TP}{TP+FP} \quad (2)$$

$$recall = \frac{TP}{TP+TFN} \quad (3)$$

$$F1-score = \frac{2 \times precision \times recall}{precision+recall} \quad (4)$$

Pada tahap *evaluation* juga dilakukan evaluasi dengan menggunakan *cross validation*. Tujuan dari evaluasi dengan menggunakan *cross validation* ini untuk melihat apakah model optimal dan tidak mengalami *overfitting* [15]. Teknik ini melibatkan pembagian data training menjadi beberapa *fold* dan melakukan training serta evaluasi secara bergantian pada setiap *fold* [16]. Dalam penelitian ini menggunakan *cross validation* dengan 10 folds. Pada setiap iterasi, satu *fold* digunakan sebagai data uji sementara 9 *fold* lainnya digunakan sebagai data latih. Proses ini diulangi sebanyak 10 kali, dengan setiap *fold* digunakan sekali sebagai data uji. *Cross validation* berkontribusi pada penilaian model dengan cara mengurangi bias dan varians. Dengan menggunakan beberapa *fold* sebagai data uji, metode ini membantu mengurangi bias yang mungkin terjadi jika menggunakan satu set data uji. Selain itu, *cross validation* membantu dalam mengurangi varians hasil model dengan memastikan bahwa model diuji pada berbagai sub-set data yang berbeda. Hasil rata-rata dari 10 iterasi memberikan estimasi yang lebih akurat tentang hasil kinerja model pada data yang tidak terlihat.

6) *Deployment*

Pada tahap *deployment* merupakan tahapan dimana hasil dari model *decision tree* yang telah dibuat kemudian diimplementasikan dan diintegrasikan ke dalam sistem berbasis website. Cara kerja dari sistem ini yaitu masyarakat atau tenaga medis mengisi kan data diri dan juga keluhan, kemudian sistem yang telah terintegrasi dengan model *decision tree* akan mengeluarkan hasil diagnosa berupa kemungkinan penyakit yang diderita.

B. Alat dan Bahan Penelitian

Penelitian ini menggunakan beberapa alat dalam dan bahan untuk mendukung proses analisis data. Alat dan bahan tersebut berperan penting dalam memastikan penelitian dapat dilaksanakan dengan baik dan menghasilkan data yang akurat serta valid.

1) *Google Colab*

Pada penelitian ini sebagai alat untuk mengolah data dan pembuatan model menggunakan Google Colab. Google Colab sendiri memberikan akses dan penggunaan yang mudah yang dapat membantu dalam pembuatan model yang pada penelitian ini dengan mudah [17].

2) *Visual Studio Code*

Pada penelitian ini menggunakan Visual Studio Code dalam proses pengembangan sistem diagnosis penyakit umum berdasarkan keluhan. Alat ini digunakan karena berbagai macam kemudahan dan juga fitur-fitur yang diberikan oleh Visual Studio Code dapat sangat membantu dalam proses pengembangan sistem [18].

3) *Python*

Pada penelitian ini menggunakan bahasa pemrograman Python untuk membangun dan mengevaluasi performa model. Proses ini menggunakan library yang telah disediakan oleh bahasa pemrograman Python. Library yang digunakan pada penelitian dijelaskan pada Tabel II [19]:

TABEL II
LIBRARY PYTHON

Library	Fungsi
Pandas	Untuk memanipulasi dan analisis data dalam bentuk dataframe, seperti membaca file csv, menampilkan data, dan memeriksa nilai yang hilang.
Matplotlib	Untuk membuat visualisasi grafik, seperti plotting table dan menampilkan distribusi data
Sklearn	Menyediakan algoritma data mining seperti <code>train_test_split</code> , <code>cross_val_score</code> , <code>StratifiedKfold</code> , <code>GridsearchCV</code> , <code>DecisionTreeClassifier</code>
Seaborn	Untuk membuat visualisasi data statistik yang lebih menarik dan informatif berdasarkan visual
Numpy	Memberikan sebuah fungsi perhitungan numeric, dan membangun struktur data array multi dimensi
Math	Untuk menghitung rumus matematika
Joblib	Untuk menyimpan model (<i>serialize</i>) dan memuat (<i>deserialize</i>) model machine learning

III. HASIL DAN PEMBAHASAN

A. Business Understanding

Penerapan data mining pada penelitian ini berhubungan dengan beberapa permasalahan yang terjadi seperti kurangnya kesadaran masyarakat tentang kondisi kesehatan, termasuk berbagai macam penyakit umum, serta distribusi dokter yang tidak merata di berbagai wilayah di Indonesia yang berdampak pada kualitas pelayanan kesehatan masyarakat.

B. Data Understanding

Data yang digunakan merupakan hasil rekam medik pasien yang berasal dari Rumah Sakit Islam Jemursari Surabaya dengan total 9934 data. Atribut yang digunakan seperti keluhan (ku), ringkasan pengaduan sakit (RPS), dan hasil diagnosa penyakit (Diagnosis). Dataset yang digunakan dalam penelitian ini dapat dilihat pada tabel III.

TABEL III
DATASET AWAL

ku	RPS	Diagnosis
batuk berdahak,kering	-	ispa
NYERI ULU HATI	MUAL MUNTAH	GASTRITIS + COLIC ABD
DEMAM,LEMAS	MUAL MUNTAH MUNTAH TERUS, DIARE	DENGUE FEVER + GASTRITIS + VOMITING + KU LEMAH
BATUK NGEKEL	TENGGOROKAN SAKIT, DEMAM	ISPA
kontrol	demam masih, diare masih, batuk juga	gea non dehidrasi; ispa
kontrol	demam -, badan lemes, perut sakit	dhf grade 1

C. Data Preparation

Data preparation adalah tahap kritis dalam analisis data yang melibatkan beberapa langkah *pre-processing* sehingga dataset benar-benar bersih. Dalam penelitian ini, penghapusan data yang tidak lengkap atau tidak relevan dilakukan untuk memastikan keandalan dan validitas data yang digunakan dalam model. Proses ini melibatkan beberapa langkah penting seperti menghapus data dengan atribut yang tidak relevan, data yang menunjukkan inkonsistensi, dan data dengan banyak nilai yang hilang.

1) Data Cleaning

Dalam tahap *data cleaning*, proses pertama yang dilakukan adalah mengidentifikasi dan menghapus data dalam kolom 'ku' yang kosong untuk meningkatkan kualitas dataset. Setelah dilakukan proses *data cleaning* dengan menghapus data dalam kolom 'ku' yang kosong jumlah data menjadi sebesar 9758 data. Hasil dari penghapusan data dalam kolom 'ku' yang kosong dapat dilihat pada tabel IV.

TABEL IV
DROP VALUE KOSONG

ku	RPS	Diagnosis
batuk berdahak,kering	-	ispa
NYERI ULU HATI	MUAL MUNTAH	GASTRITIS + COLIC ABD

DEMAM.LEMAS	MUAL MUNTAH MUNTAH TERUS, DIARE	DENGUE FEVER + GASTRITIS + VOMITING + KU LEMAH
BATUK NGEKEL kontrol	TENGGOROKAN SAKIT, DEMAM demam masih, diare masih, batuk juga	ISPA gea non dehidrasi; ispa

Proses kedua adalah menghapus data pada kolom ‘ku’ dengan value hanya ‘Kontrol’ untuk meningkatkan kualitas dataset. Dalam tahapan ini total jumlah dataset menjadi sebesar 4919 data. Hasil dari penghapusan data dalam kolom ‘ku’ dengan value hanya ‘Kontrol’ dapat dilihat pada tabel V.

TABEL V
DROP VALUE HANYA KONTROL

ku	RPS	Diagnosis
batuk berdahak,kering	-	ispa
NYERI ULU HATI	MUAL MUNTAH	GASTRITIS + COLIC ABD
DEMAM.LEMAS	MUAL MUNTAH MUNTAH TERUS, DIARE	DENGUE FEVER + GASTRITIS + VOMITING + KU LEMAH
BATUK NGEKEL	TENGGOROKAN SAKIT, DEMAM	ISPA
BATUK KERING	-	ISPA

Proses ketiga melakukan penghapusan data pada kolom ‘assesment’ yang memiliki lebih dari satu diagnosis untuk meningkatkan spesifikasi analisis dataset. Proses ini dilakukan dengan memilih data yang hanya memiliki satu diagnosis untuk memastikan konsistensi data. Diagnosis yang menunjukkan lebih dari satu hasil dalam kolom ‘assessment’ dihapus untuk memastikan bahwa setiap entri data memiliki satu diagnosis yang konsisten. Hal ini dilakukan untuk menghindari ambiguitas dan memastikan keandalan data dalam analisis lebih lanjut. Dalam tahapan ini total jumlah dataset menjadi sebesar 3335 data. Hasil dari penghapusan data dalam kolom ‘assesment’ yang memiliki lebih dari satu diagnosis dapat dilihat pada tabel VI.

TABEL VI
DROP VALUE ASSESMENT GANDA

ku	RPS	Diagnosis
batuk berdahak,kering	-	ispa
BATUK KERING	-	ISPA
BATUK NGEKEL	TENGGOROKAN SAKIT, DEMAM	ISPA
PANAS MULAI KEMARIN, MUAL	-	OBS FEBRIS
BATUK KERING	-	ISPA

Proses selanjutnya adalah menggabungkan kolom ‘rps’ dengan kolom ‘ku’ menjadi kolom ‘gejala’. Hal ini karena kedua kolom tersebut merupakan kolom yang sama-sama menjelaskan keluhan dan kondisi kesehatan pasien. Hasil dari penggabungan kolom ‘rps’ dengan kolom ‘ku’ menjadi kolom ‘gejala’ dapat dilihat pada tabel VII.

TABEL VII
MENGGABUNGKAN KOLOM

Gejala	Diagnosis
batuk berdahak,kering	Isa
BATUK KERING	ISPA
BATUK NGEKEL. TENGGOROKAN SAKIT, DEMAM	ISPA
PANAS MULAI KEMARIN, MUAL	OBS FEBRIS
BATUK KERING	ISPA

Terakhir, proses ketiga yaitu menghapus data dalam kelas dengan frekuensi kurang dari 30 untuk memastikan setiap kelas memiliki representasi yang memadai dalam dataset. Penghapusan data tertentu memang dapat mempengaruhi distribusi kelas dalam dataset. Hal ini penting untuk menghindari bias yang mungkin timbul akibat ketidakseimbangan data. Hasil akhir dari tahap Data Cleaning menghasilkan jumlah dataset baru yaitu menjadi sebanyak 1180 data dengan 5 kelas penyakit yaitu Dermatitis, Gastritis, ISK, ISPA, dan Typhoid Fever.

2) Normalisasi Bahasa

Proses normalisasi dalam konteks ini mengacu pada pengidentifikasian dan menormalisasikan bahasa daerah atau bahasa non-formal ke dalam bahasa Indonesia atau bahasa medis yang memudahkan dalam pemilihan keluhan yang tepat untuk diagnosis medis. Hasil dari normalisasi data dapat dilihat pada tabel VIII.

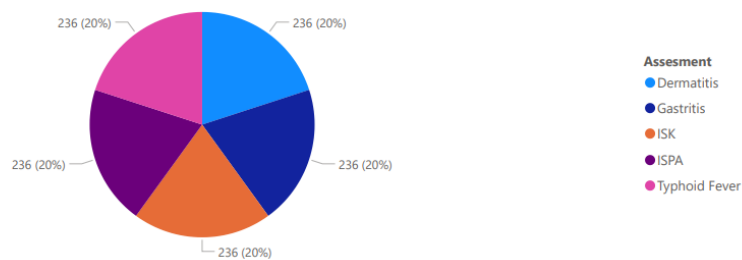
TABEL VIII
MENGABUNGKAN KOLOM

Kata Non Medis	Kata Medis
Nggreges	Demam
Badan Pegal-pega;	Badan Sakit
Nggliyeng	Pusing
Kepala Puyeng	Pusing
Dada Terasa Ampek	Sesak
Mencret	Diare

3) Random Over Under Sampling

Terdapat beberapa metode yang dapat digunakan untuk mengatasi masalah ketidak seimbangan data dalam kelas salah satunya adalah metode *over-sampling* dan *under-sampling* [20]. Pada tahap selanjutnya adalah menerapkan metode *Random Over-Under Sampling (ROUS)* kedalam dataset dengan total data keseluruhan dari kelima label yakni 1180. Kemudian mengambil nilai rata-rata dari kelima label yaitu menjadi 236 data untuk setiap label. Hasil penerapan Random Over Under Sampling dapat dilihat pada gambar 2.

Pie Chart by Assesment



Gambar 2. Dataset ROUS

Penggunaan metode *Random Over-Under Sampling* dalam penelitian ini dilakukan untuk menangani ketidakseimbangan data dan memastikan bahwa model dapat membuat prediksi yang akurat untuk setiap kelas penyakit. Dengan menyamakan jumlah data untuk setiap kelas menjadi 236 data per label, bias yang disebabkan oleh ketidakseimbangan data dapat diatasi dan performa model secara keseluruhan dapat ditingkatkan. Metode ini membantu dalam menciptakan dataset yang seimbang, yang penting untuk menghasilkan model machine learning yang andal dan efektif.

4) Transformasi Data

Pada tahap transformasi dataset, bertujuan untuk merubah kolom 'ku' agar menjadi atribut yang efektif untuk model *Decision Tree* dengan mempertimbangkan probabilitas munculnya keluhan tertentu terkait klasifikasi label yang diinginkan. Semula data keluhan menjadi satu dalam kolom 'ku' kemudian data dipecah menjadi beberapa atribut sehingga dataset memiliki beberapa kolom berdasarkan keluhan yang telah dipecah. Transformasi data dilakukan untuk meningkatkan kualitas, interpretabilitas, dan struktur data. Pemecahan kolom 'ku' menjadi beberapa atribut membantu dalam mengidentifikasi pola yang lebih spesifik dan relevan. Dengan data yang lebih terstruktur, model dapat membuat keputusan yang lebih akurat. Hasil dari transformasi data dapat dilihat pada tabel IX.

TABEL IX
TRANSFORMASI DATA

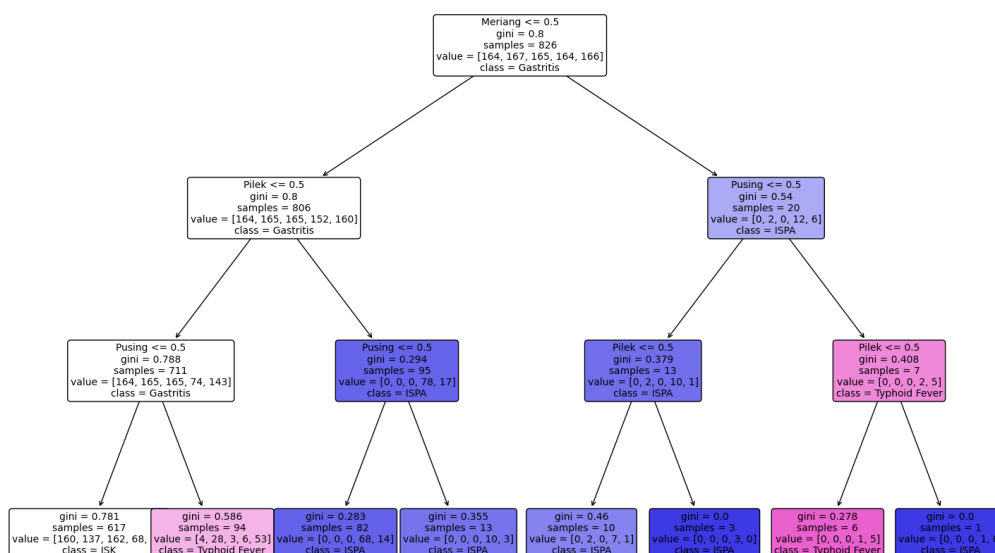
Diagnosis	Kulit Gatal-gatal	Muncul Bintik Merah	Kulit Bengkak	Pusing
Dermatitis	1	0	0	0
Dermatitis	1	0	0	0
Dermatitis	1	0	0	0
Dermatitis	1	1	0	0
Dermatitis	0	0	1	0

Penghapusan data yang tidak relevan juga berdampak pada hasil model. Dengan menggunakan data yang lebih bersih dan representatif, model Decision Tree yang digunakan dalam penelitian ini dapat membuat prediksi yang lebih akurat. Proses normalisasi dan transformasi data juga membantu dalam meningkatkan interpretabilitas dan struktur data, sehingga model dapat membuat keputusan yang lebih tepat. Sebagai hasilnya, model Decision Tree dengan algoritma CART mampu menangani data dengan lebih baik dan menghasilkan pohon keputusan yang mudah dipahami oleh praktisi medis .

D. Modeling

Pada tahap ini, model klasifikasi yang dipilih mengidentifikasi pola informasi yang berguna bagi masyarakat dan dokter. Pola klasifikasi yang dihasilkan oleh teknik data mining ini bertujuan untuk memprediksi penyakit umum berdasarkan gejala yang teridentifikasi. Salah satu algoritma yang digunakan adalah *Classification and Regression Tree* (CART), yang dikenal sebagai model klasifikasi dengan struktur pohon keputusan. Kelebihan dari algoritma CART adalah algoritma tersebut dapat dengan mudah menangani outliers [21]. Prosesnya melibatkan transformasi kumpulan data menjadi pohon keputusan dengan menetapkan serangkaian aturan pada setiap cabang, sehingga memperlancar dan meningkatkan efisiensi pengambilan keputusan yang melibatkan banyak atribut yang kompleks [22]. Struktur pohon keputusan ini terdiri dari root node, internal node, dan leaf node. Hasil pengolahan data pada algoritma CART dapat dilihat pada gambar 3.

Decision Tree With Selected Features (Pusing, Pilek, Meriang)

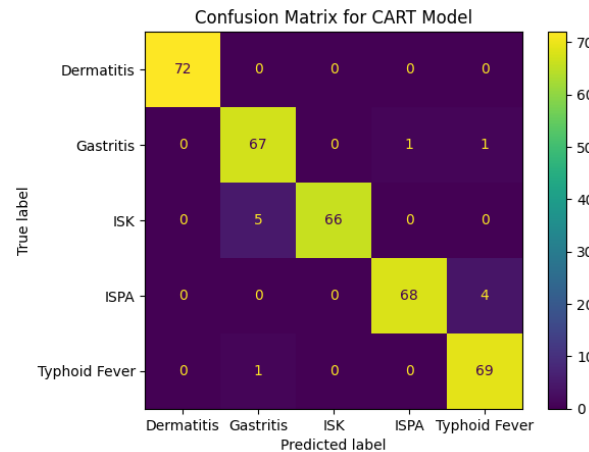


Gambar 3. Classification Tree dalam prediksi penyakit umum

Hasil ini merupakan visualisasi pengolahan data pada algoritma CART. Pohon keputusan dimulai dengan fitur 'Meriang' sebagai *root node*. Pada *node* ini, jika 'Meriang' memiliki nilai lebih besar dari 0.5, maka akan menuju ke cabang kanan. Dalam kasus ini, nilai 'Meriang' adalah 1 atau lebih besar dari 0.5, sehingga perhitungan akan dilanjutkan ke cabang kanan. Pada cabang kanan, fitur selanjutnya yang dievaluasi adalah 'Pusing'. Jika 'Pusing' memiliki nilai lebih besar dari 0.5, maka perhitungan akan dilanjutkan ke cabang kanan. Dalam kasus ini nilai 'Pusing' adalah 1 atau lebih besar dari 0.5, sehingga perhitungan akan dilanjutkan ke cabang kanan. Pada *node* ketiga, fitur selanjutnya yang dievaluasi adalah 'Pilek'. Jika 'Pilek' memiliki nilai lebih besar dari 0.5, maka perhitungan akan dilanjutkan ke cabang kanan. Dalam kasus ini nilai 'Pilek' adalah 1 atau lebih besar dari 0.5, sehingga perhitungan akan dilanjutkan ke cabang kanan. Berdasarkan tahapan *node* tersebut *leaf node* menghasilkan prediksi yaitu kelas 'ISPA'. Pohon keputusan ini menunjukkan bahwa dengan input 'Pusing', 'Pilek', dan 'Meriang' dimana masing-masing memiliki nilai 1, model menghasilkan prediksi yaitu 'ISPA'.

E. Evaluation

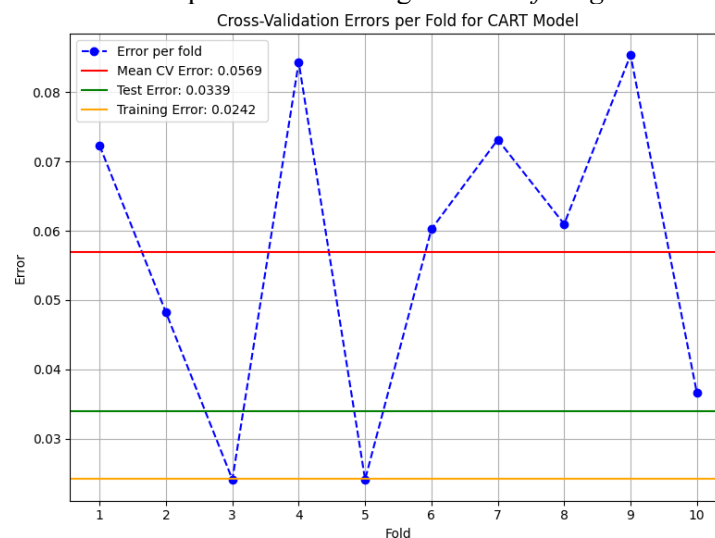
Setelah memperoleh pola klasifikasi dari algoritma CART, langkah selanjutnya adalah melakukan evaluasi komparatif terhadap algoritma ini. Evaluasi dilakukan menggunakan *Confusion Matrix*, yang memberikan informasi mengenai perbandingan antara hasil prediksi model dan hasil klasifikasi sebenarnya. Proses evaluasi ini meliputi pengukuran nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Hasil evaluasi model yang telah dibangun dengan menggunakan algoritma CART dapat dilihat pada gambar 4.



Gambar 4. Confusion Matrix pada algoritma CART

Berdasarkan gambar 4 model algoritma CART menunjukkan hasil yang bagus dengan menunjukkan hasil *accuracy* yang tinggi sebesar 96.61 %, *precision* 96.76 %, *recall* 96.61 %, *f1-score* 96,62 % dengan menggunakan test size = 0.3 dan random state = 90.

Pada tahap *evaluation* juga dilakukan evaluasi dengan menggunakan cross validation dengan tujuan untuk mengvaluasi kualitas model dan melihat apakah model mengalami *overfitting*.



Gambar 5. Cross Validation pada algoritma CART

Dapat dilihat pada gambar 5 nilai error per *fold* menunjukkan performa model pada setiap subset data yang digunakan untuk validasi. Jika model mengalami *overfitting*, nilai error per *fold* akan tinggi pada subset data yang tidak digunakan untuk training dan rendah pada subset data yang digunakan untuk training. Nilai error per *fold* relatif stabil pada semua subset data, dengan rata-rata error 5.69 %. Hal ini menunjukkan bahwa model tidak terlalu sensitif terhadap perubahan data dan dapat bekerja dengan baik pada data yang baru baru. Nilai error training yang lebih rendah daripada nilai error test. Jika model mengalami *overfitting*, nilai error training akan jauh lebih rendah daripada nilai error test. Pada gambar 5, nilai error training adalah 2.42 % dimana nilai tersebut lebih rendah daripada nilai error test yaitu 3.39 %, tetapi selisihnya yang tidak terlalu signifikan dengan selisih 0.97 %. Berdasarkan hal tersebut menunjukkan bahwa model tidak mengalami *overfitting*. Model juga tidak terlalu fokus pada data training dan dapat menggeneralisasi dengan baik pada data baru.

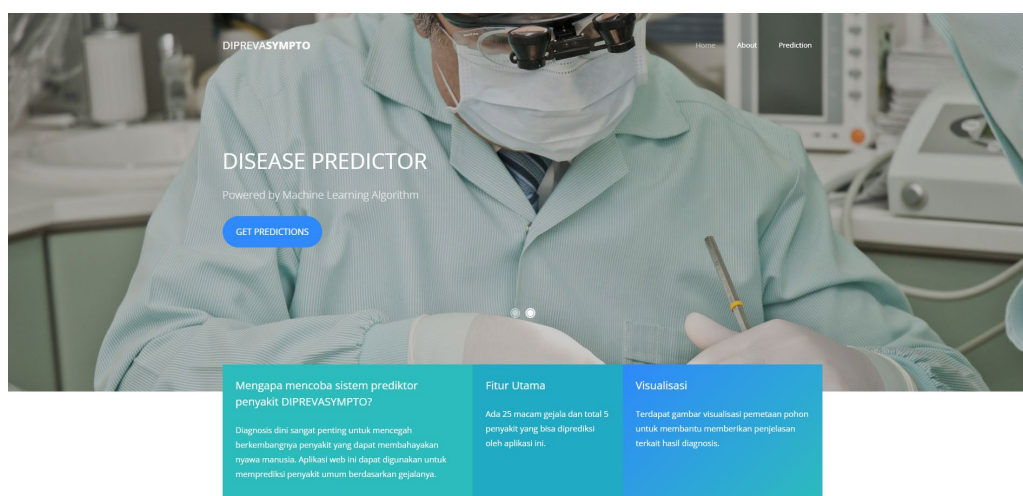
Untuk memberikan gambaran yang lebih komprehensif mengenai hasil penelitian ini, hasil penelitian ini

dibandingkan dengan beberapa penelitian terkait yang telah dilakukan sebelumnya. Penelitian serupa yang berjudul *Application of Machine Learning With The Binary Decision Tree Model In Determining The Classification of Dental Disease* [10] mengevaluasi penggunaan metode decision tree untuk klasifikasi penyakit berdasarkan gejala-gejala yang dialami pasien. Penelitian tersebut mencapai akurasi sebesar 92%. Jika dibandingkan, penelitian ini menunjukkan bahwa algoritma CART dapat mencapai akurasi yang lebih tinggi sebesar 96.61%, menunjukkan peningkatan sebesar 4.61%. Selain itu, penelitian yang telah dilakukan dengan judul *Implementasi Data Mining Dalam Klasifikasi Hasil Diagnosa Pasien BPJS Menggunakan Algoritma CART* [11] yang membahas implementasi data mining dalam klasifikasi hasil diagnosa pasien BPJS menggunakan algoritma CART mencatat akurasi sebesar 69%. Performa yang dicapai dalam penelitian ini jauh lebih tinggi, menekankan bahwa dengan data dan metode yang tepat, algoritma CART dapat memberikan hasil klasifikasi yang sangat baik. Kedua perbandingan ini menegaskan keefektifan model yang digunakan dalam penelitian ini dan kontribusinya dalam meningkatkan kualitas klasifikasi penyakit berdasarkan gejala.

Berdasarkan penelitian yang telah dilakukan, ditemukan beberapa kelebihan dan keterbatasan dari algoritma CART. Kelebihan algoritma CART meliputi kemudahan interpretasi yang membantu praktisi medis dalam memahami hasil diagnosis, serta fleksibilitas dalam menangani data numerik dan kategoris yang cocok untuk dataset rekam medis pasien. Efektivitas pruning dalam menghindari overfitting juga merupakan salah satu kelebihan penting. Namun, terdapat juga beberapa keterbatasan, seperti rentan terhadap overfitting jika tidak diatur dengan baik, pemilihan atribut yang tidak selalu optimal, dan kesulitan dalam menangani data yang sangat tidak seimbang. Oleh karena itu, penting untuk melakukan validasi model, seperti cross-validation, dan melakukan pruning yang tepat untuk menghindari kompleksitas berlebihan.

F. Deployment

Dalam pembangunan website ini digunakan bahasa pemrograman HTML, CSS, Js, dan Python. Website ini juga menggunakan joblib, dan jinja2 yang digunakan untuk menghubungkan website dengan model yang sudah ada. Website ini memiliki dua bagian halaman utama yaitu home dan prediction. Halaman home merupakan halaman utama yang berisikan informasi tentang website tersebut. Kemudian halaman prediction merupakan halaman untuk diagnosis penyakit dimana dalam halaman ini berisikan form nama, umur, dan juga form berupa *checkbox* yang dapat diisi oleh pasien sesuai dengan keluhan-keluhan yang dirasakan. Setelah pasien mengisi data-data tersebut kemudian halaman prediction akan mengolah data tersebut dengan menggunakan model *decision tree* yang sudah dihubungkan dengan menggunakan joblib sebagai penghubung antar model dan website. Halaman prediction juga menampilkan nilai probabilitas untuk hasil diagnosa penyakit berdasarkan hasil perhitungan model *decision tree*. Setelah hasil diagnosa keluar, pasien juga bisa membuat medical report berupa file pdf berdasarkan hasil diagnosis penyakit dan data-data yang sudah diinputkan di awal. Hasil akhir pada penelitian ini ditampilkan dalam bentuk website yang dapat dilihat pada gambar 6 dan gambar 7.



Gambar 6. Halaman Home

DIPREVASYMPTO
Home About Prediction

Irfan
Age
21

☐ Kulit Gatal-gatal
☐ Kencing Tidak Lancar
☐ Pilek
☐ Nyeri Ulu Hati
☐ Tenggorokan Sakit
☐ Diare
☐ Sesak
☒ Batuk
☐ Nyeri Daerah Panggul
☐ Perut Sakit
☐ Pusing
☐ Muntah
☐ Perut Kembung
☒ Demam
☐ Muncul Bintik Merah
☐ Badan Lemas
☐ Tenggorokan Gatal
☐ Nyeri Telan
☐ Badan Sakit
☐ Kencing Sakit
☐ Mual
☐ Meriang
☐ Kulit, bengkak
☐ Sendawa
☐ Lidah Pahit

Predict Export PDF

Prediction: ISPA

Description: Infeksi Saluran Pernapasan Akut adalah infeksi yang mempengaruhi saluran pernapasan bagian atas, termasuk hidung, tenggorokan, dan paru-paru.

Treatment: Istirahat, minum banyak cairan, dan obat pereda nyeri.

Probabilities:

Dermatitis: 0.00%
Gastritis: 0.00%
ISK: 0.00%

Gambar 7. Halaman Prediction

IV. KESIMPULAN

Dari hasil penelitian didapatkan berdasarkan hasil evaluasi dengan menggunakan *confusion matrix* menunjukkan bahwa algoritma CART memiliki performa yang baik pada evaluasi *confusion matrix* pada nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Pada model CART memiliki hasil nilai *accuracy* 96.61 %, *precision* 96.76 %, *recall* 96.61 %, dan *f1-score* 96,62 %. Model decision tree dengan algoritma CART menunjukkan performa yang baik dalam hal stabilitas dan generalisasi terhadap data baru. Berdasarkan hasil evaluasi dengan menggunakan *cross validation* menunjukkan bahwa model CART tidak mengalami *overfitting*, dengan nilai error per *fold* yang stabil dengan nilai rata-rata error 5.69 % dan selisih yang tidak terlalu signifikan antara nilai error training 2.42 % dan error test 3.39 % dengan nilai selisih sebesar 0.97 %. Hal ini menunjukkan bahwa model CART memiliki kemampuan yang baik dalam memprediksi data baru dengan performa yang konsisten.

DAFTAR PUSTAKA

- [1] Y. Yanuardi, "Rancang Bangun Aplikasi Diagnosa Penyakit Umum Berbasis Android Pada Klinik Citra Raya Medika," *JIKA (Jurnal Inform.*, vol. 3, no. 1, pp. 9–17, 2019, doi: 10.31000/jika.v3i1.2035.
- [2] S. Sitorus *et al.*, *Ilmu Kesehatan Masyarakat*. 2020.
- [3] R. F. F. Erika, "Meningkatkan Kesadaran Masyarakat Dalam Pemeriksaan Kesehatan Melalui Metode Penyuluhan Ceramah Di Desa Rambung Sialang Tengah," *Jukeshum (Jurnal Pengabd. Masyarakat)*, vol. 2, no. 2, pp. 170–178, 2022.
- [4] Puji Wulansari, *POLA PEMANFAATAN JAMKESMAS PADA MASYARAKAT KELURAHAN BUGANGAN KECAMATAN SEMARANG TIMUR*. 2013.
- [5] Kemenkes RI, *Profil Kesehatan Indonesia 2021*. 2022.
- [6] L. W. S. Amna, Wahyuddin S, I Gede Iwan Sudipa, Tri Andi E. Putra, Ahmad Jurnaidi Wahidin, Wara Alfa Syukrilla, Anindya Khrisna Wardhani, Nono Heryana, Tutuk Indriyani, *Data Mining*, vol. 2, no. January 2013. 2005. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [7] F. S. N. Khamidah, D. P. Hapsari, and H. Nugroho, "Implementasi Fuzzy Decision Tree Untuk Prediksi Gagal Ginjal Kronis," *INTEGER J. Inf. Technol.*, vol. 3, no. 1, pp. 19–28, 2018, doi: 10.31284/j.integer.2018.v3i1.155.
- [8] A. P. Ayudhitama and Utomo Pujiyanto, "Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer," *J. Inform. Polinema*, vol. 6, no. 2, pp. 1–9, 2020, doi: 10.33795/jip.v6i2.274.
- [9] L. T. Handayani, *Pedoman Dan Standar Etik Penelitian Dan Pengembangan Kesehatan Nasional*, vol. 10, no. 1. 2018. [Online]. Available: www.litbang.kemkes.go.id
- [10] M. Ula, F. T. T. Anjani, A. F. Ulva, I. Sahputra, and A. Pratama, "Application of Machine Learning With the Binary Decision Tree Model in Determining the Classification of Dental Disease," *J. Informatics Telecommun. Eng.*, vol. 6, no. 1, pp. 170–179, 2022, doi: 10.31289/jite.v6i1.7341.
- [11] N. Tou and P. M. Endraswari, "Implementasi Data Mining Dalam Klasifikasi Hasil Diagnosa Pasien Bpjs Menggunakan Algoritma Cart," *JIKA (Jurnal Inform.*, vol. 6, no. 2, p. 170, 2022, doi: 10.31000/jika.v6i2.6164.
- [12] Suryani, D. Rahmadani, A. A. Muzafar, A. Hamid, R. Annisa, and Mustakim, "Analisis Perbandingan Algoritma C4.5 dan CART untuk Klasifikasi Penyakit Stroke," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 197–206, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [13] N. W. Wardani *et al.*, "Penerapan Data Mining Untuk Klasifikasi Penjualan Barang Terlaris Menggunakan Metode Decision Tree C4.5," *J. Teknol. Inf. dan Komput.*, no. 2, pp. 268–279, 2022.
- [14] W. I. Rahayu, C. Prianto, and E. A. Novia, "Perbandingan Algoritma K-Means dan Naive Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan pada PT. Pertamina (Persero)," *J. Tek. Inform.*, vol. 13, no. 2, pp. 1–8, 2021.
- [15] W. A. Firmansyah, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [16] S. E. Herni Yulianti, Oni Soesanto, and Yuana Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," *J. Math. Theory Appl.*, vol. 4, no. 1, pp. 21–26, 2022, doi: 10.31605/jomta.v4i1.1792.
- [17] R. O. Felani, "Analisis Prilaku Pengguna e-learning menggunakan Algoritma K-Means Clustering," *Jusikom J. Sist. Komput. Musirawas*, vol. 7,

- no. 1, pp. 61–73, 2022, doi: 10.32767/jusikom.v7i1.1538.
- [18] Ismail Setiawan, “Komparasi Kinerja Integrated Development Environment (IDE) Dalam Mengeksekusi Perintah Python,” *SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 52–59, 2022, doi: 10.54259/satesi.v2i1.784.
- [19] A. C. Nurcahyo and C. Gudiato, “Implementasi Data Mining Metode K-Means Menggunakan Framework Python Dalam Mengelompokkan Pegawai Berdasarkan Data Presensi,” vol. 8, pp. 1334–1343, 2024, doi: 10.30865/mib.v8i3.7716.
- [20] A. Syukron and A. Subekti, “Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit,” *J. Inform.*, vol. 5, no. 2, pp. 175–185, 2018, doi: 10.31311/ji.v5i2.4158.
- [21] N. I. Prabawati, Widodo, and M. F. Duskarnaen, “Kinerja Algoritma Classification and Regression Tree (Cart) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta Available at : Available at :,” *J. Pinter*, vol. 3, no. 2, pp. 139–145, 2019.
- [22] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.