

ANALISIS PREDIKSI *CUSTOMER CHURN* NASABAH BANK DENGAN ALGORITMA *RANDOM FOREST*

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

ANINDITA PUSPA AYU PRAYOGI
NPM: 2113030045

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI
2025

Skripsi oleh:

ANINDITA PUSPA AYU PRAYOGI
NPM: 2113030045

Judul:

**ANALISIS PREDIKSI *CUSTOMER CHURN* NASABAH BANK DENGAN
ALGORITMA RANDOM FOREST**

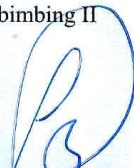
Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Prodi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal: 1 Juli 2025

Pembimbing I


Sucipto, M.Kom
NIDN.0721029101

Pembimbing II


Arie Nugroho, S.Kom., MM., M.Kom
NIDN. 0712108103

Skripsi oleh:

ANINDITA PUSPA AYU PRAYOGI

NPM: 2113030045

Judul:

**ANALISIS PREDIKSI *CUSTOMER CHURN* NASABAH BANK DENGAN
ALGORITMA *RANDOM FOREST***

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Prodi Sistem Informasi FTIK UN PGRI Kediri

Tanggal : 11 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : Sucipto, M.Kom

[.....]

2. Penguji 1 : Aidina Ristyawan, M.Kom

[.....]

3. Penguji 2 : Arie Nugroho, S.Kom., M.M., M.Kom [.....]



Mengetahui,
Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Sulistiono, M.Si.
NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama	: Anindita Puspa Ayu Prayogi
Jenis Kelamin	: Perempuan
Tempat/tgl. lahir	: Kediri/ 17 Maret 2003
NPM	: 2113030045
Fak/Jur./Prodi.	: FTIK/ S1 Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 1 Juli 2025

Yang Menyatakan



ANINDITA PUSPA AYU P
NPM: 2113030045

MOTTO

“Maka sesungguhnya bersama kesulitan ada kemudahan, sesungguhnya bersama kesulitan ada kemudahan”

(QS. Al – Insyirah: 5-6)

Abstrak

Anindita Puspa Ayu Prayogi: Analisis Prediksi *Customer Churn* Nasabah Bank dengan Algoritma *Random Forest*, Skripsi, Sistem Informasi, FTIK UN PGRI Kediri, 2025

Kata Kunci: *customer churn*, *Random Forest*, klasifikasi, *feature importance*, prediksi nasabah, *data mining*

Churn atau berpindahnya nasabah ke bank lain merupakan permasalahan serius dalam dunia perbankan yang kompetitif, karena berdampak langsung terhadap pendapatan dan loyalitas jangka panjang nasabah. Oleh karena itu, penting bagi perusahaan untuk memiliki strategi prediktif yang mampu mengidentifikasi potensi churn sejak dini. Berbagai penelitian sebelumnya telah menerapkan algoritma klasifikasi seperti Naive Bayes dan ID3, namun masih terdapat ruang untuk meningkatkan akurasi prediksi dalam konteks data yang lebih kompleks.

Penelitian ini bertujuan untuk meningkatkan akurasi model prediksi *churn* nasabah bank dengan menerapkan algoritma *Random Forest* dalam tiga konfigurasi berbeda: model *default*, model dengan *tuning parameter*, dan model dengan *tuning* serta *feature selection* berdasarkan nilai *feature importance*. Dataset yang digunakan berasal dari sumber publik di Kaggle yang memuat 10.000 data nasabah dengan 13 fitur. Penelitian ini menggunakan pendekatan CRISP-DM yang mencakup eksplorasi data, preprocessing, pembangunan model, hingga evaluasi performa menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

Hasil penelitian menunjukkan bahwa model *Random Forest* dengan *tuning parameter* menghasilkan performa terbaik dengan akurasi sebesar 86,7% dan nilai *F1-score* yang lebih seimbang antara kelas *churn* dan *non-churn* dibandingkan dua konfigurasi lainnya. Selain itu, hasil analisis *feature importance* mengidentifikasi beberapa fitur utama yang paling berpengaruh terhadap *churn*, yaitu usia (*Age*), jumlah produk (*NumOfProducts*), saldo (*Balance*), estimasi gaji (*EstimatedSalary*), dan skor kredit (*CreditScore*).

Kesimpulan dari penelitian ini menunjukkan bahwa optimasi parameter dapat meningkatkan performa model klasifikasi *churn* menggunakan *Random Forest*. Meskipun demikian, penelitian ini memiliki keterbatasan karena hanya menggunakan satu jenis algoritma, sehingga belum dapat dibandingkan dengan metode lain secara langsung. Selain itu, pendekatan penanganan data tidak seimbang yang digunakan masih terbatas dan belum mengeksplorasi metode *balancing* yang lebih optimal tanpa mengorbankan akurasi. Penelitian selanjutnya disarankan untuk membandingkan beberapa algoritma klasifikasi serta menerapkan teknik penyeimbangan data yang lebih canggih agar model dapat lebih andal dan generalisasi pada data dari berbagai sumber.

KATA PENGANTAR

Puji Syukur Kami panjatkan kehadiran Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul “Analisis Prediksi *Customer Churn* Nasabah Bank dengan Algoritma *Random Forest*” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Jurusan Sistem Informasi FTIK UN PGRI Kediri. Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus- tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
3. Sucipto, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
4. Sucipto, M.Kom. dan Arie Nugroho, S.Kom, MM., M.Kom. selaku Pembimbing skripsi yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
5. Seluruh dosen prodi Sistem Informasi Universitas Nusantara PGRI Kediri yang telah menyampaikan ilmu dan memotivasi mahasiswa.
6. Kedua orang tua dan keluarga saya yang telah mendukung dalam do’a, dana, dan motivasi dalam penyelesaian skripsi ini.
7. Teman-teman yang selalu membantu memberi semangat dalam penyelesaian skripsi ini.
8. Ucapan terima kasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu dalam penyelesaian skripsi ini.

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan tegur sapa, kritik, dan saran-saran, dari berbagai pihak sangat diharapkan. Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan, meskipun hanya ibarat setitik air bagi samudra yang luas.

Kediri, 1 Juli 2025



ANINDITA PUSPA AYU P

NPM: 2113030045

DAFTAR ISI

Halaman Sampul	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan.....	iv
Halaman Motto.....	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah.....	3
C. Rumusan Masalah	4
D. Tujuan Penelitian	4
E. Manfaat Penelitian	5
BAB II KAJIAN PUSTAKA.....	6
A. Kajian Teori	6
1. <i>Data Mining</i>	6
2. CRISP DM	7
3. <i>Machine Learning</i>	8
4. Klasifikasi	9
5. Prediksi <i>Churn</i> Nasabah.....	10
6. Algoritma <i>Random Forest</i>	11
7. <i>Feature Importance</i> dalam <i>Random Forest</i>	15
8. Evaluasi Kinerja Model.....	17
9. <i>Feature Selection</i> Menggunakan <i>Threshold</i>	19
B. Kajian Penelitian Terdahulu.....	20
C. Kerangka Berpikir.....	21
BAB III METODE PENELITIAN	23
A. <i>Business Understanding</i>	24
1. <i>Determine Business Objectives</i>	24

2.	<i>Asses Situation</i>	24
3.	<i>Determine Data mining Goals</i>	24
4.	<i>Produce Project Plan</i>	25
B.	<i>Data Understanding</i>	25
1.	<i>Collect Initial Data</i>	26
2.	<i>Describe Data</i>	26
3.	<i>Explore Data</i>	26
C.	<i>Data Preparation</i>	27
1.	<i>Data Selection</i>	27
2.	<i>Data Preprocessing</i>	27
3.	<i>Data Transformation</i>	27
D.	<i>Modeling</i>	28
1.	<i>Select Modeling Technique</i>	28
2.	<i>Generate Test Design</i>	28
3.	<i>Build Model</i>	29
E.	<i>Evaluation</i>	31
1.	<i>Evaluate Result</i>	31
2.	<i>Determine Next Steps</i>	32
F.	<i>Deployment</i>	32
BAB IV HASIL DAN PEMBAHASAN		34
A.	<i>Data Understanding</i>	34
1.	<i>Collect Initial Data</i>	34
2.	<i>Describe Data</i>	34
3.	<i>Explore Data</i>	37
B.	<i>Data preparation</i>	48
1.	<i>Penghapusan Fitur</i>	49
2.	<i>Encoding Variabel Kategorikal</i>	50
3.	<i>Pemisahan Fitur dan Label</i>	51
4.	<i>Pembagian Data Training dan Testing</i>	51
5.	<i>Feature Selection</i>	52
C.	<i>Modeling</i>	53
1.	<i>Model Default Random Forest</i>	53
2.	<i>Model Random Forest dengan Tuning Parameter</i>	54
3.	<i>Model Random Forest dengan Tuning dan Feature Selection</i>	55
4.	<i>Ilustrasi Perhitungan Manual Random Forest</i>	56
D.	<i>Evaluation</i>	62

1.	Evaluasi Performa Model <i>Default Random Forest</i>	62
2.	Evaluasi Model <i>Random Forest</i> dengan <i>Parameter tuning</i>	64
3.	Evaluasi Model <i>Random Forest</i> dengan <i>Parameter Tuning</i> dan <i>Feature Selection</i>	66
4.	Perbandingan Model <i>Random Forest</i>	67
5.	<i>Feature Importance</i>	69
E.	Analisis Faktor-Faktor yang Memengaruhi <i>Churn</i>	73
1.	<i>Age</i> (26.69%).....	73
2.	<i>NumOfProducts</i> (14.45%).....	74
3.	<i>Balance</i> (13.99%).....	75
4.	<i>EstimatedSalary</i> (12.64%)	76
5.	<i>CreditScore</i> (12.31%)	77
F.	Pembahasan.....	79
BAB V PENUTUP		82
A.	Kesimpulan	82
B.	Saran.....	83
DAFTAR PUSTAKA		85
LAMPIRAN		89

DAFTAR GAMBAR

Gambar 2.1 Kerangka Berpikir.....	22
Gambar 3.1 Alur CRISP-DM	23
Gambar 4.1 Distribusi <i>Churn</i>	37
Gambar 4.2 Boxplot Identifikasi <i>Outlier</i>	39
Gambar 4.3 <i>Heatmap</i> Korelasi.....	40
Gambar 4.4 Distribusi Fitur Age terhadap Target	41
Gambar 4.5 Distribusi Fitur <i>IsActiveMember</i> terhadap Target.....	42
Gambar 4.6 Distribusi Fitur <i>Balance</i> terhadap Target.....	43
Gambar 4.7 Distribusi Fitur <i>NumOfProducts</i> terhadap Target	43
Gambar 4.8 Distribusi Fitur <i>Gender</i> terhadap Target	44
Gambar 4.9 Distribusi Fitur <i>Geography</i> terhadap Target	45
Gambar 4.10 Distribusi Fitur <i>HasCreditCard</i> terhadap Target	46
Gambar 4.11 Distribusi Fitur <i>EstimatedSalary</i> terhadap Target.....	47
Gambar 4.12 Distribusi Fitur <i>CreditScore</i> terhadap Target.....	47
Gambar 4.13 Distribusi Fitur <i>Tenure</i> terhadap Target.....	48
Gambar 4.14 Pohon Keputusan 1	60
Gambar 4.15 Pohon Keputusan 2	60
Gambar 4.16 Pohon Keputusan 3	60
Gambar 4.17 <i>Feature Importance</i>	70

DAFTAR TABEL

Tabel 4.1 Deskripsi Atribut.....	34
Tabel 4.2 Lanjutan Deskripsi Atribut	35
Tabel 4.3 Sampel Data	35
Tabel 4.4 Informasi Data	36
Tabel 4.5 Statistik Deskriptif Fitur	38
Tabel 4.6 Perbandingan Fitur Sebelum dan Sesudah Penghapusan.....	49
Tabel 4.7 Transformasi Fitur <i>Gender</i> Sebelum dan Sesudah <i>Encoding</i>	50
Tabel 4.8 Transformasi Fitur <i>Geography</i> Sebelum dan Sesudah <i>Encoding</i>	50
Tabel 4.9 Fitur Terpilih.....	52
Tabel 4.10 Sampel Data <i>Bootstrap</i>	56
Tabel 4.11 Prediksi Pohon	61
Tabel 4.12 <i>Classification Report</i> Model 1	63
Tabel 4.13 <i>Confusion Matrix</i> Model 1	63
Tabel 4.14 <i>Classification Report</i> Model 2	64
Tabel 4.15 <i>Confusion Matrix</i> Model 2.....	64
Tabel 4.16 <i>Classification Report</i> Model 3	66
Tabel 4.17 <i>Confusion Matrix</i> Model 3	66
Tabel 4.18 Perbandingan Performa.....	67
Tabel 4.19 Perbandingan Performa Penelitian Terdahulu	79

DAFTAR LAMPIRAN

Lampiran 1 <i>Script</i> Python Prediksi <i>Customer Churn</i> Menggunakan Algoritma <i>Random Forest</i>	89
Lampiran 2 Kartu Bimbingan Skripsi.....	91
Lampiran 3 Lembar Revisi	92
Lampiran 4 Surat Keterangan Bebas Similarity.....	93
Lampiran 5 Bukti Halaman Awal Cek Similarity.....	94
Lampiran 6 Bukti Submit Artikel/LOA/Terbit	95

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Dalam dunia bisnis, mempertahankan pelanggan adalah aspek kunci bagi keberlanjutan dan pertumbuhan perusahaan. Perusahaan harus memahami kebutuhan dan keinginan pelanggan untuk menjalin hubungan yang baik (Yudiana et al., 2023). Ketika kebutuhan pelanggan terpenuhi, mereka akan merasa puas dan cenderung loyal pada perusahaan (Yudiana et al., 2023). Loyalitas pelanggan memberikan keuntungan bukan hanya lewat transaksi berulang, tetapi juga melalui rekomendasi positif yang memperkuat reputasi perusahaan. Oleh karena itu, perusahaan di berbagai sektor berupaya menciptakan pengalaman memuaskan bagi pelanggan mereka. Untuk bersaing dalam persaingan bisnis yang semakin kompetitif, strategi kepuasan pelanggan harus efektif (Rohana, 2020). Perusahaan perlu mengembangkan strategi baru seiring meningkatnya ekspektasi pelanggan, serta menganalisis kebiasaan mereka guna memahami potensi loyalitas. Karena, upaya mempertahankan pelanggan dianggap jauh lebih baik dibandingkan mencari pelanggan baru (Rohana, 2020).

Meskipun berbagai upaya dilakukan, banyak perusahaan masih menghadapi tantangan dalam mempertahankan pelanggan. Di tengah persaingan pasar yang ketat, perubahan preferensi konsumen, dan kemajuan teknologi, pelanggan memiliki banyak opsi untuk beralih ke produk atau layanan lain. Oleh karena itu, perusahaan perlu menerapkan strategi adaptasi, distribusi, promosi, dan riset menyeluruh untuk memahami perilaku dan kebutuhan pelanggan (D. K. Putri et al., 2024). Ini menuntut perusahaan agar lebih terbuka dalam memahami dan mengantisipasi kebutuhan pelanggan guna menjaga hubungan jangka panjang.

Di sektor perbankan, tantangan mempertahankan nasabah menjadi lebih signifikan. Dalam lingkungan yang penuh persaingan, bank tidak hanya harus menawarkan produk yang menarik, tetapi juga harus memastikan bahwa

nasabah merasa dihargai dan terlayani dengan baik. Karena persaingan tersebut, pelanggan memiliki kemungkinan untuk berpindah ke bank lain. Hal ini disebut dengan *churn*, dimana nasabah memilih untuk meninggalkan bank untuk ke bank lain karena nasabah merasa bank lain dapat memberikan layanan atau produk yang lebih baik (Clementine, 2022). Oleh sebab itu, bank harus mampu mengidentifikasi dan memprediksi faktor yang mendorong perpindahan nasabah, karena *churn* dapat menyebabkan kerugian besar. Ini mendorong bank untuk menggunakan analisis data guna memprediksi perilaku nasabah dan mengembangkan strategi lebih efektif dalam mempertahankan nasabah (Irmanda et al., 2019).

Dalam era kemajuan teknologi dan peningkatan *volume data* yang tersedia, *Artificial Intelligence* (AI) telah memainkan peran penting dalam analisis data, di mana teknik *data mining* dan *machine learning* menjadi pilihan populer untuk prediksi dalam berbagai bidang (Hanifa et al., 2017). Aktivitas manusia menghasilkan berbagai macam data dengan berbagai ukuran, yang tidak akan berguna jika tidak diolah dengan tepat (Nugroho et al., 2021). Dengan memanfaatkan teknik *data mining* dan *machine learning*, *dataset* yang dihasilkan di masa lalu dapat diolah menjadi informasi dan pengetahuan yang bermanfaat bagi manusia untuk mendukung aktivitas mereka di masa depan (Nugroho et al., 2021). *Data mining* adalah salah satu bidang dalam *Artificial Intelligence*, yang memungkinkan perusahaan untuk menganalisis pola dan tren yang kompleks dalam perilaku nasabah (Nugroho, 2022). Dengan menerapkan *data mining*, perusahaan dapat memahami faktor-faktor yang memengaruhi keputusan pelanggan untuk berpindah ke bank lain.

Berbagai algoritma, seperti Naive Bayes dan ID3, telah digunakan dalam penelitian sebelumnya untuk mengidentifikasi pola perilaku nasabah (Clementine, 2022). Meskipun hasil dari algoritma-algoritma ini menunjukkan akurasi yang cukup baik, masih ada potensi untuk meningkatkan akurasi model prediksi *churn* dengan pendekatan yang lebih optimal. Dalam penelitian ini, algoritma *Random Forest* dipilih karena memiliki keunggulan dalam menangani data yang kompleks, mampu mengurangi *overfitting*, serta menyediakan interpretasi fitur melalui nilai *feature importance*.

Random Forest merupakan salah satu metode pembelajaran mesin yang paling banyak digunakan untuk masalah prediksi (Hui et al., 2023). Algoritma ini, yang termasuk dalam kategori *ensemble learning*, menggabungkan hasil dari banyak *decision tree* untuk memberikan prediksi yang lebih stabil dan akurat. *Random Forest* juga tidak terlalu rentan terhadap masalah *overfitting* karena model ini menciptakan beberapa pohon klasifikasi menggunakan *subsampel* dan prediktor yang dipilih secara acak (Hui et al., 2023). Klasifikasi sendiri merupakan proses identifikasi pola atau fungsi tertentu yang digunakan untuk membedakan satu kelompok data dari kelompok lainnya, sehingga setiap objek dapat diidentifikasi dan dimasukkan ke dalam kategori yang sesuai berdasarkan kriteria yang telah ditentukan (R. W. Putri et al., 2022).

Fokus utama dari penelitian ini adalah meningkatkan akurasi model klasifikasi *customer churn* dibandingkan penelitian sebelumnya dengan menerapkan konfigurasi yang lebih optimal pada algoritma *Random Forest*. Untuk mendukung hal tersebut, penelitian ini membangun dan membandingkan tiga versi model *Random Forest*, yaitu model *default*, model dengan *tuning parameter*, dan model dengan *tuning* serta *feature selection* berdasarkan nilai *feature importance*. Selain itu, dilakukan analisis eksplorasi data (EDA) secara mendalam guna memahami pola perilaku nasabah, serta mengidentifikasi fitur-fitur paling berpengaruh yang berpotensi dijadikan dasar dalam strategi retensi nasabah oleh pihak bank.

B. Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. *Dataset* yang digunakan adalah *dataset* publik dari *Kaggle* yang berisi data nasabah bank dengan total 10000 *record*, terdiri dari 13 fitur dan 1 target yaitu *churn* dan *non-churn*.
2. Penelitian difokuskan pada penggunaan algoritma *Random Forest* sebagai metode utama untuk melakukan klasifikasi *customer churn*.
3. Model *Random Forest* yang dibangun terdiri dari tiga konfigurasi,

yaitu model *default*, model dengan *tuning parameter*, dan model dengan *tuning* serta *feature selection* berdasarkan *threshold* nilai *feature importance*.

4. Proses implementasi dan analisis dilakukan menggunakan bahasa pemrograman Python.
5. Evaluasi performa model menggunakan metrik klasifikasi, yaitu *accuracy*, *precision*, *recall*, *f1 score*, dan *confusion matrix*.
6. Analisis data dilakukan melalui eksplorasi data awal (EDA) dan analisis *feature importance* untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap *churn*.
7. Tahap deployment dalam penelitian ini dibatasi pada pembuatan dan eksekusi model klasifikasi menggunakan Python serta penyusunan laporan hasil penelitian dalam bentuk skripsi.

C. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana performa algoritma *Random Forest* dalam memprediksi *customer churn* nasabah bank menggunakan berbagai konfigurasi model (*default*, *tuning*, dan *tuning* dengan *feature selection*)?
2. Apakah model *Random Forest* yang dibangun mampu mencapai akurasi lebih baik dari hasil penelitian terdahulu?
3. Apa saja fitur yang paling berpengaruh terhadap perilaku *churn* nasabah berdasarkan analisis *feature importance* dan eksplorasi data?

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Meningkatkan akurasi model klasifikasi *customer churn* nasabah bank dibandingkan penelitian sebelumnya dengan menerapkan konfigurasi model *Random Forest* yang optimal.

2. Membangun dan membandingkan tiga konfigurasi model *Random Forest*, yaitu model *default*, model dengan *tuning parameter*, dan model dengan *tuning* serta *feature selection*.
3. Mengidentifikasi dan menganalisis fitur-fitur paling berpengaruh terhadap *customer churn* menggunakan *feature importance* dari algoritma *Random Forest*.

E. Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memberikan model prediksi *churn* yang lebih akurat dan stabil untuk membantu bank mengidentifikasi nasabah yang berisiko *churn*.
2. Mengidentifikasi fitur-fitur utama yang memengaruhi keputusan nasabah untuk berhenti menggunakan layanan bank, sehingga bank dapat menyusun strategi retensi berbasis data.
3. Memberikan gambaran perbandingan performa dari tiga pendekatan konfigurasi model *Random Forest*, sehingga dapat digunakan sebagai acuan bagi pengembangan model serupa di masa depan.
4. Memperkaya literatur dan penerapan *data mining* di bidang perbankan, khususnya dalam prediksi *churn* dengan pendekatan *Random Forest*.
5. Menjadi dasar bagi pengembang sistem dalam merancang sistem prediktif *churn* yang aplikatif dan berbasis data nyata.

DAFTAR PUSTAKA

- Abdellatif, A., Abdellatef, H., Kanesan, J., Chow, C.-O., Chuah, J. H., & Gheni, H. M. (2022). Improving the heart disease detection and patients' survival using supervised infinite feature selection and improved weighted random forest. *IEEE Access*, 10, 67363–67372.
- AlSagri, H., & Ykhlef, M. (2020). Quantifying feature importance for detecting depression using random forest. *International Journal of Advanced Computer Science and Applications*, 11(5).
- Baita, A., Prasetyo, I. A., & Cahyono, N. (2023). Hyperparameter tuning on random forest for diagnose Covid-19. *JIKO (Jurnal Informatika Dan Komputer)*, 6(2).
- Clementine, M. (2022). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan ID3. *Jurnal Processor*, 17(1), 9–18.
- Dhanya, R., Paul, I. R., Akula, S. S., Sivakumar, M., & Nair, J. J. (2020). F-test feature selection in Stacking ensemble model for breast cancer prediction. *Procedia Computer Science*, 171, 1561–1570.
- Feblian, D., & Daihani, D. U. (2017). IMPLEMENTASI MODEL CRISP-DM UNTUK MENENTUKAN SALES PIPELINE PADA PT X. *JURNAL TEKNIK INDUSTRI*, 6(1). <https://doi.org/10.25105/jti.v6i1.1526>
- Gede Iwan Sudipa, I., Andi Putra, T. E., Jurnaidi Wahidin, A., Alfa Syukrilla, W., Khrisna Wardhani, A., Heryana, N., Indriyani, T., Willyanto Santoso Tutuk Indriyani, L., & Willyanto Santoso, L. (2023). *DATA MINING* (1st ed.). www.globaleksekutifteknologi.co.id
- Hanifa, T. T., Al-Faraby, S., & Informatika, F. (2017). Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging. *Vol*, 4, 3210–3225.
- Hui, S. H., Khai, W. K., XinYing, C., & Wai, P. W. (2023). Prediction of customer churn for ABC Multistate Bank using machine learning algorithms/Hui Shan Hon...[et al.]. *Malaysian Journal of Computing (MJoC)*, 8(2), 1602–1619.
- Irmanda, H. N., Astriratma, R., & Afrizal, S. (2019). Perbandingan metode

- jaringan syaraf tiruan dan pohon keputusan untuk prediksi churn. *JSI: Jurnal Sistem Informasi (E-Journal)*, 11(2).
- Karsito, K., & Susanti, S. (2019). Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia. *Jurnal SIGMA*, 9(3), 43–48.
- Lu, M., & Li, F. (2020). Survey on lie group machine learning. *Big Data Mining and Analytics*, 3(4), 235–258.
- Makariou, D., Barrieu, P., & Chen, Y. (2021). A random forest based approach for predicting spreads in the primary catastrophe bond market. *Insurance: Mathematics and Economics*, 101, 140–162.
- Mehmet Akturk. (2020). *Churn for Bank Customers*. Kaggle. <https://www.kaggle.com/datasets/mathchi/churn-for-bank-customers>
- Muhamad Malik Matin, I. (2023). *Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware*. *MULTINETICS*, 9 (1), 43–50.
- Mulia, C., & Kurniasih, A. (2023). Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve bayes dan Logistic Regression. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya*, 4(2), 552–559.
- Nila, U., Firliana, R., & **Sucipto**, S. (2023). Analisis Data Transaksi Penjualan Produk Pertanian Menggunakan Algoritma FP-Growth. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 7(1), 175–183.
- Nugroho**, A. (2022). Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan. *JSITIK: Jurnal Sistem Informasi Dan Teknologi Informasi Komputer*, 1(1), 41–49.
- Nugroho**, A., Fanani, A. Z., & Shidik, G. F. (2021). Evaluation of Feature Selection Using Wrapper For Numeric Dataset With Random Forest Algorithm. *2021 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 179–183. <https://doi.org/10.1109/iSemantic52711.2021.9573249>
- Nugroho**, A., & Husin, A. (2022). Analisis Performa Random Forest Menggunakan Normalisasi Atribut. *SISTEMASI: Jurnal Sistem Informasi*,

11(1), 186–196.

- Orlenko, A., & Moore, J. H. (2021). A comparison of methods for interpreting random forest models of genetic association in the presence of non-additive interactions. *BioData Mining*, 14, 1–17.
- Patel, S. (2023). A survey on Diabetes Prediction Models Using Data Mining Techniques: issues and challenges. *International Journal of Scientific Research in Science, Engineering and Technology*, 263–267. <https://doi.org/10.32628/IJSRSET23103208>
- Putri, D. K., Fachri, F. N., & Andini, F. T. (2024). Pemasaran dan Riset Pemasaran Global: Konsep, Manfaat, dan Tantangan. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah*, 9(1).
- Putri, R. W., Ristyawan, A., & Muzaki, M. N. (2022). Comparison Performance of K-NN and NBC Algorithm for Classification of Heart Disease. *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem Dan Komputer*, 2(2), 143. <https://doi.org/10.32503/jtecs.v2i2.2708>
- Rianti, A., Majid, N. W. A., & Fauzi, A. (2023). CRISP-DM: Metodologi Proyek Data Science. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 106–114.
- Rizky, M. H., Faisal, M. R., Budiman, I., Kartini, D., & Abadi, F. (2024). Effect of Hyperparameter Tuning Using Random Search on Tree-Based Classification Algorithm for Software Defect Prediction. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 18(1), 95–106.
- Rohana, T. (2020). Pengaruh kepuasan terhadap loyalitas pelanggan. *Jurnal Ilman: Jurnal Ilmu Manajemen*, 8(1), 28–32.
- Schröer, C., Kruse, F., & Marx Gómez, J. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, 14(1).
- Sucipto**, Dwi Prasetya, D., & Widiyaningtyas, T. (n.d.). *An Evaluation of the Impact of Dataset Size on Classification Performance in the Cognitive Bloom's Taxonomy*. <https://orcid.org/0000-0003-3412-002X>

- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 7, 60134–60149.
- Wiyono, S., Apriliani, D., Abidin, T., & Dairoh, D. (2019). Feature selection with the random forest packages to predict student performance. *Systemic: Information System and Informatics Journal*, 5(1), 1–5.
- Yuan, X., Liu, S., Feng, W., & Dauphin, G. (2023). Feature importance ranking of random forest-based end-to-end learning algorithm. *Remote Sensing*, 15(21), 5203.
- Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1–20.