



SKRIPSI

**STUDI KOMPARATIF KINERJA LIGHTGBM DAN
CATBOOST PADA DATA CHURN PELANGGAN
STREAMING MUSIK**

AKGE NINOV ROYANA

NPM 21081010097

DOSEN PEMBIMBING

Yisti Vita Via, S.ST. M.Kom.

Chrystia Aji Putra, S.Kom, M.T

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

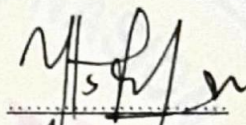
STUDI KOMPARATIF KINERJA LIGHTGBM DAN CATBOOST PADA DATA CHURN PELANGGAN STREAMING MUSIK

Oleh :
AKGE NINOV ROYANA
NPM. 21081010097

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 13 Juni 2025,

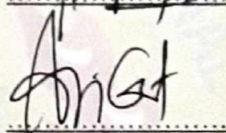
Menyetujui

Yisti Vita Via, S.ST., M.Kom.
NIP. 19860425 202121 2 001



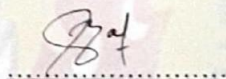
(Pembimbing I)

Chrystia Aji Putra, S.Kom, M.T
NIP. 19861008 202121 1 001



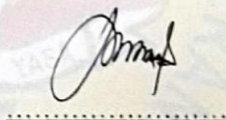
(Pembimbing II)

Made Hanindia Prami Swari, S.Kom, M.Cs
NIP. 19890205 201803 2 001



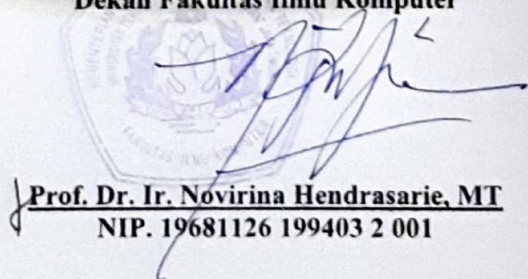
(Ketua Penguji)

Achmad Junaidi, S.Kom., M.Kom
NPT. 3 7811 04 0199 1



(Anggota Penguji)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

STUDI KOMPARATIF KINERJA LIGHTGBM DAN CATBOOST PADA DATA CHURN PELANGGAN STREAMING MUSIK

Oleh :
AKGE NINOV ROYANA
NPM. 21081010097



Menyetujui,
Koordinator Program Studi Informatika
Fakultas Ilmu Komputer

Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 2021212 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini :

Nama : Akge Ninov Royana
NPM : 21081010097
Program : Sarjana(S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 13 Juni 2025

Yang Membuat Pernyataan



Akge Ninov Royana
NPM. 21081010097

Halaman sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Akge Ninov Royana / 21081010097
Judul Skripsi : Studi Komparatif Kinerja LightGBM dan CatBoost
pada Data Churn Pelanggan Streaming Musik
Dosen Pembimbing : 1. Yisti Vita Via, S.ST., M.Kom..
2. Chrystia Aji Putra, S.Kom, M.T

Industri layanan streaming musik mengalami pertumbuhan pesat dalam beberapa tahun terakhir. Namun, persaingan yang ketat menyebabkan tingginya tingkat pelanggan berhenti berlangganan (*churn*), yang berdampak langsung pada penurunan pendapatan perusahaan. Oleh karena itu, kemampuan memprediksi pelanggan yang berpotensi *churn* menjadi krusial untuk strategi retensi yang efektif. Penelitian ini membandingkan performa dua model machine learning, LightGBM dan CatBoost, dalam memprediksi *churn* pelanggan layanan streaming musik. Metodologi yang digunakan mengacu pada CRISP-DM, dengan dataset Streaming Subscription Churn Model dari Kaggle. Tahapan mencakup preprocessing, normalisasi, dan rekayasa fitur, serta pembagian data dengan rasio 60:40, 70:30, dan 80:20 untuk pelatihan dan pengujian. Hasil menunjukkan bahwa model CatBoost dengan rasio 70:30 memberikan performa terbaik, dengan akurasi 84.98%, AUC 0.9411, recall 0.85, precision 0.8559, dan F1-score 0.8533. Berdasarkan hasil ini, model CatBoost terbukti lebih efektif dalam memprediksi pelanggan yang kemungkinan akan berhenti berlangganan dibandingkan LightGBM.

Kata kunci: LightGBM, CatBoost, Churn, Evaluasi Kinerja, Streaming Musik.

Halaman sengaja dikosongkan

ABSTRACT

Student Name / NPM : Akge Ninov Royana / 21081010097
Thesis Title : A Comparative Study of LightGBM and CatBoost
Performance on Music Streaming Customer Churn
Data
Advisors : 1. Yisti Vita Via, S.ST., M.Kom.
2. Chrystia Aji Putra, S.Kom, M.T

The music streaming service industry has experienced rapid growth in recent years. However, intense competition has led to a high rate of customer churn, which directly impacts company revenue. Therefore, the ability to predict customers who are likely to churn is crucial for effective retention strategies. This study compares the performance of two machine learning models, LightGBM and CatBoost, in predicting customer churn in music streaming services. The methodology follows the CRISP-DM framework, using the Streaming Subscription Churn Model dataset from Kaggle. The process includes preprocessing, normalization, and feature engineering, with data splits of 60:40, 70:30, and 80:20 for training and testing. The results show that the CatBoost model with a 70:30 split achieves the best performance, with an accuracy of 84.98%, AUC of 0.9411, recall of 0.85, precision of 0.8559, and F1-score of 0.8533. Based on these results, the CatBoost model proves to be more effective than LightGBM in predicting customers who are likely to churn.

Keywords: LightGBM, CatBoost, Churn, Performance Evaluation, Music Streaming

Halaman sengaja dikosongkan

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan kasih dan karunia-Nya sehingga skripsi dengan judul “**Studi Komparatif Kinerja LightGBM dan CatBoost pada Data Churn Pelanggan Streaming Musik**” dapat diselesaikan dengan baik.

Ucapan terima kasih penulis sampaikan kepada Ibu Yisti Vita Via, S.ST., M.Kom. selaku Dosen Pembimbing Utama dan Bapak Chrystia Aji Putra, S.Kom., M.T. selaku Dosen Pembimbing Kedua, atas bimbingan, arahan, dan dukungan yang telah diberikan selama proses penyusunan skripsi ini.

Penulis juga menyampaikan penghargaan dan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer.
3. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku Koordinator Skripsi yang telah membantu dalam proses administrasi.
4. Ibu Retno Mumpuni, S.Kom., M.Sc. dan Ibu Henni Endah Wahanani, S.T., M.Kom., selaku Dosen Penguji Seminar Proposal atas saran dan masukan yang diberikan.
5. Ibu Made Hanindia Prami Swari, S.Kom., M.Cs. dan Bapak Achmad Junaidi, S.Kom., M.Kom., selaku Dosen Penguji Seminar Hasil atas kritik yang membangun.
6. Bapak Firza Prima Aditiawan, S.Kom., M.T.I., selaku Dosen Wali yang telah memberikan dukungan selama masa studi.
7. Seluruh Bapak/Ibu Dosen Fakultas Ilmu Komputer yang telah membekali penulis dengan ilmu dan pengalaman.
8. Keluarga tercinta — Papi, Mami, dan Mbak Sela — atas kasih sayang, doa, dan dukungan yang tak ternilai selama proses studi ini.
9. Denis Lizard Sambawo Dimara, atas dukungan, semangat, dan kebersamaan yang berarti sebelum dan selama proses penyusunan skripsi.

10. Fildzah Attira Khair dan Rosalena Brigita Limbong, sahabat yang telah memberikan semangat, bantuan, dan waktu selama masa pengerjaan skripsi ini.
11. Seluruh teman-teman serta keluarga besar yang turut mendoakan dan memberikan dukungan secara langsung maupun tidak langsung.
12. Dan semua pihak yang tidak dapat disebutkan satu per satu, yang telah membantu penulis dalam bentuk apa pun selama proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga karya ini dapat memberikan manfaat bagi pembaca dan menjadi pengalaman berharga bagi penulis.

Surabaya, 13 Juni 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vii
ABSTRAK	ix
ABSTRACT	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xix
DAFTAR TABEL	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Layanan Streaming Musik	11
2.3 <i>Churn</i>	11
2.4 Prediksi	12
2.5 Prediksi <i>Churn</i>	12
2.6 <i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i>	13
2.7 <i>Exploratory Data Analysis (EDA)</i>	14
2.8 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	15
2.9 <i>Min-Max Scaling</i>	16
2.10 <i>Encoding</i>	17
2.11 <i>GridSearchCV</i>	18
2.12 <i>Light Gradient Boosting Machine (LightGBM)</i>	18
2.13 <i>CatBoost</i>	21

2.14	AUC-ROC	26
2.15	SHAP Analysis.....	26
2.16	Confusion Matrix.....	27
BAB III DESAIN DAN IMPLEMENTASI SISTEM.....		31
3.1	Metode Penelitian.....	31
3.2	Business Understanding.....	31
3.3	Data Understanding	31
3.3.1	Pengumpulan data.....	32
3.3.2	Deskripsi data	32
3.3.3	Exploratory Data Analysis (EDA).....	34
3.4	Data Preparation	35
3.4.1	Data Cleaning.....	35
3.4.2	Normalisasi data	35
3.4.3	Feature Engineering.....	36
3.5	Modeling.....	37
3.5.1	Pemisahan data	38
3.5.2	Synthetic Minority Oversampling Technique (SMOTE).....	38
3.5.3	Inisialisasi Hyperparameter Tuning	38
3.5.4	LightGBM	39
3.5.5	CatBoost	46
3.6	Evaluation	52
3.7	Pemilihan Model Terbaik.....	53
3.8	Skenario pengujian	54
BAB IV PENGUJIAN DAN ANALISA.....		55
4.1	Pengumpulan Data	55
4.2	Data Cleaning	55
4.2.1	Penanganan Nilai Kosong	56
4.2.2	Penanganan Data Duplikasi.....	56
4.2.3	Penghapusan Kolom Tidak Dipakai	57
4.3	Exploratory Data Analysis (EDA).....	58
4.4	Preprocessing Lanjutan.....	75
4.4.1	Normalisasi.....	75

4.4.2	<i>Feature Engineering</i>	76
4.5	Pembagian Data	77
4.5.1	Pembagian Data 60:40	78
4.5.2	Pembagian Data 70:30	79
4.5.3	Pembagian Data 80:20	80
4.6	<i>Modeling</i>	81
4.6.1	<i>LightGBM</i>	81
4.6.2	<i>CatBoost</i>	82
4.7	Pemilihan Parameter dan Model Terbaik	84
4.7.1	<i>LightGBM</i>	84
4.7.2	<i>CatBoost</i>	85
4.8	Evaluasi Model	86
4.8.1	<i>LightGBM</i>	86
4.8.2	<i>CatBoost</i>	87
4.9	Hasil Skenario Pengujian.....	89
4.10	Model Terbaik dan <i>Analisis SHAP</i>	95
BAB V PENUTUP		99
5.1	Kesimpulan	99
5.2	Saran	99
DAFTAR PUSTAKA		101

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2. 1 Tahapan <i>CRISP-DM</i>	13
Gambar 2. 2 Distribusi Sampel Minoritas (Asli) Dan Sampel Sintetis	16
Gambar 2. 3 Alur <i>LightGBM</i>	19
Gambar 2. 4 Alur <i>CatBoost</i>	23
Gambar 3. 1 Alur Penelitian <i>CRISP-DM</i>	31
Gambar 3. 2 <i>Data Understanding</i>	32
Gambar 3. 3 <i>Data Preparation</i>	35
Gambar 3. 4 <i>Modeling</i>	37
Gambar 3. 5 <i>Decision Tree LightGBM</i>	45
Gambar 3. 6 <i>Decision Tree CatBoost</i>	51
Gambar 4. 1 <i>Dataset</i>	55
Gambar 4. 2 <i>DataFrame</i> Tanpa Nilai Kosong.....	56
Gambar 4. 3 Jumlah Data Sebelum Dan Sesudah Penghapusan Duplikat.....	57
Gambar 4. 4 <i>DataFrame</i> Setelah Penghapusan Kolom Tidak Relevan.....	58
Gambar 4. 5 Distribusi Data Pelanggan <i>Churn</i> Dan <i>Non-Churn</i>	59
Gambar 4. 6 Distribusi Data Target Berdasarkan Usia.....	61
Gambar 4. 7 Distribusi Data Target Berdasarkan Tanggal Pendaftaran.....	62
Gambar 4. 8 Distribusi Target Berdasarkan Waktu Penggunaan Per Minggu	63
Gambar 4. 9 Distribusi Data Target Berdasarkan Durasi Rata-Rata Sesi.....	65
Gambar 4. 10 Distribusi Data Target Berdasarkan Frekuensi Melewatkan Lagu .	66
Gambar 4. 11 Distribusi <i>Churn</i> Berdasarkan Tipe Langganan.....	68
Gambar 4. 12 Distribusi <i>Churn</i> Berdasarkan Rencana Pembayaran	70
Gambar 4. 13 Distribusi <i>Churn</i> Frekuensi Jeda Langganan	72
Gambar 4. 14 Distribusi <i>Churn</i> Berdasarkan Metode Pembayaran.....	73
Gambar 4. 15 Distribusi <i>Churn</i> Berdasarkan Frekuensi Permintaan <i>Customer Service</i>	75
Gambar 4. 16 Hasil Normalisasi Pada Kolom Numerik.....	76
Gambar 4. 17 Hasil <i>Encoding</i> Pada Kolom Kategorikal	77
Gambar 4. 18 Hasil Pembagian Data Rasio 60:40.....	79
Gambar 4. 19 Hasil Pembagian Data Rasio 70:30.....	80

Gambar 4. 20 Hasil Pembagian Data Rasio 80:20	81
Gambar 4. 21 Proses <i>Training</i> Model <i>LightGBM</i>	82
Gambar 4. 22 Proses <i>Training</i> Model <i>CatBoost</i>	84
Gambar 4. 23 Parameter Terbaik <i>LightGBM</i> Pada Pembagian Data 60:40	85
Gambar 4. 24 Parameter Terbaik <i>LightGBM</i> Pada Pembagian Data 70:30	85
Gambar 4. 25 Parameter Terbaik <i>LightGBM</i> Pada Pembagian Data 80:20	85
Gambar 4. 26 Parameter Terbaik <i>CatBoost</i> Pada Pembagian Data 60:40.....	85
Gambar 4. 27 Parameter Terbaik <i>CatBoost</i> Pada Pembagian Data 70:30.....	86
Gambar 4. 28 Parameter Terbaik <i>CatBoost</i> Pada Pembagian Data 80:20.....	86
Gambar 4. 29 <i>Confusion Matrix</i> <i>LightGBM</i> Pada Pembagian Data 60:40.....	90
Gambar 4. 30 <i>Confusion Matrix</i> <i>CatBoost</i> Pada Pembagian Data 60:40	91
Gambar 4. 31 <i>Confusion Matrix</i> <i>LightGBM</i> Pada Pembagian Data 70:30.....	92
Gambar 4. 32 <i>Confusion Matrix</i> <i>CatBoost</i> Pada Pembagian Data 70:30	93
Gambar 4. 33 <i>Confusion Matrix</i> <i>LightGBM</i> Pada Pembagian Data 80:20.....	94
Gambar 4. 34 <i>Confusion Matrix</i> <i>CatBoost</i> Pada Pembagian Data 80:20	95
Gambar 4. 35 Visualisasi <i>SHAP</i>	96

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	27
Tabel 3. 1 Atribut Dataset	32
Tabel 3. 2 Contoh Data Pelanggan Layanan Streaming (1).....	36
Tabel 3. 3 Hasil Normalisasi.....	36
Tabel 3. 4 Contoh Data Pelanggan Layanan Streaming (2).....	37
Tabel 3. 5 Hasil <i>Encoding</i>	37
Tabel 3. 6 Contoh Data Pelanggan Layanan Streaming (3).....	40
Tabel 3. 7 Gradien Semua Pelanggan	41
Tabel 3. 8 Gradien Sebelum Dikoreksi.....	41
Tabel 3. 9 Gradien Setelah Dikoreksi	42
Tabel 3. 10 Hasil Penggabungan 2 Fitur.....	42
Tabel 3. 11 Contoh Data Pelanggan Layanan Streaming (4).....	47
Tabel 3. 12 Urutan Pelanggan Berdasarkan <i>Churn</i> Dan Gender	48
Tabel 3. 13 Hasil <i>Ordered TS</i>	49
Tabel 3. 14 Distribusi <i>Churn</i> Berdasarkan Gender.....	50
Tabel 3. 15 Distribusi <i>Churn</i> Berdasarkan Usia	51
Tabel 3. 16 Contoh <i>Confusion Matrix</i>	53
Tabel 3. 17 Skenario Pengujian	54
Tabel 4. 1 Hasil Uji Skenario.....	89

Halaman ini sengaja dikosongkan