

**OPTIMASI *RANDOM FOREST* TERHADAP DATA *TELCO CUSTOMER CHURN* MENGGUNAKAN *FIREFLY ALGORITHM***

**SKRIPSI**



**Abril Muhammad Fikar Wijaya  
1910511061**

**INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
2024**

**OPTIMASI *RANDOM FOREST* TERHADAP DATA *TELCO CUSTOMER CHURN* MENGGUNAKAN *FIREFLY ALGORITHM***

**SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana  
Komputer**



**Abril Muhammad Fikar Wijaya  
1910511061**

**INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
2024**

## PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan sumber yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Abril Muhammad Fikar Wijaya  
NIM : 1910511061  
Tanggal : 22 Januari 2024  
Judul Skripsi : *Optimasi Random Forest Terhadap Data Telco Customer Churn Menggunakan Firefly Algorithm*

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

Jakarta, 22 Januari 2024

Yang menyatakan,



Abril Muhammad Fikar Wijaya

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK  
KEPENTINGAN AKADEMIS**

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

Nama : Abril Muhammad Fikar Wijaya

NIM : 1910511061

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan karya ilmiah saya kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exchange Royalty Free Right*) untuk dipublikasikan dengan judul:

**Optimasi *Random Forest* terhadap *Data Telco Customer Churn*  
Menggunakan *Firefly Algorithm***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasi skripsi saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta data sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 21 Januari 2024

Yang Menyatakan,




Abril Muhammad Fikar Wijaya


## LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

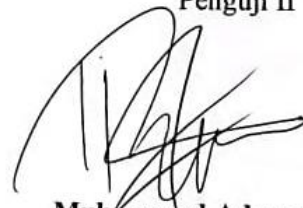
Nama : Abril Muhammad Fikar Wijaya  
NIM : 1910511061  
Program Studi : S1 Informatika  
Judul Tugas Akhir : Optimasi *Random Forest* Terhadap Data *Telco Customer Churn* Menggunakan *Firefly Algorithm*

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

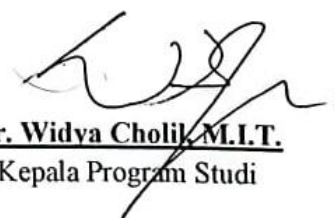
  
Bayu Hananto, S.Kom., M.Kom.  
Penguji I

  
Catur Nugrahaeni Puspita Dewi,  
S.Kom., M.Kom.  
Penguji II

  
Iin Ernawati, S.Kom., M.Si.  
Pembimbing I

  
Muhammad Adrezo, S.Kom.,  
M.Sc.  
Pembimbing II

  
  
Prof. Dr. Ir. Suprivanto, ST.,  
M.Sc., IPM.  
Dekan

  
Dr. Widya Cholik M.I.T.,  
Kepala Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 12 Januari 2024

# OPTIMASI RANDOM FOREST TERHADAP DATA *TELCO CUSTOMER CHURN* MENGGUNAKAN *FIREFLY ALGORITHM*

Abril Muhammad Fikar Wijaya

## ABSTRAK

*Customer Churn* adalah persentase pelanggan yang telah berhenti atau beralih menggunakan produk/layanan secara berkala. Telekomunikasi industri mengalami tingkat *churn* tahunan rata-rata sebesar 30-35 %, dan perolehan pelanggan baru 5-10 kali lebih banyak mahal dibandingkan mempertahankan yang sudah ada. Prediksi *churn* dapat digunakan untuk membantu perusahaan mengidentifikasi *churners* lebih awal sebelum terjadi perpindahan pelanggan. Tujuan dari penelitian ini adalah untuk melakukan klasifikasi dengan metode *Random Forest* serta dikombinasikan dengan *Firefly Algorithm* untuk meningkatkan akurasi. Hasil evaluasi klasifikasi menggunakan *confusion matrix* didapatkan nilai akurasi sebesar 80,48%, nilai *precision* sebesar 77%, nilai *recall* sebesar 82,7%, dan nilai *f1-score* sebesar 79,71% sebelum dilakukan optimasi. Sedangkan setelah dilakukan optimasi menggunakan *Firefly Algorithm* didapatkan nilai akurasi sebesar sebesar 82,08%, *precision* sebesar 77,7%, *recall* sebesar 86,2% dan *F1 Score* sebesar 81,6%.

**Kata kunci :** *Customer Churn*, Klasifikasi, *Random Forest*, *Firefly Algorithm*.

**OPTIMIZING RANDOM FOREST FOR TELCO CUSTOMER CHURN  
DATA USING THE FIREFLY ALGORITHM**

**Abril Muhammad Fikar Wijaya**

**ABSTRACT**

*Customer churn is the percentage of customers who have stopped or switched using a product/service periodically. The telecommunications industry experiences an average annual churn rate of 30-35%, and acquiring new customers is 5-10 times more expensive than retaining existing ones. Predicting churn can be used to help companies identify churners earlier before customer defection occurs. The objective of this research is to perform classification using the Random Forest method combined with the Firefly Algorithm to enhance accuracy. The classification evaluation results using the confusion matrix show an accuracy of 80,48%, precision of 77%, recall of 82,7%, and an F1-score of 79,71% before optimization. After optimization using the Firefly Algorithm, the accuracy increased to 82,08%, precision to 77,7%, recall to 86,2%, and F1-score to 81,6%.*

**Keyword :** *Customer Churn, Classification, Random Forest, Firefly Algorithm.*

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Yang Maha Esa atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan baik dan dapat menyelesaikan jenjang pendidikan sarjana S1. Penulisan skripsi ini bertujuan untuk memenuhi salah satu persyaratan untuk memperoleh gelar Sarjana Komputer, Jurusan Informatika. Penulisan skripsi dengan judul “OPTIMASI *RANDOM FOREST* TERHADAP DATA *TELCO CUSTOMER CHURN* MENGGUNAKAN *FIREFLY ALGORITHM*” yang dilakukan sejak April 2023 ini tidak terlepas dari bantuan beberapa pihak-pihak lain. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya kepada:

1. Allah Yang Maha Esa, yang telah memberikan kesehatan serta ilmu kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik.
2. Orang tua dan keluarga, yang selalu mendoakan serta mendukung peneliti sehingga dapat menyelesaikan skripsi dengan baik.
3. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM selaku Dekan Fakultas Ilmu Komputer UPN Veteran Jakarta beserta jajarannya.
4. Ibu Dr. Widya Cholil, M.I.T. selaku Kepala Program Studi Informatika UPN Veteran Jakarta.
5. Ibu Iin Ernawati, S.Kom., M.Si dan Bapak Muhammad Adrezo, S.Kom., M.Sc. selaku Dosen Pembimbing Skripsi yang membantu penulis dalam penyusunan skripsi dan memberikan saran dengan baik.
6. Bapak Hamonangan Kinantan P., S.T, MT selaku Dosen Pembimbing Akademik yang membantu penulis selama duduk di bangku perkuliahan.
7. Sahabat saya, Rana Zahra Mahfuzhah, Muhammad Dwirangga Widodo dan Zahra Shabrina Humaira yang telah menemani saya saat suka dan duka dalam menyusun tugas akhir.
8. Sahabat saya, Muhammad Taufiqurahman, Hafidz Ashabi, Farhan Akmal Muhammad, Mutiara Savina, dan Nisyanti Yahya yang telah menemani saya saat suka dan duka dalam menyusun tugas akhir.
9. Teman saya, Rizqy Nabila yang telah menemani saya saat suka dan duka dalam menyusun tugas akhir.



10. Teman-teman saya, Tania Larasati Priadiputri, Muhammad Fauzan Alhafizh, Nio Dimas Radityo, dan Muhammad Fadly Alghifary yang menghibur penulis disaat sedih maupun senang.
11. Seluruh teman dekat penulis yang telah memberikan dukungan, doa, dan semangat untuk menyelesaikan skripsi ini, yang namanya tidak dapat penulis sebutkan satu per satu.
12. Bang Windah Basudara sebagai *streamer youtube* yang sudah menghibur penulis disaat saya membutuhkan *refreshing*.
13. Seluruh pihak yang terlibat dalam kelancaran pembuatan skripsi ini dan yang belum disebutkan di atas, penulis ucapkan terimakasih.

Penulis menyadari bahwa penyusunan skripsi ini jauh dari sempurna. Oleh karena itu, penulis harap kepada para peneliti selanjutnya untuk memberikan saran dan kritik yang bersifat membangun untuk kesempurnaan skripsi ini. Akhir kata penulis berharap semoga skripsi ini mampu bermanfaat untuk semua pihak.

Jakarta, 21 Januari 2024

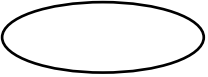
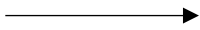


(Abril Muhammad Fikar Wijaya)

## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK .....	vi
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	x
DAFTAR SIMBOL .....	xii
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
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 Luaran yang Diharapkan .....	3
BAB 2 TINJAUAN PUSTAKA .....	4
2.1 <i>Data Mining</i> .....	4
2.2 <i>Random Forest</i> .....	5
2.3 <i>Firefly Algorithm</i> .....	8
2.4 <i>Mean Imputation</i> .....	9
2.5 <i>Imbalance Class</i> .....	10
2.6 <i>Undersampling</i> .....	10
2.7 Evaluasi .....	10
2.8 Riset yang Relevan .....	12
BAB 3 METODOLOGI PENELITIAN .....	15
3.1 Alur Penelitian .....	15
3.2 Identifikasi Masalah .....	15
3.3 Kajian Pustaka .....	16
3.4 Pengumpulan Data .....	16

3.5	Pra-Proses Data .....	16
3.6	Data Mining .....	16
3.7	Evaluasi .....	17
3.8	Waktu dan Tempat Penelitian .....	17
3.9	Alat Pendukung .....	17
3.9.1	Perangkat Keras .....	18
3.9.2	Perangkat Lunak.....	18
BAB 4 HASIL DAN PEMBAHASAN.....		20
4.1	Akuisisi Data .....	20
4.2	<i>Preprocessing Data</i> .....	25
4.2.1	<i>Data Cleaning</i> .....	26
4.2.2	Data Transformation .....	28
4.2.3	<i>Data Balancing</i> .....	34
4.2.4	<i>Data Selection</i> .....	34
4.3	<i>Data Mining</i> .....	37
4.3.1.	Klasifikasi <i>Random Forest</i> .....	38
4.3.2.	Implementasi <i>Firefly Algorithm</i> .....	46
4.4	Evaluasi .....	48
4.4.1	<i>Confusion Matrix Random Forest</i> pada Rasio 70:30.....	48
4.4.2	<i>Confusion Matrix Random Forest</i> pada Rasio 80:20.....	50
4.4.3	<i>Confusion Matrix Random Forest</i> pada Rasio 90:10.....	52
4.4.4	<i>Confusion Matrix Random Forest</i> setelah optimasi.....	53
4.5	Analisis Klasifikasi Model <i>Machine Learning</i> .....	55
BAB 5 PENUTUP .....		59
5.1	Kesimpulan.....	59
5.2	Saran .....	59
DAFTAR PUSTAKA .....		60
LAMPIRAN.....		63

## DAFTAR SIMBOL

No	Gambar	Nama	Keterangan
1		Terminal (start, end)	Menampilkan sebuah kegiatan dimulai atau kegiatan berakhir
2		Flow direction	Menampilkan alur dan arah sebuah proses
3		Process	Menampilkan sebuah proses yang dilakukan
4		Decision	Menampilkan sebuah proses yang memiliki kondisi yang harus memilih lebih dari satu proses

## DAFTAR TABEL

<b>Tabel 2.1</b> <i>Confusion Matrix</i> .....	11
<b>Tabel 2.2</b> Riset yang Relevan.....	12
<b>Tabel 3.1</b> <i>Library Package</i> .....	18
<b>Tabel 3.2</b> Jadwal Penelitian.....	19
<b>Tabel 4.1</b> Nilai Fitur dan Target.....	20
<b>Tabel 4.2</b> Sampel data variabel <i>CustomerID - Dependents</i> .....	24
<b>Tabel 4.3</b> Sampel data variabel <i>tenure – OnlineSecurity</i> .....	24
<b>Tabel 4.4</b> Sampel data variabel <i>OnlineBackup - StreamingMovies</i> .....	25
<b>Tabel 4.5</b> Sampel data variabel <i>Contract - MonthlyCharges</i> .....	25
<b>Tabel 4.6</b> Sampel data variabel <i>TotalCharges dan Churn</i> .....	25
<b>Tabel 4.7</b> Mengecek <i>Missing Value</i> .....	26
<b>Tabel 4.8</b> Mengganti nilai null dengan <i>mean</i> .....	27
<b>Tabel 4.9</b> Nilai variabel sebelum diubah menjadi <i>"No"</i> .....	29
<b>Tabel 4.10</b> Nilai variabel setelah diubah menjadi <i>"No"</i> .....	29
<b>Tabel 4.11</b> Atribut kategorik setelah dilakukan <i>Label Encoding</i> .....	30
<b>Tabel 4.12</b> Perubahan Tipe Data .....	31
<b>Tabel 4.13</b> Sampel data sebelum normalisasi.....	32
<b>Tabel 4.14</b> Sampel data setelah normalisasi.....	33
<b>Tabel 4.15</b> <i>Data Balancing</i> .....	34
<b>Tabel 4.16</b> <i>Feature Importance</i> .....	34
<b>Tabel 4.17</b> <i>Data Selection</i> .....	35
<b>Tabel 4.18</b> Pembagian data latih dan data uji.....	37
<b>Tabel 4.19</b> Data Sampel <i>Churn</i> .....	38
<b>Tabel 4.20</b> Sampel Data Fitur <i>PaperlessBilling</i> Kondisi <i>Yes</i> .....	41
<b>Tabel 4.21</b> Sampel Data Fitur <i>OnlineBackup</i> Kondisi <i>Yes</i> .....	43
<b>Tabel 4.22</b> Contoh Data Uji .....	45
<b>Tabel 4.23</b> Parameter <i>Firefly Algorithm</i> .....	47
<b>Tabel 4.24</b> Daftar <i>Hyperparameter</i> dan Rentang Nilainya .....	47
<b>Tabel 4.25</b> Perbandingan kinerja algoritma.....	57

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Tahapan Data Mining (Kamil & Cholil, 2020). .....	4
<b>Gambar 2.2</b> Struktur <i>Random Forest</i> (Hidayat et al., 2023) .....	7
<b>Gambar 3.1</b> Alur Penelitian .....	15
<b>Gambar 4.1</b> Root Node <i>PaperlessBilling</i> .....	41
<b>Gambar 4.2</b> <i>Child Node OnlineBackup</i> .....	43
<b>Gambar 4.3</b> Pohon Pertama .....	44
<b>Gambar 4.4</b> Ilustrasi Tiga Pohon .....	44
<b>Gambar 4.5</b> Hasil Klasifikasi Tiga Pohon .....	45
<b>Gambar 4.6</b> <i>Tree[0] Random Forest</i> .....	46
<b>Gambar 4.7</b> <i>Confusion Matrix Random Forest</i> pada Rasio 70:30 .....	48
<b>Gambar 4.8</b> <i>Confusion Matrix Random Forest</i> pada rasio 80:20.....	50
<b>Gambar 4.9</b> <i>Confusion Matrix Random Forest</i> pada rasio 80:20.....	52
<b>Gambar 4.10</b> <i>Confusion Matrix Random Forest</i> setelah optimasi .....	54
<b>Gambar 4.11</b> Chart perbandingan akurasi <i>Random Forest</i> .....	58