

**SEGMENTASI DAN PREDIKSI POTENSIAL PEMBELIAN
LAYANAN *PERSONAL TRAINER* BERDASARKAN
KARAKTERISTIK MENGGUNAKAN
K-MEANS CLUSTERING DAN *XGBOOST***



Tugas Akhir

Fransiska

00000060505

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**SEGMENTASI DAN PREDIKSI POTENSIAL PEMBELIAN
LAYANAN *PERSONAL TRAINER* BERDASARKAN
KARAKTERISTIK MENGGUNAKAN *K-MEANS*
CLUSTERING DAN *XGBOOST***



Tugas Akhir

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Fransiska

00000060505

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Fransiska

Nomor Induk Mahasiswa : 00000060505

Program Studi : Sistem Informasi

Tugas Akhir dengan judul:

SEGMENTASI DAN PREDIKSI POTENSIAL PEMBELIAN LAYANAN
PERSONAL TRAINER BERDASARKAN KARAKTERISTIK
MENGUNAKAN *K-MEANS CLUSTERING* DAN *XGBOOST*

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 16 Mei 2025



Fransiska

HALAMAN PENGESAHAN

Skripsi dengan judul

**SEGMENTASI DAN PREDIKSI POTENSIAL PEMBELIAN LAYANAN
PERSONAL TRAINER BERDASARKAN KARAKTERISTIK
MENGGUNAKAN K-MEANS CLUSTERING DAN XGBOOST**

Oleh

Nama : Fransiska
NIM : 00000060505
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

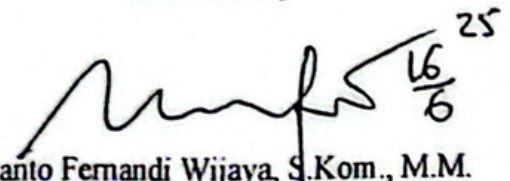
Telah diujikan pada hari Kamis, 5 Juni 2025

Pukul 08.00 s.d 10.00 dan dinyatakan

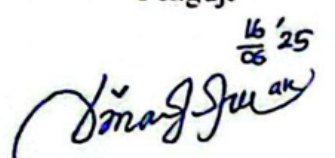
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang


Dr. Santo Fernandi Wijaya, S.Kom., M.M.
0310016902

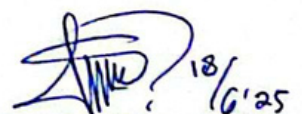
Penguji


Dinar Ajeng Kristiyanti, S.Kom, M.Kom.
0330128801

Pembimbing


Prof. Dr. Friska Natalia, S.Kom., M.T., PhD.
0306128307

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
03130588001

iii

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Fransiska

NIM : 00000060505

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

SEGMENTASI DAN PREDIKSI POTENSIAL PEMBELIAN LAYANAN
PERSONAL TRAINER BERDASARKAN KARAKTERISTIK
MENGUNAKAN *K-MEANS CLUSTERING* DAN *XGBOOST*

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 16 Mei 2025



Fransiska

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul **"Segmentasi dan Prediksi Potensial Pembelian Layanan *Personal Trainer* Berdasarkan Karakteristik menggunakan *K-Means Clustering* dan *XGBoost*"**. Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan studi dan memperoleh gelar sarjana. Dalam penyusunan tugas akhir ini, penulis menerima banyak bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Prof. Dr. Friska Natalia, S.Kom., M.T., PhD. selaku Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya tugas akhir ini.
5. Manajemen salah satu pusat kebugaran di Indonesia yang telah berkenan memberikan izin dan akses data operasional untuk keperluan tugas akhir ini. Dukungan dan kepercayaan yang diberikan sangat membantu penulis dalam memperoleh data nyata sebagai dasar analisis dalam tugas akhir ini.
6. Orang tua tercinta dan seluruh keluarga saya yang telah menjadi sumber kekuatan sepanjang perjalanan ini. Terima kasih atas doa, semangat, serta dukungan baik secara moral maupun material yang tiada henti sejak awal perkuliahan hingga penyusunan tugas akhir ini. Terima kasih telah sabar mendampingi di masa-masa sulit, menjadi tempat pulang ketika lelah, dan bahkan ikut hadir serta menemani di momen penting seperti hari sidang. Kehadiran dan kasih sayang kalian menjadi alasan kuat bagi saya untuk terus maju dan menyelesaikan perjalanan ini dengan sebaik-baiknya.

7. Teman-teman seperjuangan serta orang-orang terdekat, yang telah menjadi tempat berbagi, berdiskusi, dan saling menyemangati dalam menghadapi berbagai tantangan selama penyelesaian tugas akhir ini.
8. Diri sendiri, atas dedikasi, komitmen, dan ketekunan dalam menghadapi berbagai tantangan selama proses penyusunan tugas akhir ini. Semangat untuk terus belajar, memperbaiki diri, dan menyelesaikan setiap tahapan dengan sungguh-sungguh menjadi bagian penting dalam tercapainya penyelesaian penelitian ini. Terima kasih karena telah memilih untuk tidak menyerah, untuk tetap hadir setiap hari, dan menyelesaikan apa yang telah dimulai. Untuk semua proses panjang yang dijalani dengan air mata, kelelahan, dan hari-hari di mana rasanya ingin menyerah, tetapi tetap memilih untuk terus melangkah. *Life can be heavy, especially if you try to carry it all at once. Part of growing up and moving into new chapters of your life is about catch and release.* Ternyata, tumbuh juga tentang belajar melepaskan, dan menerima bahwa tidak harus selalu kuat setiap saat. *I want to thank me for believing in me, for doing all this hard work, for having no days off. For never quitting, and for just being me through it all.*

Penulis menyadari bahwa tugas akhir ini masih memiliki keterbatasan. Oleh karena itu, penulis sangat terbuka terhadap kritik dan saran yang membangun untuk penyempurnaan ke depannya. Semoga tugas akhir ini dapat bermanfaat bagi penulis sendiri, pihak akademisi, maupun praktisi yang berkecimpung di bidang analisis data dan industri kebugaran.

Akhir kata, semoga segala amal dan bantuan yang telah diberikan mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.

Tangerang, 16 Mei 2025



Fransiska

SEGMENTASI DAN PREDIKSI POTENSIAL PEMBELIAN LAYANAN PERSONAL TRAINER BERDASARKAN KARAKTERISTIK MENGGUNAKAN K-MEANS CLUSTERING DAN XGBOOST

(Fransiska)

ABSTRAK

Industri kebugaran di Indonesia terus berkembang seiring meningkatnya kesadaran masyarakat terhadap gaya hidup sehat. Namun, layanan pelatih pribadi masih belum dimanfaatkan secara optimal oleh seluruh anggota pusat kebugaran, sehingga menimbulkan tantangan dalam memahami perilaku pelanggan dan menyusun strategi pemasaran yang tepat sasaran. Dari perspektif sistem informasi, pendekatan analisis pelanggan umumnya dilakukan secara terpisah antara segmentasi dan prediksi, sehingga menghasilkan wawasan yang terfragmentasi dan kurang efektif.

Penelitian ini bertujuan mengembangkan pendekatan terpadu dengan menggabungkan algoritma K-Means Clustering dan Extreme Gradient Boosting dalam satu kerangka kerja data mining. K-Means digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik seperti durasi keanggotaan, frekuensi kunjungan, waktu kunjungan, dan jenis keanggotaan. Hasil segmentasi ini digunakan sebagai fitur tambahan dalam model prediksi menggunakan Extreme Gradient Boosting, yang memprediksi kemungkinan pelanggan membeli layanan pelatih pribadi. Pemilihan kedua algoritma ini didasarkan pada kemampuannya dalam mengungkap pola perilaku dan menghasilkan klasifikasi yang akurat. Seluruh proses dilakukan mengikuti tahapan Cross-Industry Standard Process for Data Mining.

Model yang dibangun menghasilkan akurasi 71,00%, Area Under the Curve sebesar 70,19%, dan F1-score sebesar 40,00% untuk kelas pembeli. Integrasi kedua algoritma ini memungkinkan pemahaman yang lebih komprehensif terhadap karakteristik pelanggan dan potensinya dalam membeli layanan, serta dapat menjadi dasar penyusunan strategi pemasaran yang lebih efektif dan berbasis data.

Kata kunci: Data Mining, K-Means Clustering, Personal Trainer, Segmentasi Pelanggan, XGBoost

***SEGMENTATION AND PREDICTION OF POTENTIAL
PERSONAL TRAINER PURCHASE
BASED ON CHARACTERISTICS
USING K-MEANS CLUSTERING AND XGBOOST***

(Fransiska)

ABSTRACT (English)

The fitness industry in Indonesia continues to grow along with increasing public awareness of healthy lifestyles. However, personal trainer services remain underutilized by many fitness center members, creating challenges in understanding customer behavior and formulating effective marketing strategies. From an information systems perspective, customer analysis approaches are often separated into segmentation and prediction processes, resulting in fragmented insights that are less effective for data-driven decision-making.

This study aims to develop an integrated analytical framework by combining the K-Means Clustering algorithm and Extreme Gradient Boosting within a unified data mining process. K-Means is applied to group customers based on characteristics such as membership duration, visit frequency, preferred workout times, and membership types. The segmentation results are then used as additional features in a prediction model built with Extreme Gradient Boosting to estimate the likelihood of purchasing personal trainer services. These algorithms were selected for their respective strengths in discovering behavior patterns and generating accurate, interpretable classification results. The entire process follows the Cross-Industry Standard Process for Data Mining methodology.

The resulting model achieved an accuracy of 71.00%, an Area Under the Curve of 70.19%, and an F1-score of 40.00% for the buyer class. Integrating both algorithms enables a more comprehensive understanding of customer characteristics and purchasing potential, serving as a foundation for more effective and personalized marketing strategies based on data insights.

Keywords: *Customer Segmentation, Data Mining, Extreme Gradient Boosting, K-Means Clustering, Personal Trainer*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
KATA PENGANTAR.....	v
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	7
1.4.1 Tujuan Penelitian	7
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan	8
BAB II LANDASAN TEORI	10
2.1 Penelitian Terkait.....	10
2.2 Teori Penelitian	14
2.2.1 <i>Gym</i>	14
2.2.2 <i>Personal Trainer (PT)</i>	15
2.2.3 Segmentasi Pelanggan.....	16
2.2.4 Prediksi Potensial Pembelian Layanan.....	17
2.3 <i>Framework</i> dan Algoritma Penelitian	17
2.3.1 <i>Machine Learning</i>	17
2.3.2 <i>Data Mining</i>	19
2.3.3 CRISP-DM.....	19

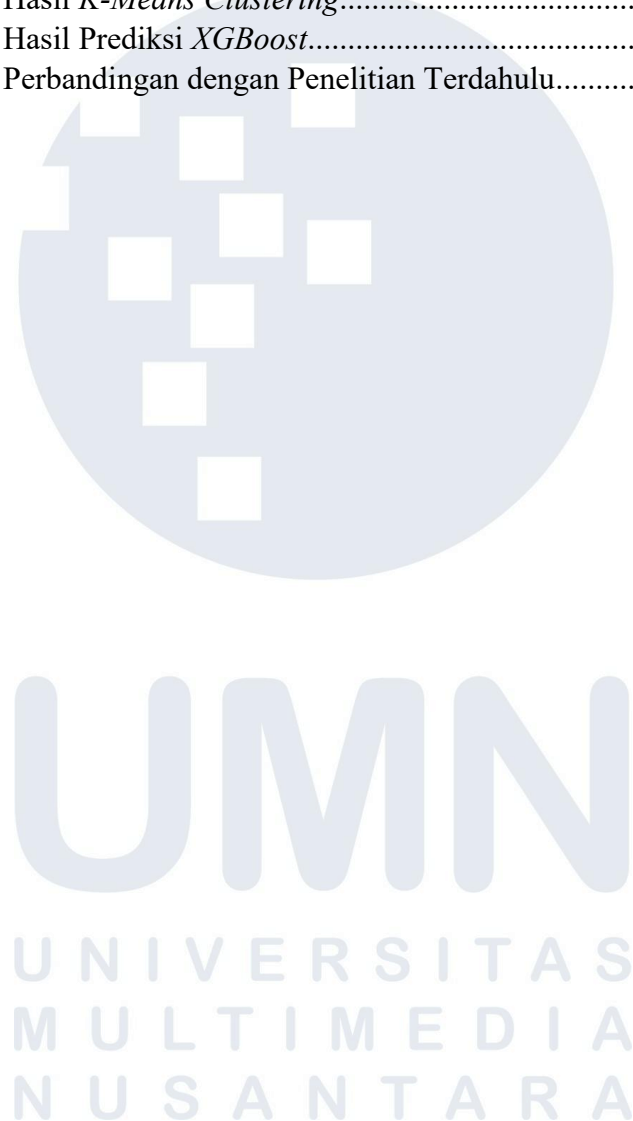
2.3.4	<i>Exploratory Data Analysis (EDA)</i>	20
2.3.5	<i>Data Preprocessing</i>	21
2.3.6	<i>Winsorizing</i>	21
2.3.7	<i>Encoding</i>	22
2.3.8	<i>Clustering</i>	23
2.3.9	<i>K-Means</i>	23
2.3.10	<i>t-SNE</i>	27
2.3.11	<i>Ensemble Learning</i>	28
2.3.12	<i>Extreme Gradient Boosting</i>	29
2.3.13	Pemisahan Data	32
2.3.14	SMOTE	33
2.3.15	<i>Hyperparameter Tuning</i>	34
2.3.16	<i>RandomizedSearchCV dan Cross-Validation</i>	35
2.3.17	Evaluasi Model	35
2.4	Tools dan Software Penelitian	41
2.4.1	Microsoft Excel	41
2.4.2	Python	41
2.4.3	Visual Studio Code	42
BAB III METODOLOGI PENELITIAN		44
3.1	Gambaran Umum Objek Penelitian	44
3.2	Metode Penelitian	44
3.2.1	Alur Penelitian	45
3.2.2	Metode Data Mining	46
3.1	Teknik Pengumpulan Data	49
3.1.1	Populasi dan Sampel	49
3.1.1	Sumber dan Proses Pengumpulan Data	49
3.1.2	Periode Data	51
3.2	Variabel Penelitian	51
3.2.1	Variabel Dependen	51
3.2.2	Variabel Independen	52
3.3	Teknik Analisis Data	54
BAB IV ANALISIS DAN HASIL PENELITIAN		58

4.1	<i>Business Understanding</i>	58
4.2	<i>Data Understanding</i>	59
4.3	<i>Data Preparation</i>	71
4.4	<i>Modeling</i>	78
4.4.1	Segmentasi Pelanggan (<i>K-Means Clustering</i>).....	80
4.4.2	Prediksi Potensi Pembelian (<i>XGBoost</i>)	90
4.5	<i>Evaluation</i>	97
4.6	<i>Deployment</i>	102
4.7	Hasil Analisis dan Diskusi	103
BAB V SIMPULAN DAN SARAN		109
5.1	Simpulan	109
5.2	Saran	109
DAFTAR PUSTAKA		111
LAMPIRAN		117



DAFTAR TABEL

Tabel 2 1 Tabel Penelitian Terdahulu	10
Tabel 2 2 Kriteria Interpretasi Nilai AUC.....	41
Tabel 3 1 Tabel Perbandingan Metode <i>Data Mining</i>	46
Tabel 3 2 Tabel Ringkasan <i>Dataset</i>	50
Tabel 3 3 Tabel Perbandingan <i>Tools</i>	54
Tabel 3 4 Tabel Perbandingan Bahasa Pemrograman.....	56
Tabel 4 1 Tabel Hasil <i>K-Means Clustering</i>	103
Tabel 4 2 Tabel Hasil Prediksi <i>XGBoost</i>	104
Tabel 4 3 Tabel Perbandingan dengan Penelitian Terdahulu.....	106



DAFTAR GAMBAR

Gambar 1. 1 Proyeksi Nilai Pasar Industri <i>Health and Fitness Club</i> di Asia Tenggara Tahun 2025-2030	1
Gambar 1. 2 Proyeksi Pertumbuhan Nilai Pasar Aplikasi Kebugaran Global Tahun 2018-2035	2
Gambar 2. 1 <i>Machine Learning</i>	17
Gambar 2. 2 CRISP-DM	19
Gambar 2. 3 <i>Exploratory Data Analysis</i> (EDA)	20
Gambar 2. 4 <i>K-Means</i>	24
Gambar 2. 5 t-SNE	27
Gambar 3. 1 Alur Penelitian	45
Gambar 3. 2 Alur Pengumpulan Data	49
Gambar 4. 1 <i>Workflow</i> Penentuan Karakteristik Pelanggan Berdasarkan Tinjauan Pustaka	59
Gambar 4. 2 Tampilan Paramount	60
Gambar 4. 3 <i>Flowchart</i> Tahapan <i>Data Understanding</i>	60
Gambar 4. 4 <i>Load Data Sales Gym</i>	60
Gambar 4. 5 Info Data Awal <i>Sales Gym</i>	61
Gambar 4. 6 Pemilihan Kolom yang Relevan dan <i>Rename</i> Kolom	62
Gambar 4. 7 Pengecekan Duplikat dan <i>Handling</i> Duplikat pada <i>Sales Gym</i>	62
Gambar 4. 8 Pengecekan <i>Missing Value</i> pada <i>Sales Gym</i>	63
Gambar 4. 9 <i>Handling Missing Value</i> pada <i>Sales Gym</i>	64
Gambar 4. 10 <i>Export Clean Data Sales Gym</i>	64
Gambar 4. 11 Kode untuk <i>Merge Data</i> dan Informasi Awal Data	66
Gambar 4. 12 Tampilan <i>Merge Data</i>	67
Gambar 4. 13 Identifikasi Data Duplikat	67
Gambar 4. 14 Hasil Analisis Univariat	68
Gambar 4. 15 Hasil Analisis Bivariat (Kategori)	68
Gambar 4. 16 Hasil Analisis Bivariat (Numerik)	69
Gambar 4. 17 Identifikasi <i>Missing Value</i>	69
Gambar 4. 18 Identifikasi <i>Outlier</i>	70
Gambar 4. 19 <i>Flowchart</i> Tahapan <i>Data Preparation</i>	71
Gambar 4. 20 <i>Handling</i> Data Duplikat	71
Gambar 4. 21 <i>Handling Missing Value</i>	72
Gambar 4. 22 Jumlah <i>Missing Value</i> setelah <i>Handling Missing Value</i>	72
Gambar 4. 23 Distribusi Kolom Kategorikal setelah <i>Handling Missing Value</i> dan Duplikat	73
Gambar 4. 24 <i>Handling Outlier</i>	73
Gambar 4. 25 <i>Boxplot</i> setelah <i>Handling Outlier</i>	74
Gambar 4. 26 <i>Encoding</i>	75
Gambar 4. 27 Hasil <i>Encoding</i>	75
Gambar 4. 28 <i>Feature Engineering</i>	76

Gambar 4. 29 Konversi Tipe Data	77
Gambar 4. 30 Tampilan Data setelah <i>Data Preparation</i>	77
Gambar 4. 31 Ekspor Data Bersih.....	78
Gambar 4. 32 Info data yang siap untuk di <i>Modeling</i>	79
Gambar 4. 33 Pemilihan Fitur Segmentasi	80
Gambar 4. 34 Sebaran Data Awal Segmentasi	81
Gambar 4. 35 <i>Scaling Data</i>	82
Gambar 4. 36 <i>Heatmap</i> Korelasi antar Fitur Segmentasi	82
Gambar 4. 37 <i>Elbow Method</i>	83
Gambar 4. 38 <i>Modeling K-Means</i>	84
Gambar 4. 39 Interpretasi Hasil <i>K-Means Clustering</i>	84
Gambar 4. 40 Distribusi Jumlah <i>Churn</i> per <i>Cluster</i>	85
Gambar 4. 41 Distribusi Pembelian Layanan <i>Personal Trainer</i> per <i>Cluster</i>	86
Gambar 4. 42 Visualisasi <i>K-Means Clustering</i>	86
Gambar 4. 43 Distribusi Jumlah <i>Member</i> per <i>Cluster</i>	89
Gambar 4. 44 Membuat Target dan Fitur untuk Prediksi	90
Gambar 4. 45 <i>Split Data Training</i> dan <i>Testing</i>	91
Gambar 4. 46 Distribusi Data sebelum SMOTE.....	91
Gambar 4. 47 Distribusi Data setelah SMOTE.....	92
Gambar 4. 48 Pengecekan <i>Feature Importance</i>	92
Gambar 4. 49 <i>Hyperparameter Tuning</i> dan <i>Cross Validation</i>	93
Gambar 4. 50 Hasil <i>Hyperparameter Tuning</i> dan <i>Cross Validation</i>	94
Gambar 4. 51 Latih Model Final.....	95
Gambar 4. 52 Prediksi Probabilitas.....	95
Gambar 4. 53 <i>Feature Importance XGBoost</i>	96
Gambar 4. 54 Evaluasi <i>K-Means Clustering</i>	97
Gambar 4. 55 Evaluasi <i>XGBoost (Train Data)</i>	98
Gambar 4. 56 Evaluasi <i>XGBoost (Test Data)</i>	98
Gambar 4. 57 <i>Confusion Matrix XGBoost</i>	99
Gambar 4. 58 <i>ROC Curve XGBoost</i>	101
Gambar 4. 59 <i>Top 10 Potential PT Buyers</i>	102

UNIVERSITAS
MULTIMEDIA
NUSANTARA

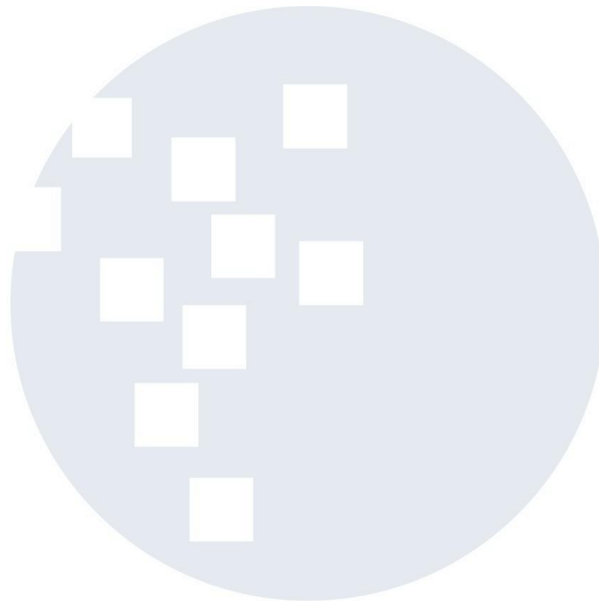
DAFTAR RUMUS

Rumus 2 1 Rumus Euclidean	24
Rumus 2 2 Rumus <i>Objective Function K-Means</i>	25
Rumus 2 3 Rumus <i>Objective Function XGBoost</i>	30
Rumus 2 4 Rumus <i>Silhouette Coefficient</i>	36
Rumus 2 5 Rumus <i>Davies-Bouldin Index</i>	37
Rumus 2 6 Rumus <i>Calinski-Harabasz</i>	37
Rumus 2 7 Rumus Akurasi	39
Rumus 2 8 Rumus Presisi	39
Rumus 2 9 Rumus <i>Recall</i>	39
Rumus 2 10 Rumus <i>F1-Score</i>	39
Rumus 2 11 Rumus <i>True Positive Rate</i>	40
Rumus 2 12 Rumus <i>False Positive Rate</i>	40
Rumus 2 13 Rumus AUC.....	40



DAFTAR LAMPIRAN

Lampiran A Turnitin <i>Similarity Report</i>	117
Lampiran B Form Konsultasi Bimbingan	125



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA