

Klasifikasi Pasien Penderita Penyakit Liver Dengan Metode Support Vector Machine

1st Alfonsus Yoga Dewa Saputra

Sistem Informasi

Telkom University

Surabaya, Indonesia

alfonsusyoga@student.telkomuniversity.ac.id

2nd Sri Hidayati

Sistem Informasi

Telkom University

Surabaya, Indonesia

srihidayati@telkomuniversity.ac.id

3rd Berlian Rahmi Lidiawaty

Sistem Informasi

Telkom University

Surabaya, Indonesia

berlianerel@telkomuniversity.ac.id

Abstrak — Deteksi dini terhadap penyakit liver sangat penting karena keterlambatan diagnosis dapat menurunkan efektivitas pengobatan. Penelitian ini bertujuan untuk membangun model klasifikasi berbasis algoritma Support Vector Machine (SVM) guna mendeteksi penyakit liver secara awal. Dataset yang digunakan berasal dari UCI Machine Learning Repository, mencakup data pasien dengan berbagai atribut, seperti hasil tes laboratorium dan gejala klinis. Metodologi yang digunakan mengikuti tahapan CRISP-DM, meliputi pemahaman bisnis, eksplorasi data, persiapan data, pembuatan model, evaluasi, dan penerapan. Setelah melalui proses preprocessing, data digunakan untuk melatih dan menguji model SVM dengan berbagai jenis kernel dan teknik oversampling. Pada tahap awal, model terbaik diperoleh dari kombinasi kernel linear dan metode SMOTE+TOMEK pada skenario pembagian data 90:10, menghasilkan akurasi 64,41%. Optimasi lebih lanjut dilakukan melalui penyesuaian hyperparameter menggunakan GridSearch, kernel sigmoid, dan teknik ADASYN. Strategi ini berhasil meningkatkan akurasi sebesar 13,56% hingga mencapai 77,97%. Hasil penelitian menunjukkan bahwa optimasi parameter dan penyeimbangan distribusi kelas secara signifikan meningkatkan performa model. Sebagai pelengkap, dikembangkan dashboard berbasis web untuk menyajikan hasil klasifikasi dan analisis data pasien secara visual dan interaktif.

Kata kunci— Liver, Support Vector Machine, Klasifikasi, CRISP-DM.

I. PENDAHULUAN

Penyakit liver merupakan penyakit dalam yang muncul dan dapat mempengaruhi fungsi vital tubuh manusia [1]. Ketika terjadi kerusakan pada hati, kinerjanya akan terganggu, yang akan muncul berbagai masalah kesehatan. Mulai dari peradangan hingga keadaan yang serius seperti gagal hati. Menurut Organisasi Kesehatan Dunia (WHO), penyakit liver menyebabkan kematian yang mencapai dua juta orang setiap tahunnya dan pada tahun 2019 merupakan korban tertinggi kematian akibat penyakit liver [2]. Kondisi tersebut harus diperhatikan secara khusus karena hati mempunyai peran sebagai penyaring utama tubuh dari racun dan zat berbahaya yang masuk kedalam perut. Ada beberapa penyakit yang tergolong ke dalam penyakit hati yaitu hepatitis, sirosis, kanker hati atau hepatoma, abses hati,

kolesistitis dan perlemakan hati non alkoholik [3]. Penyebab dari penyakit tersebut antara lain mengkonsumsi alkohol berlebihan, yang dapat merusak sel-sel hati secara perlahan. Selain itu penggunaan obat-obatan terlarang, paparan racun, serta pola makan yang buruk yang dapat menyebabkan obesitas dan penumpukan lemak pada organ hati [4].

Permasalahan yang biasanya terjadi adalah sulitnya mengenali penyakit liver sejak dini, bahkan ketika penyakit tersebut sudah menyebar. Padahal mengetahui adanya gejala penyakit liver sejak dini ini sangat diperlukan, agar penderita dapat melakukan pengobatan dengan tepat. Dengan diagnosa adanya penyakit liver lebih awal dapat meningkatkan tingkat kelangsungan hidup pasien. Oleh karena itu, pada penelitian ini ingin mengetahui apakah model Support Vector Machine memiliki akurasi yang tinggi untuk mengklasifikasikan data penyakit liver agar penyakit liver bisa dideteksi sejak dini [5].

Proses klasifikasi pasien penderita penyakit liver merupakan langkah yang penting dalam diagnosis dini dan perencanaan perawatan medis. Dengan melakukan klasifikasi, tenaga medis dapat mendiagnosa pasien yang memiliki tanda-tanda yang jelas terhadap penyakit liver sebelum dampak yang lebih buruk muncul. Mengetahui tahap awal sangat penting karena penyakit liver tidak menunjukkan gejala yang jelas hingga tahap lanjut [6]. Dengan klasifikasi yang sistematis, tenaga medis dapat lebih cepat mengetahui langkah pertama perawatan berdasarkan data prediksi. Selain itu, klasifikasi membantu dalam mengelompokkan pasien dengan kondisi serupa, yang dapat mendukung penelitian lebih lanjut tentang pola penyakit liver dan metode pengobatan yang efektif.

Penelitian ini mengklasifikasi data pasien penderita penyakit liver untuk mendiagnosa penyakit liver berdasarkan variable tertentu. Variabel dapat berupa hasil laboratorium seperti total bilirubin, direct bilirubin, alkaline, alamine, aspartate, total protiens, albumin, rasio albumin dan globulin. Dimana alanine transaminase, aspartate, alkaline, albumin, dan diliturbin merupakan variable berupa enzim yang ada di manusia, jika kandungan enzim tersebut tinggi maka seorang tersebut dapat terindikasi penyakit liver [7]. Penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* dalam klasifikasi data kesehatan pasien berdasarkan variable yang dipilih. Metode ini dapat digunakan untuk mengelompokkan pasien mengidentifikasi pola yang

menunjukkan kecenderungan diagnosa penyakit liver. Metode *Support Vector Machine (SVM)* adalah metode klasifikasi untuk membagi dua kelas berdasarkan pembagian hyperplane [8]. SVM merupakan metode klasifikasi untuk data linier dan tidak linier. SVM bekerja dengan cara memetakan data training, kemudian mencari hyperplane pemisah yang optimal berdasarkan support vector [9].

Hasil dari penelitian ini adalah dashboard dari hasil klasifikasi data pasien penyakit liver untuk mendiagnosa penyakit liver agar dapat membantu tenaga medis maupun pasien untuk lebih cepat dan tepat dalam mendiagnosa penyakit liver. Sehingga dilakukanlah analisis dan pembuatan dashboard evaluasi prediksi dengan mengelompokkan data pasien penyakit liver. Data pasien penyakit liver dikelompokkan berdasarkan kandungan-kandungan yang dimiliki oleh pasien dengan menerapkan data mining dan algoritma klasifikasi. Hasil ini juga dapat memberikan rekomendasi untuk instansi terkait diagnosa awal penyakit liver yang akurat dan terarah.

II. KAJIAN TEORI

A. Liver

Hati atau seringkali disebut liver merupakan salah satu organ tubuh manusia dimana memiliki peran yang penting bagi kehidupan manusia. Terdapat pada sisi kanan manusia, yang memiliki ciri-ciri seperti berat sekitar 1500 gram, warna coklat kemerahan, dan memiliki dua bagian besar yang disebut lobus kanan dan kiri [5]. Liver memiliki fungsi yang cukup banyak untuk keberlangsungan hidup manusia. Mulai dari membersihkan darah, memproduksi kebutuhan nutrisi, karbohidrat, lemak, dan protein bagi semua sel tubuh manusia [6]. Dari banyaknya fungsi liver, tidak sedikit juga penyakit yang dapat menyerang salah satu organ tubuh manusia ini. Walaupun memiliki kemampuan meregenerasi yang baik, liver juga mudah untuk terkena berbagai macam penyakit dan gangguan fungsi.

B. Data Mining

Data mining dapat diartikan sebagai proses yang menerapkan ilmu statistik, matematis, artificial intelligence, dan machine learning untuk menemukan pengetahuan berupa informasi yang berguna serta informasi yang saling berhubungan dari beberapa basis data yang besar [10]. Sedangkan menurut [11] Data mining adalah proses pencarian data untuk membentuk sebuah model yang mampu mengenali pola pada data lain yang belum ada dalam database.

C. Support Vector Machine

Support Vector Machine (SVM) adalah salah satu metode pembelajaran mesin yang bekerja dengan menggunakan hipotesis berupa fungsi linear pada ruang fitur berdimensi tinggi [12]. Model ini dilatih menggunakan algoritma yang berbasis teori optimasi, dengan fungsi utama Menentukan garis pemisah (hyperplane) yang paling tepat untuk membagi data menjadi dua kelompok secara maksimal. SVM memiliki fungsi untuk mencari hyperplane atau fungsi pemisah terbaik untuk memisahkan dua label pada ruang input [13]. Hyperplane dapat berwujud garis pada dua dimensi. Sementara itu, SVM non-linear menggunakan pendekatan

kernel trick, Kernel pada SVM non linear terdiri dari sigmoid, polynomial, RBF. Fungsi kernel mengubah data input berdimensi rendah menjadi ruang berdimensi lebih tinggi. Dengan demikian, data dapat dipisahkan secara linier [12]. Dalam hal ini proses klasifikasi metode Naïve Bayes dapat dirumuskan pada rumus 1.

$$f(x) = W^T x + b = 0 \quad (1)$$

Keterangan:

W = Vektor bobot normal terhadap hyperplane

x = Vektor fitur (data input)

b = Bias (intersepsi)

W^T = Transpos dari vector bobot w

D. Synthetic Minority Oversampling (SMOTE)

Synthetic Minority Over-sampling atau SMOTE merupakan teknik penyeimbang pada kelas suatu dataset yang sering digunakan. Penggunaan teknik SMOTE dapat digunakan ketika salah satu kelas dalam dataset memiliki jumlah sampel yang jauh lebih sedikit dibandingkan kelas lainnya. Metode ini bekerja dengan membuat data sintetis berdasarkan kemiripan antar data dari kelas minoritas, sehingga proporsi kelas tersebut dalam dataset menjadi lebih seimbang [14]. Data latih sintetis tersebut dibentuk dengan mengacu pada tetangga terdekat (nearest neighbor). Diusulkan oleh Chawla pada tahun 2002 dengan menciptakan sampel sintetis baru berdasarkan interpolasi data minoritas yang sudah ada. Dengan teknik ini, SMOTE dapat meminimalisir masalah overfitting yang sering terjadi pada metode oversampling sederhana [15].

E. Tomek Links

Tomek Links adalah sebuah metode undersampling yang dibuat pertama kali oleh Ivan Tomek pada tahun 1976. Metode ini dapat bekerja dengan menghilangkan data pada kelas paling banyak yang dianggap memiliki kesamaan karakteristik (borderline) [16]. Tomek Links juga berperan dalam pembersihan data dari noise atau data yang dianggap *misclassify*.

F. Adaptive Synthetic Sampling

Adaptive Synthetic Sampling (ADASYN) merupakan metode penyeimbangan data yang dirancang untuk mengatasi permasalahan ketidakseimbangan distribusi kelas dalam dataset. Teknik ini bekerja dengan cara menghasilkan sampel sintetis baru secara adaptif untuk kelas minoritas, berdasarkan distribusi lokal data dan tingkat kesulitan klasifikasinya [16].

G. Confusion Matrix

Confusion matrix menurut Han dan Kamber dalam [17] merupakan alat yang digunakan untuk mengevaluasi seberapa baik sebuah model klasifikasi mampu membedakan data atau tuple yang berasal dari kelas yang berbeda.

TABEL 1
(CONFUSION MATRIX)

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Rumus yang sering digunakan dalam evaluasi meliputi akurasi, presisi, dan *recall*, dan *F1-Score*. Berikut merupakan rumus dari evaluasi *Confusion Matrix*:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

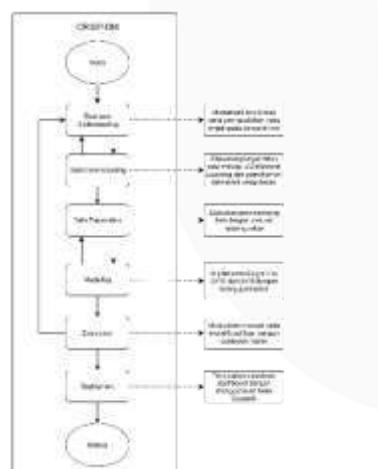
$$F1 \text{ score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (5)$$

H. Streamlit

Streamlit merupakan sebuah framework *open-source* yang memungkinkan para pengembang Data Sains dan *Machine Learning* untuk membuat sebuah dashboard berupa website yang interaktif dengan cepat dan mudah menggunakan pemrograman Python [18].

III. METODE PENELITIAN

Penelitian ini menggunakan metode Cross Industry Standard Process for Data Mining (CRISP-DM) yang terdiri dari enam tahapan utama, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran. Metode ini diterapkan dalam proses klasifikasi pasien yang menderita penyakit liver dengan memanfaatkan algoritma SVM. Pendekatan pada penelitian ini yaitu menggunakan pendekatan kuantitatif dengan memanfaatkan data pasien penderita penyakit liver. Dalam menerapkan CRISP-DM dilakukanlah beberapa tahapan seperti gambar 1 di bawah ini.



GAMBAR 1
(ALUR CRISP-DM PADA PENELITIAN)

Pada tahap *Business Understanding*, ditetapkan tujuan utama penelitian, yaitu pengembangan model *klasifikasi* berbasis *Support Vector Machine (SVM)* untuk mendeteksi penyakit liver. Jika *dataset* yang diperoleh tidak mencerminkan definisi awal atau tidak mencakup kelas penyakit yang ditargetkan, maka dilakukan iterasi untuk menyesuaikan kembali tujuan atau cakupan penelitian.

Tahap *Data Understanding* difokuskan pada eksplorasi *dataset* untuk mengetahui karakteristik penting

seperti distribusi kelas, variabel data, dan kelengkapan data. Jika ditemukan masalah seperti ketidakseimbangan kelas maka perlu dilakukan resampling kembali guna menyesuaikan *dataset* yang akan digunakan.

Tahap *Data Preparation* bertujuan mempersiapkan *dataset* agar siap digunakan dalam pelatihan model. Proses ini melibatkan *preprocessing* untuk pembersihan data dan dilakukan pengisian data kosong menggunakan nilai rata-rata atau median, normalisasi variabel numerik agar berada dalam skala yang sama.

Tahap *Modelling* dilakukan dengan menerapkan algoritma *Support Vector Machine (SVM)*. Algoritma ini digunakan untuk membuat model klasifikasi yang mampu memprediksi apakah seorang pasien memiliki penyakit liver atau tidak, berdasarkan data yang telah melalui tahap pemrosesan. *Support Vector Machine (SVM)* sendiri merupakan salah satu metode machine learning yang menggunakan hipotesis berupa fungsi linier pada ruang fitur dengan dimensi tinggi..

Tahap evaluasi dilakukan untuk menilai performa model dengan menggunakan beberapa metrik, yaitu akurasi, presisi, *recall*, dan *F1-score*. Jika performa yang dicapai belum sesuai harapan, maka dilakukan iterasi kembali ke tahap *Business Understanding* untuk meninjau ulang definisi tujuan atau parameter *dataset* yang digunakan.

Tahap terakhir, *Deployment*, melibatkan integrasi model ke dalam prediksi penyakit liver berbasis web agar pengguna dapat melakukan cek penyakit liver dan memperoleh hasil deteksi secara otomatis. Sistem ini dikembangkan dengan *framework* berbasis web agar mudah diakses dan dimanfaatkan langsung oleh dokter atau pengguna umum.

IV. HASIL DAN PEMBAHASAN

A. Business Understanding

Penelitian ini dilatarbelakangi oleh permasalahan yang dihadapi oleh masyarakat dalam mengetahui penyakit liver secara dini. Proses identifikasi secara manual, baik melalui observasi visual langsung maupun konsultasi dengan ahli, bersifat subjektif, memakan waktu, serta sangat bergantung pada pengalaman individu. Keterlambatan dalam mendeteksi penyakit liver berdampak pada tingkat keberhasilan pengobatan dan dapat memperburuk kondisi pasien.

Sebagai solusi dari permasalahan tersebut, penelitian ini bertujuan untuk membangun model klasifikasi berbasis teknologi yang dapat membantu dokter atau pasien dalam mendiagnosis penyakit liver secara otomatis dan cepat. Sistem yang dikembangkan menggunakan pendekatan *Support Vector Machine (SVM)* sebagai metode klasifikasi pasien penyakit liver.

Sistem ini dirancang untuk melakukan deteksi dini terhadap kondisi kesehatan pasien terkait penyakit liver dengan mengklasifikasikan mereka ke dalam dua kategori, yaitu: terdiagnosis menderita penyakit liver atau tidak terdiagnosis penyakit liver. Proses klasifikasi ini dilakukan secara otomatis berdasarkan analisis terhadap variabel-variabel input yang berhubungan dengan faktor risiko dan indikator medis penyakit liver.

B. Data Understanding

Data yang digunakan dalam penelitian ini berasal dari sumber sekunder, yaitu ILDP (Indian Liver Patient Dataset) yang tersedia di UCI Machine Learning. Dataset tersebut memuat 583 jumlah data dengan total 11 variabel. Rincian variabel yang digunakan disajikan pada Tabel 2 dibawah merupakan keterangan, jenis data dan penjelasan dari setiap variabel penelitian yang digunakan pada data set pasien penderita penyakit liver (Nurhuda et al., 2024) (Baiq Nurul Azmi et al., 2023).

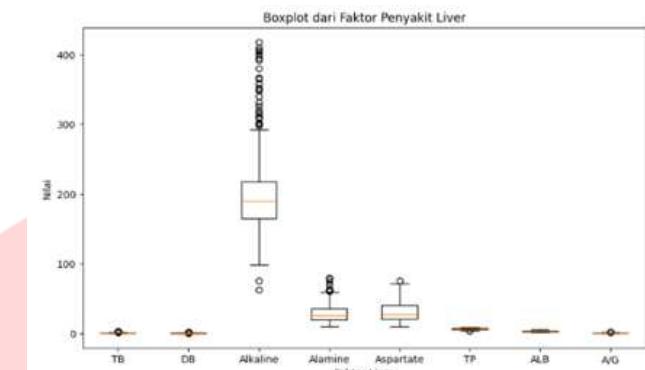
TABEL 2
(VARIABEL PENELITIAN)

Variabel	Keterangan	Jenis Data	Penjelasan
Age	Umur pasien	Numerik	Umur pasien
Gender	Jenis kelamin	Kategorik	Jenis kelamin pasien
TB	Total Bilirubin	Numerik	Jumlah pigmen kuning yang dihasilkan dari proses pemecahan sel darah merah dan dikeluarkan ke dalam empedu oleh sel-sel hati.
DB	Direct Bilirubin	Numerik	Kandungan pigmen yang terkonjugasi dalam darah dalam bentuk larut air yang dikeluarkan dari hati
Alkphos	Alkhphos Alkaline Phosphatase	Numerik	Enzim yang diproduksi di hati dan saluran empedu, dan berperan dalam memecah protein dalam tubuh
Sgpt	Sgpt Alamine Aminotransferase	Numerik	Enzim yang terdapat di hati dan berfungsi mengubah protein menjadi energi untuk sel-sel hati
Sgot	Sgot Aspartate Aminotransferase	Numerik	Enzim yang terdapat di hati dan membantu mencerna protein
TP	Total Proteins	Numerik	Jumlah protein dalam darah yang diproduksi oleh hati, sel plasma, kelenjar limfe, limpa, dan sumsum tulang
ALB	Albumin	Numerik	Protein yang dihasilkan oleh hati dan memiliki peran penting dalam mengatur keseimbangan cairan dalam tubuh.
A/G Rasio	Rasio Albumin dan Globulin	Numerik	Perbandingan jumlah albumin dan globulin dalam darah
Class	Terindikasi penyakit Liver/Tidak terindikasi penyakit Liver	kategorik	Terditi dari beberapa indikasi penyakit liver dan tidak terindikasi penyakit liver pada pasien

C. Data Preparation

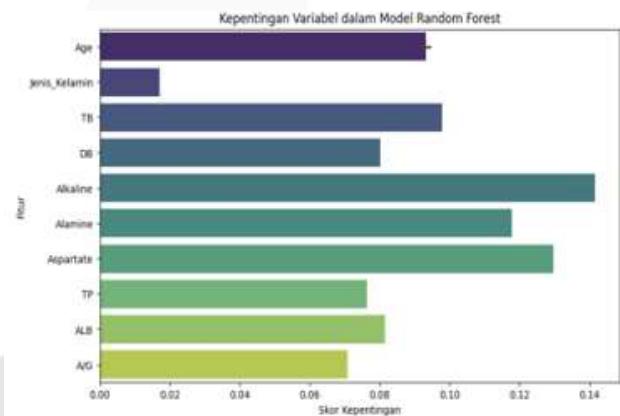
Pada tahap ini, seluruh variabel yang digunakan akan melalui proses preprocessing data. Langkah pertama adalah memeriksa apakah terdapat nilai yang hilang (missing value). Hasil pemeriksaan menunjukkan adanya 4 data yang hilang pada variabel Rasio Albumin dan Globulin (A/G). Untuk mengatasi hal tersebut, dilakukan imputasi dengan

memasukkan nilai median ke dalam 4 data yang hilang, karena variabel ini berskala numerik. Pemilihan median dilakukan karena lebih tahan terhadap outlier dibandingkan dengan nilai rata-rata (mean). Selanjutnya data dicek apakah ada yang duplikat, outlier pada data Indian Liver Patient Dataset. Dari hasil perhitungan data yang duplikat sebanyak 13 observasi. Selanjutnya dilakukan penghapusan, pada boxplot terlihat pada masing-masing variabel faktor penyakit liver terdapat data outlier pada gambar 2 berikut.



GAMBAR 2
(BOXPLOT VARIABEL DARI FAKTOR PENYAKIT LIVER)

Langkah berikutnya adalah melakukan feature selection untuk mengidentifikasi variabel-variabel yang berkontribusi terhadap kemungkinan seseorang mengidap penyakit liver. Nilai kepentingan masing-masing variabel ditampilkan pada Gambar 3 berikut



GAMBAR 3
(KEPENTINGAN VARIABEL)

Berdasarkan Gambar 3 terlihat bahwa nilai *feature importance* untuk variabel jenis kelamin mengalami penurunan yang signifikan. Hal ini mengindikasikan bahwa variabel tersebut memiliki pengaruh yang rendah dalam menentukan apakah seseorang menderita penyakit liver atau tidak. Oleh karena itu, variabel jenis kelamin dikeluarkan dari tahap analisis selanjutnya. Pada variabel umur juga akan dikeluarkan, karena pada penelitian ini ingin berfokus pada variabel yang memiliki keterkaitan dengan kandungan-kandungan yang ada pada organ liver. Setelah variabel-variabel yang relevan berhasil diidentifikasi, proses berikutnya adalah melakukan eksplorasi data. Tahap awal eksplorasi ini dilakukan dengan meninjau statistik deskriptif dari variabel-variabel independen yang digunakan.

Pada tahap selanjutnya yaitu melakukan eksplorasi variabel dependen pada data yang dihasilkan. Setelah data dibagi menjadi data pelatihan dan pengujian dengan berbagai

skenario 90:10, 80:20, dan 75:25. Dilakukan berbagai skenario dengan tujuan untuk mencari hasil nilai akurasi terbaik dalam setiap pembagian data skenario. Ditemukan bahwa distribusi data tidak seimbang. Ketidakseimbangan ini dapat menyebabkan model klasifikasi menghasilkan akurasi yang tinggi, namun hanya mampu mengklasifikasikan satu kategori dengan baik, sementara kategori lainnya diabaikan. Hal ini menunjukkan bahwa performa model sebenarnya kurang optimal. Oleh karena itu, sebelum dilakukan analisis lebih lanjut, akan diterapkan teknik Oversampling untuk menyeimbangkan distribusi kelas dalam dataset. Untuk melakukan oversampling digunakan berbagai teknik yaitu SMOTE, kombinasi antara TOMEK LINK dan SMOTE, dan ADASYN. Dan perbandingan setiap teknik dapat dilihat pada Tabel 3 berikut.

TABEL 3
(TABEL HASIL DARI SPLITTING DAN IMBALANCE DATA)

Imbalance	Skenario	Sebelum Imbalance		Setelah Imbalance	
		Liver	Tidak Liver	Liver	Tidak Liver
Smote	90:10	371	153	371	371
	80:20	329	137	329	329
	75:25	307	130	307	307
Tomek+Smote	90:10	371	153	343	343
	80:20	329	137	311	311
	75:25	307	130	284	284
Adasyn	90:10	371	153	371	344
	80:20	329	137	329	303
	75:25	307	130	307	288

D. Modelling

Pada tahap ini dilakukan proses perancangan dan pelatihan model klasifikasi menggunakan metode Support Vector Machine (SVM) untuk mendeteksi pasien penyakit liver pada adata yang tersedia. Pada penelitian ini akan digunakan empat fungsi kernel yang berbeda yaitu *linear*, *polynomial*, *sigmoid*, dan *radial basis function* (RBF). Setelah dilakukan dilakukan perhitungan SVM selanjutnya akan perhitungan SVM dengan hyperparameter untuk meningkatkan nilai akurasi data. Berdasarkan Tabel 4, SVM dengan fungsi kernel *Linear* yang menggunakan skenario 90% : 10% dengan teknik imbalance SMOTE+TOMEK menghasilkan nilai akurasi terbesar, dengan nilai 64,41%

TABEL 4
(KEBAIKAN MODEL SVM DENGAN SMOTE+TOMEK)

Skenario	Kernel SVM	Accuracy	Precision	Recall
(90:10)	Linear	64,41%	96,15%	55,56%
	Polynomial	49,15%	100%	33,33%
	RBF	54,24%	95%	42,22%
	Sigmoid	61,02%	92,31%	53,33%
(80:20)	Linear	59,83%	97,62%	47,13%
	Polynomial	44,44%	100%	25,29%
	RBF	55,56%	94,87%	42,53%

	Sigmoid	58,97%	88,24%	51,72%
(75:25)	Linear	61,64%	96,49%	50,46%
	Polynomial	49,32%	100%	32,11%
	RBF	56,85%	91,07%	46,79%
	Sigmoid	54,11%	82,81%	48,62%

TABEL 5
(KEBAIKAN MODEL SVM DENGAN SMOTE)

Skenario	Kernel SVM	Accuracy	Precision	Recall
(90:10)	Linear	61,02%	95,83%	51,11%
	Polynomial	49,15%	100%	33,33%
	RBF	54,24%	95%	42,22%
	Sigmoid	61,02%	92,31%	53,33%
(80:20)	Linear	58,97%	97,56%	45,98%
	Polynomial	44,44%	100%	25,29%
	RBF	55,56%	94,87%	42,53%
	Sigmoid	58,97%	88,24%	51,72%
(75:25)	Linear	60,96%	98,15%	48,62%
	Polynomial	47,95%	100%	30,28%
	RBF	57,53%	92,73%	46,79%
	Sigmoid	54,79%	82,09%	50,46%

TABEL 6
(KEBAIKAN MODEL SVM DENGAN ADASYN)

Skenario	Kernel SVM	Accuracy	Precision	Recall
(90:10)	Linear	61,02%	95,83%	51,11%
	Polynomial	44,07%	100%	26,67%
	RBF	54,24%	95%	42,22%
	Sigmoid	62,71%	92,59%	55,56%
(80:20)	Linear	58,97%	97,56%	45,98%
	Polynomial	41,88%	100%	21,84%
	RBF	56,41%	95%	43,68%
	Sigmoid	52,14%	94,29%	37,93%
(75:25)	Linear	59,59%	98,08%	46,79%
	Polynomial	46,58%	100%	28,44%
	RBF	57,53%	94,34%	45,87%
	Sigmoid	54,79%	82,09%	50,46%

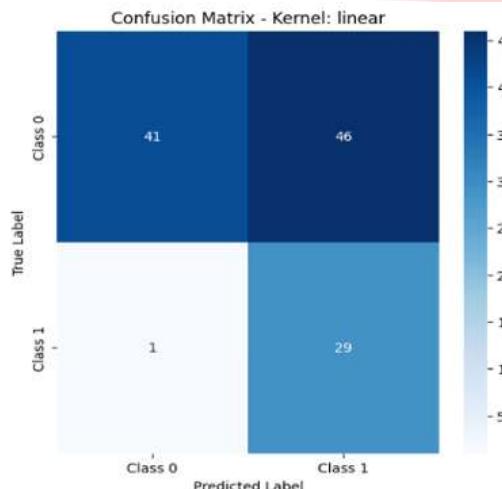
Tabel 7 menunjukkan pengaturan SVM menggunakan hyperparameter untuk meningkatkan hasil akurasi pada SVM yang sebelumnya masuk kurang tinggi. Pada penggunaan tunning Gridsearch dapat dihasilkan 3 hyperparameter yaitu C dengan nilai 100, gamma 0.1, dan kernel sigmoid. Parameter C berfungsi sebagai pengatur regularisasi, sedangkan parameter kernel menentukan tipe kernel yang dipakai dalam model. Sementara itu, parameter gamma berkaitan dengan koefisien yang digunakan dalam fungsi kernel.

TABEL 7
(PARAMETER TERBAIK PADA SVM TUNNING PARAMETER)

Hyperparameter	Values	Parameter
C	[0.1, 1, 10, 100]	100
Gamma	[0.01, 0.1, 1]	0.1
Kernel	['linear', 'rbf', 'sigmoid', 'poly']	Sigmoid

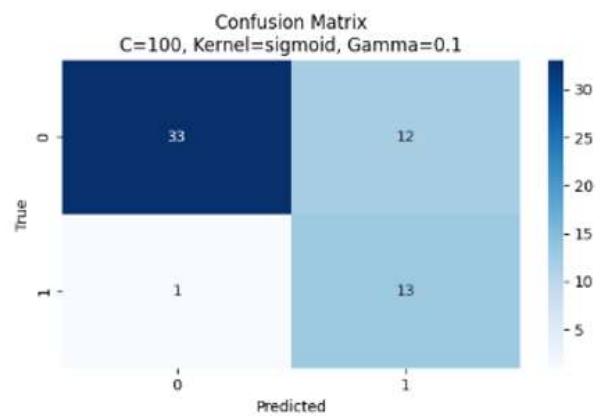
E. Evaluation

Gambar 4 menunjukkan hasil confusion matrix menggunakan model Support Vector Machine (SVM) dengan kernel linear. Hasil ini mengindikasikan bahwa model cukup baik dalam mengenali kelas positif, karena kesalahan klasifikasi terhadap kelas ini sangat kecil. Namun, model kurang akurat dalam mengenali kelas negatif, terbukti dari tingginya jumlah kesalahan pada false positive. Oleh karena itu, perlu dilakukan evaluasi lebih lanjut seperti penggantian kernel, tuning parameter, atau penerapan teknik penyeimbangan data untuk meningkatkan performa model secara keseluruhan.



GAMBAR 4
(CONFUSION MATRIX KERNEL LINEAR)

Gambar 5 menunjukkan confusion matrix dari model terbaik yang diperoleh setelah proses tuning dilakukan menggunakan GridSearch. Berdasarkan matriks tersebut, terlihat bahwa model memiliki kinerja klasifikasi yang cukup seimbang, dengan nilai true positive dan true negative yang tinggi. Kemampuan model dalam mengidentifikasi pasien yang benar-benar menderita penyakit liver cukup baik, terlihat dari jumlah false negative yang rendah. Selain itu, jumlah false positive juga menurun dibandingkan dengan model sebelumnya, menandakan bahwa prediksi model tidak condong hanya pada satu kelas. Peningkatan ini mencerminkan peningkatan sensitivitas dan spesifitas model, yang merupakan faktor penting dalam mendukung sistem diagnosis di bidang medis.



GAMBAR 5
(CONFUSION MATRIX TERBAIK GRIDSEARCH)

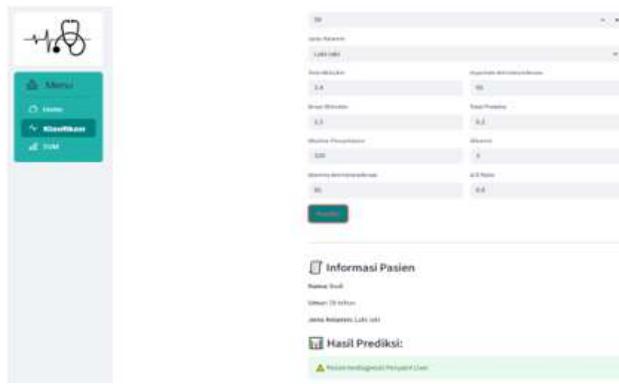
E. Deployment

Model SVM yang telah dikembangkan dan dilatih sebelumnya dalam penelitian ini dimuat secara lokal ke dalam sistem dan digunakan untuk proses klasifikasi. Model tersebut memiliki program yang dirancang untuk mengelompokan 2 kelas yaitu pasien terdiagnosis penyakit liver dan pasien tidak terdiagnosis penyakit liver. Setelah dimasukkan, sistem akan melakukan preprocessing terhadap data yang sudah dimasukkan agar sesuai dengan model yang dibuat. Kemudian hasil prediksi akan keluar dalam bentuk label kelas beserta data pendukung yang sudah dimasukkan sebelumnya. Gambar 6 dibawah memperlihatkan tampilan antarmuka pada halaman Klasifikasi dini penyakit liver.



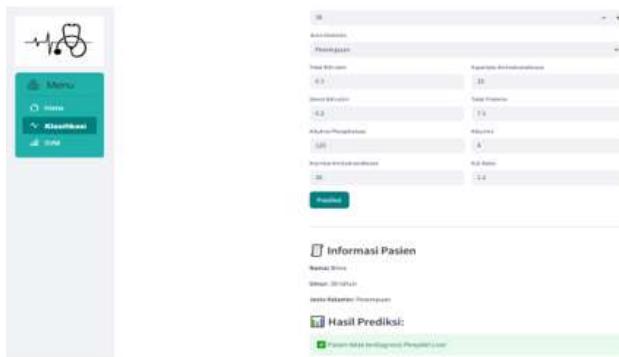
GAMBAR 6
(HALAMAN PREDIKSI PASIEN PENYAKIT LIVER)

Gambar 7 dibawah menunjukkan salah satu contoh data pasien yang terdiagnosis penyakit liver. Sebagian besar nilai pada data tersebut berada di atas atau mendekati ambang batas abnormal yang menunjukkan adanya gangguan fungsi hati. Setelah tombol “Prediksi” ditekan, sistem secara otomatis menampilkan peringatan berwarna kuning dengan keterangan “Pasien terdiagnosis Penyakit Liver”.



GAMBAR 7
(HASIL PREDIKSI PASIEN TERDIAGNOSIS PENYAKIT LIVER)

Gambar 8 dibawah menunjukkan salah satu contoh data pasien yang tidak terdiagnosis penyakit liver. Semua parameter tetap berada dalam rentang fisiologis normal. Ketika tombol “Prediksi” ditekan, antarmuka menampilkan ikon centang hijau beserta keterangan “Pasien tidak terdiagnosis Penyakit Liver”. Prediksi negatif ini menegaskan bahwa model mampu menghindari false positive pada profil biokimia yang normal dan menunjukkan bahwa integrasi model ke dalam aplikasi memberikan spesifisitas yang memadai untuk deteksi dini penyakit liver.



GAMBAR 8
(HASIL PREDIKSI PASIEN TIDAK TERDIAGNOSIS PENYAKIT LIVER)

V. KESIMPULAN

Algoritma Support Vector Machine (SVM) berhasil digunakan untuk mengklasifikasikan pasien dengan penyakit liver berdasarkan sejumlah parameter medis seperti kadar bilirubin total, enzim hati, protein dalam darah, serta rasio albumin terhadap globulin. Awalnya, kombinasi kernel linear dan teknik SMOTE+TOMEK menghasilkan akurasi sebesar 64,41%, namun performa model meningkat signifikan hingga 77,97% setelah dilakukan penyetelan hyperparameter menggunakan GridSearch serta penerapan kernel sigmoid dan metode ADASYN. Penanganan ketidakseimbangan data melalui teknik resampling seperti SMOTE, SMOTE+TOMEK, dan terutama ADASYN, memberikan kontribusi besar terhadap peningkatan kinerja model, terutama dalam mengenali kelas minoritas dengan lebih baik, yang tercermin dari peningkatan recall dan F1-score. Model yang dikembangkan kemudian diterapkan ke dalam dashboard menggunakan Streamlit, yang memungkinkan input data pasien secara langsung dan menampilkan hasil klasifikasi secara otomatis dan real-time. Sistem ini

dirancang dengan tampilan yang interaktif dan mudah dipahami, sehingga membuktikan bahwa model klasifikasi berbasis machine learning tidak hanya efektif secara teknis, tetapi juga dapat diterapkan secara praktis untuk mendukung deteksi dini penyakit liver.

REFERENSI

- [1] D. A. Sodikin, E. T. Is’ad, R. Prayoga, and A. N. I. Purwanto, “545-Article Text-2292-2-10-20240123,” *Jurnal Prosiding Seminar Nasional Rekayasa, Sains dan Teknologi*, vol. 3 No. 1, 2024.
- [2] H. Devarbhavi, S. K. Asrani, J. P. Arab, Y. A. Nartey, E. Pose, and P. S. Kamath, “Global burden of liver disease: 2023 update,” Aug. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.jhep.2023.03.017.
- [3] A. Desiani, “Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati,” *SMKOM*, vol. 7, no. 2, pp. 104–110, Jul. 2022, doi: 10.51717/simkom.v7i2.96.
- [4] A. Abdussomad, I. Kurniawan, and A. Wibowo, “Prediksi Kemungkinan Penyakit Liver menggunakan Algoritma Klasifikasi Naive Bayes,” *Technologia : Jurnal Ilmiah*, vol. 15, no. 3, p. 506, Jul. 2024, doi: 10.31602/tji.v15i3.15288.
- [5] F. Lia, D. Cahyanti, F. Sarasati, W. Astuti, and E. Firasari, “KLASIFIKASI DATA MINING DENGAN ALGORITMA MACHINE LARNING UNTUK PREDIKSI PENYAKIT LIVER,” 2023. [Online]. Available: <https://ojs.uniska-bjm.ac.id/index.php/JIT>
- [6] H. Tandra, *Hati-Hati dengan Hati*. Jakarta: Gramedia Pustaka Utama, 2024. [Online]. Available: <https://books.google.co.id/books?id=IV8wEQAAQBAJ>
- [7] A. Surandy, “Tes Fungsi Hati, Langkah Awal Menentukan Kesehatan Hati,” Rumah Sakit Pondok Indah Group. Accessed: Dec. 18, 2024. [Online]. Available: <https://www.rspondokindah.co.id/id/news/tes-fungsi-hati-langkah-awal-menentukan-kesehatan-hati>
- [8] S. A. Naufal, A. Adiwijaya, and W. Astuti, “Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray,” *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, p. 162, Feb. 2020, doi: 10.30865/jurikom.v7i1.2014.
- [9] M. Amir Nugraha, M. Itqan Mazdadi, A. Farmadi, and T. Hamongan Saragih, “PENYEIMBANGAN KELAS SMOTE DAN SELEKSI FITUR ENSEMBLE FILTER PADA SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT LIVER,” vol. 10, no. 6, pp. 1273–1284, 2023, doi: 10.25126/jtiik.2023107234.
- [10] E. luthfi and U. Amikom, *Algoritma Data Mining*. Penerbit Andi, 2009. [Online]. Available: <https://books.google.co.id/books?id=-Ojclag73O8C>
- [11] M. A. Rizqian, B. Irawan, and P. Wahyuningsih, “Implementasi Algoritma KNN Untuk Klasifikasi

- Jurusan Siswa Kelas XI di SMA Al Irsyad Tegal," *JCRD: Journal of Citizen Research and Development*, 2024.
- [12] A. H. Pratama, *BELAJAR MUDAH DAN SINGKAT MACHINE LEARNING: Panduan Praktis dengan Studi Kasus, Kode Program, dan Dataset*. Penerbit Andi, 2024. [Online]. Available: <https://books.google.co.id/books?id=pFMOEQAAQBAJ>
- [13] "Penjelasan Lengkap Algoritma Support Vector Machine (SVM)," Trivusi. Accessed: Dec. 18, 2024. [Online]. Available: <https://www.trivusi.web.id/2022/04/algoritma-svm.html>
- [14] R. M. Aldani, A. Bhirawa, and U. Pradema Sanjaya, "FROM DATA IMBALANCE TO PRECISION: SMOTE-DRIVEN MACHINE LEARNING FOR EARLY DETECTION OF KIDNEY DISEASE OPTIMASI KLASIFIKASI DATA TIDAK SEIMBANG PADA DATASET MEDIS PADA KASUS PENYAKIT GINJAL KRONIS DENGAN TEKNIK SMOTE," *JURNAL INOVTEK POLBENG*, vol. 10, no. 1, Mar. 2025.
- [15] A. A. Arifiyanti and E. D. Wahyuni, "SMOTE: METODE PENYEIMBANG KELAS PADA KLASIFIKASI DATA MINING," vol. 15, Feb. 2020, [Online]. Available: <https://www.cs.technion.ac.il/~shani/CS-734-2020/SMOTE.pdf>
- [16] D. Heksaputra, R. Gernowo, and R. R. Isnanto, "Over-Under Sampling Approach with Adaptive Synthetic and Tomek Links Methods to Handle Data Imbalance in Sentence Classification on Halal Assurance Certificate Documents," *Fusion: Practice and Applications*, vol. 19, no. 2, pp. 194–210, 2025, doi: 10.54216/FPA.190215.
- [17] R. Susetyoko, W. Yuwono, and E. Purwantini, "JIP (Jurnal Informatika Polinema) Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner," 2022.
- [18] J. Surya and F. H. Aminuddin, *Pemrograman MySQL Database With Streamlit Python*. 2024. Accessed: Jan. 14, 2025. [Online]. Available: https://www.google.co.id/books/edition/Pemrograman_SQL_Database_With_Streamli/TkkREQAAQBAJ?hl=id&gbpv=1