

PENERAPAN DATA MINING UNTUK PREDIKSI PENYAKIT HEPATITIS C MENGGUNAKAN ALGORITMA NAÏVE BAYES

Alfina Damayanti^{1,*}, Gusmelia Testiana²

^aProgram Studi Sistem Informasi, Fakultas Sains dan Teknologi UIN Raden Fatah Palembang

¹damaflaviayanti@gmail.com, ²gusmeliatestiana_uin@radenfatah.ac.id

ABSTRAK

Hepatitis sendiri merupakan penyakit peradangan pada hati (liver) yang dapat disebabkan oleh faktor genetik, infeksi virus, alkohol, dan obat-obatan. Berdasarkan laporan global oleh Organisasi. Secara global, diperkirakan 58 juta orang memiliki infeksi virus hepatitis C kronis, dengan sekitar 1,5 juta infeksi baru terjadi setiap tahun. Diperkirakan ada 3,2 juta remaja dan anak-anak dengan infeksi hepatitis C kronis. Pada tahun 2019 WHO memperkirakan sekitar 290.000 orang meninggal karena hepatitis C, sebagian besar karena sirosis dan karsinoma hepatoseluler (kanker hati primer). Pada Penelitian ini data pasien dimanfaatkan untuk digali informasi atau pengetahuan baru didalamnya Proses penggalian informasi baru dari sekumpulan data dikenal dengan data mining. Maka dari itu, dengan adanya penelitian ini bertujuan untuk memprediksi data terkait menggunakan algoritma Naïve Bayes untuk mengetahui apakah pasien tersebut menderita penyakit Hepatitis C atau tidak. Hasil yang diperoleh dari pengujian data hepatitis dengan kategori *0=Donor Darah, 0s=dugaan Donor Darah, 1=Hepatitis, 2=Fibrosis, 3=Cirrhosis). Hasil penelitian menunjukkan bahwa klasifikasi dengan algoritma Naïve Bayes dapat diterapkan dalam mengklasifikasi data penyakit Hepatitis C menghasilkan nilai akurasi sebesar 85,71% dalam kategori Good Classification.

Keywords: Data Mining,;Algoritma Naïve Bayes ;Hepatitis C.

ABSTRACT

*Hepatitis itself is an inflammatory disease of the liver (liver) that can be caused by genetic factors, viral infections, alcohol, and drugs. Based on global reports by the Organization. Globally, an estimated 58 million people have chronic hepatitis C virus infection, with about 1.5 million new infections occurring each year. There are an estimated 3.2 million adolescents and children with chronic hepatitis C infection. In 2019 the WHO estimated about 290,000 people died from hepatitis C, mostly due to cirrhosis and hepatocellular carcinoma (primary liver cancer). In this study, patient data is used to explore new information or knowledge in it: The process of extracting new information from a set of data is known as data mining. Therefore, this study aims to predict related data using the Naïve Bayes algorithm to find out whether the patient is suffering from Hepatitis C disease or not. Results obtained from hepatitis data testing with categories *0=Blood Donation, 0s=suspected Blood Donation, 1=Hepatitis, 2=Fibrosis, 3=Cirrhosis). The results showed that classification with the Naïve Bayes algorithm can be applied in classifying Hepatitis C disease data resulting in an accuracy value of 93.61% in the Good Classification category.*

Keywords: Data Mining,; Algoritma Naïve Bayes ; Hepatitis C.

1. Pendahuluan

Hepatitis merupakan suatu penyakit peradangan hati yang umumnya disebabkan oleh virus. Hepatitis sendiri merupakan penyakit peradangan pada hati (liver) yang dapat disebabkan oleh faktor genetik, infeksi virus, alkohol dan obat-obatan. Menurut laporan global dari Organisasi Kesehatan Dunia (WHO), diperkirakan 58 juta orang di seluruh dunia menderita infeksi hepatitis C kronis, dengan sekitar 1,5 juta infeksi baru setiap tahun. Diperkirakan ada 3,2 juta remaja dan anak-anak dengan infeksi hepatitis C kronis. Pada tahun 2019, WHO

DOI: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>

memperkirakan sekitar 290.000 orang meninggal karena hepatitis C, terutama karena sirosis dan karsinoma hepatoseluler (kanker hati primer). (WHO, 2022). Virus hepatitis C (VHC) merupakan salah satu virus penyebab hepatitis dan dianggap sebagai virus penyebab hepatitis yang paling mematikan. Kebanyakan orang yang terinfeksi virus hepatitis C tidak menunjukkan gejala. Banyak orang tidak menyadari bahwa mereka telah tertular virus hepatitis C sampai hati mereka rusak parah. (Alhawaris, 2019).

Dengan pesatnya perkembangan teknologi, penggunaan sistem informasi yang terkomputerisasi semakin meluas di berbagai bidang, termasuk bidang medis dan kesehatan. Sektor kesehatan telah mampu menghasilkan sejumlah besar data dan jumlah ini akan terus bertambah. Jumlah data yang meningkat ini memerlukan metode otomatis untuk mengekstrak data ini jika perlu (Milovic & Milovic, 2012). Jumlah data pasien dapat diolah dengan menggunakan teknik data mining. Data mining adalah solusi yang memungkinkan kami menemukan konten informasi tersembunyi dalam bentuk pola dan aturan dari kumpulan data besar dengan cara yang dapat dipahami (Latu Handarko, 2015). Penggunaan teknologi ini dapat diterapkan dalam memprediksi pasien yang terinfeksi hepatitis C untuk mengidentifikasi pasien secara cepat pada tahap awal. Deteksi Pasien Hepatitis C Menggunakan Anti-HCV Anti-HCV adalah salah satu tes yang dilakukan untuk memeriksa antibodi HCV dalam serum pasien. Antibodi ini terbentuk dalam serum saat pasien terinfeksi virus hepatitis C. Deteksi dini bisa dilakukan dengan rapid test dan hasilnya terlihat setelah 15 menit.

Dalam ilmu komputer, data mining merupakan ilmu yang dapat membantu memprediksi pasien hepatitis C. Clinical data mining adalah penerapan teknik data mining untuk mengungkap data medis dan klinis. Dengan metode ini, kondisi masa depan pasien dapat diprediksi berdasarkan data pasien lain dan data observasi masa lalu. Salah satu metode prediksi adalah klasifikasi.

Di bidang ilmu komputer, beberapa penelitian telah dilakukan untuk memprediksi penyakit hepatitis C dengan teknik data mining menggunakan studi algoritma Decision Tree C.45 berjudul Menerapkan Teknik Data Mining untuk Mengklasifikasikan Pasien Terduga Infeksi Virus Hepatitis C oleh algoritma Safdari et al. yaitu SVM, Nave Bayes, Decision Tree, Random Forest, Logistic Regression dan ANN.

Pada penelitian ini Penulis menggunakan algoritma Naïve Bayes. Algoritma Naïve Bayes dipilih karena menghasilkan tingkat nilai error yang lebih rendah dengan data yang relative banyak. Naïve Bayes juga cocok untuk data multiclass prediction, yaitu dataset yang memiliki banyak atribut.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan untuk memprediksi apakah pasien berpeluang besar menderita penyakit Hepatitis C atau tidak.

2. Metode

Pendekatan utama dalam penelitian ini yaitu pendekatan kualitatif dan pendekatan kuantitatif. Dalam penelitian ini metode yang digunakan yaitu metode penelitian kuantitatif. Metode penelitian kuantitatif adalah metode penelitian yang menggunakan filsafat positivisme, untuk meneliti populasi atau sampel tertentu dengan pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan (Sugiyono, 2017). Tujuan dari penelitian ini adalah melakukan klasifikasi dan evaluasi model algoritma Naïve Bayes untuk mengetahui akurasi algoritma Naïve Bayes dalam memprediksi penyakit Hepatitis C.

2.1 Metode Pengumpulan Data

Dalam mendukung penelitian ini dilakukan pengumpulan data. Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang akan digunakan menggunakan metode pengumpulan data sekunder. Data Sekunder adalah data yang diperoleh berdasarkan penelitian sebelumnya, artinya peneliti tidak harus mengambil datanya langsung ke lapangan melainkan sudah dikumpulkan dan diolah oleh peneliti lain (Suryani & Hendryadi, 2018). Data dari penelitian ini diambil dari situs kaggle <https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset>. Jumlah sampel yang digunakan dalam penelitian ini adalah sebesar 586 data yang dibagi menjadi data training dan data testing. Data ini terdiri dari 13 atribut dengan rincian 12 sebagai atribut dan 1 sebagai kelas atau kategori.

2.2 Algoritma Naïve Bayes

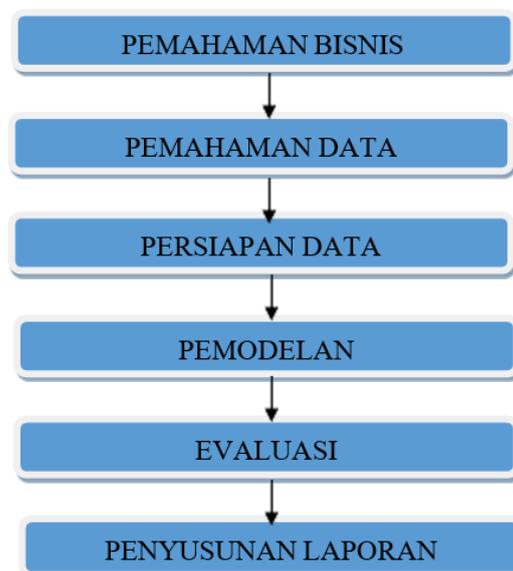
Algoritma Naïve Bayes merupakan klasifikasi yang mempresentasikan setiap kelas objek berdasarkan kesimpulan atau rekapitulasi probabilistik dan menemukan kemungkinan besar kelas yang sesuai untuk tiap objek-objek yang akan ditentukan kelasnya dari objek objek uji yang ada berdasarkan atribut atribut atau variabel yang telah diketahui nilai-nilainya (Arhami & Nasir, 2020). Algoritma Naïve Bayes dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Algoritma Naïve Bayes sangat cocok untuk melakukan klasifikasi pada dataset bertipe nominal (Suntoro, 2019). Teorema Bayes memiliki bentuk umum sebagai berikut:

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

X adalah Data dengan class yang belum diketahui, H adalah Hipotesis data X merupakan suatu class spesifik, $P(H|X)$ adalah Probabilitas hipotesis H berdasar kondisi X (posteriori probability), $P(H)$ adalah Probabilitas hipotesisi H (pror probability) dan $P(H|X)$ adalah Probabilitas X berdasar kondisi pada hipotesis H, $P(X)$ adalah Probabilitas dari X.

2.3 Tahapan Data Mining

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





DOI: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>

Gambar 1. Tahapan Data Mining

Tahapan pada Gambar 1 merupakan proses data mining dalam penelitian ini.

1. Fase pemahaman bisnis, juga dikenal sebagai fase pemahaman penelitian, berisi tujuan dan persyaratan proyek yang jelas yang berhubungan dengan seluruh perusahaan atau entitas penelitian, mengubah tujuan dan kendala menjadi perumusan definisi masalah penambangan data, Mempersiapkan Strategi untuk mencapai tujuan tersebut tercapai.
2. Fase pemahaman data dengan mengumpulkan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data untuk mendapatkan wawasan awal, menilai kualitas data, dan memilih subset yang menarik jika sesuai.
3. Fase persiapan data menyiapkan data mentah awal untuk kumpulan data akhir yang digunakan di semua fase berikutnya.
4. Fase Pemodelan Pada fase ini, teknik pemodelan yang tepat dipilih dan diterapkan. Kalibrasi pengaturan model untuk mengoptimalkan hasil. Seringkali ada beberapa teknik berbeda yang tersedia untuk masalah data mining yang sama. Jika perlu, kembali ke fase persiapan data untuk menyesuaikan bentuk data sesuai dengan kebutuhan spesifik teknik penambangan data Anda.
5. Selama Tahap Evaluasi, model atau model yang diajukan selama Tahap Modeling dievaluasi kualitas dan efektivitasnya sebelum dikerahkan untuk digunakan di lapangan, menentukan tercapai atau tidaknya tujuan yang ditetapkan. Aspek kunci dari bisnis atau masalah penelitian belum ditangani secara memadai dan menggunakan hasil penambangan data.
6. Tahapan penyusunan laporan disertasi adalah mendokumentasikan apa yang relevan dengan pekerjaan penelitian yang dilakukan. Mencatat hasil penelitian dan menerjemahkannya ke dalam argumentasi yang disajikan dalam bentuk laporan dan dapat digunakan sebagai literatur.

3. Hasil dan Pembahasan

Pengolahan data mining merupakan hasil utama dalam penelitian ini. Proses pengolahan data mining dilakukan dengan tahap sesuai Cross Industry Standard Process for Data Mining (CRISP-DM). Berikut merupakan uraian setiap tahapan pengolahan data mining yang dilakukan.

3.1 Pemahaman Bisnis

Fase pemahaman bisnis, juga dikenal sebagai fase pemahaman penelitian bertujuan untuk memahami permasalahan dan menetapkan tujuan dilakukannya data mining. Sebagaimana permasalahan yang telah diuraikan dalam latar belakang, maka penerapan data mining pada penelitian ini bertujuan untuk mendapatkan hasil prediksi penyakit Hepatitis C untuk mengetahui apakah pasien tersebut menderita Hepatitis C atau tidak dengan menggunakan algoritma Naïve Bayes, yaitu menggunakan perhitungan probabilitas dan untuk mengetahui akurasi dari algoritma tersebut.

3.2 Pemahaman Data

Tahap kedua adalah pemahaman data. Tahap ini bertujuan untuk memahami karakteristik data yang akan diolah. Adapun data yang akan digunakan dalam proses data mining adalah sampel data dari situs kaggle sebanyak 615 data.

DOI: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>

Category	Age	Sex	ALB	ALP	ALT	AST
0=Blood Donor	32	m	38.5	52.5	7.7	22.1
0=Blood Donor	32	m	38.5	70.3	18	24.7
0=Blood Donor	32	m	46.9	74.7	36.2	52.6
0=Blood Donor	32	m	43.2	52	30.6	22.6
0=Blood Donor	32	m	39.2	74.1	32.6	24.8
0=Blood Donor	32	m	41.6	43.3	18.5	19.7
0=Blood Donor	32	m	46.3	41.3	17.5	17.8
0=Blood Donor	32	m	42.2	41.9	35.8	31.1
0=Blood Donor	32	m	50.9	65.5	23.2	21.2

Gambar 2. Dataset Awal

Adapun rincian atributnya adalah sebagai berikut.

No.	Atribut	Skala Pengukuran
1	Category (Diagnosa)	0=Blood Donor (0=donor darah) 3=Cirrhosis (penyakit hati stadium akhir) 2=Fibrosis (jaringan parut fibrotik) 1=Hepatitis (1=peradangan hati) 0s=Suspect Blood Donor (0s=dugaan Donor Darah)
2	Age (Usia)	[32-77]
3	Sex (Jenis Kelamin)	[m = Male f = Female]
4	ALB (jumlah albumin)	[14,9-82,2]
5	ALP (jumlah alkaline phosphatase)	[11,3-416,6]
6	ALT (jumlah alanin transaminase)	[0,9 – 325,3]
7	AST (jumlah aspartat)	[10,6 – 324]
8	BIL (jumlah bilirubin)	[0,8 – 209]
9	CHE (jumlah kolinesterase)	[1,42 -16,41]
10	CHOL (jumlah kolesterol)	[1,43 – 9,67]
11	CREA (jumlah kreatin)	[8 – 1079,1]
12	GGT (jumlah gamma-glutamyl transferase)	[4,5 – 650,9]
13	PROT (jumlah protein)	[44,8 – 86,5]

3.3 Persiapan Data

Dataset penyakit Hepatitis C memiliki beberapa atribut yang teridentifikasi sebagai

DOI: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>

missing value, sehingga data tersebut harus di proses terlebih dahulu dengan menghilangkan missing value. Proses membersihkan data disebut dengan preprocessing data. Data yang diperoleh dari dataset Kaggle terdiri dari 615 data, dari data tersebut diantaranya 29 data teridentifikasi sebagai data missing value pada atribut ALB, ALP, BIL, CHE, dan GGT. Kemudian data missing value dihilangkan untuk memvalidasi data sehingga data yang diproses tanpa missing value adalah 586 data.

Selanjutnya, setelah tahap preprocessing data, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Dalam hal ini Penulis menggunakan 579 data sebagai data latih (data training) dan 7 data sebagai data uji (data testing).

3.4 Pemodelan

Tools yang Microsoft Excel digunakan untuk melakukan perhitungan manual menggunakan, sedangkan untuk pengujiannya juga menggunakan Software Rapid Miner.

Pada penelitian ini akan diterapkan klasifikasi dengan menggunakan algoritma yang telah dipilih yaitu Naïve Bayes. Jumlah sampel pasien penyakit Hepatitis C yang dijadikan sampel untuk proses perhitungan algoritma Naïve Bayes untuk mengetahui pasien tersebut terklasifikasi menderita penyakit Hepatitis tipe C atau tidak yang dilihat berdasarkan probabilitas dari setiap kelas dan probabilitas setiap kejadian perkelas dan keakurasian yang dihasilkan.

Tahap-tahap pemodelan data menggunakan algoritma Naïve Bayes adalah sebagai berikut:

a. Menghitung Probabilitas Setiap Kelas

Dari data yang digunakan terdapat 5 class atau kategori, yaitu 0=Donor Darah, 0s=dugaan Donor Darah, 1=Hepatitis, 2=Fibrosis, 3=Cirrhosis. Dari data uji yang digunakan, class 0=Donor Darah berjumlah 520 data, class 0s=dugaan Donor Darah berjumlah 7 data, class 1=Hepatitis berjumlah 19 data, class 2=Fibrosis berjumlah 11 data, dan class 3=Cirrhosis berjumlah 22 data. Tabel 2 berikut merupakan hasil perhitungan probabilitas dari setiap kelas.

PROBABILITAS CLASS		
CLASS/CATEGORY		
0=Blood Donor		0,898100173
0s=suspect Blood Donor		0,01208981
1=Hepatitis		0,032815199
2=Fibrosis		0,018998273
3=Cirrhosis		0,037996546
Total Probabilitas		1

Gambar 3. Probabilitas Tiap Class/Kategori

b. Menghitung Probabilitas Atribut Nominal

Pada dataset yang digunakan, terdapat satu atribut yang bersifat nominal, yaitu atribut sex. Menghitung nilai probabilitas dari atribut sex dengan cara menghitung jumlah kejadian atau atribut dalam suatu kategori dibagi dengan kategori yang ada. Nilai Probabilitasnya dapat dilihat pada Gambar 4 berikut.

PROBABILITAS SEX		
m	f	Total
0,603846154	0,396153846	1
0,857142857	0,142857143	1
0,842105263	0,157894737	1
0,636363636	0,363636364	1
0,681818182	0,318181818	1

Gambar 4. Probabilitas Atribut Sex

DOI: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>

c. Menghitung Mean dan Standar Deviasi Atribut Numerik

Dari 12 atribut pada dataset Hepatitis C, 11 diantaranya adalah atribut yang bersifat numerik. Berbeda dengan atribut nominal, atribut numerik harus dihitung mean (rata-rata) dan standar deviasi setiap atribut barulah hasil dari mean dan standar deviasi tersebut dikalikan untuk mendapatkan prediksi penyakit Hepatitis C.

Hasil mean pada tiap atribut numerik dapat dilihat pada Gambar 5 berikut.

CLASS/CATEGORY	Age	MEAN									
		ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
0=Blood Donor	47,19231	42,18884615	68,15846154	26,74557692	26,50615385	8,584615385	8,384711538	5,501288462	79,19615385	29,16692308	72,06423077
0s=suspect Blood Donor	57,57143	24,4	107,3	102,1142857	71	4,685714286	7,48	4,451428571	61,71428571	151,5142857	53,91428571
1=Hepatitis	40,15789	43,78947368	41,86315789	21,24736842	83,55263158	16,68421053	9,327368421	5,163684211	74,39473684	107,1421053	74,61578947
2=Fibrosis	50,54545	43,36363636	37,36363636	18,89090909	85,99090909	14,45454545	8,331818182	4,846363636	69,76363636	63,02727273	74,25454545
3=Cirrhosis	52,95455	32,72727273	95,47272727	8,640909091	118,9954545	64,09090909	3,556363636	3,987272727	164,1227273	144,1590909	69,50454545

Gambar 5. Mean Tiap Atribut Numerik

Berikutnya standar deviasi setiap atribut numerik pada Gambar 6.

CLASS/CATEGORY	Age	STANDAR DEVIASI									
		ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
0=Blood Donor	9,661194	5,054980052	17,95461216	14,5217311	10,66509806	6,125051519	1,886990323	1,062496639	14,50785398	25,00588336	4,553944423
0s=suspect Blood Donor	11,07335	10,56077649	52,82155494	119,836909	53,31775814	3,113105936	4,642901392	1,869005338	52,85424702	133,7685118	11,71130915
1=Hepatitis	11,14681	3,583588322	24,46176632	15,7792187	75,48855225	14,95939924	2,684187943	1,536840954	21,54221885	127,4451007	5,013898811
2=Fibrosis	12,2993	3,264130122	8,539469858	19,39213524	50,72798941	6,89004552	1,415074685	0,695029103	10,26394395	36,17994171	5,273018801
3=Cirrhosis	7,581163	4,322497274	82,7155607	8,260385042	83,4898082	64,62412132	2,005372654	0,987921974	240,5769908	155,8654446	8,080869665

Gambar 6. Standar Deviasi Tiap Atribut

d. Mengalikan Semua Atribut Perkelas

Dari nilai probabilitas, mean, dan standar deviasi yang telah diperoleh dapat digunakan untuk memprediksi atau menentukan class atau kategori data testing dengan cara mengalikan setiap atribut perkelas.

Category	Sex	Age	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
?	f	64	24	102,8	2,9	44,4	20	1,54	3,02	63	35,9	71,3

Gambar 7. Contoh Data Testing

Dari salah satu contoh data Testing pada Gambar 7, kemudian dilakukan perhitungan manual menggunakan Microsoft Excel. Hasil perhitungan data testing dapat dilihat pada Gambar 8 sebagai berikut:

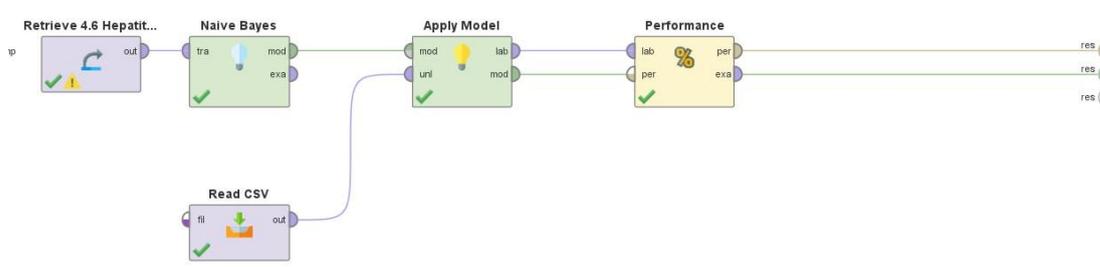
Category	Sex	Age	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT	TOTAL
?	f	64	24	102,8	2,9	44,4	20	1,54	3,02	63	35,9	71,3	
0=Blood Donor	0,396154	0,006224013	4,22981E-07	0,002276369	0,007063078	0,007319826	0,004999922	5,60967E-07	0,001656915	0,030127549	0,074218536	0,181800649	2,31881E-31
0s=suspect Blood D	0,142857	0,085607063	0,122616585	0,054508293	0,018367227	0,042607856	6,99408E-12	0,036039183	0,16235726	0,054855887	0,016346669	0,012871076	3,02135E-26
1=Hepatitis	0,157895	0,001231865	1,2021E-14	0,000162823	0,02599031	0,035095579	0,098225958	5,38482E-05	0,045996258	0,064992756	0,025861202	0,115079108	1,63411E-35
2=Fibrosis	0,363636	0,034385016	1,1498E-16	4,30619E-27	0,045908864	0,028606292	0,07953901	3,31954E-11	0,000479934	0,080681522	0,037812373	0,126953029	3,99026E-66
3=Cirrhosis	0,318182	0,017348448	0,003256496	0,043533001	0,085654309	0,0196568	0,031164764	0,102530853	0,153793593	0,021560639	0,019730182	0,133613861	3,68017E-17
Prediksi													3,68017E-17

Gambar 8. Perhitungan Data Testing Secara Manual

Berdasarkan hasil pe probabilitas terbesar dari hasil klasifikasinya. Dimana hasil perhitungan dari contoh data testing pada Tabel 6. Kelas 3=Cirrhosis memiliki nilai probabilitas terbesar yaitu 3,69017E-17.

Setelah didapatkan hasil pengujian menggunakan excel, selanjutnya pengujian menggunakan software Rapid Miner. Berikut Gambar 9 pemodelan di Rapid Miner.

DOI: <https://doi.org/10.52362/jmijayakarta.v3i2.1098>



Gambar 9. Pemodelan di Rapid Miner

Gambar 9 merupakan model yang terbentuk pada tools Rapid Miner yang digunakan untuk melakukan pengujian pada data uji (testing). Perbandingan data pada pengujian ini yaitu 579 data training dan 7 data testing.

Row No.	Category	prediction(C...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	Age	Sex	ALB
1	0=Blood Donor	0=Blood Donor	1.000	0.000	0.000	0.000	0.000	39	m	45.700
2	0=Blood Donor	0=Blood Donor	1.000	0.000	0.000	0.000	0.000	52	m	40.700
3	1=Hepatitis	0=Blood Donor	0.807	0.000	0.001	0.192	0.000	50	f	40
4	3=Cirrhosis	3=Cirrhosis	0.004	0.000	0.000	0	0.996	64	f	24
5	2=Fibrosis	2=Fibrosis	0	0.000	0.166	0.834	0.000	40	m	39
6	0=Blood Donor	0=Blood Donor	1.000	0.000	0.000	0.000	0.000	47	m	48
7	3=Cirrhosis	3=Cirrhosis	0.000	0.000	0.000	0.000	1.000	74	m	23

Gambar 10. Prediksi Klasifikasi Setiap Kelompok

3.5 Evaluasi

Tahap evaluasi bertujuan untuk mengetahui performa akurasi dari model yang telah dibangun. Evaluasi dilakukan pada Rapid Miner dengan menambahkan fungsi confusion matrix. Gambar 10 berikut merupakan hasil evaluasi yang dilakukan.

accuracy: 85.71%

	true 0=Blood Donor	true 1=Hepatitis	true 3=Cirrhosis	true 2=Fibrosis	true 0s=suspect Blood...	class precision
pred. 0=Blood Donor	3	1	0	0	0	75.00%
pred. 1=Hepatitis	0	0	0	0	0	0.00%
pred. 3=Cirrhosis	0	0	2	0	0	100.00%
pred. 2=Fibrosis	0	0	0	1	0	100.00%
pred. 0s=suspect Bloo...	0	0	0	0	0	0.00%
class recall	100.00%	0.00%	100.00%	100.00%	0.00%	

Gambar 11. Pemodelan di Rapid Miner

Hasil evaluasi menunjukkan bahwa akurasi dari algoritma Naïve Bayes sebesar 85,71%. Artinya akurasi algoritma Naïve Bayes dalam menentukan atau mengklasifikasi pasien Hepatitis C pada kategori Good Classification.

3.6 Penyusunan Laporan

Tahapan penyusunan laporan disertasi adalah mendokumentasikan apa yang relevan dengan pekerjaan penelitian yang dilakukan. Mencatat hasil penelitian dan menerjemahkannya ke dalam argumentasi yang disajikan dalam bentuk laporan dan dapat digunakan sebagai literatur.

4. Kesimpulan

Berdasarkan hasil rangkaian tahapan penelitian yang dilakukan, dapat disimpulkan bahwa proses data mining menggunakan algoritma Naïve Bayes dengan teknik klasifikasi dapat diterapkan dalam pengelompokan pasien untuk mengetahui apakah pasien tersebut akan menderita penyakit Hepatitis C atau tidak, dimana hasil akurasi yang diperoleh sebesar 85,71% yang berada dalam kategori good classification.

Referensi

- Adiba, F. (2021). Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naive Bayes. *Indonesian Journal of Fundamental Sciences*, 7(1), 18–28.
- Alhawaris. (2019). Hepatitis C: Epidemiologi, Etiologi, dan Patogenitas. *Jurnal Sains Dan Kesehatan*, 2(2), 139–150. <https://doi.org/10.25026/jsk.v2i2.132>
- Andika, L. A., Amalia, P., & Azizah, N. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. 2(1), 34–41.
- Daniel, T. (2005). *Discovering Knowledge in Data*.
- Hakim, S. H. F., Cholissodin, I., & Widodo, A. W. (2017). Seleksi Fitur Dengan Particle Swarm Optimization Untuk Pengenalan Pola Wajah Menggunakan Naive Bayes (Studi Kasus Pada Mahasiswa Universitas Brawijaya Fakultas Ilmu Komputer Gedung A). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(10), 1045–1057.
- Khomsah, S. (n.d.). *Prediksi Harapan Hidup Penderita Hepatitis Kronik Menggunakan Metode- Metode Klasifikasi*.
- Lanini, S., Easterbrook, P. J., Zumla, A., & Ippolito, G. (2016). Hepatitis C: global epidemiology and strategies for control. In *Clinical Microbiology and Infection* (Vol. 22, Issue 10, pp. 833–838). Elsevier B.V. <https://doi.org/10.1016/j.cmi.2016.07.035>
- Latu Handarko, J. (2015). Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis. *UJM*, 4(2). <http://journal.unnes.ac.id/sju/index.php/ujm>
- Milovic, B., & Milovic, M. (2012). Prediction and Decision Making in Health Care using Data Mining. *International Journal of Public Health Science (IJPHS)*, 1(2), 69–78.
- Mustopa, A. (2021). *Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization*. 2.
- R. A. Fajar and G. Testiana. (2023). *PENERAPAN DATA MINING UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER MENGGUNAKAN ALGORITMA NAÏVE BAYES*.
- Safdari, R., Deghatipour, A., Gholamzadeh, M., & Maghooli, K. (2022). Applying data mining techniques to classify patients with suspected hepatitis C virus infection. *Intelligent Medicine*. <https://doi.org/10.1016/j.imed.2021.12.003>
- Studi, P., & Informatika, M. (n.d.). *Lis Saumi Ramdhani*.
- Tinggi, S., Pati, T., Korespondensi, P., & Virus, H. C. (2022). *KLASIFIKASI HEPATITIS C VIRUS MENGGUNAKAN ALGORITMA C4 . 5 CLASSIFICATION OF HEPATITIS C VIRUS USING ALGORITHM C4 . 5*. 13(2), 43–48. <https://doi.org/10.34001/jdpt.v12i2>
- WHO. (2022). *Hepatitis C*. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>. 5/08/2022