



## KLASIFIKASI PENYAKIT HEPATITIS MENGGUNAKAN ALGORITMA CATBOOST

Syarif Muhammad Azdi Razi<sup>1</sup>, Barry Ceasar Octariadi<sup>2</sup>, Alda Cendekia Siregar<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer,  
Universitas Muhammadiyah Pontianak, Jl Jend Ahmad Yani No.111, (0561)764571, Indonesia  
<sup>1</sup>[201220102@unmuhpnk.ac.id](mailto:201220102@unmuhpnk.ac.id), <sup>2</sup>[barry.caesaro@unmuhpnk.ac.id](mailto:barry.caesaro@unmuhpnk.ac.id),  
<sup>3</sup>[alda.siregar@unmuhpnk.ac.id](mailto:alda.siregar@unmuhpnk.ac.id)

### Abstrak

Sekarang ini, isu kesehatan menjadi satu di antara permasalahan yang sangat kompleks di dunia modern yang terus berkembang. Salah satu penyakit yang menjadi ancaman kesehatan dunia ialah hepatitis. Diagnosis penyakit hepatitis sangat diperlukan untuk mengontrol, merawat, dan memulihkannya. Penggunaan *Machine learning* bersama dengan data klinis telah meningkatkan identifikasi dan klasifikasi penyakit secara signifikan, yang menunjukkan betapa efektifnya *Machine learning* dalam menangani berbagai masalah kesehatan. *Machine learning* dapat meningkatkan diagnosis hepatitis pada tahap awal, yang memungkinkan intervensi yang cepat dan hasil yang lebih baik pada pasien. Penelitian ini membangun model identifikasi menggunakan algoritma *machine learning* yaitu *Catboost*. Model dilatih menggunakan data set dari *UCI Machine Learning Repository*. Metode penelitian ini terdiri dari identifikasi masalah, pengumpulan data, proses *data mining*, pengujian, dan kesimpulan. Pengujian menggunakan *confusion matrix* yang berisi *accuracy*, *precision recall*, *f-1 score* dan skor AUC/ROC. Hasil dari pengujian tersebut didapatkan pada model pertama yang berisi *dataset training* 70% dan *testing* 30% terjadi peningkatan skor *accuracy*, *precision*, *recall*, *f1-score* dan skor AUC dari 0.77, 0.58, 0.56, 0.78 meningkat tinggi menjadi 0.946, 0.956, 0.949, 0.945, 0,981. Kemudian pada model kedua *dataset training* 80% dan *testing* 20% juga terjadi peningkatan dari 0.79, 0.60, 0.58, 0.59, 0.78 meningkat menjadi 0.960, 0.958, 0.958, 0.958, 0,991. **Kata kunci:** Hepatitis, *Data Mining*, *Machine Learning*, *Catboost*.

### Abstract

Nowadays, health issues have become one of the most complex problems in the ever-evolving modern world. One of the diseases that pose a threat to global health is hepatitis. Diagnosis of hepatitis disease is necessary to control, treat and recover from it. The use of ML together with clinical data has significantly improved the identification and classification of diseases, which shows how effective ML is in addressing various health issues. ML can improve the diagnosis of hepatitis at an early stage, which enables rapid intervention and better patient outcomes. This study builds an identification model using the machine learning algorithm *Catboost*. The model was trained using data sets from the *UCI Machine Learning Repository*. This research method consists of problem identification, data collection, data mining

### Article History:

Received: June 2025  
Reviewed: June 2025  
Published: June 2025

Plagiarism Checker No 234  
Prefix DOI:  
10.8734/Kohesi.v1i12.365  
Copyright: Author  
Publish by: Kohesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



process, testing, and conclusions. Testing uses a confusion matrix that contains accuracy, precision recall, f-1 score and AUC/ROC score. The results of the test were obtained in the 70% training dataset there was an increase in the accuracy score, precision, recall, f1-score and AUC score from 0.77, 0.58, 0.56, 0.56, 0.78 increasing to 0.946, 0.956, 0.949, 0.945, 0.981. Then on the 80% training dataset, there was an increase in accuracy, precision, recall, f1-score and AUC score from 0.79, 0.60, 0.58, 0.59, 0.78 increasing to 0.960, 0.958, 0.958, 0.958, 0.991.  
**Keywords:** Hepatitis, Data Mining, Machine Learning, Catboost.

## 1. PENDAHULUAN

Sekarang ini, isu kesehatan menjadi satu di antara permasalahan yang sangat kompleks di dunia modern yang terus berkembang. Kompleksitas ini dibuktikan dengan isu kesehatan yang terus mengalami perubahan dan evolusi seiring berjalannya waktu. Berbagai pengaruh seperti perkembangan teknologi medis, perubahan gaya hidup, demografi populasi, dan masalah sosial dan ekonomi yang terkait. Saat kesehatan tidak dijaga dengan benar, berbagai masalah kesehatan dapat muncul. Tentu saja hal ini dapat berdampak buruk pada kualitas hidup manusia. Salah satu penyakit yang menjadi ancaman kesehatan dunia ialah hepatitis [1].

Hepatitis merupakan penyakit yang disebabkan oleh penyebaran virus yang dapat merusak dan melemahkan sel-sel yang menginfeksi jaringan hati sehingga membuatnya tidak berfungsi [2]. Penggunaan alkohol berlebihan, autoimun, obat-obatan, atau racun menjadi penyebab hepatitis [3]. Akan tetapi, penyebab paling umum dari penyakit ini disebabkan oleh infeksi virus yang dikenal sebagai virus hepatitis. Hepatitis mempunyai 5 tipe berdasarkan tipe virus yaitu hepatitis A, B, C, D, dan E [4]. Menurut informasi dari Organisasi Kesehatan Dunia (WHO), pada tahun 2023, terdapat 350 juta orang di dunia yang mengalami infeksi hepatitis. Setiap harinya 3.000 orang meninggal disebabkan oleh hepatitis yang artinya setiap tiga puluh detik terdapat satu kematian hepatitis [5].

Diagnosis penyakit hepatitis sangat diperlukan untuk mengontrol, merawat, dan memulihkannya [6]. Diagnosis tradisional tergantung terhadap pada serangkaian tes laboratorium, pencitraan medis, dan evaluasi riwayat pasien. Akan tetapi, metode tradisional ini membutuhkan waktu yang lama dalam memproses kumpulan data yang berukuran besar tersebut [7][8]. Selain itu, banyak faktor yang perlu dipertimbangkan dalam mendiagnosis karena merupakan tugas yang cukup berat [2]. Maka, di sinilah kelebihan dari *machine learning*, yaitu menawarkan solusi yang menjanjikan untuk mengatasi tantangan ini.

Penggunaan *machine learning* bersama dengan data klinis telah meningkatkan identifikasi dan klasifikasi penyakit secara signifikan, yang menunjukkan betapa efektifnya *machine learning* dalam menangani berbagai masalah kesehatan [9][10]. *Machine learning* dapat meningkatkan diagnosis hepatitis pada tahap awal, yang memungkinkan intervensi yang cepat dan hasil yang lebih baik pada pasien [11]. *Machine learning* juga dapat membantu dokter dalam pengambilan keputusan karena efisiensi dan keakuratannya [12]. Salah satu algoritma *machine learning* untuk melakukan klasifikasi adalah *Catboost*.

Algoritma *Catboost* adalah kerangka kerja yang kuat dan sangat efisien yang banyak digunakan dalam aplikasi *machine learning* [13]. Ini melampaui algoritma *gradient boosting* klasik, terutama untuk kumpulan data yang rumit dengan banyak variabel kategori. Kelebihan dari algoritma *Catboost* adalah kemampuannya yang secara efektif mengelola interaksi fitur sambil mengurangi *overfitting*, sehingga menghasilkan kekuatan prediksi yang luar biasa [14].



Kapabilitas metode *Catboost* dalam klasifikasi penyakit telah teruji dan terbukti dari penelitian Cendi & Wibowo, penelitian ini melakukan komparasi 3 metode *Ensemble Learning* yaitu *Bagging*, *Boosting*, dan *Stacking* menggunakan 3 *dataset* dalam memprediksi diabetes. *Dataset* yang dipergunakan berasal dari *Frankfurt Hospital Diabetes*, *Sylhet Hospital Diabetes*, dan *Pima Indians Diabetes*. Hasil yang didapatkan dari ketiga *dataset*, diperoleh bahwa metode *Boosting* melampaui metode *Bagging* dan *Stacking*. Pada *dataset* 1, model *Gradient Boosting*, *Extreme Gradient Boosting*, dan *Catboosting* memperoleh akurasi tertinggi sebesar 81.82 %. Pada *dataset* 2, model *Light Gradient Boosting* memperoleh akurasi tertinggi sebesar 99.25%. Sedangkan pada *dataset* 3 model *Light Gradient Boosting* dan *Cat Boosting* memperoleh akurasi tertinggi sebesar 100% [15].

Berdasarkan latar belakang yang telah dipaparkan, maka peneliti mengangkat penelitian ini dengan judul “Klasifikasi Penyakit Hepatitis Menggunakan Metode *CatBoost*”. Diharapkan dengan penelitian ini mampu memberikan solusi yang akurat sesuai permasalahan yang diangkat.

## 2. METODE

Metode penelitian yang digunakan pada penelitian ini dibuat untuk memberikan gambaran sistematis dan jelas bagaimana penelitian ini akan dilakukan. Metode penelitian ini dapat dilihat pada diagram alir yang dipaparkan pada gambar 1 di bawah ini:



Gambar 1. Metode Penelitian

### 2.1 Studi Literatur

Pada langkah ini, dilakukan pencarian yang terkait dengan penelitian skripsi. Studi literatur dilakukan untuk mengumpulkan bahan referensi. Pencarian literatur, seperti jurnal penelitian terdahulu, data-data, dan buku, dalam penelitian ini bertujuan untuk memperbanyak analisis data, menguatkan landasan teori, dan mendukung argumentasi dalam penelitian ini.



## 2.2 Identifikasi Masalah

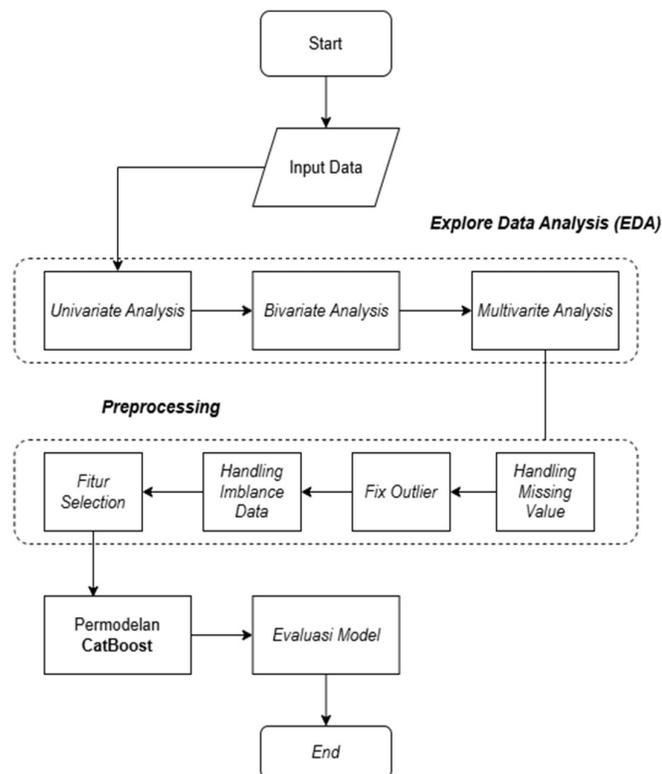
Pada langkah ini, penelitian diawali dengan mengidentifikasi masalah yang ada, kemudian membuat rumusan masalah agar penelitian lebih terarah. Kemudian, membuat tujuan penelitian yang berdasarkan rumusan masalah yang telah dibuat [16]. Permasalahan pada penelitian ini ialah bagaimana cara memprediksi penyakit hepatitis menggunakan *data mining*, dari permasalahan tersebut peneliti membuat model terbaik yang bisa mengklasifikasi agar dapat memprediksi penyakit hepatitis menggunakan *machine learning*.

## 2.3 Pengumpulan Data

Pengumpulan data dalam penelitian melibatkan pemanfaatan dari berbagai sumber data yang ada. Data yang didapatkan langsung oleh peneliti dari hasil penelitiannya seperti observasi, wawancara, ataupun eksperimen, dikategorikan sebagai data primer, sedangkan data sekunder adalah data yang diperoleh secara tidak langsung dari dokumen, jurnal, buku, dan informasi lainnya yang terkait dengan topik penelitian [17][18]. Sumber data yang dipergunakan dalam penelitian ini adalah data sekunder yang didapatkan dari *Machine Learning Repository UCI (University California, Invene)* dengan web <https://archive.ics.uci.edu/dataset/46/hepatitis> [19].

## 2.4 Proses Data Mining

Proses *data mining* merupakan suatu metode untuk menjelajahi data dan mengidentifikasi pola atau informasi yang bermanfaat dari *dataset* tersebut [20]. Prosesnya dapat dilihat pada Gambar 2. di bawah ini:



Gambar 2. Proses Data Mining



- 1) *Exploratory Data Analysis (EDA)*: Setelah mendapatkan *dataset*, langkah pertama adalah pengolahan *data mining* adalah dimulai dengan pengolahan *Exploratory Data Analysis (EDA)*. Tahap memahami data digunakan untuk menganalisis setiap fitur *dataset* EDA, yang terdiri dari *Univariate Analysis*, *Bivariate Analysis*, dan *Multivariate Analysis* [21].
- 2) *Preprocessing*: *Preprocessing* dilakukan untuk memastikan kualitas data dan mengekstrak informasi berguna dari kumpulan data, sehingga mempengaruhi efisiensi model dan memastikan kualitas data untuk model [22]. Pada tahap ini kita menyiapkan *dataset* sebelum dilatih oleh model dengan membersihkan *missing value* dengan mengubahnya menjadi nilai *mean*, kemudian menangani *outlier*, *handling data imbalance* dan fitur *selection*.
- 3) *Pemodelan CatBoost*: Setelah dilakukan *preprocessing*, selanjutnya pada penelitian ini dilakukan membuat model dengan algoritma *CatBoost*.
- 4) *Evaluation Model*: Kemudian setelah dilakukan pengecekan performa dari pemodelan *CatBoost*, peneliti menilai hasil akurasi tersebut untuk memastikan apakah hasil tersebut sesuai dengan harapan peneliti.

## 2.5 Pengujian

Dalam penelitian ini, berbagai tes dilakukan dengan menggunakan *Matrix Confusion*, skor AUC dan ROC:

- 1) *Pengujian 1*: Pada pengujian pertama ini dilakukan *splitting* data sebesar 70% data training dan 30% data testing kemudian dilakukan pengujian sebelum melakukan persiapan data dan sesudah melakukan persiapan data.
- 2) *Pengujian 2*: Pada pengujian kedua ini dilakukan *splitting* data sebesar 80% data training dan 20% data testing kemudian dilakukan pengujian sebelum melakukan persiapan data dan sesudah melakukan persiapan data.

## 2.6 Kesimpulan

Kemudian setelah analisis hasil pengujian yang sudah dilakukan maka didapatkan suatu kesimpulan. Kesimpulan ini ialah hasil dari evaluasi dari semua data yang telah diolah.

## 3. HASIL DAN DISKUSI

### 3.1 *Exploratory Data Analysis (EDA)*

Pada hasil ini terdapat tiga yang pertama *univariate analysis*. Analisis ini dilakukan pada fitur-fitur di *dataset* penyakit hepatitis yang mana hanya satu fitur yang terlibat. Hasil visualisasi pada *univariate analysis* dapat dilihat pada Gambar 3.

	count	mean	std	min	25%	50%	75%	max
age	155	41.2	12.56587835	7.0000	32.0000	39.0000	50.0000	78.0000
bilirubin	149	1.427516779	1.212149033	0.3000	0.7000	1.0000	1.5000	8.0000
alk_phosphate	126	105.3253968	51.50810878	26.0000	74.2500	85.0000	132.2500	295.0000
sgot	151	85.89403974	89.65088974	14.0000	31.5000	58.0000	100.5000	648.0000
albumin	139	3.817266187	0.651523083	2.1000	3.4000	4.0000	4.2000	6.4000
protime	88	61.85227273	22.87524374	0.0000	46.0000	61.0000	76.2500	100.0000

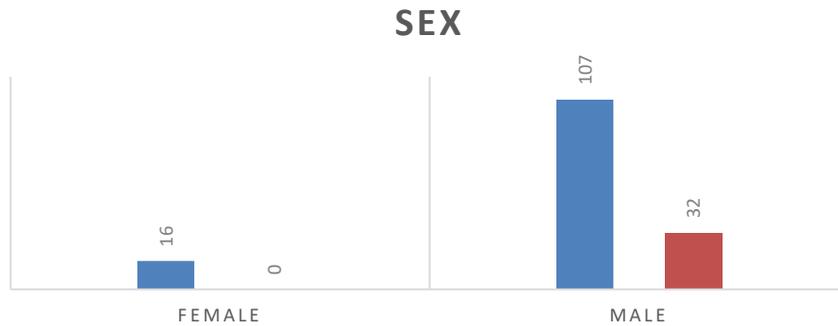
Gambar 3. Deskripsi fitur numerik

Pada *dataset* awal sebelum proses menangani fitur-fitur kategorik yang merubah fitur-fitur tersebut menjadi numerik, fitur-fitur pada Gambar 3 adalah fitur awal numerik pada *dataset*. Pada gambar tersebut dapat dilihat nilai rata-rata, standar deviasi, min, dan maks pada fitur.

Kedua *Bivariate Analysis* yang mana analisis ini dilakukan kepada dua variabel untuk mengetahui korelasi antar keduanya. Pada penelitian ini analisis dilakukan terhadap fitur dan label yang mana dapat memberitahukan seberapa terkaitnya fitur tersebut terhadap label.

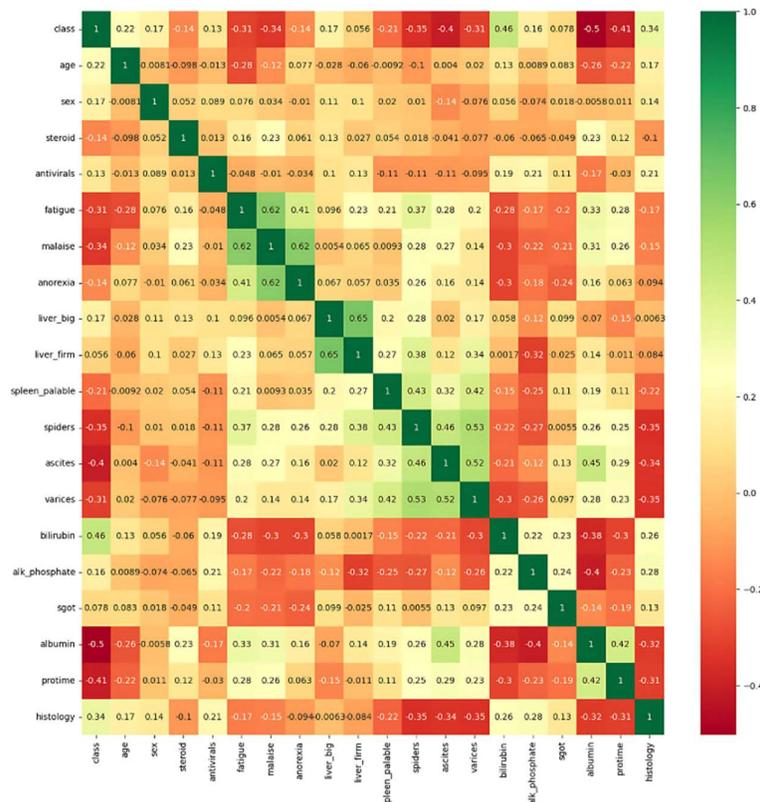


Salah satu contoh analisis pada penelitian ini yaitu antara sex dan label dapat dilihat pada Gambar 4.



Gambar 4. Antara sex dan label

Pada Gambar 4 dapat dilihat bahwa jenis kelamin memiliki pengaruh terkena hepatitis atau tidak hepatitis. Perbandingan antara wanita dan pria yang terkena hepatitis, 0% dari total wanita terkena hepatitis, 21% dari total pria terkena hepatitis. Berdasarkan data tersebut pria memiliki resiko lebih tinggi dibandingkan wanita. Bar berwarna *orange* menandakan terkena hepatitis. Bar warna biru muda menandakan tidak terkena hepatitis.



Gambar 5. Korelasi antar fitur dengan heatmap

Ketiga *Multivariate analysis*, dapat dilihat pada Gambar 5 dapat dilihat korelasi antar fitur-fitur dengan menggunakan *heatmap*, semakin menjauh dari 0 maka semakin besar korelasi antar fitur, nilai yang di bawah 0 disebut *negative correlation* dan *positive correlation* jika nilai berada di atas 0. Gambar 5 menunjukkan beberapa fitur dengan korelasi pada penyakit hepatitis yang sangat mendekati 0 yaitu *antivirals*, *liver\_firm*, dan *sgot*.



### 3.2 Hasil Preprocessing

#### 3.2.1 Missing Value

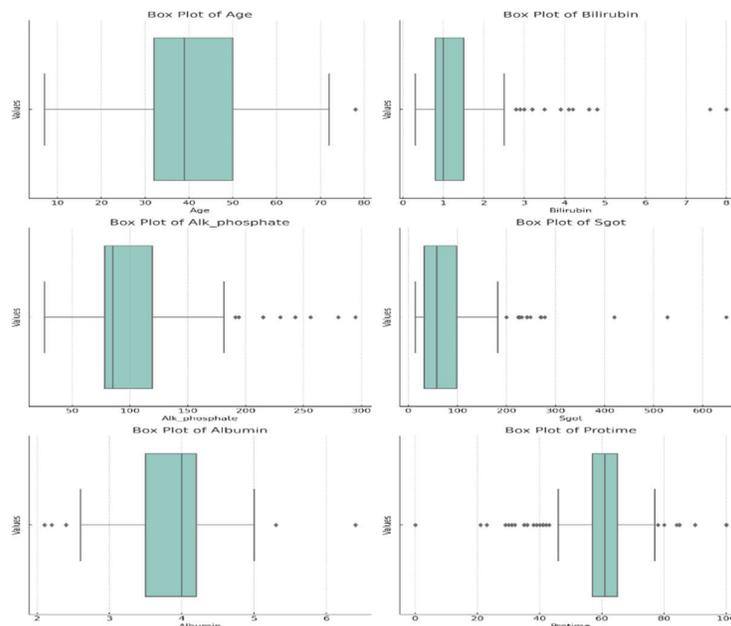
Pada tahap pertama dilakukan pengecekan *missing value* pada *dataset*, setelah dianalisis terdapat banyak sekali *missing value* ada 16 fitur yang memiliki *missing value*. Hal ini diharuskan untuk melakukan penanganan agar data menjadi lebih baik. Pada penelitian ini dilakukan penanganan *missing value* dengan merubah data yang hilang menggunakan modus untuk data yang bertipe kategorik seperti *steroid*, *fatigue*, dan lain-lain, kemudian merubah data yang hilang menggunakan *mean* untuk tipe data numerikal seperti *bilirubin*, *alk\_phosphate*, dan seterusnya dapat dilihat pada Gambar 6.

```
class      0
age        0
sex        0
steroid    1
antivirals 0
fatigue    1
malaise    1
anorexia   1
liver_big  10
liver_firm 11
spleen_palable 5
spiders    5
ascites    5
varices    5
bilirubin  6
alk_phosphate 29
sgot       4
albumin    16
protime    67
histology  0
dtype: int64
```

Gambar 6. Data missing value

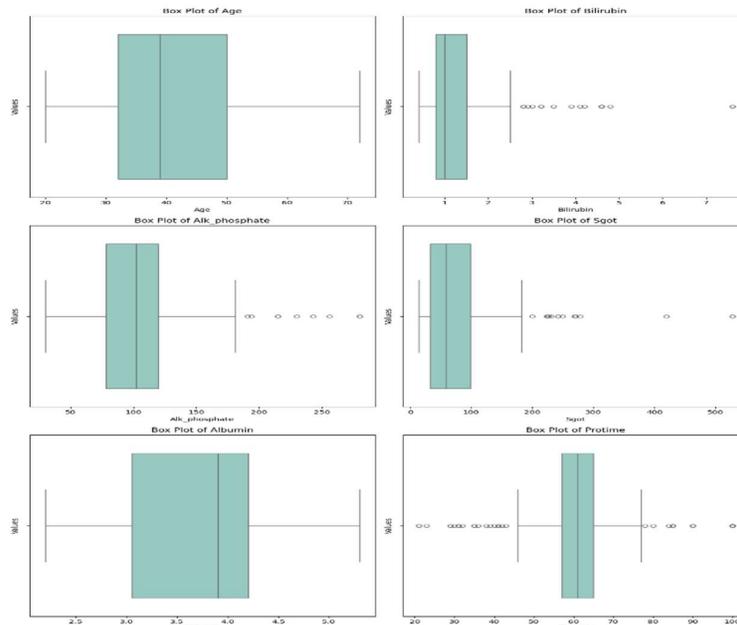
#### 3.2.2 Outlier

Kemudian dilakukan penanganan *outlier* yang mana jika terdapat *outlier* pada data bisa mempengaruhi kualitas data yang berdampak pada akurasi model.



Gambar 7. Data outlier

Gambar diatas adalah *box plot* yang digunakan untuk melihat *outlier*, dapat dilihat terdapat beberapa fitur yang memiliki *outlier* yang cukup ekstrim seperti *alk\_phosphate*, *sgot*, *protime*. Oleh karena itu peneliti melakukan penanganan *outlier*.



Gambar 8. Penanganan *outlier* menggunakan *capping* dengan persentil

Gambar 8 adalah hasil melakukan *capping* dengan persentil yang mana tidak membuang terlalu banyak *outlier*, disebabkan *dataset* pada penelitian ini relatif kecil yang mana jika dilakukan penghapusan *outlier* terlalu membuang banyak data dan informasi penting. Penanganan *outlier* menggunakan *capping* dengan persentil untuk mempertahankan informasi berharga dari data dan tidak merubah nilai asli agar tidak menurunkan akurasi dari pemodelan *catboost*.

### 3.2.3 Imbalance Data

Label tidak seimbang pada *dataset* menyebabkan kemampuan prediksi model akan rendah pada label kelas data yang berjumlah kecil dan baik dengan label kelas dengan jumlah yang besar. Setelah mendapatkan informasi pada tahap EDA yang menunjukkan label kelas pada data sangat tidak seimbang, sehingga pada tahap data *preprocessing* ketidakseimbangan pada kelas ditangani dengan *Resampling* data.

Pada penelitian ini dilakukan *resampling* data dengan menggunakan *Oversampling SMOTE Method*. Sehingga, kelas label yang minoritas akan bertambah jumlahnya untuk mengatasi ketidakseimbangan data yang cukup besar pada label *dataset*. Awalnya jumlah kelas label 1 yang berjumlah 32 menjadi 123. Setelah melalui *oversampling*, persentase perbandingannya menjadi seimbang yaitu 50%.

### 3.2.4 Seleksi Fitur

Pada tahap akhir dilakukan seleksi fitur. Setelah mengetahui fitur-fitur yang mana saja yang mempunyai korelasi yang sangat rendah dibandingkan fitur lainnya melalui *multivariate analysis*, fitur-fitur tersebut akan dihapus untuk meningkatkan performa model. Fitur yang dihapus adalah *antivirals*, *liver\_firm*, dan *sgot*. Fitur-fitur tersebut setelah dilakukan analisis mempunyai korelasi yang rendah pada variabel target. Fitur-fitur pada *dataset* setelah dilakukan seleksi berjumlah 16 dari 19 fitur.



### 3.3 Hasil Membangun Model

Penelitian membangun dua model, model pertama dengan data *training* 70% dan *testing* 30% lalu model kedua data *training* 80% dan *testing* 20% pelatihan model terbagi menjadi dua tahap yaitu; tahap awal dan tahap akhir. Sebelum model dilatih dengan data yang telah melalui persiapan data. Model akan dilatih terlebih dahulu untuk membandingkan performa model pada tahap ini dengan tahap berikutnya.

Tabel 1. Evaluasi model tahap awal

Model	Accuracy	Precision	Recall	F1-score	AUC
Model Pertama	0.77	0.58	0.56	0.56	0,78
Model Kedua	0.74	0.60	0.58	0.59	0,78

Pada tabel 1 dapat dilihat performa model yang dilatih dengan data 70% *training* dan 30 % *testing* (model pertama) memiliki skor *accuracy* dan skor AUC cukup tinggi, tetapi skor *precision*, *recall* dan *f1-score* yang rendah. Skor *precision*, *recall* dan *f1-score* yang rendah tersebut memberikan informasi bahwa model masih banyak melakukan kesalahan klasifikasi label kelas 1 yang berarti seseorang terkena hepatitis. Kemudian performa model yang dilatih dengan data 80% *training* dan 20 % *testing* (model kedua) menunjukkan skor akurasi sedikit lebih rendah dibandingkan dengan model kedua tetapi skor *precision*, *recall* dan *f1-score* sedikit meningkat.

Setelah melakukan pelatihan model tahap awal dilanjutkan dengan tahap akhir. Tahap ini merupakan tahapan pelatihan model yang dilakukan menggunakan *dataset* yang telah dilakukan optimalisasi dengan melalui persiapan data. Penyebab performa model dalam memprediksi dipengaruhi oleh *dataset* yang telah dilakukan berbagai proses dalam persiapan data seperti menangani *missing value*, menangani *outlier*, menangani data *imbalance*, dan seleksi fitur.

Tabel 2. Evaluasi model tahap akhir

Model	Accuracy	Precision	Recall	F1-score	AUC
Model Pertama	0.946	0.956	0.949	0.945	0.981
Model Kedua	0.960	0.958	0.958	0.958	0.991

Pada tabel 2 dapat dilihat bahwa kedua model mendapatkan peningkatan yang tinggi setelah dilatih dengan menggunakan *dataset* yang telah dilakukan optimalisasi data. Pada model pertama dapat kita bandingkan evaluasi model sebelum dan setelah dilakukan optimalisasi data yang mana pada tahap awal skor *accuracy*, *precision*, *recall*, *f1-score* dan skor AUC hanya sebesar 0.77, 0.58, 0.56, 0.56, 0.78 meningkat tinggi menjadi 0.946, 0.956, 0.949, 0.945, 0,981. Hal tersebut dapat disimpulkan bahwa optimalisasi data sangat berpengaruh terhadap performa model.

Performa model Kedua juga mengalami peningkatan yang sangat tinggi dibandingkan sebelumnya yang mana skor *accuracy*, *precision*, *recall*, *f1-score* menjadi 0.960, 0.958, 0.958, 0.958, 0,991. Jika dibandingkan dengan model pertama, performa model ini sedikit lebih baik.



#### 4. KESIMPULAN

Berdasarkan hasil analisis dan pengujian, model *Catboost* dapat dibangun dan mampu mengklasifikasi pasien terkena penyakit hepatitis atau tidak. Kinerja model *Catboost* meningkat signifikan melalui persiapan data yang efektif, dibuktikan pada dua skenario pelatihan. Dalam model pertama dengan data *training* 70% dan *testing* 30%, skor akurasi, presisi, *recall*, *f1-score*, dan AUC meningkat dari 0.77, 0.58, 0.56, 0.56, 0.78 menjadi 0.946, 0.956, 0.949, 0.945, dan 0.981. Pada model kedua dengan data *training* 80% dan *testing* 20%, terjadi peningkatan serupa, dari 0.79, 0.60, 0.58, 0.59, 0.78 menjadi 0.960, 0.958, 0.958, 0.958, dan 0.991 yang mana ini merupakan hasil terbaik.

#### REFERENSI

- [1] D. C. P. Buani, "Prediksi Penyakit Hepatitis Menggunakan Algoritma Naïve Bayes Dengan Seleksi Fitur Algoritma Genetika," *Evolusi: Jurnal Sains dan Manajemen*, vol. 6, no. 2, pp. 1-5, 2018, doi: 10.31294/evolusi.v6i2.4381.
- [2] I. I. Ahmed, D. Y. Mohammed, and K. A. Zidan, "Diagnosis of hepatitis disease using machine learning techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 3, pp. 1564-1572, 2022, doi: 10.11591/ijeecs.v26.i3.pp1564-1572.
- [3] P. Mehta, L. M. Grant, and A. Kumar Reddy Reddivari, "Viral Hepatitis," StatPearls Publishing. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK554549/>
- [4] G. Suwardika, "Pengelompokan Dan Klasifikasi Pada Data Hepatitis Dengan Menggunakan Support Vector Machine (SVM), Classification And Regression Tree (Cart) Dan Regresi Logistik Biner," *Journal of Education Research and Evaluation*, vol. 1, no. 3, p. 183, 2017, doi: 10.23887/jere.v1i3.12016.
- [5] W. H. O. (WHO), *World Hepatitis Day*, World Health Organization (WHO).
- [6] K. Maoulaei, H. Sharifi, A. A. Bahaadinbeigy, Kambiz, Haghdoost, and N. Nasiri, "Machine learning for prediction of viral hepatitis: A systematic review and meta-analysis," *Int J Med Inform*, vol. 179, p. 105243, 2023, doi: 10.1016/j.ijmedinf.2023.105243.
- [7] D. Zhang, X. Liu, M. Shao, Y. Sun, Q. Lian, and H. Zhang, "The value of artificial intelligence and imaging diagnosis in the fight against COVID-19," *Pers Ubiquitous Comput*, vol. 27, no. 3, pp. 783-792, 2023, doi: 10.1007/s00779-021-01522-7.
- [8] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor," *IEEE Access*, vol. 8, pp. 55135-55144, 2020, doi: 10.1109/ACCESS.2020.2978629.
- [9] C. Krittanawong *et al.*, "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Sci Rep*, vol. 10, no. 1, pp. 1-11, 2020, doi: 10.1038/s41598-020-72685-1.
- [10] M. Sayadi, V. Varadarajan, E. Gozali, and M. Sadeghi, "Effective factors in diagnosing the degree of hepatitis C using machine learning," *Frontiers in Health Informatics*, vol. 12, 2023, doi: 10.30699/fhi.v12i0.440.
- [11] A. Alizargar, Y. L. Chang, and T. H. Tan, "Performance Comparison of Machine Learning Approaches on Hepatitis C Prediction Employing Data Mining Techniques," *Bioengineering*, vol. 10, no. 4, 2023, doi: 10.3390/bioengineering10040481.
- [12] A. S. Ahuja, "The impact of artificial intelligence in medicine on the future role of the physician," *PeerJ*, vol. 2019, no. 10, 2019, doi: 10.7717/peerj.7702.
- [13] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00369-8.
- [14] M. Hu, H. Peng, X. Zhang, L. Wang, and J. Ren, "Building gender-specific sexually transmitted infection risk prediction models using CatBoost algorithm and NHANES data," *BMC Med Inform Decis Mak*, vol. 24, no. 1, pp. 1-8, 2024, doi: 10.1186/s12911-024-02426-1.



- [15] L. M. Cendani and A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *Jurnal Masyarakat Informatika*, vol. 13, no. 1, pp. 33-44, 2022, doi: 10.14710/jmasif.13.1.42912.
- [16] R. Ahsana, R. Rohmat Saedudin, and V. P. Widartha, "Perbandingan Akurasi Algoritma Adaboost Dan Algoritma Lightgbm Untuk Klasifikasi Penyakit Diabetes," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 9738-9748, 2021.
- [17] Supriyono and F. Erida, "Sistem Pakar Penegakan Diagnosa Penyakit Hipertensi Dengan Inferensi Forward Chaining Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Kesehatan Medika Udayana*, vol. 8, no. 02, pp. 207-221, 2022, doi: 10.47859/jmu.v8i02.283.
- [18] H. A. Chandra, "Particle Swarm Optimization Pada Metode Knn Euclidean Distance Berbasis Variasi Jarak Untuk Penilaian," *Technologia: Jurnal Ilmiah*, vol. 9, no. 1, p. 59, 2018, doi: 10.31602/tji.v9i1.1103.
- [19] G. Gong, "Hepatitis," UCI Machine Learning Repository. [Online]. Available: <https://archive.ics.uci.edu/dataset/46/hepatitis>
- [20] I. N. Abrar, A. Abdullah, and S. Sucipto, "Liver Disease Classification Using the Elbow Method to Determine Optimal K in the K-Nearest Neighbor (K-NN) Algorithm," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 218-228, 2023, doi: 10.32736/sisfokom.v12i2.1643.
- [21] G. Szabo, G. Polatkan, P. Oscar Boykin, and A. Chalkiopoulos, *Social Media Data Mining and Analytics*. 2018.
- [22] C. Y, P. Kiran, and M. P B, "The Novel Method for Data Preprocessing CLI," *Advances in Intelligent Systems and Technologies*, pp. 117-120, 2022, doi: 10.53759/aist/978-9914-9946-1-2\_21.