



PREDIKSI *CHURN* NASABAH PERBANKAN MENGGUNAKAN METODE *RANDOM FOREST* DAN *LOGISTIC REGRESSION* BERBASIS OPTIMASI *HYPERPARAMETER*

MUTHI'A INSANI



**PROGRAM STUDI AKTUARIA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul "*Prediksi Churn Nasabah Perbankan Menggunakan Metode Random Forest dan Logistic Regression Berbasis Optimasi Hyperparameter*" adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juni 2026

Muthi'a Insani
G5402221033

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

MUTHI'A INSANI. Prediksi *Churn* Nasabah Perbankan Menggunakan Metode *Random Forest* dan *Logistic Regression* Berbasis Optimasi *Hyperparameter*. Dibimbing oleh BUDI SUHARJO dan NUR AGUSTIANI.

Churn nasabah perbankan merupakan kondisi ketika nasabah mengakhiri hubungan bisnis secara menyeluruh sehingga berpotensi mengancam stabilitas finansial perbankan. Oleh karena itu, diperlukan strategi prediksi *churn* nasabah yang akurat. Penelitian ini membandingkan kinerja dua metode analisis untuk memprediksi *churn* yaitu *random forest* dan *logistic regression* yang dioptimasi melalui *hyperparameter tuning*. Menggunakan 10,000 observasi nasabah dari platform Kaggle, strategi *random undersampling* dan *cluster undersampling* diterapkan sebagai *data balancing*. Evaluasi model menggunakan AUC, *recall*, *precision*, dan *confusion matrix*, sedangkan interpretasi fitur memanfaatkan *feature importance*, *odds ratio*, dan SHAP. Hasil menunjukkan *random forest* memiliki kinerja prediksi lebih baik dengan AUC 0.8579 pada *random undersampling* 70%, sementara *logistic regression* mencapai AUC 0.8334 pada *random undersampling* 30%. Skenario *cluster undersampling* terbukti paling efektif memaksimalkan *recall*. Analisis mengidentifikasi jumlah produk, usia, dan status keaktifan sebagai prediktor utama. Nasabah di atas 46 tahun dengan lebih dari dua produk namun pasif paling berisiko *churn* sehingga retensi pada segmen ini perlu diprioritaskan.

Kata kunci: *churn*, *hyperparameter tuning*, *logistic regression*, perbankan, *random forest*.

ABSTRACT

MUTHI'A INSANI. Bank Customer Churn Prediction Using Random Forest and Logistic Regression Methods Based on Hyperparameter Optimization. Supervised by BUDI SUHARJO and NUR AGUSTIANI.

Customer churn in banking is a condition in which customers completely terminate their business relationships, potentially threatening the financial stability of banks. Therefore, an accurate customer churn prediction strategy is required. This study compares the performance of two analytical methods for predicting churn, namely random forest and logistic regression, optimized through hyperparameter tuning. Utilizing 10,000 customer observations from the Kaggle platform, random undersampling and cluster undersampling strategies were applied for data balancing. Model evaluation utilized AUC, recall, precision, and a confusion matrix, while feature interpretation employed feature importance, odds ratios, and SHAP. The results indicate that random forest exhibited better predictive performance with an AUC of 0.8579 at 70% random undersampling, while logistic regression achieved an AUC of 0.8334 at 30% random undersampling. The cluster undersampling scenario proved to be the most effective in maximizing recall. The analysis identified the number of products, age, and activity status as the primary predictors. Passive customers over 46 years old with more than two products are at the highest risk of churning; thus, retention efforts for this segment must be prioritized.

Keywords: *banking*, *churn*, *hyperparameter tuning*, *logistic regression*, *random forest*.



© Hak Cipta milik IPB, tahun 2026¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PREDIKSI *CHURN* NASABAH PERBANKAN MENGGUNAKAN METODE *RANDOM FOREST* DAN *LOGISTIC REGRESSION* BERBASIS OPTIMASI *HYPERPARAMETER*

MUTHI'A INSANI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Aktuaria

**PROGRAM STUDI AKTUARIA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Prediksi *Churn* Nasabah Perbankan Menggunakan Metode *Random Forest* dan *Logistic Regression* Berbasis Optimasi *Hyperparameter*

Nama : Muthi'a Insani
NIM : G5402221033

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Dr. Ir. Budi Suharjo, M.S.

Pembimbing 2:
Nur Agustiani, M.Si.

Diketahui oleh

Ketua Program Studi:
Dr. Ir. I Gusti Putu Purnaba, DEA.
NIP. 196512181990021001

Tanggal Ujian:
18 Juni 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah *Subhanahu wa Ta'ala* atas segala rahmat dan karunia-Nya, sehingga karya ilmiah ini berhasil diselesaikan dengan baik. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan September 2025 sampai bulan Juni 2026 ini ialah analisis risiko dan prediksi *churn* nasabah perbankan, dengan judul “Prediksi *Churn* Nasabah Perbankan Menggunakan Metode *Random Forest* dan *Logistic Regression* Berbasis Optimasi *Hyperparameter*”.

Penyusunan karya ilmiah ini tentunya tidak lepas dari bimbingan, arahan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Dr. Ir. Budi Suharjo, M.S. dan Nur Agustiani, M.Si. selaku dosen pembimbing, serta Fendy Septyanto, M.Si. selaku dosen penguji yang telah senantiasa meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, kritik, serta saran yang sangat berharga dari awal penelitian hingga selesainya karya ilmiah ini.
2. Papa dan Mama sebagai kedua orang tua tercinta, atas untaian doa yang tiada henti, serta kakak dan adik yang selalu menghibur dan memberikan semangat kepada penulis dalam setiap kondisi.
3. Dosen dan seluruh staf civitas akademika Program Studi Aktuaria IPB University atas segala ilmu, fasilitas, dan bantuan operasional yang diberikan selama masa perkuliahan.
4. Pemilik NIM G6401221013, terima kasih telah bersedia menemani perjalanan skripsi penulis dari awal hingga selesai, serta atas *support* dan canda tawa yang selalu diberikan di setiap prosesnya.
5. Rekan-rekan seperjuangan dan sahabat terdekat penulis: Warintil, Warintil x Hehehe, Komplotan Cegil (Blekping), dan MBL yang selalu memberikan kebersamaan, keceriaan, serta motivasi yang tak pernah putus di masa-masa sulit penyusunan tugas akhir ini.
6. Diri sendiri, terima kasih atas kerja keras, keteguhan, kesabaran, dan perjuangan yang luar biasa untuk bertahan dan menyelesaikan seluruh rangkaian perkuliahan dan skripsi ini dengan baik.

Penyelesaian karya ilmiah ini merupakan sebuah perjalanan panjang yang memberikan banyak pembelajaran berharga. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juni 2026

Muthi'a Insani



DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 Perbankan	3
2.2 <i>Churn</i>	3
2.3 <i>Exploratory Data Analysis</i>	4
2.4 <i>Data Preprocessing</i>	4
2.5 <i>Random Forest</i>	6
2.6 <i>Logistic Regression</i>	7
2.7 <i>Optimasi Hyperparameter</i>	10
2.8 <i>Shapley Additive Explanations</i>	12
III METODE	13
3.1 Sumber Data	13
3.2 Prosedur Analisis	14
IV HASIL DAN PEMBAHASAN	17
4.1 Karakteristik Umum Data	17
4.2 <i>Data Splitting</i>	19
4.3 Uji Multikolinearitas	19
4.4 <i>Data Balancing</i>	20
4.5 <i>Analisis Hyperparameter Tuning Model Random Forest</i>	21
4.6 <i>Analisis Hyperparameter Tuning Model Logistic Regression</i>	23
4.7 Evaluasi Performa Model pada Data Uji	24
4.8 Interpretasi Variabel Prediktor Pengaruh <i>Churn</i>	30
4.9 Interpretasi Variabel Prediktor dengan Pendekatan SHAP <i>Value</i>	34
V SIMPULAN DAN SARAN	38
5.1 Simpulan	38
5.2 Saran	38
DAFTAR PUSTAKA	39
LAMPIRAN	41
RIWAYAT HIDUP	50

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	<i>Confusion matrix</i>	11
2	Tabulasi kategori setiap variabel	13
3	Proporsi nasabah <i>churn</i> setiap kategori variabel	18
4	Proporsi data setelah <i>splitting</i>	19
5	Hasil uji multikolinearitas	19
6	Proporsi data sebelum dan setelah <i>undersampling</i>	20
7	Perbandingan nilai AUC data <i>original</i> pada <i>random forest</i>	21
8	Perbandingan nilai AUC Skenario 1 (S1) pada <i>random forest</i>	21
9	Perbandingan nilai AUC Skenario 2 (S2) pada <i>random forest</i>	21
10	Perbandingan nilai AUC Skenario 3 (S3) pada <i>random forest</i>	21
11	Perbandingan nilai AUC Skenario 4 (S4) pada <i>random forest</i>	22
12	Perbandingan performa <i>hyperparameter tuning random forest</i>	22
13	Perbandingan nilai AUC seluruh skenario pada <i>logistic regression</i>	23
14	Perbandingan performa <i>hyperparameter tuning logistic regression</i>	23
15	Evaluasi performa model pada data uji	24
16	Perbandingan nilai <i>mean decrease gini</i> pada <i>random forest</i>	31
17	Estimasi parameter dan <i>odds ratio</i> pada <i>logistic regression S1</i>	31
18	Perbandingan nilai <i>odds ratio</i> pada <i>logistic regression</i>	33

DAFTAR GAMBAR

1	Diagram alur prosedur analisis data	14
2	Proporsi <i>churn</i> nasabah	17
3	Perbandingan <i>confusion matrix</i> pada data uji <i>random forest</i>	26
4	Perbandingan <i>confusion matrix</i> pada data uji <i>logistic regression</i>	29
5	ROC <i>curve</i> skenario terbaik pada kedua metode	30
6	SHAP <i>importance</i> pada (a) <i>random forest</i> dan (b) <i>logistic regression</i>	35
7	SHAP <i>summary plot</i> pada (a) <i>random forest</i> dan (b) <i>logistic regression</i>	36

DAFTAR LAMPIRAN

1	Visualisasi distribusi variabel kategorik dan numerik	41
2	Sintaks program r untuk pemodelan prediksi <i>churn</i>	42



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.