

Prediksi Churn Pelanggan dengan Machine Learning: Studi Perbandingan Random Forest dan Gradient Boosting di SAS Viya

Yohanes Wiliam Hadiprojo

Information System

Multimedia Nusa Tenggara University

yohanes.wiliam@student.umn.ac.id

Abstract— Customer Churn merupakan masalah yang signifikan bagi bank, karena mempertahankan nasabah yang sudah ada lebih hemat biaya dibandingkan dengan mengakuisisi nasabah baru. Penelitian ini menggunakan teknik machine learning untuk memprediksi churn nasabah dan mengidentifikasi faktor-faktor utama yang mempengaruhi churn dalam basis nasabah bank. Dengan menggunakan dataset churn yang komprehensif, kami melakukan persiapan data, analisis data eksploratif (EDA), dan menerapkan model Random Forest dan Gradient Boosting untuk memprediksi hasil churn. Analisis kami menunjukkan bahwa usia, jumlah produk, dan status keanggotaan aktif merupakan prediktor churn yang paling signifikan, dengan nasabah yang lebih tua dan mereka yang menggunakan beberapa produk cenderung tidak akan meninggalkan bank. Kedua model menunjukkan akurasi prediksi yang tinggi, dengan Gradient Boosting sedikit mengungguli Random Forest. Berdasarkan temuan ini, kami merekomendasikan strategi retensi yang ditargetkan, promosi penggunaan berbagai produk, dan inisiatif keterlibatan nasabah yang lebih baik untuk mengurangi tingkat churn. Penelitian ini memberikan wawasan yang dapat ditindaklanjuti bagi bank untuk mengembangkan program retensi yang efektif dan meningkatkan loyalitas nasabah, sehingga dapat meningkatkan profitabilitas. Penelitian di masa depan dapat mengeksplorasi faktor-faktor tambahan dan model yang lebih canggih untuk lebih menyempurnakan prediksi churn dan strategi pencegahannya.

Keywords— Customer churn, Machine learning, Random Forest, Gradient Boosting, Predictive modeling, Bank customer retention, Churn predictors, Data analysis, Customer engagement, Financial services

I. LATAR BELAKANG & PEMAHAMAN BISNIS

Dalam perkembangan lanskap perbankan yang semakin kompetitif, retensi pelanggan menjadi faktor yang dominan dalam penentuan kesuksesan berkelanjutan. Karena pengakuisisian nasabah baru menjadi hal yang sangat dibutuhkan dan lebih mahal dibandingkan mempertahankan nasabah yang telah ada, semakin banyak bank yang memiliki pendekatan basis data dalam melakukan prediksi dan pencegahan perpindahan nasabah. Penelitian ini ditujukan untuk memanfaatkan teknik analitik canggih dengan machine learning algorithm dalam menindaklanjuti tantangan prediksi perpindahan nasabah dalam sektor perbankan.

Penelitian di bidang data telah mengakui terhadap pentingnya penerapan kerangka kerja terstruktur dalam penanganan tugas analisis yang kompleks secara efektif. Salah satu kerangka kerja yang umum digunakan dengan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining), yang mencakup fase berbeda: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, & Deployment.

Dalam konteks memprediksi customer churn di industri perbankan, setiap fase metodologi CRISP-DM memiliki relevansi yang signifikan. Fase Pemahaman Bisnis melibatkan pemahaman yang komprehensif tentang domain masalah dan mendefinisikan tujuan bisnis. Dalam penelitian

ini, tujuannya adalah untuk mengembangkan model prediktif yang dapat secara akurat mengidentifikasi nasabah yang berisiko melakukan churning dari bank. Memahami pendorong utama dan indikator churn, seperti skor kredit, informasi demografis, dan perilaku perbankan, sangat penting untuk merumuskan strategi retensi yang efektif.

Literatur mengenai prediksi churn nasabah di sektor perbankan memberikan wawasan berharga mengenai metodologi dan teknik yang digunakan oleh para peneliti dan praktisi. Berbagai penelitian telah menunjukkan kemampuan algoritma machine learning, seperti random forest dan gradient boosting, pada beberapa feature seperti demografis nasabah, data transaksional, dan account attributes. Platform SAS Viya telah memberikan banyak peluang bagi organisasi dengan big data dan cloud computing dalam prediksi churn yang terukur dan efisien.

II. TINJAUAN TEORITIS

Pada industri perbankan, kemampuan prediksi dan memitigasi perpindahan nasabah menjadi daya tarik terhadap lembaga keuangan. Customer churn yang mengacu terhadap fenomena dimana pelanggan melakukan pemutusan hubungan terhadap bank yang dapat menimbulkan kehilangan pendapatan, peningkatan biaya akuisisi, dan penurunan reputasi merek. Dalam mengatasi hal tersebut metodologi sains data dan teknik analitik canggih semakin diminati bank dalam pengembangan model prediksi churn yang proaktif.

Inti dari prediksi churn nasabah terletak pada pemanfaatan algoritma machine learning, yang memungkinkan bank untuk memanfaatkan data nasabah dalam jumlah yang sangat besar untuk mengidentifikasi pola dan indikator potensi churn. Fitur-fitur utama seperti skor kredit, informasi demografis, perilaku perbankan, dan atribut akun berfungsi sebagai input penting untuk model prediktif ini. Analisis fitur-fitur ini memungkinkan bank untuk mendapatkan wawasan tentang perilaku dan preferensi nasabah, sehingga memungkinkan strategi retensi yang ditargetkan dan pengalaman nasabah yang dipersonalisasi.

Salah satu tujuan utama dari penelitian ini adalah untuk membandingkan kinerja dari dua algoritma machine learning yang banyak digunakan: Random Forest dan Gradient Boosting. Algoritma ini menawarkan keunggulan yang berbeda dalam hal akurasi prediksi, kemampuan interpretasi model, dan efisiensi komputasi. Dengan mengevaluasi dan membandingkan kinerja model-model ini menggunakan data perbankan dunia nyata, penelitian ini berupaya memberikan wawasan berharga tentang efektivitas pendekatan pemodelan prediktif yang berbeda untuk prediksi churn nasabah.

Selain perbandingan algoritmik, penelitian ini juga menekankan pentingnya persiapan data dan rekayasa fitur dalam proses prediksi churn. Langkah-langkah preprocessing seperti penanganan nilai yang hilang, pengkodean variabel kategorikal, dan standarisasi fitur numerik sangat penting untuk memastikan ketahanan dan keandalan model prediktif. Lebih lanjut, penelitian ini mengeksplorasi penggunaan SAS Viya, sebuah platform analitik canggih, untuk mengimplementasikan dan menerapkan model-model ini dalam skala besar.

Dengan mengadopsi kerangka kerja data science yang komprehensif, yang mencakup fase-fase seperti pemahaman bisnis, persiapan data, pemodelan, evaluasi, dan penerapan, penelitian ini bertujuan untuk memberikan pendekatan holistik terhadap prediksi churn pelanggan di sektor perbankan. Melalui analisis empiris dan wawasan teoritis, studi ini berkontribusi pada pengetahuan yang berkembang tentang analisis prediktif di bidang keuangan, yang menawarkan implikasi praktis bagi bank yang ingin meningkatkan strategi retensi nasabah dan mendorong pertumbuhan berkelanjutan di pasar yang semakin kompetitif.

III. METHODOLOGY

metodologi penelitian yang digunakan dalam melakukan analisis perbandingan model Random Forest dan Gradient Boosting untuk memprediksi customer churn pada sektor perbankan dengan menggunakan SAS Viya. Metodologi ini mencakup beberapa fase utama, termasuk pengumpulan data, preprocessing data, pemilihan model, pelatihan model, evaluasi model, dan interpretasi hasil.

Dataset yang digunakan dalam penelitian ini terdiri dari catatan nasabah bank dan atribut terkait, yang diperoleh dari [berikan sumber data atau jelaskan bagaimana dataset dikumpulkan]. Setiap catatan dalam dataset mewakili contoh nasabah yang unik, dengan fitur-fitur yang mencakup informasi demografis, perilaku perbankan, dan atribut rekening. Dataset berisi kolom-kolom berikut:

1. Row Number: Sesuai dengan nomor catatan (baris) dan tidak berpengaruh pada output.
2. CustomerId: Berisi nilai acak dan tidak berpengaruh pada nasabah yang meninggalkan bank.
3. Surname: Nama belakang nasabah, yang tidak berpengaruh pada keputusan mereka untuk meninggalkan bank.
4. CreditScore: Merupakan skor kredit nasabah, yang berpotensi mempengaruhi perpindahan nasabah, karena nasabah dengan skor kredit yang lebih tinggi memiliki kemungkinan yang lebih kecil untuk meninggalkan bank.

5. Geography: Menunjukkan lokasi nasabah, yang dapat mempengaruhi keputusan mereka untuk meninggalkan bank.
6. Gender: Menunjukkan jenis kelamin nasabah, yang menarik untuk dieksplorasi terkait perannya dalam perpindahan nasabah.
7. Age: Menunjukkan usia nasabah, yang relevan karena nasabah yang lebih tua umumnya lebih kecil kemungkinannya untuk meninggalkan bank mereka dibandingkan nasabah yang lebih muda.
8. Tenure: Mengacu pada jumlah tahun nasabah telah menjadi nasabah bank, dengan nasabah yang lebih tua biasanya lebih loyal dan kecil kemungkinannya untuk meninggalkan bank.
9. Balance: Mewakili saldo di rekening nasabah, yang berfungsi sebagai indikator signifikan dari perputaran nasabah, karena nasabah dengan saldo yang lebih tinggi cenderung tidak akan keluar dibandingkan dengan nasabah yang memiliki saldo yang lebih rendah.
10. NumOfProduct: Menunjukkan jumlah produk yang telah dibeli nasabah melalui bank.
11. HasCrCard: Menunjukkan apakah nasabah memiliki kartu kredit atau tidak, yang relevan karena nasabah yang memiliki kartu kredit lebih kecil kemungkinannya untuk meninggalkan bank.
12. IsActiveMember: Menunjukkan apakah nasabah merupakan anggota aktif, karena nasabah yang aktif cenderung tidak akan meninggalkan bank.
13. Estimated Salary: Mewakili perkiraan gaji nasabah, yang berfungsi sebagai indikator lain dari perputaran nasabah, karena nasabah dengan gaji yang lebih rendah lebih mungkin untuk keluar dibandingkan dengan mereka yang memiliki gaji yang lebih tinggi.
14. Exited: Variabel biner yang menunjukkan apakah nasabah keluar dari bank atau tidak.

1. DATA PREPARATION

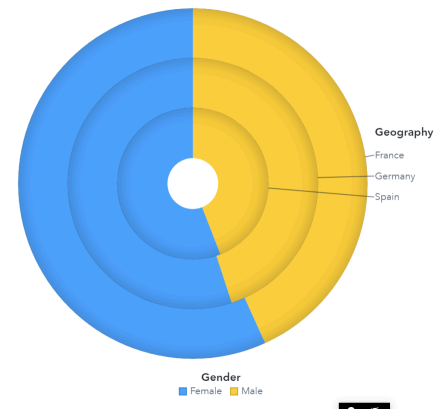
Table	Profile	Metadata
CHURN		
The session table is not current to the plan. Run the plan to update the table.		
Row...	Cust...	Geo...
63	1570...	Jeffrey
64	1575...	Pirozzi
65	1559...	Jackson
66	1578...	Hamm...
67	1569...	Brownless
68	1564...	Chibugo
69	1563...	Glauert
70	1575...	Pisano
71	1570...	Konova...
72	1562...	McKee
73	1581...	Palermo
74	1577...	Ballard
75	1577...	Wallace
76	1578...	Cavenagh
77	1561...	Hu
78	156A...	Blair

Sebelum pelatihan model, dataset menjalani beberapa langkah prapemrosesan untuk memastikan kualitas data dan kompatibilitas dengan algoritme pembelajaran mesin yang dipilih. Langkah-langkah prapemrosesan ini meliputi:

1. Handling missing value: Setiap nilai yang hilang dalam dataset diperhitungkan menggunakan teknik yang sesuai seperti imputasi rata-rata atau imputasi prediktif.
2. Menstandarisasi Variabel Numerik: Variabel numerik seperti Exited, Geography, Row Number distandarisasi agar memiliki nilai rata-rata 0 dan deviasi standar 1 untuk memastikan penskalaan yang seragam di seluruh fitur.

2. EDA

churn rate by geography and gender
Exited

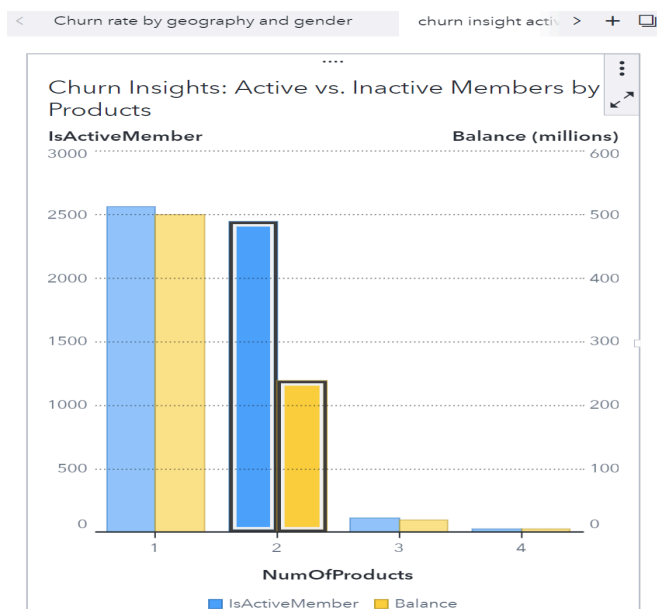


Fase analisis data eksplorasi melibatkan visualisasi dan analisis tingkat churn berdasarkan geografi dan gender untuk mengungkap pola dan wawasan. Kumpulan data mencakup tiga wilayah geografis: Prancis, Jerman, dan Spanyol. Tingkat churn divisualisasikan menggunakan bagan donat, yang mengungkapkan beberapa wawasan utama. Untuk setiap geografi, segmen bagan dibedakan dengan dua warna: biru mewakili pelanggan pria dan kuning mewakili pelanggan wanita. Lingkaran luar menggambarkan tingkat churn total untuk setiap geografi, sedangkan lingkaran dalam membedakan tingkat churn berdasarkan jenis kelamin.

Analisis ini menyoroti perbedaan yang mencolok dalam tingkat churn di seluruh wilayah geografis dan gender. Di Prancis, tingkat churn rate perempuan adalah 55,0%, dengan 448 orang yang keluar, dibandingkan dengan tingkat churn rate laki-laki sebesar 44,1%, dengan 366 orang yang keluar. Jerman menunjukkan tren yang sama, dengan tingkat churn rate perempuan sebesar 59,9% (231 orang keluar) dan

tingkat churn rate laki-laki sebesar 43,2% (182 orang keluar). Di Spanyol, tingkat keluar masuknya karyawan perempuan mencapai 55,5%, sementara tingkat keluar masuknya karyawan laki-laki mencapai 43,7%, dengan 350 karyawan laki-laki yang keluar. Namun, jumlah pasti individu perempuan yang keluar di Spanyol tidak dapat ditranskrip karena adanya potensi kesalahan.

Temuan ini menunjukkan bahwa gender memainkan peran penting dalam perpindahan pelanggan, dengan pelanggan wanita secara konsisten menunjukkan tingkat perpindahan yang lebih tinggi daripada pelanggan pria di seluruh wilayah. Selain itu, perbedaan regional juga terlihat jelas, dengan Jerman menunjukkan tingkat churn tertinggi secara keseluruhan. Wawasan ini menunjukkan perlunya bank mengembangkan strategi retensi yang ditargetkan yang memperhitungkan variasi gender dan geografis untuk mengurangi churn secara efektif. EDA awal ini memberikan gambaran yang berharga, dan analisis lebih lanjut dapat mempelajari faktor-faktor tambahan yang mempengaruhi churn untuk menginformasikan strategi retensi yang lebih bernuansa.



Dalam analisis data eksplorasi kami, kami memeriksa tingkat churn berdasarkan geografi dan jenis kelamin dengan menggunakan visualisasi diagram batang yang berjudul "Tingkat Churn Berdasarkan Geografi dan Jenis Kelamin Menggunakan Anggota Aktif yang Tidak Aktif Berdasarkan Produk." Visualisasi ini memberikan wawasan yang berharga tentang hubungan antara churn pelanggan, penggunaan produk, dan status aktivitas anggota.

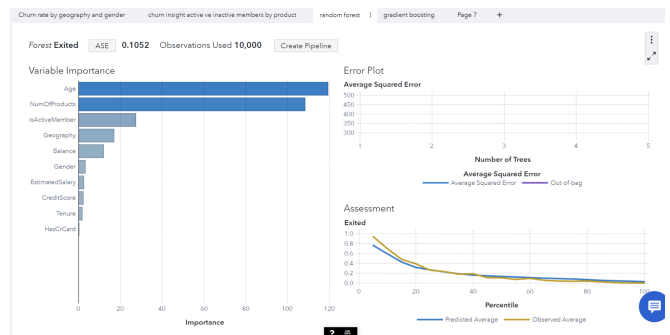
Sumbu x pada diagram batang menunjukkan jumlah produk yang digunakan oleh nasabah bank, mulai dari 1 (menunjukkan nasabah yang hanya menggunakan satu produk) hingga 4+ (menunjukkan nasabah yang menggunakan empat produk atau lebih). Sumbu y di sebelah kiri menunjukkan jumlah anggota, yang membedakan antara anggota yang aktif dan tidak aktif, sedangkan sumbu y di sebelah kanan menunjukkan saldo (dalam jutaan) yang terkait dengan setiap jumlah produk.

Beberapa wawasan penting dapat diambil dari visualisasi ini. Pertama, nasabah dengan lebih banyak produk cenderung memiliki saldo yang lebih tinggi, yang mengindikasikan korelasi positif antara jumlah produk yang digunakan nasabah dengan saldo keseluruhan. Kedua, tingkat churn rate lebih tinggi pada nasabah yang memiliki lebih sedikit produk, terutama nasabah yang hanya memiliki 1 atau 2 produk. Hal ini menunjukkan bahwa nasabah yang menggunakan lebih sedikit produk bank lebih cenderung meninggalkan bank.

Untuk mengurangi churn, bank harus fokus untuk mempertahankan nasabah yang memiliki banyak produk, karena nasabah ini tidak hanya memiliki saldo yang lebih tinggi tetapi juga lebih kecil kemungkinannya untuk berpindah bank. Pendekatan yang ditargetkan ini dapat membantu bank meningkatkan loyalitas nasabah dan mengurangi tingkat churn secara keseluruhan.

Meskipun visualisasi ini memberikan gambaran hubungan antara penggunaan produk dan churn, analisis lebih lanjut dapat mengeksplorasi faktor-faktor lain yang mempengaruhi churn untuk mengembangkan strategi retensi yang lebih komprehensif.

3. MODELING



Random Forest

Pemodelan: Random Forest

Model Random Forest diterapkan pada dataset churn untuk memprediksi churn pelanggan. Visualisasi yang disediakan menawarkan wawasan tentang kinerja model dan pentingnya berbagai fitur dalam memprediksi churn pelanggan.

1. Pentingnya Variabel

Diagram batang "Variable Importance" menunjukkan fitur mana yang paling berpengaruh dalam memprediksi apakah pelanggan akan keluar (churn). Dalam bagan ini:

- Age: Fitur yang paling penting, yang menunjukkan bahwa usia pelanggan merupakan prediktor kuat terjadinya churn. Hal ini sejalan dengan pemahaman bahwa nasabah yang lebih tua umumnya lebih kecil kemungkinannya untuk meninggalkan bank dibandingkan dengan nasabah yang lebih muda.
- Num Of Product: Fitur terpenting kedua. Nasabah yang memiliki lebih banyak produk cenderung tidak akan berpindah ke produk lain, yang menunjukkan bahwa keterlibatan dengan produk sangat penting untuk retensi.
- Is Active Members: Fitur terpenting ketiga. Anggota aktif lebih kecil kemungkinannya untuk berpindah ke bank lain, yang mengindikasikan bahwa keterlibatan dengan layanan bank merupakan faktor yang signifikan.
- Geography: Fitur ini juga berkontribusi pada prediksi, menunjukkan bahwa perbedaan regional mempengaruhi tingkat churn.
- Balance, Gender, Estimated Salary: Fitur-fitur ini memiliki tingkat kepentingan yang moderat, menunjukkan bahwa status keuangan dan karakteristik demografis berperan dalam churn nasabah.
- CreditScore, Tenure, HasCrCard: Fitur-fitur ini memiliki tingkat kepentingan yang lebih rendah, yang menunjukkan bahwa meskipun fitur-fitur ini berkontribusi pada prediksi, namun tidak sepenting faktor-faktor di atas.

2. Plot Kesalahan

"Error Plot" menunjukkan Average Squared Error (ASE) di seluruh jumlah pohon yang berbeda dalam Random Forest. Hal ini membantu dalam memahami stabilitas dan akurasi model:

- ASE relatif rendah (0,1052), menunjukkan bahwa model memiliki akurasi prediksi yang baik.
- Kesalahan menurun dan stabil seiring dengan bertambahnya jumlah pohon, yang menunjukkan bahwa menambahkan lebih banyak pohon akan meningkatkan kinerja model hingga titik tertentu.

3. Assessment Plot

"Assessment Plot" membandingkan tingkat churn rata-rata yang diprediksi dengan tingkat churn rata-rata yang diamati pada persentil yang berbeda:

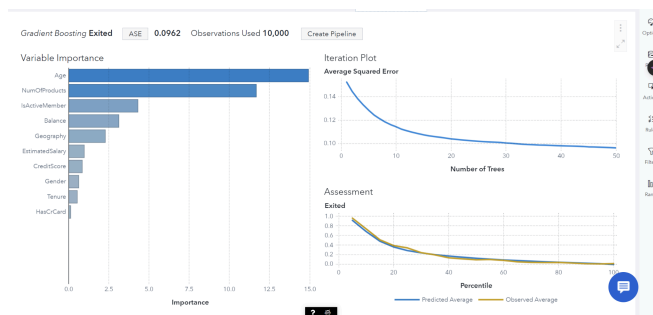
- Plot menunjukkan kecocokan yang erat antara nilai yang diprediksi dan nilai yang diamati, yang mengindikasikan bahwa model secara akurat menangkap distribusi tingkat churn.
- Bentuk kurva menunjukkan bahwa model berkinerja baik di berbagai tingkat risiko churn yang diprediksi.

4. Insight

Hasil dari model Random Forest memperkuat beberapa wawasan dari analisis data eksplorasi sebelumnya:

- Keterlibatan Pelanggan: Fitur-fitur seperti jumlah produk yang digunakan dan status keanggotaan aktif sangat penting untuk memprediksi churn, yang menekankan pentingnya keterlibatan pelanggan dalam strategi retensi.
- Faktor Demografis: Usia dan geografi merupakan prediktor yang signifikan, menunjukkan bahwa strategi retensi harus disesuaikan dengan kelompok usia dan wilayah yang berbeda.
- Faktor Keuangan: Meskipun saldo dan perkiraan gaji penting, namun tingkat kepentingannya yang lebih rendah dibandingkan dengan fitur-fitur keterlibatan menunjukkan bahwa status keuangan saja bukanlah prediktor terkuat dari churn.

Dengan memanfaatkan wawasan ini, bank dapat mengembangkan strategi yang ditargetkan untuk memitigasi churn, dengan fokus pada peningkatan keterlibatan nasabah, menyesuaikan layanan dengan segmen demografis, dan mempertimbangkan perbedaan regional.



Gradient Boosting

Model Gradient Boosting diterapkan pada dataset churn untuk memprediksi churn pelanggan. Visualisasi yang disediakan menyoroti kinerja dan pentingnya berbagai fitur dalam memprediksi apakah seorang pelanggan akan berpindah.

1. Pentingnya Variabel

Diagram batang "Variable Importance" mengidentifikasi fitur-fitur yang paling berpengaruh dalam memprediksi churn pelanggan. Dalam bagan ini:

- Age: Fitur yang paling penting, menunjukkan bahwa usia pelanggan secara signifikan mempengaruhi kemungkinan churn. Hal ini konsisten dengan pengamatan bahwa pelanggan yang lebih tua umumnya lebih loyal.
- NumOfProduct: Fitur terpenting kedua. Nasabah yang memiliki lebih banyak produk bank cenderung memiliki tingkat churn yang lebih rendah, menggarisbawahi nilai dari penjualan silang layanan bank.
- IsActiveMember: Fitur terpenting ketiga. Anggota yang aktif cenderung tidak berpindah ke bank lain, sehingga memperkuat pentingnya keterlibatan pelanggan.
- Balance: Fitur ini juga penting. Saldo akun yang lebih tinggi dikaitkan dengan tingkat churn yang lebih rendah.
- Geography: Lokasi geografis memengaruhi tingkat churn, menunjukkan perbedaan regional dalam perilaku pelanggan.
- EstimatedSalary, CreditScore, Gender, Tenue, HasCrCard: Fitur-fitur ini memiliki tingkat kepentingan yang lebih rendah namun tetap berkontribusi pada prediksi. Faktor keuangan dan demografi, meskipun kurang berpengaruh secara individual, masih berperan dalam prediksi churn secara keseluruhan.

2. Iteration Plot

"Iteration Plot" menunjukkan Average Squared Error (ASE) di berbagai jumlah pohon dalam model Gradient Boosting:

- ASE relatif rendah (0,0962), menunjukkan akurasi prediksi yang tinggi.
- Kesalahan berkurang seiring bertambahnya jumlah pohon, dan menjadi stabil ketika model menambahkan lebih banyak pohon, yang menunjukkan model yang disetel dengan baik dengan keseimbangan yang baik antara bias dan varians.

3. Assessment Plot

"Assessment Plot" membandingkan tingkat churn rata-rata yang diprediksi dengan tingkat churn rata-rata yang diamati pada persentil yang berbeda:

Plot menunjukkan keselarasan yang erat antara nilai yang diprediksi dan yang diamati, yang mengindikasikan bahwa model secara akurat menangkap distribusi tingkat churn.

Kurva tersebut menunjukkan kinerja prediktif yang kuat di berbagai tingkat risiko churn yang diprediksi, dengan deviasi minimal antara rata-rata yang diprediksi dan yang diamati.

Insight

Hasil dari model Gradient Boosting memperkuat beberapa wawasan dari analisis data eksplorasi sebelumnya dan model Random Forest:

- Keterlibatan Pelanggan: Fitur-fitur seperti jumlah produk yang digunakan dan status keanggotaan aktif sangat penting untuk memprediksi churn, menyoroti pentingnya keterlibatan pelanggan dalam strategi retensi.
- Faktor Demografis: Usia dan geografi adalah prediktor yang signifikan, yang menunjukkan perlunya strategi retensi yang disesuaikan berdasarkan segmen demografis.
- Faktor Keuangan: Saldo dan perkiraan gaji merupakan hal yang penting, mendukung gagasan bahwa nasabah yang stabil secara finansial cenderung tidak berpindah.

Dengan memanfaatkan wawasan ini, bank dapat mengembangkan strategi yang lebih efektif untuk memitigasi churn, dengan fokus pada peningkatan keterlibatan nasabah, menyediakan layanan yang dipersonalisasi berdasarkan faktor demografis, dan mempertimbangkan perbedaan regional.

IV. CONCLUSION

Penelitian ini bertujuan untuk memprediksi churn nasabah di bank dengan menggunakan model machine learning, dengan fokus pada identifikasi faktor-faktor utama yang mempengaruhi churn dan memberikan wawasan yang dapat ditindaklanjuti untuk pencegahan. Hasil analisis menunjukkan bahwa usia, jumlah produk, dan status keanggotaan aktif merupakan prediktor churn yang paling signifikan. Nasabah yang lebih tua dan mereka yang menggunakan beberapa produk bank cenderung tidak akan meninggalkan bank, menyoroti pentingnya strategi retensi yang ditargetkan untuk nasabah yang lebih muda dan penjualan silang produk keuangan. Model Random Forest dan Gradient Boosting menunjukkan akurasi prediksi yang tinggi, sehingga memperkuat keandalan temuan ini. Berdasarkan wawasan ini, disarankan agar bank

mengembangkan program retensi yang disesuaikan, mempromosikan penggunaan berbagai produk, dan meningkatkan keterlibatan nasabah melalui layanan yang dipersonalisasi dan program loyalitas. Menerapkan strategi-strategi ini dapat secara signifikan mengurangi tingkat churn, sehingga meningkatkan retensi nasabah dan profitabilitas secara keseluruhan. Penelitian di masa depan dapat mengeksplorasi faktor tambahan dan model yang lebih canggih untuk lebih menyempurnakan prediksi churn dan strategi pencegahannya.

REFERENCES

- [1] <https://towardsdatascience.com/understanding-random-forest-58381c0602d2>
- [2] <https://medium.com/@janjummy/bank-customer-churn-analysis-fc34df7ab3a8>