

**ANALISIS PERFORMA ALGORITMA CATBOOST DALAM
KLASIFIKASI DIAGNOSIS KANKER PROSTAT
MENGUNAKAN DATA FITUR MORFOLOGI**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh

IZZA DHINILLAH

09040222057

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2026

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Izza Dhinillah
NIM : 09040222057
Program Studi : Matematika
Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "Analisis Performa Algoritma CatBoost dalam Klasifikasi Diagnosis Kanker Prostat Menggunakan Data Fitur Morfologi". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 17 Juni 2026

Yang menyatakan,



Izza Dhinillah

NIM. 09040222057

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : Izza Dhinillah
NIM : 09040222057
Judul proposal skripsi : Analisis Performa Algoritma CatBoost dalam
Klasifikasi Diagnosis Kanker Prostat Menggunakan
Data Fitur Morfologi

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Aris Fanani, M.kom
NIP. 198701272014031002

Pembimbing II



Ahmad Lybab, M.Si
NIP. 198111182009121003

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Dr. Yuniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

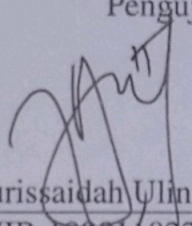
Skripsi oleh

Nama : Izza Dhinillah
NIM : 09040222057
Judul Skripsi : Analisis Performa Algoritma CatBoost dalam
Klasifikasi Diagnosis Kanker Prostat Menggunakan
Data Fitur Morfologi

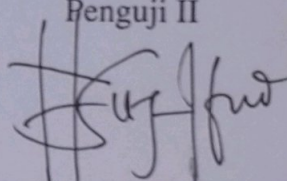
Telah dipertahankan di depan Tim Penguji
pada tanggal 17 Juni 2026

Mengesahkan,
Tim Penguji

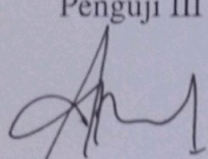
Penguji I


Nurissaidah Ulinnuha, M. kom
NIP. 199011022014032004

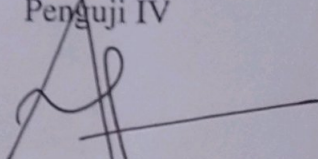
Penguji II


Dr. Yuniar Farida, M.T
NIP. 197905272014032002

Penguji III


Aris Fanani, M.kom
NIP. 198701272014031002

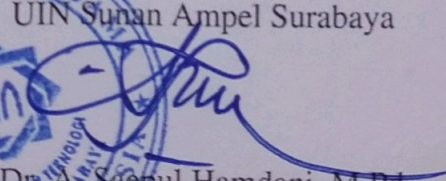
Penguji IV


Ahmad Lubab, M.Si
NIP. 198111182009121003

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya




Dr. A. Saepul Hamdani, M.Pd
NIP. 196507312000031002

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : 122A DHINILLAH
NIM : 09040232057
Fakultas/Jurusan : FAKULTAS SAINS DAN TEKNOLOGI / MATEMATIKA
E-mail address : dhinillahj33a@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Skripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

ANALISIS PERFORMA ALGORITMA CATBOOST DALAM
KLASIFIKASI DIAGNOSIS KANKER PROSTAT MENGGUNAKAN
DATA FITUR MORFOLOGI

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 17 Juni 2026

Penulis



(122A DHINILLAH)
nama terang dan tanda tangan

ABSTRAK

Analisis Performa Algoritma CatBoost dalam Klasifikasi Diagnosis Kanker Prostat Menggunakan Data Fitur Morfologi

Kanker prostat merupakan salah satu jenis kanker yang sering terjadi pada pria. Perbedaan karakteristik morfologi sel, seperti ukuran, bentuk, tekstur, dan struktur inti sel, dapat digunakan untuk membedakan jaringan prostat yang bersifat jinak (benign) dan ganas (malignant). Banyaknya fitur morfologi yang digunakan menyebabkan proses analisis secara manual menjadi kurang efisien sehingga diperlukan metode klasifikasi yang mampu mengenali pola data secara akurat. Penelitian ini bertujuan untuk menganalisis performa algoritma CatBoost dalam klasifikasi kanker prostat berdasarkan fitur morfologi serta mengidentifikasi fitur yang paling berpengaruh menggunakan analisis feature importance. Dataset yang digunakan terdiri atas 100 data pasien dengan 8 variabel independen berupa fitur morfologi sel dan 1 variabel dependen berupa kelas kanker prostat, yaitu jinak dan ganas. Tahapan penelitian meliputi preprocessing data, penanganan missing value menggunakan median, pembagian data menggunakan metode 5-Fold Cross Validation, pemodelan menggunakan algoritma CatBoost, serta evaluasi model menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa model CatBoost menghasilkan performa terbaik pada kombinasi parameter learning rate 0,2 dan depth 4 dengan nilai akurasi sebesar 92%. Hasil analisis feature importance menunjukkan bahwa fitur Perimeter merupakan fitur yang paling berpengaruh dalam proses klasifikasi, diikuti oleh fitur-fitur morfologi lainnya yang turut berkontribusi dalam membedakan kelas kanker prostat jinak dan ganas. Hasil penelitian ini menunjukkan bahwa algoritma CatBoost memiliki kemampuan yang sangat baik dalam mengklasifikasikan kanker prostat berdasarkan karakteristik morfologi sel.

Kata kunci: CatBoost, Kanker Prostat, Fitur Morfologi, Klasifikasi, Feature Importance, Machine Learning.

ABSTRACT

Prostate cancer is one of the most common types of cancer affecting men. Differences in cellular morphological characteristics, such as size, shape, texture, and nuclear structure, can be used to distinguish between benign and malignant prostate tissues. The large number of morphological features involved makes manual analysis less efficient, thereby requiring a classification method capable of accurately identifying patterns within the data. This study aims to analyze the performance of the CatBoost algorithm in prostate cancer classification based on morphological features and to identify the most influential features using feature importance analysis. The dataset used consists of 100 patient records with 8 independent variables representing cellular morphological features and 1 dependent variable representing prostate cancer classes, namely benign and malignant. The research stages include data preprocessing, handling missing values using median imputation, data partitioning using the 5-Fold Cross-Validation method, model development using the CatBoost algorithm, and model evaluation using a confusion matrix. The results show that the CatBoost model achieved the best performance with a learning rate of 0.2 and a depth of 4, resulting in an accuracy of 92%. Feature importance analysis revealed that the Perimeter feature was the most influential feature in the classification process, followed by other morphological features that also contributed to distinguishing between benign and malignant prostate cancer classes. These findings indicate that the CatBoost algorithm demonstrates excellent capability in classifying prostate cancer based on cellular morphological characteristics.

Keywords: CatBoost, Prostate Cancer, Morphological Features, Classification, Feature Importance, Machine Learning.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
ABSTRAK	xiv
ABSTRACT	xv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	10
1.3. Tujuan Penelitian	10
1.4. Batasan Masalah	10
1.5. Manfaat Teoritis	11
1.6. Manfaat Praktis	12

1.7. Sistematika Penulisan	12
II TINJAUAN PUSTAKA	14
2.1. Kanker Prostat	14
2.2. Machine Learning (ML)	18
2.3. Algoritma CatBoost	19
2.4. K-Fold Cross Validation	26
2.5. Evaluasi Model	26
2.6. <i>Feature Importance</i>	28
2.7. Integrasi Keilmuan TwinTower	30
III METODE PENELITIAN	35
3.1. Jenis Penelitian	35
3.2. Sumber Data	35
3.3. Variabel Penelitian	36
3.3.1. Variabel Independen (X)	36
3.3.2. Variabel Dependen (Y)	36
3.4. Metode Analisis Data	37
3.5. Tahapan Penelitian	37
3.5.1. Input Data	38
3.5.2. Pra Pemrosesan Data	38
3.5.3. Pembagian Data	38
3.5.4. Penerapan Klasifikasi Algoritma CatBoost	39
3.5.5. <i>Analisis Feature Importance</i>	40
3.5.6. Evaluasi Model	40
3.5.7. Skenario Uji Coba	40
IV HASIL DAN PEMBAHASAN	42
4.1. Deskripsi Data	42
4.2. Preprocessing	44

4.3. K-Fold Cross validation	45
4.4. Proses Klasifikasi	47
4.5. Analisis Feature Importance	54
4.6. Evaluasi Model	56
4.7. Skenario Uji Coba	58
4.8. Pembahasan	61
4.9. Integrasi Keilmuan TwinTower	64
V PENUTUP	67
5.1. Kesimpulan	67
5.2. Saran	67
DAFTAR PUSTAKA	68
LAMPIRAN	75

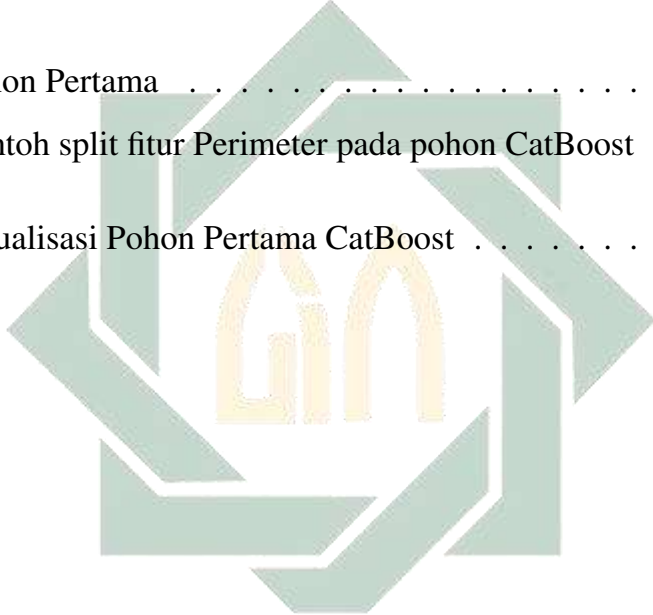
UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

3.1	Fitur Variabel Independen (X) pada Dataset Kanker Prostat .	36
3.2	Variabel Dependen (Y) pada Dataset Kanker Prostat	36
3.3	Skenario Uji Coba	41
4.1	Statistik Deskriptif Data	42
4.2	Data Kanker Prostat	44
4.3	Data Kanker Prostat	45
4.4	Pembagian Data	45
4.5	Nilai Residual dan LogLoss	49
4.6	Data Sampel yang Digunakan dalam Perhitungan Manual . .	51
4.7	Nilai Feature Importance pada Model CatBoost	54
4.8	Confusion Matrix	56
4.9	Hasil Uji Coba Model CatBoost	58
4.10	Hasil Uji Coba Model CatBoost	59
4.11	Hasil Uji Coba Model CatBoost	60
4.12	Hasil Kombinasi Parameter Algoritma CatBoost	61
4.13	Perbandingan Hasil Penelitian	64

DAFTAR GAMBAR

2.1	K-Fold Cross Validation	26
2.2	Confusion Matrix	27
3.1	Flowchart Penelitian	37
3.2	Flowchart CatBoost	39
4.1	Pohon Pertama	50
4.2	Contoh split fitur Perimeter pada pohon CatBoost	55
5.1	Visualisasi Pohon Pertama CatBoost	75



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR PUSTAKA

- Abdurrosyid, R., Teguh, A., and Almais, W. (2025). Deteksi dini diabetes menggunakan machine learning dengan metode pca dan xgboost. *JEPIN*, 11(1):51–56.
- Aditiya, A., Umbara, F. R., and Sabrina, P. N. (2025). Klasifikasi indeks standar pencemaran udara menggunakan algoritma catboost dengan teknik balancing data random undersampling. *Jurnal Algoritma*, pages 1751–1760.
- Afriani, B. N. S. and Pratiwi, D. (2024). Karakteristik sel kanker dan perubahan morfologi pada metastasis. *Jurnal Kesehatan Masyarakat*, 8(2):100–108.
- Alfando, R. H. (2023). Klasifikasi Teks Berita Berbahasa Indonesia Menggunakan Machine Learning dan Deep Learning: Studi Literatur. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(1):681–686.
- Amalina, N., Hilmy, R., and Akrom, M. (2024). Implementasi qsvm dalam klasifikasi biner pada kasus kanker prostat. *Jurnal Ilmiah NERO*, 9(2):119–126.
- Amin, R., Ilyas, M., Muis, M., Zainuddin, A. A., Johan, M. P., and Latief, N. (2023). Akurasi Diagnostik Nilai ADC MRI dalam Memprediksi Tumor Jaringan Lunak Jinak dan Ganas Berdasarkan Hasil Pemeriksaan Histopatologi. *Jurnal Medika Udayana*, 12(7):2–5.
- Ananda, H. R., Sartika, S. D., and Thamrin, R. (2024). Things That Have a Relationship with Cancer Prostat in Patients in Several Locations in The Asia , Amerika and Europe Region from 2014 To 2021. *Bosowa Medical Journa*, 2(2):56–60.
- Arisandi, Riza Rizky Robbi Warsito, B. H. A. R. (2022). Aplikasi naïve bayes classifier (nbc) pada klasifikasi status gizi balita stunting dengan pengujian k-fold cross validation 1,2,3. *Jurnal Gaussian*, 11:130–139.

- Austin, Y. S. e. a. (2024). Klasifikasi penyakit alzheimer dari mri menggunakan convnext. *JTIK*, 11(6):1223–1232.
- Blesyova, N. and Hasan, F. N. (2024). Analisis sentimen terhadap bea cukai menggunakan support vector machine (svm) dan k-fold cross validation. *Jurnal Mahasiswa Teknik Informatika*, 8(6):12051–12056.
- Cai, Y., Yuan, Y., and Zhou, A. (2024). Predictive slope stability early warning model based on CatBoost. *Scientific Reports*, pages 1–14.
- Christian, J., Ernawati, I., and Chamidah, N. (2022). Implementasi algoritma catboost untuk prediksi pembatalan hotel. *SENAMIKA*, pages 814–824.
- Christina, Shintia Sanchia, H. N. R. (2022). Kanker Prostat : Risiko dan Pencegahannya Prostat Cancer : Risk and Prevention. *Jurnal Kedokteran dan Kesehatan*, 1(2):73–81.
- Fitria, E. Y. and Mutijarsa, K. (2023). Research survey on artificial intelligence methods for detecting cyberattack technology threats. *JTIK*, 10(6).
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Hero, S. K. (2021). Faktor Risiko Kanker Payudara. *Jurnal Medika Hutama*, 03(01):3–8.
- Hidayat, T. (2020). Gambaran klinis kanker prostat pada pasien lansia. *Jurnal Urologi Indonesia*, 5(2):85–90.
- Istianto, A. F., Hadiana, A. I., and Umbara, F. R. (2023). Prediksi Curah Hujan Menggunakan Metode Categorical Boosting (CatBoost). *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 7(4):2930–2937.

- Jasman, T. Z., Fadhlullah, M. A., Pratama, A. L., and Rismayani, R. (2022). Analisis algoritma gradient boosting, adaboost dan catboost dalam klasifikasi kualitas air. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(2):392–402.
- Limet, O. B. and Prabowo (2025). Pemilihan fitur untuk prediksi kematangan kelapa sawit menggunakan explainable ai (xai) / feature selection for oil palm ripeness prediction using explainable ai. *Jurnal Algoritma*, 7(2):137–150.
- Maharani, F. D. (2026). Analisis Prediksi Pola Konsumsi Menggunakan Algoritma Machine Learning Berbasis Gradient Boosting. *Jurnal Data Sains*, 1(1):1–11.
- Mandala, M. B., Witanti, W., and Komarudin, A. (2025). Prediksi Penyakit Kanker Payudara Menggunakan Algoritma Synthetic Minority Oversampling Technique dan Categorical Boosting Classifier. *Bulletin of Information Technology*, 7(1):616–625.
- Mawardi, A. B., Pradini, R. S., Haris, M. S., and Boosting, G. (2025). Komparasi Algoritma Boosting untuk Prediksi Gangguan Tidur. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3).
- Munawaroh, L. (2022). Klasifikasi citra mammografi pada kasus tumor jinak dan ganas berbasis ciri tekstur dan morfologi menggunakan metode multy-layer perceptron. Skripsi, Universitas Islam Negeri Walisongo Semarang.
- Murtiningsih, D. A., Sari, B. W., and Fajri, I. N. (2025). Comparison of Light Gradient Boosting Machine , eXtreme Gradient Boosting , and CatBoost with Balancing and Hyperparameter Tuning for Hypertension Risk Prediction on Clinical Dataset. *Journal of Applied Informatics and Computing (JAIC)*, 9(5):2753–2763.
- Muzakir, A., Desiani, A., and Amran, A. (2023). Klasifikasi penyakit kanker prostat menggunakan naive bayes dan k-nn. *Komputika*, 12(1):73–79.
- Pahlevi, O., Ayu, D., Wulandari, N., Rahayu, L. K., Leidiyana, H., and Handrianto, Y. (2024). Model klasifikasi risiko stunting pada balita

- menggunakan algoritma catboost classifier. *Bulletin of Computer Science Research*, 6(4):414–421.
- Prasetya, J. (2022). Penerapan Klasifikasi Naive Bayes Dengan Algoritma Random Oversampling Dan Random Undersampling Pada Data Tidak Seimbang Cervical Cancer Risk Factors. *Jurnal Matematika*, 2:11–22.
- Pratama, R. and Nugroho, A. (2023). Gambaran histopatologi prostat normal dan kanker prostat. *Jurnal Urologi Indonesia*, 30(1):15–22.
- Pratiwi, C. A. and Irhamah (2018). Metode fuzzy support vector machine (fsvm) - genetic algorithm. *Jurnal Sains dan Seni*, 7(2).
- Putra, A. R. and Wijaya, H. (2022). Peran mutasi gen brca1 dan brca2 terhadap risiko kanker prostat. *Jurnal Biomedik Indonesia*, 5(1):45–53.
- Putri, R. A. and Handayani, S. (2021). Analisis morfologi sel dalam penentuan keganasan. *Jurnal Patologi Indonesia*, 4(2):90–98.
- Rahman, M. F. (2023). Metastasis kanker prostat ke tulang dan penanganannya. *Jurnal Onkologi Indonesia*, 9(1):25–33.
- Rahmasari, I., Kurniadewi, R., Dyah, L., and Arini, D. (2021). Tingkat pengetahuan mahasiswa tentang kanker payudara di apikes citra medika surakarta. *Jurnal Keperawatan Duta Medika*, 1(2):42–48.
- Rani, P. M. and Zufria, I. (2025). Predicting Student Stress Levels Based on Lifestyle Factors Using the Catbost Algorithm Corresponding Author :. *ZERO: Jurnal Sains, Matematika dan Terapan*, 9(1):264–276.
- Rizkallah, L. W. (2025). Enhancing the performance of gradient boosting trees on regression problems. *Journal of Big Data*, 12(1).
- Rosyid, I. F. and Pramaditya, H. (2025). Visual interpretation of machine learning models (random forest) for lung cancer risk classification using explainable artificial intelligence (shap lime). *Journal Teknik Informatika (JUTIF)*, 6(4):2187–2206.
- Sabili, N. L. and Umbara, F. R. (2024). Klasifikasi Penyakit Diabetes Menggunakan Algoritma Categorical Boosting dengan Faktor Risiko Diabetes. *Jurnal Mahasiswa Teknik Informatika*, 8(6):11391–11398.

- Saleh, M. R. A., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., Azar, A. T., and Shaikh, A. (2023). Enhancing Breast Cancer Detection and Classification Using Advanced Multi-Model Features and Ensemble Machine Learning Techniques. *Life*, 13(10).
- Saputra, P. A., Irawan, S. S., Rahmadden, R., Prianto, R., and Hidayat, T. (2025). Prediksi dan analisis pola perubahan iklim menggunakan algoritma gradient boosting. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2).
- Sari, N. and Lestari, D. (2023). Pengaruh faktor lingkungan dan gaya hidup terhadap kejadian kanker. *Jurnal Kesehatan Masyarakat*, 18(1):33–40.
- Suryaputri Cantika Okzen, R. M. (2026). Analysis of Gradient Boosted Trees Algorithm in Breast Cancer Classification. *Journal of Applied Informatics and Computing (JAIC)*, 10(1):605–618.
- Suyanto (2020). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Informatika, Bandung.
- Syamballa, M. T., Khomsah, S., Setiya, Y., and Nur, R. (2024). Implementasi algoritma catboost dan shapley additive explanations (shap) dalam memprediksi popularitas game indie pada platform steam. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(4).
- Ulfaningtyas, k., Norahmawati, E., Anita, K. W., Angelina, A., and Seputra, K. P. (2021). The Clinicopathological Profile Of Prostate Adenocarcinoma At Dr. Saiful Anwar Public Hospital Malang From 2015 To 2019: A Retrospective Study. *Majalah Kesehatan*, 8:87–99.
- Wibowo, H. (2022). Pendekatan terapi komprehensif pada kanker prostat. *Jurnal Kedokteran Indonesia*, 8(3):120–128.
- Wibowo, M. T. (2025). Memprediksi Kekambuhan Kanker Payudara Berdasarkan Karakteristik Sel dengan Menggunakan Algoritma C4 . 5. *Journal of Systems, Information Technology, and Electronics Engineering*, 5(1):1–9.

World Health Organization (2022). Global cancer report 2022. <https://www.who.int/publications/i/item/9789240036728>, Diakses 3 Mei 2026.

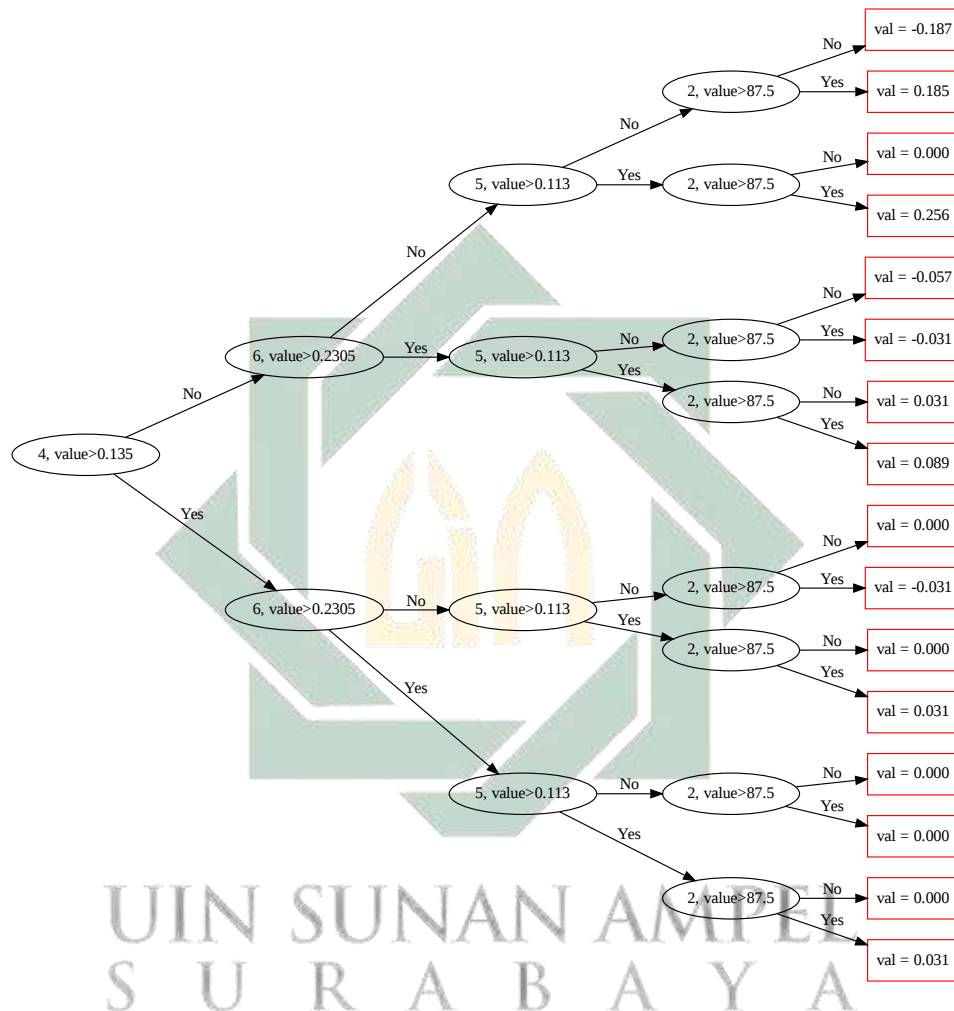
Wulandari, D. and Prasetyo, B. (2022). Stres oksidatif dan kerusakan dna pada paparan polusi lingkungan. *Jurnal Ilmu Lingkungan*, 20(3):210–218.

Yuhana, U. L. and Purwarianti, A. (2022). Tuning Hyperparameter pada Gradient Boosting untuk Klasifikasi Soal Cerita Otomatis. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 8(1):134–139.



UIN SUNAN AMPEL
S U R A B A Y A

LAMPIRAN



Gambar 5.1: Visualisasi Pohon Pertama CatBoost