



**IMPLEMENTASI CHURN PREDICTION DI INDUSTRI  
TELEKOMUNIKASI DENGAN METODE *LOGISTIC REGRESSION* DAN  
*CORRELATION-BASED FEATURE SELECTION***

**SKRIPSI**

**DHEA LAKSMI PRIANTO**

**1710511025**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2022**



**IMPLEMENTASI CHURN PREDICTION DI INDUSTRI  
TELEKOMUNIKASI DENGAN METODE *LOGISTIC REGRESSION* DAN  
*CORRELATION-BASED FEATURE SELECTION***

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Mendapatkan Gelar Sarjana  
Komputer**

**DHEA LAKSMI PRIANTO**

**1710511025**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA**

**2022**

## PERNYATAAN ORISINALITAS

Tugas skripsi ini adalah hasil karya sendiri, dan semua sumber dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Dhea Laksmi Prianto

NIM : 1710511025

Tanggal : 10 Desember 2021

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

Jakarta, 10 Desember 2021  
Yang Menyatakan,



(Dhea Laksmi Prianto)

## **PERNYATAAN PERSETUJUAN PUBLIKASI**

Sebagai civitas academica Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Dhea Laksmi Prianto

NIM : 1710511025

Fakultas : Ilmu Komputer

Program Studi : Informatika

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

### **IMPLEMENTASI CHURN PREDICTION DI INDUSTRI TELEKOMUNIKASI DENGAN METODE *LOGISTIC REGRESSION* DAN *CORRELATION-BASED FEATURE SELECTION***

Dengan hak bebas royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan kata (basis data), 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.

Dibuat di : Jakarta  
Pada Tanggal : 10 Desember 2021  
Yang Menyatakan,



(Dhea Laksmi Prianto)



## LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa Tugas Akhir berikut:

Nama : Dhea Laksmi Prianto

NIM : 1710511025


Program Studi : S1-Informatika

Judul Tugas Akhir : Implementasi Churn Prediction di Industri Telekomunikasi Dengan Metode Logistic Regression dan Correlation Based-Feature Selection

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Dr. Ermatita, M.Kom.  
Penguji Utama



Bayu Hananto, S.Kom., M.Kom.  
Penguji Lembaga



Iis Ernawati, S.Kom., M.Si.  
Pembimbing 1



Nurul Chamidah, S.Kom., M.Kom.  
Pembimbing 2



Dr. Ermatita, M Kom.  
Dekan



Yuni Widiastiwi, S.Kom., M.Si.  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 13 Desember 2021



**IMPLEMENTASI CHURN PREDICTION DI INDUSTRI  
TELEKOMUNIKASI DENGAN METODE *LOGISTIC REGRESSION* DAN  
*CORRELATION-BASED FEATURE SELECTION***

**DHEA LAKSMI PRIANTO**

**Abstrak**

Industri telekomunikasi menjadi salah satu industri yang semakin kompetitif dan telah menumbuhkan ketertarikan akan prediksi churn untuk pelanggan. Prediksi churn tentang bagaimana mendeteksi pelanggan yang memiliki kecenderungan untuk meninggalkan layanan. Prediksi ini merupakan salah satu upaya perusahaan untuk mempertahankan pelanggan didalam *Customer Relationship Management (CRM)*. Beberapa penulis mengemukakan bahwa metode *logistic regression* memiliki pemodelan dan hasil performa yang bagus untuk diaplikasikan untuk data prediktif. Dataset diambil dari salah satu perusahaan telekomunikasi di Amerika bernama Orange yang tersedia di situs Kaggle yang kemudian diolah untuk menganalisis performa prediksi churn menggunakan *data mining* dengan teknik pemilihan *correlation-based feature selection forward selection* serta algoritma machine learning *logistic regression*.

**Kata Kunci :** Prediksi *churn*, *customer relationship management (CRM)*, *logistic regression*, *data mining*, *correlation-based feature selection*, *forward selection*, machine learning

**CHURN PREDICTION IMPLEMENTATION IN TELECOMMUNICATION  
INDUSTRY WITH *LOGISTIC REGRESSION AND CORRELATION-BASED  
FEATURE SELECTION***

**DHEA LAKSMI PRIANTO**

**ABSTRACT**

Telecommunication industry became one of the industry that is becoming more competitive and is growing interest in churn prediction for customer. Churn prediction is about how to detect customer who have tendency to leave the service. This prediction is one of the company's effort to retain customers in Customer Relationship Management (CRM). Several authors suggest that *logistic regression* method has good modelling and performance results to be applied to predictive data. Dataset is taken from one of the United States' telecommunication company named Orange which available on Kaggle website then proceed to analyze the performance of churn prediction using data mining with *correlation-based feature selection forward selection* and machine learning *logistic regression* algorithm.

**Keywords :** *churn prediction, customer relationship management (CRM), logistic regression, data mining, correlation-based feature selection, forward selection, machine learning*

## KATA PENGANTAR

Puji syukur kepada kehadiran Allah SWT, sehingga penulis dapat menyelesaikan penulisan Tugas Akhir dengan judul “Implementasi Churn Prediction di Industri Telekomunikasi Dengan Metode Logistic Regression dan Correlation Based-Feature Selection”. Tugas Akhir adalah salah satu dari mata kuliah wajib yang harus ditempuh oleh setiap mahasiswa Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1).

Dalam kesempatan ini, penulis mengucapkan terima kasih atas bantuan dan dukungan, baik materi maupun non-materi yang telah diberikan kepada penulis selama kerja praktik berlangsung secara khusus kepada :

1. Ibu Dr. Ermatita, M. Kom., selaku Dekan Fakultas Ilmu Komputes Universitas Pembangunan Nasional Veteran Jakarta.
2. Ibu Iin Ernawati, S.Kom., M.Si. selaku dosen pembimbing 1 yang telah membimbing, memberikan ilmu, serta mengayomi selama penulisan Tugas Akhir.
3. Ibu Nurul Chamidah, S.Kom., M.Kom. selaku dosen pembimbing 2 yang telah membimbing, memberikan ilmu, serta nasehat yang bermanfaat selama penulisan Tugas Akhir.
4. Ibu Yuni Widiastiwi, S.Kom., M.Si. selaku Ketua Prodi Informatika
5. Edi Prianto dan Emawati selaku orang tua yang telah memberikan dukungan baik secara moril maupun materil.
6. Dony Novrianto, Decky Dwi Priantono dan Kurnia Dara Mutia selaku kakak kandung dan kakak ipar dari penulis yang telah hadir untuk mendukung dan mendorong penulis untuk terus semangat dalam pengerjaan Tugas Akhir ini.
7. Ester Friscila Lumbantobing dan Serafim Clara selaku teman-teman penulis yang telah memberikan saran yang baik, menjadi pendengar yang baik, serta menjadi tempat berbagi cerita selama masa perkuliahan penulis.



8. *Last but not least, I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for never quitting, I wanna thank me for always being a giver and tryna give more than I receive, I wanna thank me for tryna do more right than wrong, I wanna thank me for just being me at all times.*

Penulis menyadari penyusunan Tugas Akhir ini jauh dari sempurna. Maka dari itu diharapkan kritik dan saran yang membangun untuk menyempurnakan penyusunan proposal kedepannya.

Jakarta, 20 Januari 2022

A handwritten signature in black ink, appearing to read 'Hedra', with a horizontal line underneath it.

Penulis

# DAFTAR ISI

PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xiv
BAB 1_PENDAHULUAN .....	15
1.1. Latar Belakang .....	15
1.2. Perumusan Masalah.....	16
1.3. Batasan Masalah.....	16
1.4. Tujuan Penelitian.....	17
1.5. Luaran Yang Diharapkan .....	17
1.6. Kontribusi .....	17
1.7. Sistematika Penulisan.....	18
BAB 2_TINJAUAN PUSTAKA .....	20
2.1. <i>Data Mining</i> .....	20
2.1.1. Tahapan <i>Data Mining</i> .....	20
2.2. Pengertian CRM.....	23
2.2.1. Jenis – Jenis CRM.....	23

2.3. Prediksi <i>Churn</i> .....	24
2.3.1. Tipe <i>Churn</i> .....	25
2.4. <i>Logistic Regression</i> .....	25
2.5. <i>Correlation-Based Feature Selection</i> .....	26
2.5.1. Forward Selection.....	27
2.5.1.1. Sequential Forward Selection .....	28
2.6. Evaluasi Performa .....	28
2.7. Penelitian Terkait .....	30
BAB 3_Metodologi Penelitian.....	31
3.1. Kerangka Berfikir.....	31
3.1.1. Identifikasi Masalah.....	32
3.1.2. Studi Literatur .....	32
3.1.3. Deskripsi Data.....	32
3.1.4. <i>Data Preparation</i> .....	33
3.1.5. <i>Forward Selection</i> .....	35
3.1.5.1 <i>Sequential Forward Selection</i> .....	35
3.1.6. <i>Modelling</i> .....	36
3.1.7. <i>Testing dan Evaluating</i> .....	36
3.2. Tahapan Kegiatan.....	37
BAB 4_PEMBAHASAN.....	38
4.1. Deskripsi Data .....	38
4.2. Data Preparation .....	39
4.3. Data Transformation .....	39
4.4. Identifikasi Korelasi .....	40

4.5. <i>Forward Selection</i> .....	40
4.6. <i>Modelling</i> .....	43
4.7. <i>Testing and Evaluating</i> .....	45
BAB 5_KESIMPULAN.....	49
DAFTAR PUSTAKA .....	50
RIWAYAT HIDUP.....	52
LAMPIRAN.....	53
Lampiran 1. 100 Data Sample Churn Customer Orange .....	54
Lampiran 2. <i>Similarity Index</i> Tugas Akhir .....	58

## DAFTAR TABEL

Tabel 2.1. <i>Confusion Matrix</i> .....	29
Tabel 3.1. Deskripsi Data.....	32
Tabel 3.2. Contoh Tabel Transformasi .....	33
Tabel 3.3. Tahapan Kegiatan .....	37
Tabel 4.1. Frekuensi Keseluruhan Data Churn .....	38
Tabel 4.2. Sample Data Churn .....	38
Tabel 4.3. Hasil Transformasi.....	39
Tabel 4.4. Hasil Korelasi Antar Fitur.....	40
Tabel 4.5. Hasil Variabel <i>Forward Feature Selection</i> .....	41
Tabel 4.6. Hasil Taksiran Koefisien Seluruh Variabel .....	44
Tabel 4.7. <i>Confusion Matrix</i> Seluruh Data .....	46
Tabel 4.8. Hasil Taksiran Koefisien Variabel Terpilih.....	46
Tabel 4.9. <i>Confusion Matrix</i> Variabel Terpilih.....	47
Tabel 4.5. Laporan Klasifikasi .....	48

## DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Tahapan <i>Data Mining</i> .....	21
Gambar 3.1. Kerangka Berfikir.....	31