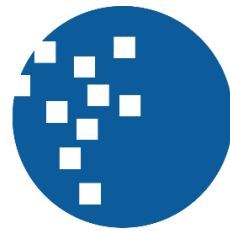


***CUSTOMER CHURN ANALYSIS PADA UNIVERSITAS
MULTIMEDIA NUSANTARA DENGAN MENGGUNAKAN
ALGORITMA DECISION TREE, RANDOM FOREST, DAN
XGBOOST***



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

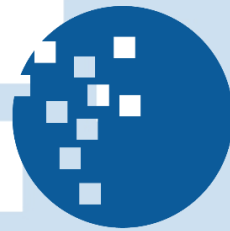
Clement Aurelio

00000043266

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

***CUSTOMER CHURN ANALYSIS PADA UNIVERSITAS
MULTIMEDIA NUSANTARA DENGAN MENGGUNAKAN
ALGORITMA DECISION TREE, RANDOM FOREST, DAN***

XGBOOST



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh

Gelar Sarjana Komputer

Clement Aurelio

00000043266

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

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Clement Aurelio

Nomor Induk Mahasiswa : 00000043266

Program studi : Sistem Informasi

Skripsi dengan judul:

CUSTOMER CHURN ANALYSIS PADA UNIVERSITAS MULTIMEDIA NUSANTARA DENGAN MENGGUNAKAN ALGORITMA DECISION TREE, RANDOM FOREST, DAN XGBOOST

merupakan hasil karya saya sendiri bukan plagiat dari karya 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 skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 14 Mei 2024



Handwritten signature of Clement Aurelio in black ink.

Clement Aurelio

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

HALAMAN PENGESAHAN

Skripsi dengan judul

CUSTOMER CHURN ANALYSIS PADA UNIVERSITAS MULTIMEDIA NUSANTARA DENGAN MENGGUNAKAN ALGORITMA DECISION TREE, RANDOM FOREST, DAN XGBOOST

Oleh

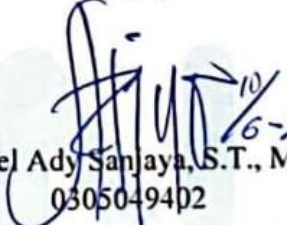
Nama : Clement Aurelio
NIM : 00000043266
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 29 Mei 2024
Pukul 08.00 s/d 10.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang


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

Penguji


Samuel Ady Sanjaya, S.T., M.T.
0305049402

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
313058001

Pembimbing


Johan Setiawan, S.Kom., M.M., M.B.A.
0317106402

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Clement Aurelio
NIM : 00000043266
Program Studi : Sistem Informasi
Jenjang : D3/S1/S2
Judul Karya Ilmiah :

CUSTOMER CHURN ANALYSIS PADA UNIVERSITAS MULTIMEDIA NUSANTARA DENGAN MENGGUNAKAN ALGORITMA DECISION TREE, RANDOM FOREST, DAN XGBOOST

Menyatakan dengan sesungguhnya bahwa 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 akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

Saya tidak bersedia, dikarenakan:

Dalam proses pengajuan penerbitan ke dalam jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

Tangerang, 11 Juni 2024



Clement Aurelio

* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji Syukur atas selesainya penulisan skripsi ini dengan judul: “*Customer Churn Analysis* pada Universitas Multimedia Nusantara dengan Menggunakan Algoritma Decision Tree, Random Forest, dan XGBoost” untuk memenuhi salah satu syarat mencapai gelar Sarjana Komputer Jurusan Sistem Informasi pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, M.A., 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. Bapak Johan Setiawan, S.Kom., M.M., M.B.A., sebagai Dosen Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Kepada Biro Informasi Akademik (BIA) yang telah memberikan izin terhadap pengambilan data untuk keperluan penelitian.
6. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga skripsi ini dapat terselesaikan.

Karya ilmiah ini masih jauh dari sempurna. Baik saran maupun kritik akan sangat diterima oleh penulis untuk memperbaiki hasil pekerjaan saya ke depannya.

Tangerang, 14 Mei 2024



Clement Aurelio

CUSTOMER CHURN ANALYSIS PADA UNIVERSITAS MULTIMEDIA NUSANTARA DENGAN MENGGUNAKAN ALGORITMA DECISION TREE, RANDOM FOREST, DAN XGBOOST

Clement Aurelio

ABSTRAK

Customer churn merupakan perpindahan pelanggan dari suatu perusahaan/bisnis ke perusahaan/bisnis lain. *Customer churn* penting untuk diketahui, terlebih untuk para pemilik bisnis yang mengandalkan loyalitas pelanggan. Fenomena churn dapat terjadi pada berbagai sektor, termasuk sektor pendidikan yang disebut sebagai *student attrition*. Sektor pendidikan yang dibahas pada penelitian ini adalah perguruan tinggi/universitas. Terjadinya *customer churn* pada universitas dapat menyebabkan kerugian, seperti akreditasi/reputasi universitas yang menurun. Akreditasi universitas yang menurun akan menyebabkan universitas kurang dipandang oleh calon mahasiswa. Penelitian ini memanfaatkan algoritma Decision Tree, Random Forest, dan XGBoost untuk membuat model yang dapat melakukan analisa terhadap faktor-faktor penyebab *churn* dan melakukan klasifikasi mahasiswa ke dalam kelas *churn* atau *not churn*. Ketiga algoritma ini digunakan karena mampu menyelesaikan masalah klasifikasi dan memiliki mekanisme kerja yang sama. XGBoost dan Random Forest merupakan algoritma yang menerapkan teknik boosting dan bagging terhadap Decision Tree. Data yang digunakan dalam penelitian ini diperoleh dari Biro Informasi Akademik (BIA) Universitas Multimedia Nusantara (UMN). Data akan diolah dan dibersihkan terlebih dahulu sebelum nantinya digunakan untuk membuat model klasifikasi. Model dengan performa terbaik akan digunakan pada tahap *deployment* untuk melakukan prediksi terhadap *unseen* data. Berdasarkan penelitian yang dilakukan, masing-masing model memiliki performa yang berbeda, baik dari segi akurasi, presisi, *recall*, dan *f1-score*. Model XGBoost memiliki performa terbaik dengan akurasi sebesar 98%, presisi sebesar 99%, *recall* sebesar 96%, dan *f1-score* sebesar 97%. Model ini memiliki performa yang lebih baik dibandingkan Decision Tree sebagai base modelnya dengan kenaikan 2% pada akurasi, 6% pada presisi, 2% pada *recall*, dan 3% pada *f1-score*. Seluruh model memiliki performa yang lebih baik pada kelas mayoritas (*not churn*) dibandingkan pada kelas minoritas (*churn*).

Kata kunci: *Churn, Student Attrition, Machine Learning*, Klasifikasi, Prediksi, Decision Tree, Random Forest, XGBoost

CUSTOMER CHURN ANALYSIS ON UNIVERSITAS MULTIMEDIA NUSANTARA USING THE DECISION TREE, RANDOM FOREST, AND XGBOOST ALGORITHM

Clement Aurelio

ABSTRACT

Customer churn is the movement of customers from one company/business to another company/business. Customer churn is important to know, especially for business owners who rely on customer loyalty. The churn phenomenon can occur in various sectors, including the education sector, which is known as student attrition. The education sector discussed in this research is colleges/universities. The occurrence of customer churn at universities can cause losses, such as decreased university accreditation/reputation. Decreased university accreditation will cause universities to be less viewed by prospective students. This research utilizes the Decision Tree, Random Forest, and XGBoost algorithms to create a model that can analyze the factors that cause churn and classify students into churn or not churn classes. These three algorithms are used because they are able to solve classification problems and have the same working mechanism. XGBoost and Random Forest are algorithms that apply boosting and bagging techniques to Decision Trees. The data used in this research was obtained from the Academic Information Bureau (BIA) of Universitas Multimedia Nusantara (UMN). The data will be processed and cleaned first before being used to create a classification model. The model with the best performance will be used at the deployment stage to make predictions on unseen data. Based on research conducted, each model has different performance, both in terms of accuracy, precision, recall and f1-score. The XGBoost model has the best performance with an accuracy of 98%, precision of 99%, recall of 96%, and f1-score of 97%. This model has better performance than Decision Tree as the base model with an increase of 2% in accuracy, 6% in precision, 2% in recall, and 3% in f1-score. All models have better performance in the majority class (not churn) than in the minority class (churn).

Keywords: *Churn, Student Attrition, Machine Learning, Classification, Prediction, Decision Tree, Random Forest, XGBoost*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA	iv
KATA PENGANTAR	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian dan Manfaat Penelitian.....	4
1.4.1. Tujuan Penelitian	4
1.4.2. Manfaat Penelitian	4
1.5. Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	6
2.1 Penelitian Terdahulu.....	6
2.2 Customer Churn	9
2.3 Framework dan Algoritma	10
2.3.1 CRISP-DM.....	10
2.3.1.1 Business Understanding.....	11
2.3.1.2 Data Understanding	11
2.3.1.3 Data Preparation	11
2.3.1.4 Modeling.....	11
2.3.1.5 Evaluation	12
2.3.1.6 Deployment.....	12

2.3.2	Klasifikasi	12
2.3.3	Oversampling	12
2.3.4	Algoritma	13
2.3.4.1	Decision Tree	14
2.3.4.2	Random Forest	15
2.3.4.3	XGBoost	17
2.3.5	Confusion Matrix	18
2.3.6	Precision	19
2.3.7	Recall	19
2.3.8	F-measure	20
2.3.9	Accuracy	20
2.4	Tools	21
2.4.1	Microsoft Excel	21
2.4.2	Visual Studio Code	21
2.4.3	Python	21
BAB III METODOLOGI PENELITIAN		23
3.1	Objek Penelitian	23
3.1.1.	Profil Perusahaan	23
3.2	Metode Penelitian	25
3.2.1.	CRISP-DM	25
3.2.1.1.	Business Understanding	26
3.2.1.2.	Data Understanding	27
3.2.1.3.	Data Preparation	27
3.2.1.4.	Modeling	29
3.2.1.5.	Evaluation	30
3.2.1.6.	Deployment	30
3.3	Teknik Pengumpulan Data	31
3.4	Teknik Pengambilan Sampel	31
3.5	Teknik Analisis Data	31
BAB IV ANALISIS DAN HASIL PENELITIAN		34
4.1	Business Understanding	34
4.2	Data Understanding	34

4.3	Data Preparation	38
4.4	Modeling	46
4.4.1	Decision Tree	47
4.4.2	Random Forest	49
4.4.3	XGBoost.....	51
4.5	Evaluation.....	53
4.5.1	Decision Tree Evaluation.....	54
4.5.2	Random Forest Evaluation.....	55
4.5.3	XGBoost Evaluation	57
4.6	Deployment	58
4.7	Results and Discussion.....	62
BAB V	SIMPULAN DAN SARAN.....	65
5.1	Kesimpulan.....	65
5.2	Saran.....	66
DAFTAR PUSTAKA	67
LAMPIRAN	71

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	6
Tabel 3. 1 Kelebihan dan Kekurangan Tools.....	32
Tabel 4. 1 Atribut Data.....	35
Tabel 4. 2 Perbandingan Hasil Akhir Tiap Model	63



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 1. 1 Churn Rates Berdasarkan Industri (Monthly)	1
Gambar 2. 1 CRISP-DM Methodology	10
Gambar 2. 2 SMOTE	13
Gambar 2. 3 Decision Tree Classifier	14
Gambar 2. 4 Random Forest Classifier	16
Gambar 2. 5 XGBoost Classifier	17
Gambar 2. 6 Confusion Matrix	18
Gambar 3. 1 Universitas Multimedia Nusantara	24
Gambar 3. 2 Alur Penelitian	26
Gambar 4. 1 Sample Data	37
Gambar 4. 2 Lanjutan Sample Data	37
Gambar 4. 3 Labeling Atribut status_mahasiswa	38
Gambar 4. 4 Cek Null Values	39
Gambar 4. 5 Drop Null Values dan Data Filtering	40
Gambar 4. 6 Sample Data Setelah Reformatting	41
Gambar 4. 7 Lanjutan Sample Data Setelah Reformatting	41
Gambar 4. 8 Cek Null Pada Data Hasil Gabungan	42
Gambar 4. 9 Distribusi Data Pada Kelas	43
Gambar 4. 10 Fill Null Value Pada Data Gabungan	43
Gambar 4. 11 Data Encoding	44
Gambar 4. 12 Splitting Data	45
Gambar 4. 13 Oversampling Imbalanced Dataset	46
Gambar 4. 14 Feature Importance Pada Decision Tree	47
Gambar 4. 15 Pembuatan Model dengan Feature Importance Decision Tree	48
Gambar 4. 16 Feature Importance Pada Random Forest	49
Gambar 4. 17 Pembuatan Model dengan Feature Importance Random Forest	51
Gambar 4. 18 Feature Importance Pada XGBoost	52
Gambar 4. 19 Pembuatan Model dengan Feature Importance XGBoost	53
Gambar 4. 20 Akurasi Model Decision Tree	54
Gambar 4. 21 Performa Rinci Model Decision Tree	54
Gambar 4. 22 Akurasi Model Random Forest	55
Gambar 4. 23 Performa Rinci Model Random Forest	56
Gambar 4. 24 Akurasi Model XGBoost	57
Gambar 4. 25 Performa Rinci Model XGBoost	57
Gambar 4. 26 Distribusi Data untuk Tiap Kelas Pada Unseen Data	59
Gambar 4. 27 Prediksi Unseen Data dengan Model XGBoost	59
Gambar 4. 28 Performa Rinci Model XGBoost Terhadap Unseen Data	60
Gambar 4. 29 Deploy Model Pada Streamlit	61
Gambar 4. 30 Output Prediksi Pada Web App	61

DAFTAR RUMUS

Rumus 2. 1 Gini Index	15
Rumus 2. 2 Rumus Precision	19
Rumus 2. 3 Rumus Recall.....	20
Rumus 2. 4 Rumus f-measure	20
Rumus 2. 5 Rumus Accuracy.....	21

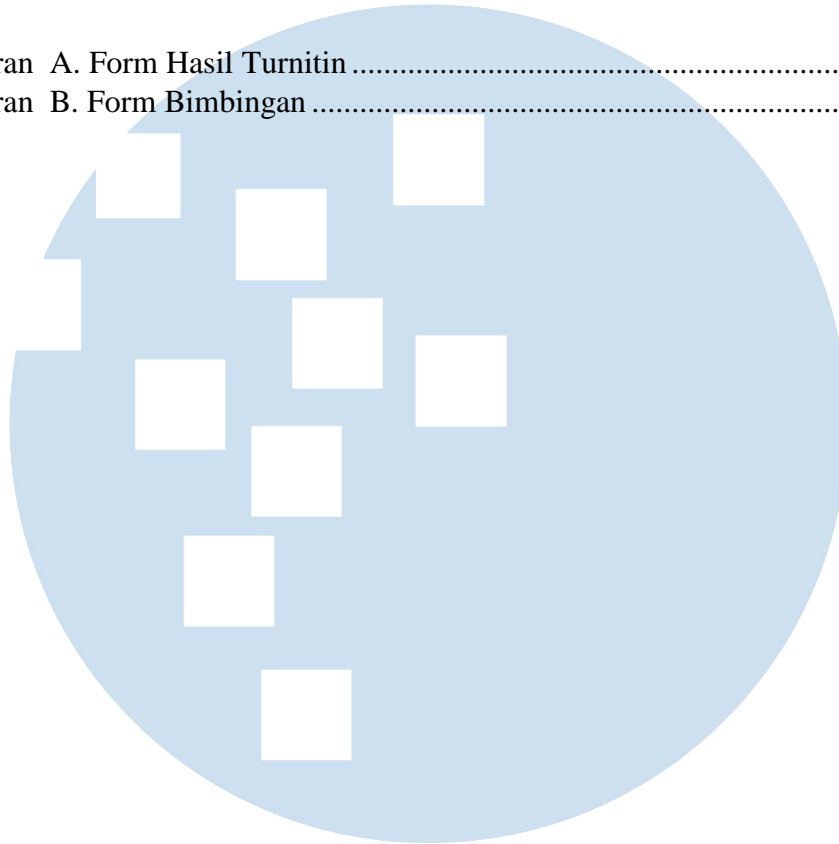


UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A. Form Hasil Turnitin	71
Lampiran B. Form Bimbingan	74



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA