



**PENGEMBANGAN MODEL PREDIKSI *CHURN* PADA
INDUSTRI TELEKOMUNIKASI DENGAN METODE
*LOGISTIC REGRESSION***

SKRIPSI

MUHAMAD ILYAS HAIKAL

2010314056

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA

FAKULTAS TEKNIK

PROGRAM STUDI S1 TEKNIK ELEKTRO

2024



**PENGEMBANGAN MODEL PREDIKSI *CHURN* PADA
INDUSTRI TELEKOMUNIKASI DENGAN METODE
*LOGISTIC REGRESSION***

SKRIPSI

MUHAMAD ILYAS HAIKAL

2010314056

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA

FAKULTAS TEKNIK

PROGRAM STUDI S1 TEKNIK ELEKTRO

2024

LEMBAR PENGESAHAN PENGUJI

Skripsi diajukan oleh :

Nama : Muhamad Ilyas Haikal

NIM : 2010314056

Program Studi : Teknik Elektro

Judul Skripsi : Pengembangan Model Prediksi *Churn* pada Industri Telekomunikasi dengan metode *Logistic regression*

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Teknik pada Program Studi Teknik Elektro, Fakultas Teknik, Universitas Pembangunan Nasional Veteran Jakarta.



Ferdianto. S.T., M.T.

Penguji Utama



Silvia Anggraeni. S.T., M.Sc., Ph.D.

Penguji Lembaga



Dr. Ir. Muchamad Oktaviandri.
S.T., M.T., IPM., ASEAN.Eng.

Plt. Dekan Fakultas Teknik



Fajar Rahayu. S.T., M.T.

Penguji I (Pembimbing)



Ir. Achmad Zuchriadi S.T.,
M.T., CEC

Ka. Prodi Teknik Elektro

Ditetapkan di : Jakarta

Tanggal Ujian : 11 Juli 2024

LEMBAR PENGESAHAN SKRIPSI
PENGEMBANGAN MODEL PREDIKSI *CHURN* PADA
INDUSTRI TELEKOMUNIKASI DENGAN METODE
LOGISTIC REGRESSION

Muhamad Ilyas Haikal

2010314056

Disetujui Oleh :

Pembimbing I



Eajar Rahayu S.T., M.T.

Pembimbing II



Ir. Achmad Zuchriadi S.T., M.T.CEC

Mengetahui,

Ketua Program Studi Teknik Elektro

Fakultas Teknik Universitas Pembangunan Nasional Veteran Jakarta



Ir. Achmad Zuchriadi S.T., M.T.CEC

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri, dengan semua sumber yang dikutip maupun dirujuk telah saya nyatakan benar.

Nama : Muhamad Ilyas Haikal

NIM : 2010314056

Program Studi : Teknik Elektro

Apabila dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka akan bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 10 July 2024

Penulis,



(Muhamad Ilyas Haikal)

HALAMAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta,
saya yang bertanda tangan dibawah ini:

Nama : Muhamad Ilyas Haikal

NIM : 2010314056

Program Studi : Teknik Elektro

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta. Hak Bebas Royalti Noneklusif (Non Exclusive Royalti Free Right) atas karya ilmiah saya yang berjudul:

Pengembangan Model Prediksi *Churn* pada Industri Telekomunikasi dengan Metode *Logistic Regression*

Beserta software yang ada (jika diperlukan) dengan Hak Bebas Royalti ini, Universitas Nasional Veteran Jakarta berhak menyimpan, mengalihkan media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 10 July 2024

Yang menyatakan,



(Muhamad Ilyas Haikal)

PENGEMBANGAN MODEL PREDIKSI *CHURN* PADA INDUSTRI TELEKOMUNIKASI DENGAN METODE *LOGISTIC REGRESSION*

Muhamad Ilyas Haikal

ABSTRAK

Seiring dengan perkembangan pengguna layanan internet dan telekomunikasi di Indonesia, jumlah perusahaan telekomunikasi juga semakin meningkat. Hal ini menyebabkan banyak pelanggan yang telah berlangganan melakukan *Churn*, yaitu pemutusan layanan jasa telekomunikasi oleh pelanggan atau perusahaan. Perusahaan lebih memilih untuk mempertahankan pelanggan karena biaya yang diperlukan lebih sedikit dibandingkan mencari pelanggan baru. Namun, kebijakan untuk mengurangi *Churn* masih belum efektif karena berbagai kendala. Penelitian ini bertujuan untuk membuat model prediksi pelanggan *Churn* yang akurat dan mengambil keputusan strategi mengurangi *Churn* yang efektif. Penelitian ini menggunakan metode prediksi model yang terdiri dari regresi logistik dengan metode (RSCV) Random Search Cross Validation dan SMOTE (*Synthetic Minority Over-sampling Technique*). Hasil dari model LRG SMOTE menunjukkan akurasi sebesar 74.2% dan recall tertinggi sebesar 0.820, menunjukkan kemampuan model dalam mengidentifikasi pelanggan yang benar-benar akan *Churn*. Analisis koefisien menunjukkan bahwa fitur seperti kontrak jangka panjang dan layanan telepon memiliki pengaruh signifikan dalam mengurangi *Churn*, sementara tagihan bulanan yang tinggi meningkatkan risiko *Churn*. Lebih lanjut, analisis *Recursive Feature Elimination* (RFE) menunjukkan bahwa fitur seperti *Total Charges*, *Monthly Charges*, dan *Tenure* merupakan prediktor utama *Churn*. Korelasi tinggi antara layanan keamanan seperti *OnlineSecurity* dan *OnlineBackup* memberikan peluang untuk strategi *cross-selling*, sementara potensi bundling layanan hiburan dapat ditemukan dari korelasi antara *StreamingTV* dan *StreamingMovies*.

Kata Kunci: *Churn Management, Regresi Logistik, RSCV, SMOTE, RFE, Cross-Selling, Bundling*

DEVELOPMENT OF A CHURN PREDICTION IN THE TELECOMMUNICATION INDUSTRY USING LOGISTIC REGRESSION

Muhamad Ilyas Haikal

ABSTRACT

With the growing number of internet and telecommunication service users in Indonesia, the number of telecommunication companies is also increasing. This growth has led to a higher incidence of Churn, where Customers terminate their telecommunication services. Companies prefer to retain Customers as it is less costly than acquiring new ones. However, current strategies to reduce Churn are still facing various challenges. This study aims to create an accurate Churn Prediction model and develop effective strategies to reduce Churn. This research employs various Prediction models, focusing on Logistic regression with the (RSCV) Random Search Cross Validation and SMOTE (Synthetic Minority Over-sampling Technique) method. The LRG SMOTE model stands out with an accuracy of 74.2% and the highest recall of 0.820, demonstrating its effectiveness in identifying Customers who are likely to Churn. Coefficient analysis indicates that features such as long-term contracts and telephone services significantly reduce Churn, while higher monthly charges increase the risk of Churn. Further, Recursive Feature Elimination (RFE) analysis identifies Total Charges, Monthly Charges, and Tenure as the main predictors of Churn. High correlations between security services like OnlineSecurity and OnlineBackup suggest opportunities for cross-selling strategies, while potential bundling strategies are indicated by correlations between StreamingTV and StreamingMovies.

Keywords: *Churn Management, Logistic regression, RSCV, SMOTE, RFE, Cross-Selling, Bundling*

KATA PENGANTAR

Puji syukur kehadiran Allah Subhanahu Wa Ta'ala yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Pengembangan Model Prediksi *Churn* pada Industri Telekomunikasi dengan metode *Logistic regression*”**. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik Elektro di Universitas Pembangunan Nasional Veteran Jakarta.

Pada kesempatan ini, penulis ingin menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Fajar Rahayu, S.T., M.T., selaku dosen pembimbing I yang telah meluangkan waktu, tenaga, dan pikiran dalam membimbing penulis sehingga penelitian ini dapat diselesaikan.
2. Bapak Achmad Zuhriadi P., S.T., M.T., CEC, selaku dosen pembimbing II sekaligus Ketua Program Studi Teknik Elektro, Fakultas Teknik Universitas Pembangunan Nasional Veteran Jakarta, yang telah memberikan banyak saran yang konstruktif serta mendukung kreativitas mahasiswa.
3. Bapak Dr. Muchamad Oktaviandri, ST., MT., IPM., ASEAN.Eng, selaku Dekan Fakultas Teknik Universitas Pembangunan Nasional Veteran Jakarta.
4. Orang tua penulis, Misar Iman dan Zaenabun, yang telah memberikan dukungan baik secara moral maupun materil.
5. Saudara-saudari penulis, Shella Irmawati, Vera Pebrianti, Vikra Rahmalia, Irena Lestari, dan Siti Sarah, yang selalu mendukung dan mendorong penulis untuk terus bersemangat dalam pengerjaan Tugas Akhir ini.
6. Teman-teman Program Studi S1 Teknik Elektro yang telah membantu dan memotivasi penulis dalam penyusunan Tugas Akhir ini.
7. *Last but not least, thank you to myself for fighting to seek knowledge and being true to myself at all times. I appreciate the hard work and determination I have to continue learning and growing to become the best*

version of myself. I thank myself for never giving up and always striving to be better every day.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan dan memiliki kekurangan baik dari segi penjelasan maupun penulisan. Oleh karena itu, penulis dengan rendah hati menerima kritik dan saran yang membangun untuk meningkatkan kesempurnaan Tugas Akhir ini.

Akhir kata, semoga amal ibadah dari semua pihak yang telah membantu dalam penulisan Tugas Akhir ini mendapat ganjaran yang berlimpah dari Allah Subhanahu Wa Ta'ala. Semoga Tugas Akhir ini dapat bermanfaat bagi para pembaca dan pengguna lainnya.

Jakarta, 10 July 2024



Muhamad Ilyas Haikal

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN PENGUJI	ii
HALAMAN PENGESAHAN PEMBIMBING	iii
HALAMAN PERNYATAAN ORISINALITAS	iv
HALAMAN PERSETUJUAN PUBLIKASI SKRIPSI	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	3
1.5 Sistematika Penulisan	3
BAB 2 TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.1.1 Analisa Penelitian Terdahulu	8
2.1.2 Kesimpulan Penelitian Terdahulu.....	10
2.2 Data Mining.....	10
2.2.1 Tahapan Data Mining	11
2.3 Prediksi <i>Churn</i>	13
2.3.1 Tipe <i>Churn</i>	13
2.4 Analisis Regression	14
2.5 Logistik Regresi	15
2.6 Random Search Cross Validation	17

2.6.1 Forward Selection	17
2.6.2 Sequential Forward Selection	18
2.7 Synthetic Minority Over-sampling Technique.....	19
2.8 Evaluasi Performa	20
2.8.1 Confusion Matrix	20
2.8.2 Akurasi	21
2.8.3 Presisi	22
2.8.4 Recall	23
2.8.5 F1 Score.....	24
BAB 3 METODE PENELITIAN	26
3.1 Kerangka Berpikir.....	26
3.2 Alur Kerja Pengolahan Data.....	27
3.2.1 Deskripsi Data.....	28
3.2.2 Data Processing.....	29
3.2.3 Exploratory Data Analysis	30
3.3 Modeling.....	32
3.4 Testing dan Evaluating	33
3.5 Analisa Data	33
3.6 Timeline Alur Penelitian	34
BAB 4 ANALISA DATA.....	35
4.1 Pengolahan Data	35
4.1.1 Prediksi <i>Churn</i>	35
4.1.2 <i>Churn</i>	36
4.1.3 Gender	37
4.1.4 Senior Citizen.....	38
4.1.5 Partner	39
4.1.6 Dependents	40
4.1.7 Phone Service.....	41
4.1.8 Paperless Billing	41
4.1.9 Internet Service	42
4.1.10 Online Security	43

4.1.11 Online Backup	44
4.1.12 Device Protection.....	45
4.1.13 Technology Support	46
4.1.14 Streaming TV and Streaming Movie.....	47
4.1.15 Multiple Lines	48
4.1.16 Contract	49
4.1.17 Payment Method.....	50
4.1.18 Tenure.....	51
4.1.19 Monthly Charges	52
4.1.20 Total Charges	53
4.2 Analisis dan Evaluasi Model.....	54
4.2.1 Regresi Logistik.....	54
4.2.2 Regresi Logistik RSCV	58
4.2.3 Regresi Logistik SMOTE.....	62
4.3 TESTING DAN EVALUATING	65
4.3.1 Pemilihan Model Terbaik Berdasarkan Testing & Evaluating	65
4.3.2 Penerapan Model Prediktif Menggunakan Koefisien Variable	66
4.4 Analisis Data	71
4.4.1 Korelasi Pada Variabel Utuh.....	71
4.4.2 Analisa Recursive Feature Elimination	73
BAB 5 PENUTUP	75
5.1 Kesimpulan.....	75
5.2 Saran	76

DAFTAR PUSTAKA

RIWAYAT HIDUP

LAMPIRAN

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Tahapan Data Mining.....	11
Gambar 2.2 Kurva Regresi Logistik.....	16
Gambar 3.1 Flowchart Alur Penelitian	26
Gambar 3.2 Pemrograman Data Cleansing.....	29
Gambar 3.3 Pemrograman Forward Selection for Correlation.....	30
Gambar 3.4 Pemrograman Target Feature	31
Gambar 3.5 Pemrograman Categorical Feature.....	31
Gambar 3.6 Pemrograman Numerical Feature	31
Gambar 3.7 Pemrograman Training Dataset	32
Gambar 3.8 Pemrograman Fitting Model dengan Logistic Regression	32
Gambar 3.9 Pemrograman Hyperparameter dengan Randomized Search Cross Validation	32
Gambar 3.10 Pemrograman Testing Dataset	33
Gambar 3.11 Pemrograman Performa dan Kinerja Model	33
Gambar 4.1 Pie Chart Perbandingan <i>Churn</i> dan Tidak <i>Churn</i>	36
Gambar 4.2 Grafik Distribusi Gender	37
Gambar 4.3 Grafik Distribusi Senior Citizen	38
Gambar 4.4 Grafik Distribusi Partner	39
Gambar 4.5 Grafik Distribusi Dependents.....	40
Gambar 4.6 Grafik Distribusi Phone Service	41
Gambar 4.7 Grafik Distribusi Paperless Billing	41
Gambar 4.8 Grafik Distribusi Internet Service	42
Gambar 4.9 Grafik Distribusi Online Security	43
Gambar 4.10 Grafik Distribusi Online Backup	44
Gambar 4.11 Grafik Distribusi Device Protection	45
Gambar 4.12 Grafik Distribusi Technology Support	46
Gambar 4.13 Grafik Distribusi Streaming TV.....	47
Gambar 4.14 Grafik Distribusi Streaming Movies.....	47
Gambar 4.15 Grafik Distribusi Multiple Lines.....	48
Gambar 4.16 Grafik Distribusi Contract.....	49
Gambar 4.17 Grafik Distribusi Payment Method.....	50
Gambar 4.18 Grafik Distribusi Tenure	51
Gambar 4.19 Grafik Distribusi Monthly Charges	52
Gambar 4.20 Grafik Distribusi Total Charges.....	53

DAFTAR TABEL

Tabel 2.1 Penelitian-Penelitian Terdahulu	5
Tabel 2.2 Confusion Matrix	20
Tabel 3.1 Deskripsi Data	28
Tabel 3.2 Contoh Tabel Transformasi.....	30
Tabel 4.1 Confusion Matrix dari Nilai Logistik Regresi.....	54
Tabel 4.2 Evaluasi Model dari Logistik Regresi	57
Tabel 4.3 Confusion Matrix dari Nilai Logistik Regresi dengan RSCV.....	58
Tabel 4.4 Evaluasi Model dari Logistik Regresi dengan RSCV	61
Tabel 4.5 Confusion Matrix dari Nilai Logistik Regresi dengan SMOTE ...	62
Tabel 4.6 Evaluasi Model dari Logistik Regresi dengan SMOTE	64
Tabel 4.7 Evaluasi dari Semua Model.....	65
Tabel 4.8 Uji Estimasi Parameter Secara Parsial pada Regresi Logistik SMOTE dengan Variabel Utuh.....	67
Tabel 4.9 Uji Korelasi pada Setiap Variabel dengan Model Regresi Logistik SMOTE pada Variabel Utuh.....	71

DAFTAR LAMPIRAN

Lampiran 1 Dokumentasi Syntax Software Python untuk analisis Regresi Logistik

Lampiran 2 Dokumentasi Syntax Software Python untuk analisis Regresi Logistik
dengan Random Search Cross Validation

Lampiran 3 Dokumentasi Syntax Software Python untuk analisis Regresi Logistik
dengan Synthetic Minority Over-sampling Technic

Lampiran 4 Lembar Konsultasi Dosen