COMPARATIVE ANALYSIS OF LOGISTIC REGRESSION, DECISION TREE, AND RANDOM FOREST FOR HIGH-VALUE CUSTOMER TRANSACTION PREDICTION

THESIS REPORT

MUHAMMAD ILHAM HAEKAL
415220101223

INFORMATICS STUDY PROGRAM
FACULTY OF COMPUTER SCIENCE
MERCU BUANA UNIVERSITY
JAKARTA
2026

COMPARATIVE ANALYSIS OF LOGISTIC REGRESSION, DECISION TREE, AND RANDOM FOREST FOR HIGH-VALUE CUSTOMER TRANSACTION PREDICTION

THESIS REPORT

Submitted as One of the Requirements to Obtain a Bachelor's Degree

MUHAMMAD ILHAM HAEKAL

415220101223

INFORMATICS STUDY PROGRAM

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS MERCU BUANA

JAKARTA

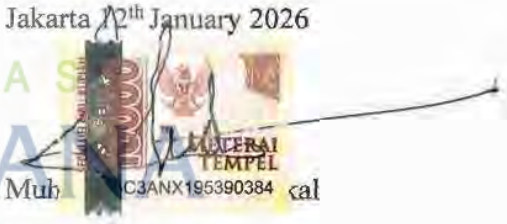
2026

OWN WORK STATEMENT PAGE

I, the undersigned below	:	
Name	:	Muhammad Ilham Haekal
Student Number (NIM)	:	41522010223
Study Program	:	Informatics Engineering

I hereby declare in truth that the Final Project titled " Comparative Analysis of Logistic Regression, Decision Tree, and Random Forest for High-Value Customer Transaction Prediction" is my own original work and does not contain any elements of plagiarism, copyright infringement, or illegal content of any kind, nor does it violate the law or the rights of any party. Should any violation of this statement be discovered in the future, I am prepared to bear all legal consequences and release Universitas Mercu Buana from any form of legal claims, and I am ready to accept the academic sanctions applicable at the university; I make this statement in truth to be used as required.

Jakarta 12th January 2026



Muh kal

UNIVERSITAS
MERCU BUANA
AKREDITASI UNGGUL

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name	: Muhammad Ilham Haekal
NIM /Student ID Number	: 41522010223
Program Studi /Study Program	: Teknik Informatika

Dengan Judul Tugas Akhir
/The title:
"Comparative Analysis of Logistic Regression, Decision Tree, and Random Forest for High-Value Customer Transaction Prediction "
telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:
/Similarity checks have been carried out with the Turnitin system on the date:

10 Januari 2026

dengan nilai persentase sebesar :
/with a percentage value of:

12%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*
File hasil cek similarity turnitin:
/Turnitin similarity report file

https://drive.google.com/file/d/1t3VYNm6YaySBKQvruvBz-BM17eCs2s_o/view?usp=drive_link

Jakarta, 10 Januari 2026
Admin Turnitin Fasilkom UMB

Agung Prawoto

Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

Fakultas Ilmu Komputer
KAMPUS MENARA BHAKTI
Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650
Telp. 021-5840816 (Hunting), Psw : 5700 Fax. 021-5840813
http://www.mercubuana.ac.id, e-mail : fasilkom@mercubuana.ac.id

ENDORSEMENT PAGE

This Final Project is submitted by	:	
Name	:	Muhammad Ilham Haekal
Student Number (NIM)	:	41522010223
Study Program	:	Informatics Engineering
Title	:	Comparative Analysis of Logistic Regression, Decision Tree, and Random Forest for High-Value Customer Transaction Prediction

Has been successfully defended in the examination on January 12, 2026, and accepted as part of the requirements to obtain a Bachelor's degree in the Informatics Engineering Study Program, Faculty of Computer Science, Universitas Mercu Buana.

Approved by:

Supervisor

Ilham Nugraha, S.Kom., M.Sc.

NIDN/NUPTK: 0307098904

Jakarta, 12 January 2026

Acknowledged by,

Dean of the Faculty of Computer Science	Head of Informatics Engineering Study Program
Dr. Bambang Jokonowo, S.Si., MTI	Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0320037002	NIDN/NUPTK: 0225067701

PREFACE

First and foremost, I offer my deepest gratitude to God Almighty for His blessings and grace, which have enabled me to complete this Undergraduate Thesis. This work is submitted in partial fulfillment of the requirements for the Bachelor's Degree in Informatics at the Faculty of Computer Science, Universitas Mercu Buana. I am fully aware that completing this journey—from my initial studies to the finalization of this thesis—would have been immensely difficult without the invaluable support and guidance from various parties. Therefore, I would like to express my sincere appreciation to:

1. Prof. Dr. Andi Adriansyah, M.Eng., as the Rector of Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S.Si., MTI, as the Dean of the Faculty of Computer Science.
3. Dr. Hadi Santoso, S.Kom., M.Kom., as the Head of the Informatics Study Program at Universitas Mercu Buana.
4. Ilham Nugraha, S.KOM., M.Sc., my thesis advisor, for providing constant guidance, motivation, and for generously dedicating time, energy, and insight to ensure this project remained well-structured and on schedule.
5. My Parents, for their unwavering support, prayers, and encouragement throughout my academic journey at Universitas Mercu Buana.

Finally, I hope that God Almighty bestows His blessings upon all those who have assisted me. May this thesis contribute meaningfully to the advancement of knowledge in the field of technology.

Jakarta, August 30, 2025

Muhammad Ilham Haekal

FINAL PROJECT PUBLICATION APPROVAL STATEMENT PAGE FOR THE UNIVERSITY REPOSITORY

As an academic member of Universitas Mercu Buana, I, the undersigned:

Name	:	Muhammad Ilham Haekal
Student Number (NIM)	:	41522010223
Study Program	:	Informatics Engineering
Title	:	Comparative Analysis of Logistic Regression, Decision Tree, and Random Forest for High-Value Customer Transaction Prediction

For the purpose of the development of science and knowledge, I hereby grant permission and agree to provide **Universitas Mercu Buana** with a **Non-exclusive Royalty-Free Right** for my scientific work titled above, including any accompanying materials (if required).

Under this Non-exclusive Royalty-Free Right, **Universitas Mercu Buana** reserves the right to store, transfer media/format, manage within a database, maintain, and publish my Final Project, provided that my name is always cited as the author/creator and as the owner of the Copyright.

I make this statement in truth.

Jakarta 12th January 2026

Mu ckal

ABSTRACT

In the era of digital commerce, the ability to predict high-value customer transactions has become a strategic asset for businesses aiming to increase revenue and customer retention. This research conducts a comparative analysis of three machine learning models—Logistic Regression, Decision Tree, and Random Forest—to classify and predict high-value transactions using the Online Retail dataset from the UCI Machine Learning Repository. The study employs key behavioral features, including Recency, Frequency, and Product Variety, derived from transactional data to model customer behavior and assess future transaction value. Furthermore, the study incorporates an extended analysis using the XGBoost classifier to evaluate the performance gains of gradient boosting techniques against the primary models. These models represent a spectrum from transparent linear methods to powerful ensemble methods, making their comparison crucial for balancing predictive power with interpretability. The performance of all three models is rigorously evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The comprehensive analysis provides guidance on the inherent trade-off between predictive performance and model explainability, assisting e-commerce businesses in selecting the most suitable algorithm for their strategic needs.

Keywords: High-Value Transactions, Logistic Regression, Decision Tree, Random Forest, Predictive Modeling, Customer Behavior, Online Retail Dataset, E-Commerce Analytics, XGBoost

ABSTRAK

Di era perdagangan digital, kemampuan untuk memprediksi transaksi pelanggan bernilai tinggi telah menjadi aset strategis bagi bisnis yang bertujuan untuk meningkatkan pendapatan dan retensi pelanggan. Penelitian ini melakukan analisis komparatif terhadap tiga model pembelajaran mesin—Regresi Logistik, Pohon Keputusan, dan Hutan Acak—untuk mengklasifikasikan dan memprediksi transaksi bernilai tinggi menggunakan dataset Ritel Online dari Repositori Pembelajaran Mesin UCI. Studi ini menggunakan fitur perilaku utama, termasuk Kekinian, Frekuensi, dan Variasi Produk, yang berasal dari data transaksional untuk memodelkan perilaku pelanggan dan menilai nilai transaksi di masa mendatang. Lebih lanjut, studi ini menggabungkan analisis yang diperluas menggunakan pengklasifikasi XGBoost untuk mengevaluasi peningkatan kinerja teknik penguatan gradien terhadap model utama. Model-model ini mewakili spektrum dari metode linier yang transparan hingga metode ensemble yang kuat, sehingga perbandingannya sangat penting untuk menyeimbangkan kekuatan prediksi dengan interpretasi. Kinerja ketiga model tersebut dievaluasi secara ketat menggunakan metrik seperti akurasi, presisi, recall, skor F1, dan ROC-AUC. Analisis komprehensif ini memberikan panduan tentang pertukaran yang melekat antara kinerja prediktif dan kemampuan menjelaskan model, membantu bisnis e-commerce dalam memilih algoritma yang paling sesuai untuk kebutuhan strategis mereka

Kata Kunci: Transaksi Bernilai Tinggi, Regresi Logistik, Pohon Keputusan, Hutan Acak, Pemodelan Prediktif, Perilaku Pelanggan, Kumpulan Data Ritel Online, Analisis E-Commerce, XGBoost

MOTTO

"And whoever fears Allah—He will make for him a way out. And will provide for him from where he does not expect."

(Qur'an, Surah At-Talaq, 65:2–3)

Every achievement I have reached stands not upon my own strength, but upon the countless silent sacrifices, relentless prayers, and unwavering love of two extraordinary souls—my parents. Their unseen struggles, their quiet resilience, and their boundless devotion have carried me through this journey.

This thesis is not merely a personal milestone; it is a living testament to their faith, their hardship, and their enduring presence in my life.

I walk forward with the knowledge that every step is a result of divine mercy and their undying support. To Allah belongs all praise; to them, my eternal gratitude.

DEDICATION

In the name of Allah, the Most Gracious, the Most Merciful. With profound gratitude and humility, I dedicate this work to:

1. **My Creator, Allah Subhanahu wa Ta'ala**

   The source of all strength, knowledge, and patience. It is by His infinite mercy and guidance that I have reached this point. When I faltered, He steadied me. When I doubted, He illuminated my path. Every page of this thesis, every line of code, every late night spent learning—was made possible only through His grace.

   O Allah, accept this effort as a small act of worship. Let it be beneficial in this world and weighty in the Hereafter.

2. **My beloved parents**

   There are no words vast enough to contain the depth of my gratitude. To the father who bore burdens so I could carry dreams, and to the mother

whose prayers became my shield when all else failed—this thesis is yours as much as it is mine.
Your sacrifices were often invisible, but never unrecognized. You gave without asking, stood silently behind every step, and planted seeds of strength and purpose within me.
May Allah bless you both with the highest rank in Paradise, as you have been my greatest blessing in this life.

3. **To my educators and mentors**

   Especially to my supervisors and lecturers in the Informatics Engineering Department, Universitas Mercu Buana—thank you for the knowledge imparted, the patience offered, and the wisdom shared. You have shaped not only my academic foundation but also my mindset as a lifelong learner. May Allah reward you for every lesson taught, and every mind enlightened.

4. **To those who walked beside me**

   Friends, peers, and companions in struggle—thank you for the conversations that sparked ideas, the laughter that lifted burdens, and the moments of shared exhaustion and triumph. Your presence, no matter how small, left a mark on this journey.

5. **To the one whose presence shaped a chapter of my life**

   Though our paths have since diverged, your support in a critical phase of this academic journey was real and meaningful. Your encouragement, even if brief, helped carry me through moments I might not have withstood alone. This work carries a quiet trace of that chapter—with respect, gratitude, and closure.

FOREWORD

In the Name of Allah, the Most Gracious, the Most Merciful. All praise is due to Allah, Lord of the worlds. It is by His mercy, guidance, and wisdom that this academic journey has reached its culmination.

This undergraduate thesis, entitled "Comparative Analysis of Logistic Regression, Decision Tree, and Random Forest for High-Value Customer Transaction Prediction" is submitted in partial fulfillment of the requirements for the degree of Bachelor of Inf ormatics Engineering at Mercu Buana University. More than an academic document, it is the result of a long and personal journey—of learning, unlearning, rethinking, doubting, and ultimately trusting the process that led to growth both as a student and as a person.

The process of completing this thesis has been filled with both challenges and moments of inspiration. There were nights of uncertainty, days of relentless revision, and countless instances where I questioned my capability to finish. Yet through every struggle, Allah Subhanahu wa Ta'ala provided ease after difficulty, and guidance amid confusion. I am deeply aware that every ounce of progress made is only by His will and infinite mercy.

I would like to take this opportunity to express my sincerest gratitude and heartfelt appreciation to the many individuals who have supported me throughout this endeavor:

First and foremost,

1. To Allah Subhanahu wa Ta'ala The One who grants knowledge to whom He wills, and the One who guides hearts. Every letter in this thesis is written under His watch, every insight achieved through His permission. I ask that He accepts this humble work as a part of my lifelong pursuit of beneficial knowledge (*'ilm an-nafi'*) and makes it a means of benefit to others.
2. To my beloved parents Thank you will never be enough. Your endless sacrifices, late-night prayers, and unwavering belief in my potential are the invisible pillars upon which this achievement stands. You gave me everything you never had, so I could become someone you never had the

chance to be. May Allah reward you with goodness in this life and eternal peace in the next.

3. To my supervisors and lecturers at the Faculty of Computer Science, Thank you for your patience, guidance, and academic integrity. Your insights have not only enriched this research, but also taught me how to think critically, write responsibly, and pursue knowledge with sincerity. I especially thank my thesis advisor, whose mentorship was both firm and nurturing, pushing me to reach higher academic standards while remaining grounded.
4. To my fellow students and peers Who became companions in struggle and celebration. Thank you for the shared deadlines, the mutual motivation, the quick code reviews, and the small but powerful gestures of solidarity. You made this journey lighter and more meaningful.
5. To the academic community and the authors of the open-access journals I referenced Thank you for your commitment to making knowledge accessible and your contributions to research. This thesis is built upon your foundations.

I am fully aware that this work is not without flaws and limitations. Every shortcoming within this thesis is my own, and I wholeheartedly welcome constructive feedback, corrections, and suggestions for future improvement.

It is my deepest hope that this research may be useful—academically, practically, or even personally—to those who read it. May it contribute, in however small a way, to the development of knowledge in the field of data science and to the continued relevance of ethical, explainable machine learning in our increasingly digital world.

May Allah make this work a source of benefit, a seed for future research, and a small act of devotion recorded in the scale of good deeds.

Jakarta, 16-01-2026

Muhammad Ilham Haekal

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

LIST OF APPENDIX