



**Komparasi Tiga *Machine Learning* Klasifikasi  
Untuk Prediksi *Churn***

**LAPORAN SKRIPSI**

**UNIVERSITAS  
MERCU BUANA**

Muhammad Hibatur Akmal  
41821110025

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2024**

**HALAMAN JUDUL**



**Komparasi Tiga *Machine Learning* Klasifikasi  
Untuk Prediksi *Churn***

**LAPORAN SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana**

**UNIVERSITAS  
MERCU BUANA**

**Muhammad Hibatur Akmal  
41821110025**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2024**

## HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : *Muhammad Hibatur Akmal*  
NIM : *41821110025*  
Program Studi : *Sistem Informasi*  
Judul Laporan Skripsi : *Komparasi Tiga Machine Learning  
Klasifikasi Untuk Prediksi Churn*

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, *2 Februari 2024*



*Muhammad Hibatur Akmal*

## LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Muhammad Hibatur Akmal  
 NIM (41821110025)  
 Judul Tugas Akhir : Komparasi Tiga *Machine Learning* Klasifikasi  
 Untuk Prediksi *Churn*

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 20 Desember 2023

Menyetujui,

Pembimbing	: Ifan Prihandi, S.Kom, M.Kom	(  )
NIDN	: 0313098901	
Ketua Penguji	: Ruci Meiyanti, Dr. S.Kom, M.Kom	(  )
NIDN	: 0304056803	
Penguji 1	: Ruci Meiyanti, Dr. S.Kom, M.Kom	(  )
NIDN	: 0304056803	
Penguji 2	: Sulis Sandiwarno, S.Kom, M.Kom	(  )
NIDN	: 0302028803	

Mengetahui,

  
**Dr. Bambang Jokonowo, S.Si., M.T.I.**  
 Dekan Fakultas Ilmu Komputer

  
**Dr. Ruci Meiyanti, M.Kom**  
 Ka.Prodi Sistem Informasi

## KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Sistem Informasi pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Prof. Dr. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S.Si., M.T.I. selaku Dekan Fakultas Ilmu Komputer.
3. Ruci Meiyanti, Dr. S.Kom, M.Kom selaku Ketua Program Studi Sistem Informasi.
4. Ifan Prihandi, S.Kom, M.Kom selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
5. Sulis Sandiwarno, S.Kom, M.Kom selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Eka Yuliawan selaku Ayah penulis.
7. Nur Asiah selaku Ibu penulis.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 2 Februari 2024

Muhammad Hibatur Akmal



## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Hibatur Akmal  
NIM : 91821110025  
Program Studi : Sistem Informasi  
Judul Laporan Skripsi : Komparasi Tiga Machine Learning  
klasifikasi Untuk Prediksi Churn

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta **U** dan **V** sebagai **S I** pemilik **S** Hak **C**ipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 2 Februari 2024

Yang menyatakan,



Muhammad Hibatur Akmal

## ABSTRAK

Nama : Muhammad Hibatur Akmal  
NIM : 41821110025  
Pembimbing TA : Ifan Prihandi, S.Kom, M.Kom  
Judul : Komparasi Tiga *Machine Learning* Klasifikasi  
Untuk Prediksi *Churn*

Tingginya *bank Customer churn* atau pindahnya nasabah dari satu bank ke bank lain dapat merugikan bank. Nasabah sebagai aset berharga yang dimiliki oleh bank harus bisa dipertahankan dengan baik. Prediksi customer churn dapat menjadi solusi permasalahan ini, karena dengan mengetahui ciri nasabah yang akan churn, bank dapat melakukan tindakan preventif agar nasabah tidak churn sekaligus mengurangi tingkat churn. Teknologi dan kumpulan data customer yang ada di bank saat ini dapat digunakan menjadi alat dan bahan dalam memprediksi churn menggunakan analisis data atau data mining. Algoritma machine learning klasifikasi seperti *KNN*, *Decision Tree*, dan *XGBoost* cukup handal untuk melakukan perbandingan hasil model dilihat dari nilai akurasi, precision, recall, dan auc score. Metode CRISP-DM yang biasa digunakan dalam data mining berperan penting sebagai metode penelitian. Dataset yang telah dipersiapkan dengan baik, kemudian diproses menggunakan teknik *log transformation*, *standardization*, dan *class imbalance handling* mampu memperbaiki setiap nilai metric model seperti yang sudah disarankan pada penelitian sebelumnya. Hasil dari penelitian ini menunjukkan bahwa *XGBoost* tanpa feature selection meraih nilai akurasi dan precision terbesar dengan nilai 86% dan 75% dibandingkan *KNN* dan *Decision Tree*. Penelitian selanjutnya dapat meningkatkan nilai recall dan F1 score yang rendah pada tiap model di penelitian ini.

Kata kunci: *Churn*, *Decision Tree*, *K-Nearest Neighbor*, *Machine Learning*, *XGBoost*.

## ABSTRACT

Name : Muhammad Hibatur Akmal  
Student Number : 41821110025  
Counsellor : Ifan Prihandi, S.Kom, M.Kom  
Title : Komparasi Tiga *Machine Learning* Klasifikasi  
Untuk Prediksi *Churn*

High bank customer churn or the movement of customers from one bank to another can be detrimental to the bank. Customers as valuable assets owned by the company must be maintained well. Predicting customer churn can be a solution to this problem, because by knowing the characteristics of customers who will churn, banks can take preventive action so that customers do not churn and at the same time reduce the churn rate. The technology and collection of customer data currently available in banks can be used as tools and materials to predict churn using data analysis or data mining. Classification machine learning algorithms such as KNN, Decision Tree, and XGBoost are reliable enough to compare model results in terms of accuracy, precision, recall, and AUC scores. The CRISP-DM method which is commonly used in data mining plays an important role as a research method. A dataset that has been well prepared, then processed using log transformation, standardization, and class imbalance handling techniques is able to improve each model metric value as suggested in previous research. The results of this research show that XGBoost without feature selection achieved the highest accuracy and precision values with values of 86% and 75% compared to KNN and Decision Tree. Further research can increase the low recall and F1 scores in each model in this study.

Keywords: *Churn, Decision Tree, K-Nearest Neighbor, Machine Learning, XGBoost.*



## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERNYATAAN KARYA SENDIRI .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRACT.....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian .....	3
1.4. Batasan Masalah .....	3
a. Data .....	3
b. Metode .....	3
c. Hasil .....	3
1.5. Manfaat Penelitian .....	4
a. Teoritis/akademis .....	4
b. Praktis. ....	4
1.6. Sistematika Penulisan .....	4
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1. Cross Industry Standard Process for Data Mining (CRISP-DM) .....	6
a. Business Understanding.....	6
b. Data Understanding.....	6
c. Data Preparation.....	6
d. Modeling .....	6
e. Evaluation .....	7
f. Deployment .....	7
2.2. Data Mining .....	7

2.3. Preprocessing .....	7
a. Label Encoding .....	7
b. One Hot Encoding.....	7
c. Z-Score Normalization.....	8
d. MinMaxScaler.....	8
e. Synthetic Minority Over-sampling Technique (SMOTE) .....	9
2.4. K-Nearest Neighbor (KNN).....	9
2.5. Decision Tree C4.5 .....	10
2.6. Extreme Gradient Boosting (XGBoost).....	11
2.7. Randomized Search Hyperparameter Tuning .....	11
2.8. K-Fold Cross Validation .....	12
2.9. Confusion Matrix .....	12
2.10. Literature Review.....	13
2.11. Analisis Literature Review.....	22
<b>BAB III METODE PENELITIAN .....</b>	<b>24</b>
3.1 Deskripsi Sumber Data .....	24
3.2 Teknik Pengumpulan Data.....	25
3.3 Diagram Alir Penelitian .....	25
3.3.1. Pengumpulan Data .....	25
3.3.2 Encoding Data.....	26
3.3.3. Split Data .....	26
3.3.4. Preprocessing Data.....	26
3.3.5 Pembuatan Model .....	27
3.3.6. Hasil Pemodelan .....	27
3.3.7. Evaluasi Hasil Model .....	27
3.3.8. Hasil Evaluasi .....	27
3.3.9. Penerapan Hasil Evaluasi .....	28
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>29</b>
4.1. HASIL.....	29
4.1.1. Pengumpulan Data .....	29
4.1.2. Encoding Data.....	30
4.1.3. Split Data .....	32
4.1.4 Preprocessing Data Training .....	32
4.1.5. Preprocessing Data Testing.....	37
4.1.6. Feature Selection.....	39

4.1.7 Pemodelan.....	41
4.1.8. Hasil Pemodelan .....	42
4.1.9. Evaluasi Hasil Model .....	44
4.1.10. Hasil Evaluasi .....	47
4.1.11. Penerapan Hasil Evaluasi.....	52
4.2. PEMBAHASAN .....	53
4.2.1 Number Of Products: .....	54
4.2.2 Is Active Member.....	54
4.2.3 Age.....	55
4.2.4 Geography.....	56
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>57</b>
5.1. Kesimpulan .....	57
5.2. Saran .....	57
<b>DAFTAR PUSTAKA.....</b>	<b>58</b>
<b>LAMPIRAN.....</b>	<b>62</b>



UNIVERSITAS  
MERCU BUANA

## DAFTAR TABEL

Tabel 1 Tabel <i>Literature Review</i> .....	13
Tabel 2 Deskripsi Kolom Dataset.....	24
Tabel 3 Hasil Pemodelan Ketiga Machine Learning .....	42
Tabel 4 Hasil Model Machine Learning Setelah Hyperparameter Tuning .....	52



## DAFTAR GAMBAR

Gambar 1. Confusion Matrix .....	12
Gambar 2 Diagram Alir Penelitian .....	25
Gambar 3 Contoh kolom dan baris dataset .....	30
Gambar 4 Informasi dataset .....	30
Gambar 5 Perubahan kolom <i>Gender</i> menggunakan Label Encoding .....	31
Gambar 6 Perubahan kolom <i>Geography</i> menggunakan One Hot Encoding.....	31
Gambar 7 Source code split train dan test data .....	32
Gambar 8 Z-Score Normalization pada data training .....	33
Gambar 9 Kolom <i>Age</i> sebelum dilakukan Log Transformation.....	33
Gambar 10 Kolom <i>Age</i> setelah dilakukan Log Transformation .....	34
Gambar 11 Source code standarisasi menggunakan MinMaxScaler .....	35
Gambar 12 Hasil standarisasi menggunakan MinMaxScaler .....	35
Gambar 13 Ketidakseimbangan kelas <i>churn</i> dan tidak <i>churn</i> .....	35
Gambar 14 Source code penanganan ketidakseimbangan kelas .....	36
Gambar 15 Hasil penanganan ketidakseimbangan kelas .....	36
Gambar 16 Penyatuan hasil preprocessing menjadi file dataset training .....	37
Gambar 17 Kolom <i>Age</i> sebelum dilakukan Log Transformation.....	37
Gambar 18 Kolom <i>Age</i> sesudah dilakukan Log Transformation .....	38
Gambar 19 Source code standarisasi menggunakan MinMaxScaler .....	39
Gambar 20 Hasil standarisasi menggunakan MinMaxScaler .....	39
Gambar 21 Penyatuan hasil preprocessing menjadi data testing .....	39
Gambar 22 Hasil Feature Selection Chi-Squared .....	40
Gambar 23 Hasil Feature Selection Pearson Correlation.....	40
Gambar 24 Hasil Feature Selection Extra Tree Classifier .....	41
Gambar 25 Source code algoritma machine learning KNN.....	41
Gambar 26 Source code algoritma machine learning Decision Tree.....	42
Gambar 27 Source code algoritma machine learning XGBoost .....	42
Gambar 28 Learning Curve KNN.....	43
Gambar 29 Learning Curve Decision Tree .....	44
Gambar 30 Learning Curve XGBoost .....	44
Gambar 31 Hyperparameter tuning KNN.....	45
Gambar 32 Validasi silang KNN .....	45
Gambar 33 Hyperparameter tuning Decision Tree .....	46
Gambar 34 Validasi silang Decision Tree .....	46



Gambar 35 Hyperparameter tuning XGBoost .....	47
Gambar 36 Validasi silang XGBoost.....	47
Gambar 37 Confusion Matrix hasil tuned KNN .....	48
Gambar 38 Confusion Matrix hasil tuned Decision Tree .....	49
Gambar 39 Confusion Matrix hasil tuned XGBoost.....	50
Gambar 40 Learning Curve Tuned KNN.....	51
Gambar 41 Learning Curve Tuned Decision Tree.....	51
Gambar 42 Learning Curve Tuned XGBoost .....	51
Gambar 43 Feature Importance hasil model XGBoost .....	53
Gambar 44 Tingkat Churn Berdasarkan NumOfProduct.....	54
Gambar 45 Tingkat Churn berdasarkan IsActiveMember .....	54
Gambar 46 Tingkat Churn berdasarkan Age .....	55
Gambar 47 Tingkat Churn Berdasarkan Geography.....	56



UNIVERSITAS  
MERCU BUANA

## DAFTAR LAMPIRAN

Lampiran 1. Biodata Penulis.....	62
Lampiran 2. Kartu Bimbingan .....	63
Lampiran 3. Curriculum Vitae .....	64
Lampiran 4. Plagiarism Check Laporan TA .....	66

