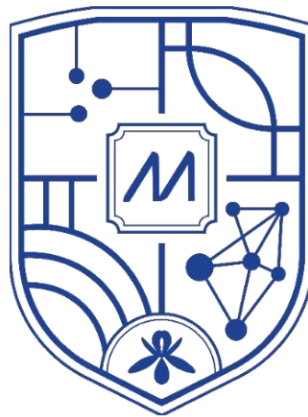


**PREDIKSI *CHURN* PELANGGAN E-COMMERCE
MENGUNAKAN *ENSEMBLE LEARNING***

SKRIPSI

Oleh:

**SHESYL YULIA NOOR DIWANI
NIM. 221122330**



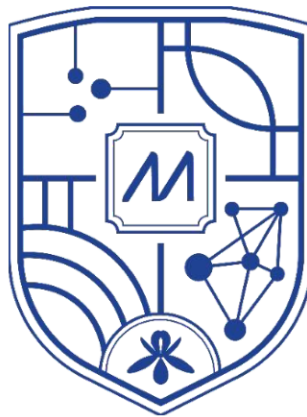
**PROGRAM STUDI S-1 SISTEM INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

E-COMMERCE CUSTOMER CHURN PREDICTION USING ENSEMBLE LEARNING

UNDERGRADUATE THESIS

By:

**SHESYL YULIA NOOR DIWANI
ID NUMBER. 221122330**



**UNDERGRADUATE PROGRAM OF INFORMATION SYSTEMS
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN
PREDIKSI *CHURN* PELANGGAN *E-COMMERCE* MENGGUNAKAN
ENSEMBLE LEARNING

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1

Oleh:

SHESYL YULIA NOOR DIWANI
NIM. 221122330

Disetujui Oleh:

Dosen Pembimbing I,


Roni Yunis, S.Kom., M.T.

Dosen Pembimbing II,


Elly, S.Kom., M.TI.

Medan, 14 Juli 2026
Diketahui dan Disahkan Oleh:

Ketua Program Studi
S1,




Tri Wulandari Ginting, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Sistem Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Shesyl Yulia Noor Diwani
NIM : 221122330

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Prediksi *Churn* Pelanggan *E-Commerce* Menggunakan *Ensemble Learning*
Tempat Penelitian : -
Alamat Tempat Penelitian : -
No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 26 Juni 2026

Saya yang membuat pernyataan,



Shesyl Yulia Noor Diwani

PREDIKSI CHURN PELANGGAN E-COMMERCE MENGGUNAKAN ENSEMBLE LEARNING

Abstrak

Prediksi churn pelanggan merupakan salah satu permasalahan penting dalam industri e-commerce karena berkaitan dengan upaya mempertahankan pelanggan dan meningkatkan keberlangsungan bisnis. Penelitian ini bertujuan membangun model prediksi churn pelanggan e-commerce menggunakan metode ensemble learning, menganalisis pengaruh penanganan class imbalance menggunakan Synthetic Minority Oversampling Technique (SMOTE), serta membandingkan kinerja Random Forest, Gradient Boosting Machine (GBM), dan Extreme Gradient Boosting (XGBoost). Penelitian menggunakan E-Commerce Churn Dataset 2025 yang diperoleh dari Kaggle dan dilaksanakan berdasarkan tahapan OSEMN (Obtain, Scrub, Explore, Model, dan Interpret). Data diproses melalui preprocessing, pembagian data latih dan data uji, penanganan class imbalance menggunakan SMOTE, one-hot encoding, pelatihan model, serta evaluasi menggunakan Confusion Matrix, Recall, dan Area Under the Curve (AUC). Hasil penelitian menunjukkan bahwa penerapan SMOTE dan penyesuaian threshold dari 0,5 menjadi 0,2 meningkatkan kemampuan deteksi churn. XGBoost menghasilkan performa terbaik dengan nilai recall sebesar 0,9388 dan AUC sebesar 0,5534, sedangkan Random Forest memperoleh recall sempurna sebesar 1,0000, namun menunjukkan bias klasifikasi yang menyebabkan hasil prediksi kurang valid. Meskipun XGBoost memberikan kompromi performa terbaik di antara ketiga model, nilai AUC seluruh model masih berada di bawah 0,60, sehingga model yang dihasilkan belum siap untuk diimplementasikan secara praktis dan lebih menunjukkan potensi awal penerapan ensemble learning untuk pengembangan penelitian selanjutnya.

Kata kunci: prediksi churn, ensemble learning, SMOTE, XGBoost, e-commerce

Abstract

Customer churn prediction is a key challenges in the e-commerce industry because it is tied to customer retention and business sustainability. This study aims to build an e-commerce customer churn prediction model using ensemble learning, analyze the impact of addressing class imbalance using the Synthetic Minority Oversampling Technique (SMOTE), and compare the performance of Random Forest, Gradient Boosting Machine (GBM), and Extreme Gradient Boosting (XGBoost). The study uses the 2025 E-Commerce Churn Dataset obtained from Kaggle and follows the OSEMN framework (Obtain, Scrub, Explore, Model, and Interpret). The data were preprocessed, split into training and test sets, balanced using SMOTE, one-hot encoded, trained, and evaluated using the Confusion Matrix, Recall, and Area Under the Curve (AUC). The results show that SMOTE and threshold adjustment from 0.5 to 0.2 improved churn detection. XGBoost achieved the best performance with a recall of 0.9388 and an AUC of 0.5534, while Random Forest obtained a perfect recall of 1.0000 but suffered from classification bias. Although XGBoost provided the best performance among the three models, the AUC values of all models remained below 0.60. Therefore, the proposed models are not yet suitable for practical implementation and demonstrate the potential of ensemble learning for future research.

Keywords: churn prediction, ensemble learning, SMOTE, XGBoost, e-commerce

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala Rahmat dan Karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "*Prediksi Churn Pelanggan E-Commerce Menggunakan Ensemble Learning*" sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Universitas Mikroskil.

Penelitian ini dilakukan sebagai bentuk penerapan ilmu *machine learning* dalam menyelesaikan permasalahan nyata di bidang *e-commerce*, khususnya dalam memprediksi pelanggan yang berpotensi melakukan *churn*. Melalui penelitian ini, penulis memperoleh banyak pengalaman dan pembelajaran, baik dalam proses pengolahan data, pembangunan model prediksi, evaluasi model, maupun penyusunan karya ilmiah.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa penyelesaiannya tidak terlepas dari bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dengan sepenuh hati penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Roni Yunis, S.Kom., M.T., selaku Dosen Pembimbing I yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, masukan, serta inspirasi kepada penulis dengan penuh kesabaran selama proses penyusunan skripsi ini sehingga dapat diselesaikan dengan baik.
2. Ibu Elly, S.Kom., M.Tl., selaku Dosen Pembimbing II yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, masukan, serta inspirasi kepada penulis dengan penuh kesabaran selama proses penyusunan skripsi ini sehingga dapat diselesaikan dengan baik.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Ibu Tri Wulandari Ginting, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Fakultas Informatika Universitas Mikroskil Medan.
6. Seluruh dosen dan staf Universitas Mikroskil yang telah memberikan ilmu, pengalaman, bantuan, dan pelayanan yang baik kepada penulis selama menjalani proses perkuliahan.

7. Kedua orang tua tersayang serta seluruh keluarga yang senantiasa memberikan perhatian, doa, dukungan, motivasi, dan kepercayaan kepada penulis dalam setiap prosesnya, sehingga penulis mampu menyelesaikan penulisan skripsi.
8. Teman-teman penulis baik di dalam maupun di luar kampus yang selalu mendukung dan turut mendoakan kelancaran proses skripsi.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan penelitian di masa mendatang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya pada bidang *machine learning* dan prediksi *churn* pelanggan.

Medan, 26 Juni 2026

Penulis,

Shesyl Yulia Noor Diwani

DAFTAR ISI

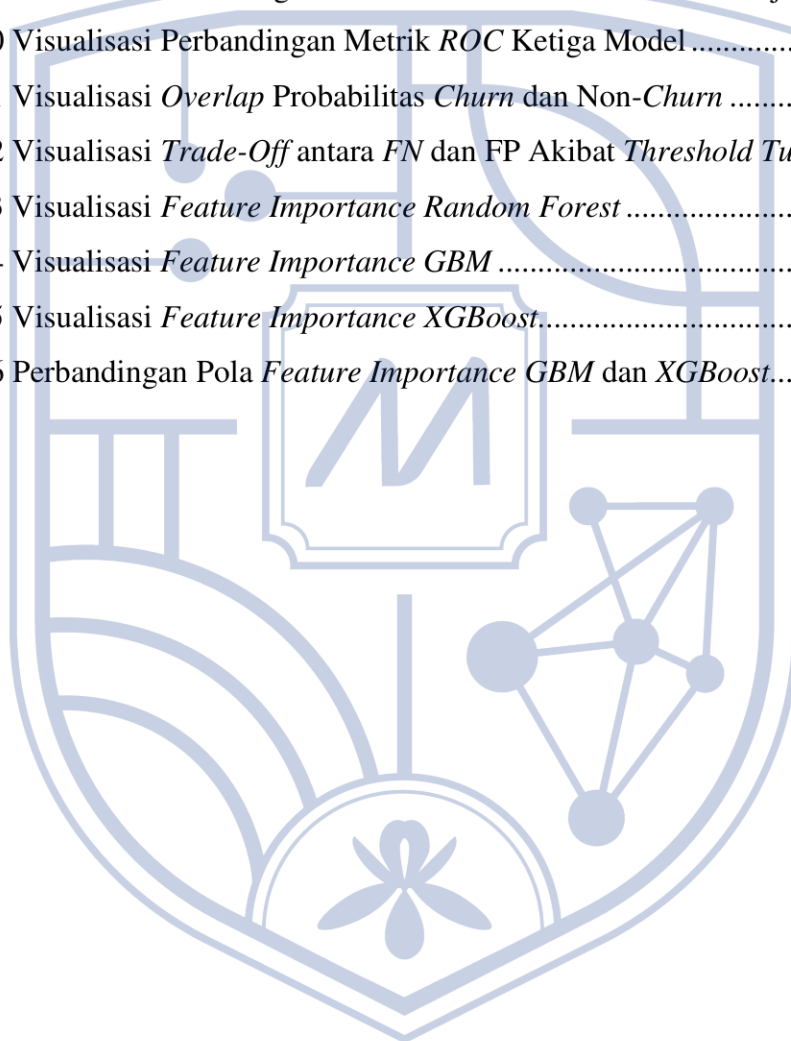
ABSTRAK.....	i
KATA PENGANTAR.....	ii
DAFTAR ISI	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL	viii
DAFTAR LAMPIRAN.....	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Manfaat	4
1.5 Ruang Lingkup.....	5
BAB II KAJIAN LITERATUR.....	6
2.1 <i>Customer Churn</i>	6
2.2 <i>Machine Learning</i>	6
2.3 <i>Class Imbalance</i>	8
2.4 <i>Ensemble Learning</i>	9
2.5 <i>Random Forest</i>	12
2.6 <i>GBM dan XGBoost</i>	14
2.6.1 <i>GBM</i>	14
2.6.2 <i>XGBoost</i>	15
2.7 <i>SMOTE</i>	16
2.8 Evaluasi Model.....	18
2.9 <i>OSEMN</i>	20
2.10 Studi Literatur	22
BAB III TAHAPAN PELAKSANAAN.....	26
3.1 Data Penelitian	26
3.2 Tahapan Penelitian	28
BAB IV HASIL DAN PEMBAHASAN	51
4.1 Hasil <i>Preprocessing</i> dan Pembagian Data	51
4.2 Hasil Pemodelan <i>Random Forest</i>	54
4.3 Hasil Pemodelan <i>GBM</i>	57

4.4 Hasil Pemodelan <i>XGBoost</i>	61
4.5 Perbandingan Performansi Model.....	66
4.5.1 Analisis Pengaruh <i>SMOTE</i> dan <i>Threshold Tuning</i> terhadap Performa Model....	67
4.5.1.1 Analisis Pengaruh <i>SMOTE</i> Terhadap Performa Model	67
4.5.1.2 Analisis Pengaruh <i>Threshold Tuning</i> Terhadap Performa Model	69
4.5.2 Perbandingan Hasil Evaluasi Model	70
4.5.3 Analisis <i>ROC Curve</i> dan Nilai <i>AUC</i>	73
4.5.4 Analisis Distribusi Probabilitas Prediksi	75
4.5.5 Analisis <i>Trade-Off Recall</i> , <i>False Negative</i> dan <i>False Positive</i>	79
4.5.6 Analisis Fitur <i>Importance</i>	82
4.5.7. Penentuan Model Terbaik.....	85
4.6 Implikasi dan Interpretasi Hasil Penelitian	88
BAB V PENUTUP	92
5.1 Kesimpulan	92
5.2 Saran.....	93
DAFTAR PUSTAKA.....	94
DAFTAR RIWAYAT HIDUP	101

DAFTAR GAMBAR

Gambar 2.1 Mekanisme <i>Bagging</i> pada <i>Ensemble Learning</i>	11
Gambar 2.2 Mekanisme <i>Boosting</i> pada <i>Ensemble Learning</i>	12
Gambar 3.1 Diagram Alur Penelitian	28
Gambar 3.2 <i>Loading Package</i>	29
Gambar 3.3 <i>Dataset</i>	30
Gambar 3.4 Struktur Data	30
Gambar 3.5 Membuat Fitur <i>Churn</i>	31
Gambar 3.6 Menghapus Variabel <i>subscription_status</i>	31
Gambar 3.7 Menghapus Variabel Tidak Relevan.....	32
Gambar 3.8 Pembuatan Fitur <i>recency</i> dan <i>customer_lifetime</i>	32
Gambar 3.9 Pembuatan Fitur <i>total_spent</i>	33
Gambar 3.10 Menghapus Fitur <i>signup_date</i> dan <i>last_purchase_date</i>	33
Gambar 3.11 Mengubah Data Kategorikal Menjadi <i>Factor</i>	33
Gambar 3.12 Mengubah Variabel <i>churn</i> Menjadi Label Kelas	34
Gambar 3.13 Struktur <i>Dataset</i> Hasil <i>Preprocessing</i>	34
Gambar 3.14 Tampilan Data Awal (<i>Head</i>) <i>Dataset</i> Hasil <i>Preprocessing</i>	35
Gambar 3.15 Memeriksa <i>Missing Value</i>	35
Gambar 3.16 Memeriksa <i>Outlier</i>	35
Gambar 3.17 Visualisasi <i>Outlier</i>	36
Gambar 3.18 Penanganan <i>Outlier</i>	37
Gambar 3.19 Visualisasi Hasil Penanganan <i>Outlier</i>	37
Gambar 3.20 Distribusi Jumlah dan Proporsi Kelas <i>Churn</i>	38
Gambar 3.21 Visualisasi Distribusi <i>Churn</i>	38
Gambar 3.22 <i>Output</i> Statistik Deskriptif Menggunakan Fungsi <i>summary()</i>	39
Gambar 3.23 Visualisasi Hasil Analisis Korelasi	40
Gambar 3.24 Visualisasi Distribusi Fitur <i>recency</i>	42
Gambar 3.25 Visualisasi Distribusi Fitur <i>customer_lifetime</i>	43
Gambar 3.26 Visualisasi Analisis Fitur <i>recency</i> dan <i>churn</i>	44
Gambar 3.27 Visualisasi Analisis Fitur <i>total_spent</i> dan <i>churn</i>	45
Gambar 3.28 Visualisasi Analisis Fitur <i>customer_lifetime</i> dan <i>churn</i>	46
Gambar 4.1 Visualisasi <i>Split</i> Data <i>Training</i> dan Data <i>Testing</i>	51

Gambar 4.2 Visualisasi Distribusi Kelas Sebelum <i>SMOTE</i>	52
Gambar 4.3 Visualisasi Distribusi Kelas Sesudah <i>SMOTE</i>	53
Gambar 4.4 Grafik <i>ROC Curve</i> Model <i>Random Forest</i>	56
Gambar 4.5 Grafik <i>ROC Curve</i> Model <i>GBM</i>	60
Gambar 4.6 Grafik <i>ROC Curve</i> Model <i>XGBoost</i>	65
Gambar 4.7 Pengaruh <i>SMOTE</i> dan <i>Threshold Tuning</i> terhadap <i>Recall</i> Model	69
Gambar 4.8 Perbandingan Nilai <i>Recall</i> , <i>Precision</i> , <i>F1-Score</i> , dan <i>AUC</i> Ketiga Model	71
Gambar 4.9 Visualisasi Perbandingan <i>AUC Cross-Validation</i> dan Data Uji	73
Gambar 4.10 Visualisasi Perbandingan Metrik <i>ROC</i> Ketiga Model	74
Gambar 4.11 Visualisasi <i>Overlap</i> Probabilitas <i>Churn</i> dan <i>Non-Churn</i>	77
Gambar 4.12 Visualisasi <i>Trade-Off</i> antara <i>FN</i> dan <i>FP</i> Akibat <i>Threshold Tuning</i>	80
Gambar 4.13 Visualisasi <i>Feature Importance</i> <i>Random Forest</i>	82
Gambar 4.14 Visualisasi <i>Feature Importance</i> <i>GBM</i>	82
Gambar 4.15 Visualisasi <i>Feature Importance</i> <i>XGBoost</i>	83
Gambar 4.16 Perbandingan Pola <i>Feature Importance</i> <i>GBM</i> dan <i>XGBoost</i>	85



DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	19
Tabel 2.2 Nilai <i>AUC</i>	20
Tabel 2.3 Studi Literatur	22
Tabel 3.1 Variabel Data Penelitian	27
Tabel 3.2 <i>Library</i> yang Digunakan.....	29
Tabel 4.1 Hasil <i>Cross-validation</i> dan <i>Hyperparameter Tuning Random Forest</i>	54
Tabel 4.2 Hasil <i>Confusion Matrix</i> Awal Model <i>Random Forest</i>	54
Tabel 4.3 Hasil <i>Threshold Tuning</i> Model <i>Random Forest</i>	55
Tabel 4.4 Hasil <i>Confusion Matrix</i> Akhir Model <i>Random Forest</i>	55
Tabel 4.5 Hasil <i>Cross-validation</i> dan <i>Hyperparameter Tuning GBM</i>	57
Tabel 4.6 Hasil <i>Confusion Matrix</i> Awal Model <i>GBM</i>	58
Tabel 4.7 Hasil <i>Threshold Tuning</i> Model <i>GBM</i>	59
Tabel 4.8 Hasil <i>Confusion Matrix</i> Akhir Model <i>GBM</i>	59
Tabel 4.9 Hasil <i>Evaluation Log Cross-Validation</i> per Iterasi <i>Boosting</i>	62
Tabel 4.10 Parameter <i>XGBoost</i> yang Digunakan.....	62
Tabel 4.11 Hasil <i>Confusion Matrix</i> Awal Model <i>XGBoost</i>	63
Tabel 4.12 Hasil <i>Threshold Tuning</i> Model <i>XGBoost</i>	63
Tabel 4.13 Hasil <i>Confusion Matrix</i> Akhir Model <i>XGBoost</i>	64
Tabel 4.14 Perbandingan <i>Recall</i> , <i>Precision</i> , <i>F1-Score</i> , dan <i>AUC</i> pada Empat Skenario Eksperimen	67
Tabel 4.15 Perbandingan Performa Akhir Ketiga Model	70
Tabel 4.16 Perbandingan Nilai <i>AUC Cross-Validation</i> dan Data Uji.....	73
Tabel 4.17 Statistik Probabilitas Prediksi <i>XGBoost</i> Berdasarkan Kelas Aktual.....	76
Tabel 4.18 Perubahan <i>Recall</i> , <i>FN</i> , dan <i>FP</i> Akibat <i>Threshold Tuning</i> pada Ketiga Model .	79
Tabel 4.19 Top 5 <i>Feature Importance</i> pada Setiap Model	83
Tabel 4.20 Penentuan Model Terbaik Berdasarkan Hasil Evaluasi Keseluruhan.....	85

DAFTAR LAMPIRAN

Lampiran 1 Link Github	100
------------------------------	-----

