



## PENERAPAN ALGORITMA RANDOM FOREST DALAM PREDIKSI CUSTOMER CHURN UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN

Oktaviana Putri Agung<sup>a\*</sup>, Fairuza Mayla Faizal<sup>b</sup>, Irenia Mascharenhas<sup>c</sup>

<sup>a</sup> Ilmu Komputer / Sistem Informasi; [oktavianaptri71@gmail.com](mailto:oktavianaptri71@gmail.com), Universitas Pamulang; Jl. Puspitek, Buaran, Kec.Pamulang, Kota Tangerang Selatan

<sup>b</sup> Ilmu Komputer / Sistem Informasi; [fairuzamayla@gmail.com](mailto:fairuzamayla@gmail.com), Universitas Pamulang; Jl. Puspitek, Buaran, Kec.Pamulang, Kota Tangerang Selatan

<sup>c</sup> Ilmu Komputer / Sistem Informasi; [ireniamascharenhas21@gmail.com](mailto:ireniamascharenhas21@gmail.com), Universitas Pamulang; Jl. Puspitek, Buaran, Kec.Pamulang, Kota Tangerang Selatan

\* Penulis Korespondensi: Oktaviana Putri Agung

### ABSTRACT

*Customer churn is a condition when customers stop using company services within a certain period. A high churn rate can negatively impact company revenue and customer loyalty, especially in telecommunications companies. This study aims to implement the Random Forest algorithm to predict customer churn in order to support customer retention strategies. The dataset used in this study is Telco Customer Churn obtained from Kaggle with a total of 7043 customer records. This research applies the CRISP-DM methodology consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The preprocessing stage includes handling missing values, transforming data types, and encoding categorical data. The modeling process uses the Random Forest Classifier algorithm with an 80:20 split between training and testing data. The results show that the Random Forest model achieved an accuracy of 77% in predicting customer churn. The most influential factors affecting churn based on feature importance are TotalCharges, tenure, MonthlyCharges, and customer contract types. Based on the research results, the Random Forest algorithm can help companies identify customers with churn potential and support more effective customer retention strategies.*

**Keywords:** Customer Churn; Data Mining; Random Forest; Classification; Customer Retention

### Abstrak

*Customer churn* merupakan kondisi ketika pelanggan berhenti menggunakan layanan perusahaan dalam periode tertentu. Tingginya tingkat customer churn dapat memberikan dampak negatif terhadap pendapatan perusahaan dan loyalitas pelanggan, khususnya pada perusahaan telekomunikasi. Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam memprediksi customer churn guna mendukung strategi retensi pelanggan. Dataset yang digunakan adalah Telco Customer Churn yang diperoleh dari Kaggle dengan jumlah data sebanyak 7043 pelanggan. Penelitian ini menggunakan metode CRISP-DM yang terdiri dari business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Tahapan preprocessing meliputi penanganan missing value, transformasi tipe data, dan encoding data kategorik. Proses pemodelan dilakukan menggunakan algoritma Random Forest Classifier dengan pembagian data training dan testing sebesar 80:20. Hasil penelitian menunjukkan bahwa model Random Forest menghasilkan accuracy sebesar 77% dalam memprediksi customer churn. Faktor yang paling mempengaruhi churn berdasarkan feature importance adalah TotalCharges, tenure, MonthlyCharges, dan jenis kontrak pelanggan. Berdasarkan hasil penelitian, algoritma Random Forest mampu membantu perusahaan dalam mengidentifikasi pelanggan yang berpotensi churn sehingga dapat digunakan sebagai dasar dalam menentukan strategi retensi pelanggan secara lebih efektif.

**Kata Kunci:** Customer Churn; Data Mining; Random Forest; Klasifikasi; Retensi Pelanggan

## 1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong meningkatnya persaingan bisnis pada berbagai sektor, termasuk industri telekomunikasi. Perusahaan telekomunikasi tidak hanya dituntut untuk memperoleh pelanggan baru, tetapi juga mempertahankan pelanggan yang telah menggunakan layanan perusahaan. Salah satu permasalahan yang sering dihadapi adalah *customer churn*, yaitu kondisi ketika pelanggan memutuskan berhenti menggunakan layanan dalam periode tertentu [1]. Tingginya tingkat *customer churn* dapat berdampak pada penurunan pendapatan perusahaan serta meningkatnya biaya akuisisi pelanggan baru.

Perkembangan teknologi digital menghasilkan data pelanggan dalam jumlah besar yang dapat dimanfaatkan untuk mendukung pengambilan keputusan bisnis. Salah satu pendekatan yang dapat digunakan adalah *data mining*. *Data mining* merupakan proses pencarian pola, hubungan, dan informasi penting dari kumpulan data yang besar menggunakan teknik statistik maupun *machine learning* [2]. Melalui *data mining*, perusahaan dapat melakukan analisis perilaku pelanggan sehingga potensi perpindahan pelanggan dapat diidentifikasi lebih awal.

Prediksi *customer churn* termasuk ke dalam teknik klasifikasi dalam *data mining*, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan data historis [2]. Berbagai algoritma klasifikasi dapat digunakan untuk memprediksi *customer churn*, seperti Decision Tree, Naïve Bayes, Support Vector Machine, dan Random Forest. Pada penelitian ini digunakan algoritma Random Forest karena memiliki kemampuan klasifikasi yang baik, mampu menangani data dalam jumlah besar, serta dapat mengurangi risiko *overfitting* melalui pembentukan banyak pohon keputusan (*decision tree*) [3]. Selain itu, Random Forest mampu memberikan informasi mengenai atribut yang paling berpengaruh terhadap prediksi melalui analisis *feature importance*.

Penelitian ini menggunakan dataset *Telco Customer Churn* yang diperoleh dari Kaggle dengan jumlah data sebanyak 7043 pelanggan. Dataset tersebut memuat informasi pelanggan seperti lama berlangganan (*tenure*), biaya bulanan (*MonthlyCharges*), total pembayaran (*TotalCharges*), jenis kontrak, serta status *churn* pelanggan. Pemanfaatan dataset ini diharapkan mampu menghasilkan model prediksi yang dapat membantu perusahaan dalam mengenali pelanggan yang berpotensi *churn*.

Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah bagaimana penerapan algoritma Random Forest dalam memprediksi *customer churn* untuk mendukung strategi retensi pelanggan. Adapun tujuan penelitian ini adalah menerapkan algoritma Random Forest untuk memprediksi *customer churn* serta memberikan rekomendasi strategi retensi pelanggan berdasarkan hasil analisis data.

## 2. TINJAUAN PUSTAKA

### 2.1. Data Mining

*Data mining* merupakan proses menemukan pola, hubungan, serta informasi penting dari kumpulan data berukuran besar menggunakan teknik statistik dan *machine learning* [4]. Pemanfaatan *data mining* memungkinkan organisasi memperoleh pengetahuan baru yang dapat digunakan untuk mendukung pengambilan keputusan secara lebih efektif [5]. Dalam bidang bisnis dan telekomunikasi, *data mining* sering dimanfaatkan untuk memahami perilaku pelanggan, melakukan segmentasi pasar, serta memprediksi kemungkinan pelanggan berhenti menggunakan layanan perusahaan.

Pada penelitian ini, *data mining* digunakan untuk menganalisis perilaku pelanggan telekomunikasi dan mendukung proses prediksi *customer churn*. Melalui teknik klasifikasi, pola dari data historis pelanggan dapat dipelajari sehingga perusahaan dapat mengidentifikasi pelanggan yang memiliki potensi *churn* lebih awal.

### 2.2. Customer Churn

*Customer churn* merupakan kondisi ketika pelanggan memutuskan berhenti menggunakan layanan atau produk yang ditawarkan perusahaan dalam periode tertentu [6]. Tingginya tingkat *customer churn* menjadi tantangan bagi perusahaan telekomunikasi karena dapat menyebabkan penurunan pendapatan serta

meningkatnya biaya untuk memperoleh pelanggan baru [7]. Oleh karena itu, perusahaan perlu memahami faktor-faktor yang memengaruhi *churn* agar strategi retensi pelanggan dapat dilakukan secara lebih efektif. Pada industri telekomunikasi, *customer churn* dipengaruhi oleh berbagai faktor seperti lama berlangganan (*tenure*), biaya layanan, kualitas pelayanan, jenis kontrak pelanggan, serta tingkat kepuasan terhadap layanan yang digunakan [6]. Prediksi *customer churn* menjadi penting karena dapat membantu perusahaan mengenali pelanggan yang berisiko berhenti menggunakan layanan sebelum keputusan tersebut terjadi.

Dalam penelitian ini, prediksi *customer churn* dilakukan menggunakan data historis pelanggan telekomunikasi untuk mengidentifikasi pelanggan yang berpotensi *churn* dan mendukung pengambilan keputusan bisnis berbasis data.

### 2.3. Klasifikasi

Klasifikasi merupakan salah satu teknik dalam *data mining* yang digunakan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan pola dari data historis [8]. Teknik ini bekerja dengan membangun model menggunakan *training data* yang telah memiliki label, kemudian model tersebut digunakan untuk memprediksi kelas pada data baru [9]. Dalam konteks *machine learning*, klasifikasi termasuk metode *supervised learning* karena memerlukan data target atau label sebagai acuan pembelajaran model.

Berbagai algoritma klasifikasi dapat digunakan dalam proses prediksi, seperti Decision Tree, Naïve Bayes, Support Vector Machine, dan Random Forest [8]. Pemilihan algoritma bergantung pada karakteristik data, tujuan penelitian, serta performa model yang dihasilkan. Teknik klasifikasi banyak digunakan dalam bidang bisnis, kesehatan, pendidikan, dan telekomunikasi karena mampu membantu proses pengambilan keputusan berbasis data.

Pada penelitian ini, klasifikasi digunakan untuk memprediksi status *customer churn* pelanggan telekomunikasi, yaitu pelanggan yang berpotensi berhenti menggunakan layanan (*churn*) dan pelanggan yang tetap menggunakan layanan (*non-churn*). Dengan adanya model klasifikasi, perusahaan dapat melakukan identifikasi risiko *churn* secara lebih awal dan menyusun strategi retensi pelanggan yang lebih efektif.

### 2.4. Random Forest

Random Forest merupakan algoritma *machine learning* berbasis *ensemble learning* yang bekerja dengan membangun banyak *decision tree* dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir [10]. Pendekatan ini bertujuan meningkatkan akurasi prediksi serta mengurangi risiko *overfitting* yang sering terjadi pada model *decision tree* tunggal [11].

Pada algoritma Random Forest, setiap *decision tree* dibangun menggunakan sampel data dan subset atribut yang dipilih secara acak. Hasil prediksi akhir diperoleh melalui mekanisme *majority voting*, yaitu kelas dengan jumlah prediksi terbanyak dari seluruh pohon keputusan akan dipilih sebagai hasil klasifikasi [10]. Mekanisme tersebut menjadikan Random Forest memiliki performa yang stabil dalam menangani data dengan jumlah besar maupun atribut yang beragam.

Random Forest banyak digunakan dalam berbagai penelitian klasifikasi karena mampu menghasilkan tingkat akurasi yang baik serta memiliki kemampuan mengidentifikasi atribut yang paling berpengaruh terhadap prediksi melalui analisis *feature importance* [11]. Dalam penelitian ini, algoritma Random Forest digunakan untuk memprediksi status *customer churn* berdasarkan data historis pelanggan telekomunikasi.

### 2.5. Penelitian Terdahulu

Penelitian mengenai prediksi *customer churn* telah banyak dilakukan menggunakan berbagai algoritma *machine learning*. Salah satu penelitian menunjukkan bahwa model klasifikasi mampu membantu perusahaan telekomunikasi mengidentifikasi pelanggan yang memiliki risiko berhenti menggunakan layanan sehingga strategi retensi pelanggan dapat dilakukan lebih awal [12]. Hasil penelitian tersebut menunjukkan bahwa penerapan *machine learning* dapat meningkatkan efektivitas pengambilan keputusan dalam pengelolaan pelanggan.

Penelitian lain membandingkan beberapa algoritma klasifikasi dalam prediksi *customer churn*, seperti Decision Tree, Logistic Regression, dan Random Forest [13]. Hasil penelitian menunjukkan bahwa Random Forest memiliki performa yang baik karena mampu menghasilkan nilai akurasi yang tinggi dan stabil dibandingkan beberapa algoritma lainnya.

Selain itu, penelitian terkait *customer churn prediction* menggunakan metode *ensemble learning* menunjukkan bahwa Random Forest tidak hanya mampu menghasilkan prediksi yang akurat, tetapi juga dapat mengidentifikasi faktor-faktor yang paling berpengaruh terhadap keputusan pelanggan melalui analisis *feature importance* [14]. Kemampuan tersebut menjadi salah satu alasan pemilihan Random Forest dalam penelitian ini.

Berdasarkan penelitian terdahulu, sebagian besar penelitian berfokus pada peningkatan performa model klasifikasi. Penelitian ini memiliki perbedaan karena selain melakukan prediksi *customer churn* menggunakan Random Forest, penelitian juga menganalisis atribut yang paling memengaruhi *churn* melalui *feature importance* serta memberikan rekomendasi strategi retensi pelanggan berdasarkan hasil model.

### 3. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode **CRISP-DM (Cross Industry Standard Process for Data Mining)** sebagai kerangka kerja dalam proses *data mining*. Metode CRISP-DM dipilih karena menyediakan tahapan penelitian yang sistematis mulai dari pemahaman masalah bisnis hingga evaluasi model yang dihasilkan. Tahapan yang digunakan dalam penelitian ini meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*.

#### 3.1 Business Understanding

Tahap *Business Understanding* bertujuan memahami permasalahan bisnis pada industri telekomunikasi terkait tingginya tingkat *customer churn*. Tingginya jumlah pelanggan yang berhenti menggunakan layanan dapat menyebabkan penurunan pendapatan perusahaan serta meningkatnya biaya untuk memperoleh pelanggan baru. Oleh karena itu, penelitian ini bertujuan membangun model prediksi *customer churn* menggunakan algoritma Random Forest guna membantu perusahaan mengidentifikasi pelanggan yang berpotensi berhenti menggunakan layanan dan mendukung strategi retensi pelanggan.

#### 3.2 Data Understanding

Penelitian ini menggunakan dataset **Telco Customer Churn** yang diperoleh dari Kaggle dengan jumlah **7043 data pelanggan**. Dataset memuat berbagai atribut pelanggan telekomunikasi seperti *tenure*, *MonthlyCharges*, *TotalCharges*, jenis kontrak (*Contract*), metode pembayaran (*PaymentMethod*), serta status *Churn* sebagai variabel target.

Pada tahap ini dilakukan eksplorasi data menggunakan fungsi `head()`, `info()`, dan `isnull().sum()` untuk memahami struktur data, tipe data, serta mengidentifikasi *missing value* pada dataset.

#### 3.3 Data Preparation

Tahap *Data Preparation* dilakukan untuk membersihkan dan menyiapkan data sebelum proses pemodelan. Tahapan preprocessing yang dilakukan meliputi penghapusan atribut `customerID` karena tidak berpengaruh terhadap prediksi, konversi atribut `TotalCharges` ke bentuk numerik, penanganan *missing value* menggunakan metode `dropna()`, transformasi variabel target *Churn* menjadi bentuk numerik (0 dan 1), serta proses *encoding* terhadap atribut kategorikal menggunakan metode `get_dummies()`.

Tahapan preprocessing bertujuan menghasilkan dataset yang bersih dan siap digunakan pada proses klasifikasi menggunakan algoritma Random Forest.

#### 3.4 Modeling

Pada tahap *Modeling*, data dipisahkan menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) menggunakan metode `train_test_split()` dengan proporsi 80:20. Model klasifikasi dibangun menggunakan algoritma **Random Forest Classifier** dari pustaka *scikit-learn*.

Algoritma Random Forest dipilih karena mampu menghasilkan performa klasifikasi yang baik, menangani data dengan jumlah atribut yang banyak, serta mengurangi risiko *overfitting* melalui pendekatan *ensemble learning*. Model yang telah dilatih menggunakan data pelatihan selanjutnya digunakan untuk melakukan prediksi terhadap data pengujian.

#### 3.5 Evaluation

---

*Penerapan Algoritma Random Forest dalam Prediksi Customer Churn Untuk Mendukung Strategi Retensi Pelanggan (Oktaviana Putri Agung)*

Tahap *Evaluation* dilakukan untuk mengukur performa model Random Forest dalam memprediksi *customer churn*. Evaluasi model dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *ROC-AUC*. Selain itu, dilakukan analisis *feature importance* untuk mengetahui atribut yang paling memengaruhi prediksi *customer churn*.

Evaluasi model bertujuan mengetahui tingkat keakuratan model sekaligus memastikan bahwa model yang dibangun mampu mendukung pengambilan keputusan bisnis terkait strategi retensi pelanggan.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Experimen

Pada penelitian ini dilakukan eksperimen menggunakan dataset *Telco Customer Churn* yang telah melalui tahap *preprocessing* dan *encoding*. Dataset kemudian dipisahkan menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) menggunakan metode *train-test split* dengan proporsi 80:20.

Model klasifikasi dibangun menggunakan algoritma **Random Forest Classifier** dari pustaka *Scikit-Learn*. Model dilatih menggunakan data pelatihan (*training data*) dan selanjutnya digunakan untuk melakukan prediksi terhadap data pengujian (*testing data*) guna mengetahui kemampuan model dalam memprediksi status *customer churn* pelanggan.

### 4.2 Evaluasi Model

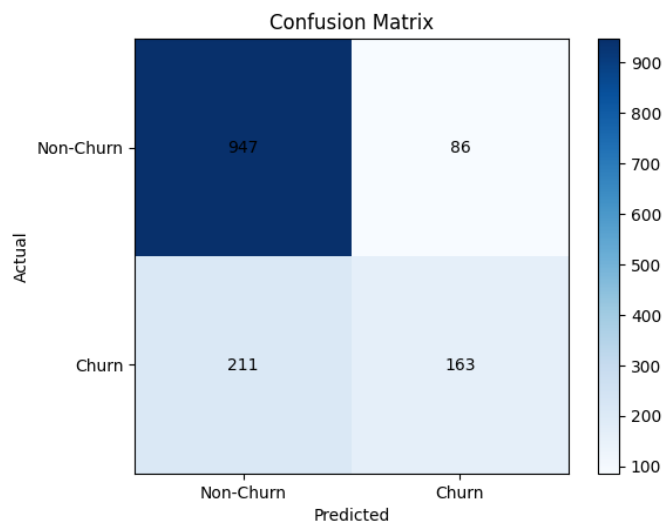
Evaluasi model dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Berdasarkan hasil pengujian, model Random Forest menghasilkan nilai *accuracy* sebesar 77%. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan pelanggan *churn* dan *non-churn*.

Tabel 1. Hasil evaluasi random Forest

No	Metrik	Nilai
1	Accuracy	77,75%
2	Precision	61%
3	Recall	47%
4	F1-Score	53%

### 4.3 Confusion Matrix

Selain menggunakan *accuracy* dan *classification report*, evaluasi model juga dilakukan menggunakan *confusion matrix* untuk mengetahui distribusi prediksi model terhadap data aktual.



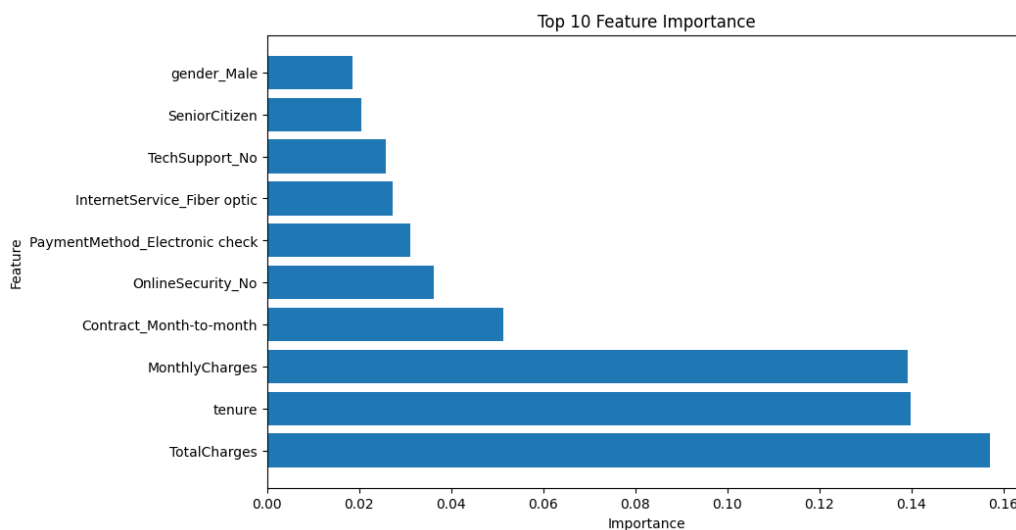
Gambar 1. Confusion Matrix Random Forest

Interpretasi *confusion matrix* menunjukkan bahwa model berhasil mengidentifikasi **947 pelanggan non-churn** dengan benar dan **163 pelanggan churn** dengan benar. Namun demikian, masih terdapat **86 pelanggan non-churn** yang diprediksi sebagai *churn* serta **211 pelanggan churn** yang tidak berhasil terdeteksi oleh model.

Hasil tersebut menunjukkan bahwa algoritma Random Forest memiliki kemampuan yang cukup baik dalam mengenali pelanggan yang tetap menggunakan layanan (*non-churn*). Namun, model masih menghadapi tantangan dalam mendeteksi seluruh pelanggan yang benar-benar berpotensi *churn*, terlihat dari jumlah pelanggan *churn* yang masih belum teridentifikasi secara optimal.

**4.4 Interpretasi Hasil (Feature Importance)**

Analisis *feature importance* dilakukan untuk mengetahui atribut yang paling memengaruhi prediksi *customer churn*. Berdasarkan hasil algoritma Random Forest, atribut dengan tingkat pengaruh tertinggi antara lain TotalCharges, tenure, MonthlyCharges, dan Contract\_Month-to-month.



Gambar 2 Feature Importance Random Forest

Interpretasi *confusion matrix* menunjukkan bahwa model berhasil mengidentifikasi 947 pelanggan non-churn dengan benar dan 163 pelanggan churn dengan benar. Namun demikian, masih terdapat 86 pelanggan non-churn yang diprediksi sebagai churn serta 211 pelanggan churn yang tidak berhasil terdeteksi oleh model. Jumlah keseluruhan data pada *confusion matrix* adalah 1407 data, yang sesuai dengan jumlah data testing hasil pembagian dataset menggunakan metode train-test split.

Hasil tersebut menunjukkan bahwa algoritma Random Forest memiliki kemampuan yang cukup baik dalam mengenali pelanggan yang tetap menggunakan layanan (*non-churn*). Namun, model masih menghadapi tantangan dalam mendeteksi seluruh pelanggan yang benar-benar berpotensi *churn*, terlihat dari jumlah pelanggan *churn* yang masih belum teridentifikasi secara optimal.

**4.5 Rekomendasi Bisnis**

Berdasarkan hasil analisis Random Forest, perusahaan dapat menerapkan strategi retensi pelanggan dengan memfokuskan perhatian pada pelanggan yang memiliki nilai TotalCharges, MonthlyCharges, dan tenure tertentu serta pelanggan dengan kontrak bulanan (*month-to-month*). Untuk mengurangi risiko churn, perusahaan dapat memberikan promosi, diskon, maupun program loyalitas kepada pelanggan yang teridentifikasi berpotensi berpindah layanan. Selain itu, peningkatan kualitas layanan dan dukungan pelanggan perlu dilakukan untuk meningkatkan kepuasan pelanggan dan mengurangi potensi perpindahan ke layanan kompetitor.

#### 4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian menggunakan algoritma Random Forest pada dataset Telco Customer Churn, model berhasil digunakan untuk memprediksi kemungkinan customer churn dengan nilai accuracy sebesar 77,75%. Hasil evaluasi menunjukkan bahwa model cukup baik dalam mengidentifikasi pelanggan non-churn, meskipun masih terdapat pelanggan churn yang belum terdeteksi secara optimal. Berdasarkan analisis feature importance, atribut TotalCharges, tenure, dan MonthlyCharges menjadi faktor yang paling berpengaruh terhadap prediksi customer churn. Hasil penelitian ini menunjukkan bahwa teknik data mining menggunakan Random Forest dapat mendukung perusahaan dalam membantu mengidentifikasi pelanggan yang berpotensi berhenti menggunakan layanan sehingga strategi retensi dapat dilakukan lebih awal. Untuk penelitian selanjutnya, disarankan menggunakan teknik penyeimbangan data (data balancing) apabila distribusi kelas target tidak seimbang, serta membandingkan performa Random Forest dengan algoritma klasifikasi lainnya untuk meningkatkan akurasi prediksi. Untuk penelitian selanjutnya, disarankan melakukan perbandingan dengan algoritma klasifikasi lain seperti **Decision Tree**, **Naïve Bayes**, atau **Support Vector Machine** guna memperoleh model dengan performa yang lebih optimal. Selain itu, penggunaan dataset yang lebih besar maupun penambahan variabel pendukung juga dapat dilakukan agar hasil prediksi *customer churn* menjadi lebih akurat.

#### Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada dosen pengampu mata kuliah yang telah memberikan arahan dan bimbingan selama pelaksanaan penelitian ini. Penulis juga mengucapkan terima kasih kepada pengelola Kaggle yang digunakan dalam penelitian serta kepada seluruh pihak yang telah mendukung penyelesaian penelitian ini.

#### DAFTAR PUSTAKA

- [1] A. Amin, S. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain, and K. Huang, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research*, vol. 94, pp. 290–301, 2019, doi: 10.1016/j.jbusres.2018.03.004.
- [2] S. Idris, A. Khan, and Y. S. Lee, "Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost based ensemble classification," *Applied Intelligence*, vol. 39, no. 3, pp. 659–672, 2013, doi: 10.1007/s10489-012-0438-0.
- [3] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [4] A. C. Berson, S. Smith, and K. Thearling, "Building Data Mining Applications for CRM," *Data Mining and Knowledge Discovery*, vol. 6, no. 1, pp. 85–102, 2002.
- [5] M. H. Dunham, "Data Mining: Introductory and Advanced Topics," *Data Mining and Knowledge Discovery*, vol. 8, no. 2, pp. 145–160, 2003.
- [6] N. Ahmad, A. Hussain, and M. S. Khan, "Predicting Customer Churn Using Machine Learning Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 1–8, 2019.
- [7] Kaggle, "Telco Customer Churn Dataset." Internet: <https://www.kaggle.com/blastchar/telco-customer-churn> [May 30, 2026].