PENERAPAN METODE NAIVE BAYES UNTUK MEMPREDIKSI CUSTOMER CHURN BERDASARKAN KINERJA APLIKASI DUOLINGO DI INDONESIA

[1]Suharni, [2]Eel Susilowati, [3]Ristiafif Naufal Reyhandera

Jurusan Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi,
Universitas Gunadarma
[1] harni@staff.gunadarma.ac.id, [2]eel@ staff.gunadarma.ac.id ,
[3] ristiafifnaufal@student.gunadarma.ac.id

ABSTRAK

Pengguna aplikasi pembelajaran daring seperti Duolingo di Indonesia terus meningkat, namun tantangan retensi tetap muncul akibat customer churn. Penelitian ini membangun model prediksi churn berbasis survei persepsi kinerja aplikasi dengan algoritma Bernoulli Naive Bayes dalam kerangka CRISP-DM. Data diperoleh dari 200 responden melalui survei online, kemudian diolah melalui data cleaning, penghapusan atribut berpotensi bocor, feature selection, encoding, hingga pembentukan variabel target. Model dievaluasi dengan 5-fold cross-validation menggunakan metrik akurasi, presisi, recall, dan f1-score. Hasilnya menunjukkan performa rata-rata akurasi 0,8600, presisi 0,8788, recall 0,8721, dan f1-score 0,8717. Analisis sensitivitas menegaskan kinerja model tetap stabil meski satu fitur dihapus, karena informasi penting tersebar di banyak variabel. Temuan ini dapat menjadi dasar strategi retensi yang lebih menyeluruh, dengan fokus pada aspek seperti ketersediaan bahasa, kejelasan progres belajar, keterbacaan ikon, status streak, dan kemudahan navigasi menu.

***Kata Kunci**: Churn, CRISP-DM, Duolingo, Naive Bayes,* Prediksi

ABSTRACT

The users of online learning applications such as Duolingo in Indonesia continue to increase, but retention challenges still arise due to customer churn. This research builds a churn prediction model based on application performance perception surveys with the Bernoulli Naive Bayes algorithm in the CRISP-DM framework. Data is obtained from 200 respondents through an online survey, then processed through data cleaning, removal of potentially leaky attributes, feature selection, encoding, to the formation of target variables. The model was evaluated with 5-fold cross-validation using accuracy, precision, recall, and f1-score metrics. The results showed an average performance of accuracy 0.8600, precision 0.8788, recall 0.8721, and f1-score 0.8717. Sensitivity analysis confirmed the model's performance remained stable even if one feature was removed, as the important information was spread across many variables. These findings can form the basis of a more comprehensive retention strategy, focusing on aspects such as language availability, clarity of learning progress, icon readability, streak status, and ease of menu navigation.

***Keywords**: Churn, CRISP-DM, Duolingo, Naive Bayes, Prediction*

PENDAHULUAN

Perkembangan platform pembelajaran daring (EdTech) seperti Duolingo terus meningkat dari tahun ke tahun karena mampu memberikan kemudahan akses dalam mempelajari bahasa. Hingga Mei 2025, Duolingo mencatat lebih dari 130 juta pengguna aktif bulanan, naik sekitar 11 persen dibandingkan tahun sebelumnya. Di Asia Tenggara, Indonesia menjadi pasar terbesar kedua setelah Vietnam dengan pertumbuhan pengguna yang bahkan mencapai enam kali lipat dalam tiga tahun terakhir. Kondisi ini menunjukkan peluang besar, tetapi juga menghadirkan tantangan dalam mempertahankan loyalitas pengguna.

Salah satu tantangan utama adalah *customer churn*, yaitu kondisi ketika pengguna berhenti menggunakan layanan. Fenomena ini penting diperhatikan karena berdampak langsung terhadap pendapatan dan keberlangsungan bisnis (Mirkovic, M., *et al.,* 2022), sementara biaya mempertahankan pengguna terbukti lebih rendah dibandingkan mencari pengguna baru (Lin, T.Y., *et al.,* 2023). Tingginya tingkat churn juga dapat menurunkan jumlah pengguna aktif serta efektivitas strategi pertumbuhan, sehingga diperlukan pendekatan analitis yang mampu memprediksi churn sejak dini untuk mendukung strategi retensi yang lebih tepat sasaran.

Naïve Bayes menjadi salah satu algoritma yang relevan untuk tugas ini karena mampu memanfaatkan data historis pengguna, seperti demografi, pola penggunaan, dan interaksi dengan fitur aplikasi, guna mengenali pola yang mengindikasikan potensi churn (Nalatissifa, H. and Pardede, H.F., 2021). Sejumlah penelitian terdahulu juga telah mengkaji prediksi churn dengan berbagai metode. Rifky Alfarez, Rianto, dan Vega Purwayoga (2024) menggunakan Naïve Bayes pada PT Hutchison 3 Indonesia dengan akurasi 91,3%, meskipun hanya berbasis data transaksi tunggal. Faladiba Muthia Nadhira dan Haryastuti Rizqi (2023) menerapkan Random Forest dengan *Super Learner* pada data Telkomsel dengan akurasi 75,99% dan deteksi churn 80,1%, namun masih terbatas dari sisi atribut. Sementara itu, Antoh et al. (2025) menunjukkan bahwa seleksi fitur dan optimisasi *hyperparameter* dapat meningkatkan akurasi algoritma C4.5 dari 74,09% menjadi 80,05%, meski terbatas pada dataset spesifik dan satu jenis metode.

Berdasarkan keterbatasan tersebut, penelitian ini mengkaji prediksi *churn* pada konteks berbeda, yaitu aplikasi Duolingo di Indonesia, dengan memanfaatkan algoritma Naïve Bayes serta data survei yang mencakup demografi, pola penggunaan, dan persepsi pengguna. Pendekatan ini diharapkan tidak hanya menghasilkan prediksi yang akurat, tetapi juga membantu menjelaskan lebih baik faktor-faktor yang membedakan pengguna churn dan non-churn, sehingga dapat menjadi landasan dalam merancang strategi retensi yang lebih tepat.

TINJAUAN PUSTAKA

Duolingo

Duolingo merupakan sebuah platform pembelajaran bahasa yang dapat diakses secara gratis, dirancang dengan antarmuka pengguna yang menarik dan interaktif (Geeks for Geeks, 2025). Inti dari pengalaman belajar di Duolingo adalah mengubah proses belajar bahasa menjadi kegiatan yang mirip dengan bermain game, yang dikenal dengan istilah gamifikasi. Pengguna diajak untuk menyelesaikan serangkaian pelajaran singkat yang disajikan dalam berbagai format latihan. Latihan-latihan ini meliputi penerjemahan kalimat, pencocokan gambar dengan kosakata yang sesuai, pengisian bagian kalimat yang

rumpang, serta latihan untuk meningkatkan kemampuan mendengar dan mengucapkan kata atau frasa dalam bahasa target (Caldwell, 2024). Penyelesaian setiap pelajaran akan diganjar dengan poin pengalaman (XP) dan "lingots," mata uang virtual dalam aplikasi yang dapat dimanfaatkan untuk memperoleh item tambahan atau mengakses fitur-fitur khusus. Untuk menjaga motivasi pengguna, Duolingo juga mengimplementasikan sistem "streak," yang mencatat dan memberikan penghargaan atas rentetan hari belajar secara berturut-turut (Mansur, 2022).

Duolingo menyediakan pilihan kursus bahasa yang sangat beragam. Pengguna dapat memilih untuk mempelajari bahasa-bahasa populer seperti Inggris, Spanyol, Prancis, dan Jerman, hingga bahasa-bahasa yang mungkin kurang umum terdengar, atau bahkan bahasa-bahasa fiktif yang berasal dari karya budaya populer, misalnya Klingon dari serial Star Trek dan High Valyrian dari serial Game of Thrones dan (Duolingo, 2022). 8 Kemudahan akses menjadi salah satu keunggulan Duolingo, karena platform ini tersedia dalam format aplikasi seluler untuk perangkat iOS dan Android, serta dapat diakses melalui versi web (Amin, 2021). Hal tersebut memungkinkan pengguna untuk belajar kapan saja dan di mana saja, asalkan terkoneksi dengan internet. Di luar platform pembelajarannya, Duolingo juga telah mengembangkan Duolingo English Test, sebuah tes kemahiran bahasa Inggris berbasis komputer yang kini telah diakui oleh ribuan institusi pendidikan di seluruh dunia sebagai alternatif dari tes standar internasional seperti TOEFL dan IELTS (Wodzak, 2022).

Customer Churn

Dalam dunia bisnis, mempertahankan pelanggan yang sudah ada seringkali lebih bernilai dan hemat biaya dibandingkan dengan terus-menerus mengakuisisi pelanggan baru (Lin et al., 2023). Salah satu metrik penting yang menjadi perhatian utama perusahaan terkait hal ini adalah customer churn. Istilah customer churn ini merujuk pada fenomena ketika pelanggan berhenti menggunakan produk atau layanan suatu bisnis selama periode waktu tertentu.

Memahami customer churn merupakan hal yang penting bagi kelangsungan dan pertumbuhan bisnis. Tingkat churn yang tinggi dapat berdampak negatif secara signifikan terhadap pendapatan dan profitabilitas perusahaan (Mirkovic et al., 2022). Kehilangan pelanggan berarti kehilangan sumber pendapatan berulang, dan biaya untuk menarik pelanggan baru pun jauh lebih tinggi daripada biaya untuk mempertahankan pelanggan yang sudah loyal (Kilimci, 2022). Lebih jauh lagi, churn yang tinggi bisa menjadi indikasi adanya masalah mendasar dalam produk, layanan, atau pengalaman pelanggan yang ditawarkan. Jika tidak segera diatasi, hal ini dapat merusak reputasi perusahaan dan mempersulit upaya untuk bersaing di pasar (Liu et al., 2024). Oleh karena itu, kemampuan untuk mengidentifikasi, mengukur, dan pada akhirnya mengurangi customer churn menjadi sangat penting.

Machine Learning Machine learning atau pembelajaran mesin merupakan cabang dari kecerdasan buatan (AI) yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya seiring waktu tanpa diprogram secara eksplisit untuk setiap tugas tertentu. Alih-alih mengikuti serangkaian instruksi yang kaku, sistem machine learning membangun model matematis berdasarkan data sampel, yang dikenal sebagai "data pelatihan," untuk membuat prediksi atau keputusan (Daqiqil Id, 2021).

Cara kerja machine learning melibatkan beberapa tahapan. Prosesnya

dimulai dengan pengumpulan dan persiapan data yang relevan dan berkualitas tinggi (Brown, 2021). Data ini kemudian digunakan untuk melatih algoritma machine learning. Selama proses pelatihan, algoritma mencari pola, hubungan, dan wawasan tersembunyi di dalam data. Berdasarkan pola yang ditemukan, algoritma membangun sebuah model prediktif atau deskriptif. Setelah model dilatih dan diuji validitasnya, model tersebut dapat digunakan untuk menganalisis data baru yang belum pernah dilihat sebelumnya dan menghasilkan output berupa prediksi, klasifikasi, atau rekomendasi (Kufel et al., 2023). Kemampuan untuk belajar dan beradaptasi inilah yang membuat machine learning sangat kuat dan serbaguna.

Naïve Bayes

Naive Bayes merupakan sekelompok algoritma klasifikasi probabilistik yang populer dalam dunia machine learning, dikenal luas karena kesederhanaan dan efektifitasnya dalam menangani berbagai kasus (Wickramasinghe and Kalutarage, 2021). Metode ini berlandaskan pada teorema Bayes, sebuah prinsip fundamental dalam teori probabilitas yang menjelaskan cara memperbarui probabilitas suatu hipotesis ketika ada bukti baru. Sebagai salah satu teknik yang dirancang untuk tugas-tugas spesifik seperti klasifikasi, Naive Bayes seringkali digunakan karena kemampuannya yang dapat memberikan hasil yang baik meskipun dengan asumsi yang sederhana.

Cara kerja Naive Bayes dalam machine learning cukup unik dan didasarkan pada asumsi utama yang memberinya nama "naive" atau lugu. Asumsi ini berarti semua fitur atau variabel masukan yang digunakan untuk membuat prediksi bersifat independen satu sama lain (Senhadji and Ahmed, 2022). Dengan kata lain, kehadiran atau ketiadaan suatu fitur tertentu tidak mempengaruhi kehadiran atau ketiadaan fitur lainnya. Proses kerja Naïve Bayes sendiri melibatkan perhitungan probabilitas. Pertama, dari data pelatihan, algoritma menghitung probabilitas awal (prior probability) dari setiap kelas. Kemudian, ia menghitung probabilitas kemunculan setiap fitur untuk setiap kelas (likelihood). Ketika data baru masuk untuk diklasifikasikan, Naive Bayes menggunakan informasi probabilitas ini, dikombinasikan dengan teorema Bayes, untuk menghitung probabilitas posterior dari setiap kelas, dan akhirnya memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi.

Bernoulli Naïve Bayes

Bernoulli Naive Bayes merupakan salah satu varian dari algoritma Naive Bayes yang dirancang khusus untuk menangani data biner (Artur, 2021), yaitu data yang hanya memiliki dua kemungkinan nilai, seperti “ya/tidak” atau “0/1”. Algoritma ini sangat berguna ketika fitur yang dianalisis merepresentasikan keberadaan atau ketiadaan suatu karakteristik dalam data. Karena kesederhanaannya dan efisiensinya dalam memproses data dengan dimensi tinggi, Bernoulli Naive Bayes sering digunakan dalam berbagai tugas klasifikasi, terutama dalam pemrosesan teks seperti analisis sentiment (Azizah, Rintyarna and Cahyanto, 2022).

Prinsip kerja Bernoulli Naive Bayes tetap berlandaskan pada teorema Bayes, dengan asumsi “naive” bahwa setiap fitur dianggap independen terhadap fitur lainnya. Namun, berbeda dengan Gaussian atau Multinomial Naive Bayes yang bekerja dengan nilai numerik atau frekuensi, Bernoulli Naive Bayes fokus pada informasi apakah suatu fitur muncul atau tidak. Dengan demikian, meskipun suatu kata hanya muncul sekali dalam dokumen, kehadirannya akan dihitung

sama pentingnya dengan kemunculan berkali-kali (Wardani, Prahutama and Kartikasari, 2020).

CRISP-DM

CRISP-DM merupakan singkatan dari Cross-Industry Standard Process for Data Mining, adalah sebuah model proses yang menyediakan kerangka kerja atau panduan langkah demi langkah yang komprehensif untuk merencanakan, melaksanakan, dan mengelola proyek-proyek yang berfokus pada analisis data dan penemuan pola untuk menghasilkan wawasan yang bernilai (Mihaela and Emilia, 2020).

Metodologi CRISP-DM menguraikan proses penggalian data atau pengembangan solusi machine learning ke dalam enam tahapan utama yang saling terkait (Schröer, Kruse and Gómez, 2021):
1. Pemahaman Bisnis (Business Understanding): Tahapan ini menjadi fondasi yang penting. Fokus utamanya adalah mendefinisikan secara jelas tujuan proyek dari perspektif bisnis, memahami kebutuhan spesifik yang ingin dicapai, dan menerjemahkannya ke dalam tujuan teknis penggalian data atau pembelajaran mesin, serta menyusun rencana awal proyek.
2. Pemahaman Data (Data Understanding): Setelah tujuan bisnis dan teknis terdefinisi, langkah berikutnya adalah mengumpulkan data awal dan melakukan eksplorasi untuk memahami isinya. Ini melibatkan identifikasi kualitas data, penemuan wawasan awal melalui statistik deskriptif dan visualisasi, serta pembentukan hipotesis mengenai data. 3. Persiapan Data (Data Preparation): Pada tahap ini beberapa hal yang mesti dilakukan, meliputi pemilihan data yang relevan, pembersihan data (menangani nilai yang hilang, outlier, atau inkonsistensi), konstruksi atribut baru (feature engineering) yang mungkin lebih informatif, penggabungan data dari berbagai sumber, dan pemformatan data agar sesuai dengan kebutuhan alat pemodelan.
4. Pemodelan (Modeling): Di tahap ini, berbagai teknik pemodelan machine learning dipilih dan diterapkan pada data yang telah dipersiapkan. Ini dapat melibatkan pemilihan algoritma klasifikasi, regresi, klastering, atau lainnya. Beberapa model akan dibangun dan parameter masing-masing model disesuaikan untuk mendapatkan performa optimal dari sudut pandang teknis.
5. Evaluasi (Evaluation): Model-model yang telah dibangun kemudian dievaluasi secara menyeluruh. Penekanannya bukan hanya pada akurasi teknis, namun juga, apakah model tersebut benar-benar menjawab tujuan bisnis yang telah ditetapkan di awal. Hasil evaluasi ini akan menentukan apakah model siap untuk diimplementasikan atau apakah perlu kembali ke tahapan sebelumnya untuk perbaikan.
6. Penerapan (Deployment): Setelah model dianggap memuaskan dan memberikan nilai bisnis, tahapan terakhir adalah penerapannya dalam lingkungan operasional. Ini berarti mengintegrasikan model ke dalam sistem bisnis yang ada, membuat laporan akhir, atau merancang proses pemantauan dan pemeliharaan model secara berkelanjutan untuk memastikan kinerjanya tetap optimal seiring waktu.

METODE PENELITIAN

Prosedur penelitian mengikuti kerangka CRISP-DM yang terdiri dari enam tahapan. Tahap pertama yakni *business understanding*, dimulai dengan merumuskan tujuan penelitian, yaitu membangun model prediksi churn berbasis survei pengguna Duolingo.

Tahap *data understanding* dilakukan melalui pengumpulan data kuesioner, pemeriksaan kelengkapan jawaban, serta eksplorasi distribusi data. Tahap *data preparation* mencakup

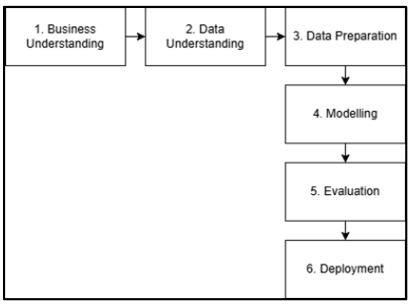
pembersihan data dari nilai kosong, imputasi sederhana, serta transformasi kategori menjadi representasi biner melalui teknik *ordinal encoding* dan *one-hot encoding*.

Tahap *modeling* dilakukan dengan menggunakan algoritma Bernoulli Naïve Bayes, yang sesuai digunakan karena seluruh fitur sudah berbentuk biner. Skema *5-fold cross-validation* diterapkan agar hasil evaluasi lebih reliabel. Selanjutnya analisis sensitivitas dilakukan melalui pendekatan *leave-one-feature-out* untuk menilai pengaruh relatif setiap faktor terhadap hasil prediksi di tahap evaluasi.

Sementara itu, tahap terakhir *deployment* berfokus pada penyusunan rekomendasi strategis berbasis data untuk mendukung retensi pengguna, sekaligus dokumentasi teknis agar penelitian dapat direplikasi atau dikembangkan lebih lanjut.

HASIL DAN PEMBAHASAN

Kerangka Umum Penelitian dengan metode CRISP-DM yang terdiri dari enam tahap yang ditunjukkan pada Gambar 1.



Gambar 1 Tahapan CRISP-DM

Penelitian ini menghasilkan aplikasi prediksi customer churn pada pengguna aplikasi Duolingo dengan algoritma Bernoulli Naïve Bayes. Proses pengolahan data dilakukan sesuai tahapan CRISP-DM, dimulai dari pembersihan data, encoding variabel kategorikal menjadi biner, hingga evaluasi model dengan validasi silang.

Hasil Persiapan Data

Tahap awal dilakukan proses pembersihan data dengan menghapus dua kolom yang tidak relevan terhadap prediksi perilaku pengguna, yaitu "Timestamp" dan "Inisial Anda". Penghapusan ini mengurangi dimensi data dari (200, 44) menjadi (200, 42). Hasil pembersihan data ditunjukkan pada Gambar 2.

```
Shape setelah cleaning: (200, 42)
```

Gambar 2 Hasil Pembersihan Data

Setelah data dibersihkan, selanjutnya dilakukan pembuatan variabel target bernama Churn yang menunjukkan status keaktifan pengguna. Penambahan variabel ini meningkatkan jumlah kolom menjadi 43 (200, 43), sebagaimana terlihat pada Gambar 3.

```
Shape setelah pembuatan variabel churn: (200, 43)
```

Gambar 3 Pembuatan Variabel Target

Adapun distribusi jumlah responden yang masuk ke dalam kategori *churn* (1) dan tidak churn (0) disajikan pada Tabel 1.

Tabel 1 Distribusi Responden Kategori Churn

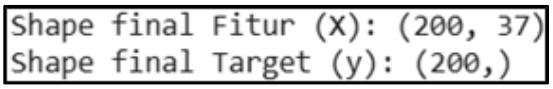
Kelas	Jumlah Responden
1 (Churn)	150
0 (Tidak Churn)	50

Setelah variabel target dibuat, proses dilanjutkan dengan menyeleksi fitur atau penghapusan atribut yang berisiko menyebabkan *data leakage* seperti pertanyaan langsung terkait alasan berhenti, dengan penyeleksian ini, jumlah

kolom menjadi 38. Perubahan ini disajikan pada Gambar 4.

Gambar 4 Hasil Seleksi Fitur

Langkah selanjutnya, dataset dipisah-kan menjadi matriks fitur (X) yang berisi variabel input dan vektor target (y) yang memuat label output (Churn), sesuai format yang dibutuhkan model machine learning. Hasil dari proses ini adalah matriks fitur (X) berukuran (200, 37) dan vektor target (y) berukuran (200,), seperti pada Gambar 5.



Gambar 5 Hasil Pemisahan Fitur dan Target

Seluruh tahapan ini, mulai dari pembersihan data hingga memisahkan fitur dan target, dilakukan guna memastikan bahwa data yang digunakan benar-benar bersih, relevan, dan bebas dari kebocoran informasi, sehingga model prediksi yang dibangun dapat menghasilkan evaluasi yang valid serta dapat dipercaya dalam menggambarkan perilaku churn pengguna.

Evaluasi Model

Pada tahap evaluasi model, data yang telah siap digunakan setelah melewati beberapa tahapan seperti pembersihan, seleksi fitur dan pembuatan kelas target, kemudian digunakan dalam proses pelatihan dan pengujian kinerja model Bernoulli Naive Bayes. Pada tahap ni model dievaluasi menggunakan 5-fold cross-validation dengan metrik akurasi, presisi, recall, dan F1-score. Hasil dari evaluasi model ini dimuat dalam Tabel 2.

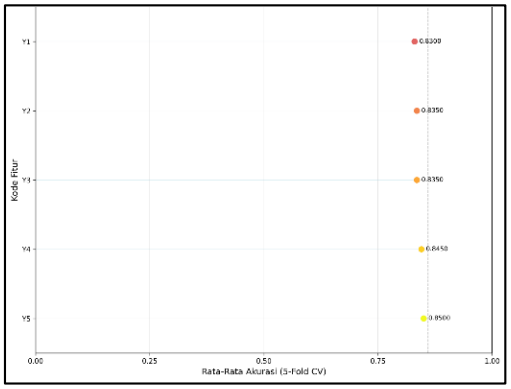
Tabel 2 Hasil Evaluasi Model dengan Cross-Validation

Metric	Mean Score	Std Deviation
Accuracy	0. 8600	0. 0583
Precision	0. 8788	0. 0570
Recall	0. 8721	0. 0554
F1_Score	0. 8717	0. 0547

Berdasarkan Tabel 2, model Bernoulli Naive Bayes setelah pra-pemrosesan mencapai akurasi rata-rata 0,8600 dengan simpangan baku 0,0583, menunjukkan prediksi benar pada 86 persen data uji dengan variasi kecil antar lipatan. Precision rata-rata 0,8788 dengan simpangan baku 0,0570, recall 0,8721 yang simpangan bakunya 0,0554, dan F1-score 0,8717 dengan nilai simpangan baku sebesar 0,0547 menunjukkan konsistensi model dalam membedakan churn dan non-churn. Nilai simpangan baku yang rendah pada seluruh metrik mengindikasikan performa model stabil dan dapat dengan baik memprediksi churn.

Analisis Sensitivitas Fitur

Untuk menguji kontribusi fitur, dilakukan analisis Leave-One-Feature-Out. Setiap fitur dihapus secara bergantian, lalu kinerja model dibandingkan dengan baseline. Hasilnya menunjukkan bahwa tidak ada satu fitur pun yang dominan mutlak, akurasi model tetap relatif stabil di kisaran 0.83 hingga 0.86. Gambar 5 menampilkan hasil leave one feature out pada 5 fitur awal yang berkaitan dengan persepsi pengguna terhadap navigasi, lama penggunaan, kualitas fitur audio, status langganan serta kecepatan pemuatan halaman profil.



Hasil analisis sensitivitas menunjuk-kan bahwa penghapusan salah satu dari fitur ini menyebabkan akurasi model berada di kisaran 0,83-0,85, sedikit lebih rendah dibandingkan akurasi awal model yang sebesar 0,86. Hal ini mengindikasikan bahwa masing-masing fitur memiliki kontribusi yang relatif kecil terhadap penurunan kinerja model ketika dihapus secara individual.

Faktor Pendorong Churn

Analisis faktor pendorong churn dilakukan dengan menghitung selisih rata-rata jawaban antara pengguna churn dan non-churn. Metode ini dipilih karena sederhana, transparan, mudah diinterpretasikan, serta sesuai untuk data ordinal. Semakin besar selisih rata-rata, semakin kuat pertanyaan tersebut menjadi pembeda. Perhitungan dilakukan dengan memetakan jawaban ke skala numerik, lalu menghitung rata-rata tiap kelompok dan menyusun daftar fitur berdasarkan selisih terbesar seperti yang ditampilkan pada Tabel 3**.** Pada tabel tersebut, kolom "*Selisih rata-rata*" memperlihatkan perbedaan jawaban kedua kelompok. Misalnya, nilai 1,94 pada pertanyaan nomor 1 menunjukkan perbedaan hampir dua tingkat pada skala survei, sedangkan nilai 0,43 pada pertanyaan nomor 10 menandakan perbedaan yang relatif kecil.

Tabel 3 Selisih Rata-rata jawaban churn dan non churn

No	Pertanyaan Lengkap	Selisih rata-rata churn dan non churn
1	Seberapa lengkap menurut Anda daftar bahasa yang tersedia "Untuk penutur Bahasa Indonesia"?	1,94
2	Seberapa jelas informasi mengenai progres belajar Anda (misalnya, Section dan Unit yang sedang aktif) ditampilkan pada halaman utama pembelajaran?	1,70
3	Seberapa jelas ikon-ikon pada menu navigasi utama di bagian bawah layar (misalnya Beranda, Latihan, Profil) dalam merepresentasikan fungsinya masing-masing?	1,28
4	Seberapa jelas informasi mengenai status streak harian Anda (termasuk kalender progres dan target streak) disajikan pada halaman Streak?	1,14
5	Seberapa sering Anda merasa 'tersesat' atau tidak yakin di mana Anda berada dalam struktur menu atau bagian aplikasi Duolingo?	0,93

Berdasarkan hasil tabel berikut, dapat disimpulkan bahwa faktor pendorong churn tidak bersumber dari satu masalah tunggal, melainkan kombinasi dari berbagai pengalaman dan penilaian negatif pengguna. Pola jawaban yang konsisten pada aspek seperti kelengkapan daftar bahasa, kejelasan informasi progres, keterbacaan ikon navigasi, kejelasan status streak, serta kemudahan dalam menelusuri

menu menunjukkan bahwa pengguna yang berhenti cenderung memiliki persepsi kurang positif secara merata terhadap berbagai elemen aplikasi. Kondisi ini memudahkan model prediktif dalam membedakan pengguna churn dan non-churn.

Rekomendasi Strategis Retensi Pengguna

Berdasarkan analisis perbedaan rata-rata jawaban antara pengguna churn dan non-churn, ditemukan sepuluh faktor utama yang berpotensi kuat membedakan perilaku pengguna. Faktor-faktor ini menjadi pijakan penting bagi pengembang Duolingo dalam menyusun strategi retensi. Beberapa strateginya antara lain menambah kelengkapan daftar bahasa dan mempermudah akses bagi penutur Bahasa Indonesia, memperjelas informasi progres belajar di halaman utama, serta meningkatkan kejernihan ikon navigasi agar lebih mudah dikenali. Strategi lain mencakup penyajian status streak harian yang lebih jelas, pengurangan kebingungan navigasi melalui penambahan penunjuk posisi atau breadcrumb, serta penyediaan instruksi dan indikator progres yang ringkas pada fitur tertentu.

Penerapan perbaikan pada aspek-aspek tersebut diharapkan mampu mengurangi pengalaman negatif pengguna dan menurunkan tingkat churn. Selain itu, model prediksi dapat dimanfaatkan untuk mengidentifikasi pengguna berisiko tinggi sejak dini, sehingga strategi retensi bisa dilakukan secara lebih proaktif.

PENUTUP

Kesimpulan

Berdasarkan hasil pengolahan dan analisis data, terdapat beberapa temuan utama. Model Bernoulli Naive Bayes yang dikembangkan mampu memprediksi churn dengan tingkat akurasi 0,8600, precision 0,8788, recall 0,8721, serta F1-score 0,8717 pada skema validasi silang 5 lipatan. Nilai metrik yang konsisten ini menunjukkan bahwa model dapat memberikan hasil klasifikasi yang andal. Hasil analisis Leave One Feature Out mengindikasikan bahwa penghapusan satu fitur tidak menyebabkan penurunan akurasi yang berarti. Temuan ini menegaskan bahwa informasi prediktif tersebar merata pada sejumlah fitur, sehingga model tidak bergantung pada satu variabel tunggal. Lebih lanjut, analisis perbedaan rata-rata jawaban antara kelompok churn dan non-churn berhasil mengidentifikasi sepuluh aspek utama yang memiliki perbedaan paling mencolok. Beberapa di antaranya adalah kelengkapan daftar bahasa, kejelasan informasi progres belajar, kejernihan ikon navigasi, kejelasan status streak, serta pengalaman pengguna ketika menavigasi menu. Visualisasi faktor-faktor tersebut memperlihatkan kecenderungan bahwa responden churn secara konsisten memberikan penilaian negatif, sedangkan non-churn lebih banyak memberikan penilaian positif.

Saran

Berdasarkan temuan penelitian, terdapat sejumlah saran yang dapat dijadikan pertimbangan untuk pengembangan lebih lanjut. Bagi pengembang aplikasi Duolingo, penting untuk meningkatkan kualitas pengalaman pengguna, khususnya pada aspek-aspek yang mendapatkan penilaian negatif dari kelompok churn. Perhatian lebih perlu diberikan pada kelengkapan dan akses daftar bahasa, kejelasan informasi progres belajar, serta kemudahan navigasi agar aplikasi lebih ramah digunakan. Bagi peneliti selanjutnya, penelitian dapat diperluas dengan menggabungkan data survei dengan data perilaku aktual yang diperoleh dari catatan aktivitas aplikasi (user log), serta menguji algoritma lain di

luar Bernoulli Naive Bayes. Pengumpulan data dengan jumlah responden yang lebih besar dan beragam juga akan memungkinkan model menangkap pola perilaku pengguna dengan lebih representatif. Sementara itu, bagi pengelola platform pembelajaran daring secara umum, pendekatan prediksi churn berbasis data survei dapat dimanfaatkan untuk mendeteksi secara dini pengguna yang berisiko berhenti. Penerapan metode serupa dapat menjadi acuan dalam menyusun strategi retensi yang lebih terarah dan sesuai dengan kebutuhan pengguna.

DAFTAR PUSTAKA

Alfarez, R. *et al.* (2024) 'PENERAPAN NAÏVE BAYES UNTUK PREDIKSI CUSTOMER CHURN (STUDI KASUS: PT HUTCHISON 3 INDONESIA)', *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, 05.

Amin, N.S. (2021) 'GAMIFICATION OF DUOLINGO IN RISING STUDENT'S ENGLISH LANGUAGE LEARNING MOTIVATION', Jurnal Bahasa Lingua Scientia, 13(2). Available at: https://doi.org/10.21274/ls.2019.11.2.219-236.

Antoh, S. *et al.* (2025) 'Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4.5', *Jurnal Sistem Informasi Bisnis*, p. 1. https://doi.org/10.21456/vol15iss1pp60-67.

Artur, M. (2021) 'Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features', in Procedia Computer Science. Elsevier B.V., pp. 564–570. https://doi.org/10.1016/j.procs.2021.06.066.

Azizah, H., Rintyarna, B.S. and Cahyanto, T.A. (2022) 'Sentimen Analisis Untuk Mengukur Kepercayaan Masyarakat Terhadap Pengadaan Vaksin Covid-19 Berbasis Bernoulli Naive Bayes', *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, 3(1), pp. 23–29. https://doi.org/10.37148/bios.v3i1.36.

Brown, S. (2021) Machine learning, explained, mitsloan.mit.edu. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

Caldwell, E. (2024) What is Duolingo? The Complete Guide to the Popular Language Learning App, duolingoguides.com. https://duolingoguides.com/what-is duolingo/

Daqiqil Id, I. (2021) Machine Learning : Teori, Studi Kasus dan Implementasi Menggunakan Python. 1st ed. Pekanbaru: UR PRESS. https://doi.org/10.5281/zenodo.5113507.

Duolingo (2022) An update is coming… to our High Valyrian course!,

blog.duolingo.com. https://blog.duolingo.com/high-valyrian-course updates/

Geeks for Geeks (2025) What Is Duolingo & How Does it Work?, geeksforgeeks.org. https://www.geeksforgeeks.org/what-is-duolingo-how-does-it-work/

Kilimci, Z.H. (2022) 'Prediction of user loyalty in mobile applications using deep contextualized word representations', Journal of Information and Telecom-munication, 6(1), pp. 43–62. https://doi.org/10.1080/24751839.2021.1981684.

Kufel, J. et al. (2023) 'What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine', Diagnostics. Multidisciplinary Digital Publishing Institute (MDPI). : https://doi.org/10.3390/diagnostics13152582.

Lin, T.Y. *et al.* (2023) 'Journal of Engineering Technology and Applied Physics Stacking Ensemble Approach for Churn Prediction: Integrating CNN and Machine Learning Models with CatBoost Meta-Learner', *Journal of Engineering Technology and Applied Physics*, 5(2), pp. 99–107. Available at: https://doi.org/10.33093/jetap.2023.5.2

Liu, X. et al. (2024) 'Customer churn prediction model based on hybrid neural networks', Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-79603-9.

Mansur, O. (2022) The habit-building research behind your Duolingo streak, blog.duolingo.com. https://blog.duolingo.com/how-duolingo-streak-builds habit/

Mihaela, C. and Emilia, T. (2020) 'BRAIN. Broad Research in Artificial Intelligence and Neuroscience Adapting CRISP-DM for Social Sciences', 11(2), pp. 99–06. https://doi.org/10.18662/brain/11.2Sup1/97.

Mirkovic, M. *et al.* (2022) 'Customer Churn Prediction in B2B Non-Contractual Business Settings Using Invoice Data', *Applied Sciences (Switzerland)*,12(10). https://doi.org/10.3390/app12105001

Nadhira Faladiba, M. and Haryastuti, R. (2023) 'Churn Analisis Pada Data Pelanggan Telekomunikasi Meng-gunakan Ensemble Learning', STATMAT (Jurnal Statistika dan Matematika) 5(1), pp. 45–54.

Nalatissifa, H. and Pardede, H.F. (2021) 'Customer Decision Prediction Using Deep Neural Network on Telco Customer Churn Data', *Jurnal Elektronika dan Telekomunikasi*, 21(2),122. https://doi.org/10.14203/jet.v21.122-127.

Schröer, C., Kruse, F. and Gómez, J.M. (2021) 'A systematic literature review on applying CRISP-DM process model', in *Procedia Computer Science*. Elsevier B.V., pp.526–534. https://doi.org/10.1016/j.procs.2021.01.199.

Senhadji, S. and Ahmed, R.A.S. (2022) 'Fake news detection using naïve Bayes and long short term memory algorithms', IAES International Journal of Artificial Intelligence, 11(2),pp.746–752. https://doi.org/10.11591/ijai.v11.i2.pp746-752.

Wardani, N.S., Prahutama, A. and Kartikasari, P. (2020) 'Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial', https://doi.org/10.14710/j.gauss.9.3.237-246.

Wickramasinghe, I. and Kalutarage, H. (2021) 'Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation', Soft Computing, 25(3), pp. 2277–2293. Available at: https://doi.org/10.1007/s00500-020-05297-6.

Wodzak, S. (2022) The Duolingo English Test: a year in review, blog.duolingo.com. https://blog.duolingo.com/det-year-in-review-2022/