Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

Analisis Perbandingan Algoritma Random Forest dan Support Vector Machine pada Prediksi *Customer Churn* Menggunakan IBM Telco *Customer Churn* Dataset

Saudurma Seven Septiana Sidabutar*, Ningsih Septi Uli Purba, Wulan Liviana Simbolon, Betharya Tampubolon, Jaya Tata Hardinata

Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas HKBP Nommensen Pematangsiantar, Pematangsiantar, Indonesia
Jl. Sangnawaluh No.4, Siopat Suhu, Kec. Siantar Timur, Kota Pematangsiantar, Sumatra Utara, Indonesia
Email: [1,*]saudurmasidabutar@gmail.com, [2]ningsihpurba71@gmail.com, [3]wulansimbolon611@gmail.com, [4]betharyatampubolon@gmail.com, [5]jayatatahardinata@uhnp.ac.id
Email Penulis Korespondensi: saudurmasidabutar@gmail.com

Abstrak- Customer churn merupakan kondisi ketika pelanggan memutuskan untuk berhenti menggunakan layanan suatu perusahaan sehingga dapat berdampak pada penurunan pendapatan dan meningkatnya biaya untuk memperoleh pelanggan baru. Pada industri telekomunikasi, kemampuan memprediksi pelanggan yang berpotensi melakukan customer churn menjadi salah satu strategi penting untuk meningkatkan efektivitas program retensi pelanggan. Penelitian ini bertujuan untuk membandingkan performa algoritma Random Forest dan Support Vector Machine (SVM) dalam memprediksi customer churn menggunakan IBM Telco Customer Churn Dataset yang terdiri dari 7.043 data pelanggan dengan 21 atribut. Proses penelitian meliputi tahap preprocessing menggunakan Orange Data Mining, visualisasi data menggunakan Microsoft Power BI, pembangunan model klasifikasi, serta evaluasi model menggunakan metode 10-Fold Cross Validation. Kinerja model dievaluasi berdasarkan nilai Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC). Hasil penelitian menunjukkan bahwa algoritma Random Forest menghasilkan performa yang lebih baik dibandingkan Support Vector Machine. Random Forest memperoleh nilai AUC sebesar 0,832, Accuracy 0,796, Precision 0,785, Recall 0,796, F1-Score 0,787, dan MCC 0,442, sedangkan Support Vector Machine memperoleh AUC 0,575, Accuracy 0,661, Precision 0,657, Recall 0,661, F1-Score 0,659, dan MCC 0,121. Hasil tersebut menunjukkan bahwa algoritma Random Forest lebih efektif dalam mengenali pola data pelanggan sehingga mampu menghasilkan model prediksi customer churn yang lebih akurat pada dataset penelitian. Penelitian ini diharapkan dapat menjadi referensi bagi perusahaan telekomunikasi dalam memilih algoritma klasifikasi yang tepat untuk mendukung pengambilan keputusan dan penyusunan strategi retensi pelanggan secara lebih efektif.

Kata Kunci: Customer Churn; Random Forest; Support Vector Machine; Orange Data Mining; Microsoft Power BI

Abstract- *Customer churn is a condition in which customers decide to discontinue the services provided by a company. In the telecommunications industry, a high customer churn rate can reduce company revenue and customer loyalty. Therefore, an accurate prediction method is needed to identify customers who are likely to churn so that preventive strategies can be implemented at an early stage. This study aims to compare the performance of the Random Forest and Support Vector Machine (SVM) algorithms in predicting customer churn using the IBM Telco Customer Churn Dataset. The research stages include data collection, data preprocessing, model development using Orange Data Mining, model evaluation through Test and Score, Confusion matrix, and Receiver operating characteristic (ROC), as well as data visualization using Microsoft Power BI. The results indicate that both algorithms are capable of classifying customer data; however, the Random Forest algorithm achieves better performance than the Support Vector Machine based on the evaluation metrics obtained. Furthermore, data visualization using Microsoft Power BI provides a clearer understanding of customer characteristics and supports the interpretation of the research findings. Therefore, Random Forest is recommended as a more effective algorithm for customer churn prediction in the telecommunications sector.*

Keywords*: Customer Churn; Random Forest; Support Vector Machine; Orange Data Mining; Microsoft Power BI*

1. PENDAHULUAN

Transformasi digital telah menjadi salah satu faktor utama yang mendorong perubahan paradigma operasional di berbagai sektor industri, termasuk industri telekomunikasi. Perkembangan teknologi informasi, komputasi awan (*cloud computing*), Internet of Things (IoT), analisis data berskala besar (*big data analytics*), serta kecerdasan buatan (*Artificial Intelligence*) memungkinkan perusahaan telekomunikasi menghasilkan dan mengelola data pelanggan dalam jumlah yang sangat besar setiap harinya. Data tersebut tidak hanya berasal dari aktivitas komunikasi pelanggan, tetapi juga mencakup informasi demografis, jenis layanan yang digunakan, lama berlangganan, pola penggunaan internet, metode pembayaran, jenis kontrak, penggunaan layanan tambahan, frekuensi interaksi dengan layanan pelanggan, hingga riwayat pembayaran [1]. Seluruh informasi tersebut merupakan aset yang sangat bernilai apabila diolah menjadi pengetahuan yang dapat mendukung proses pengambilan keputusan secara lebih efektif dan berbasis data [2].

Dalam era persaingan bisnis yang semakin kompetitif, perusahaan telekomunikasi tidak hanya dituntut untuk memperoleh pelanggan baru, tetapi juga mempertahankan pelanggan yang telah dimiliki. Strategi mempertahankan pelanggan menjadi lebih penting karena biaya memperoleh pelanggan baru umumnya lebih tinggi dibandingkan biaya mempertahankan pelanggan lama. Selain itu, pelanggan yang loyal memiliki kontribusi yang lebih besar terhadap peningkatan pendapatan perusahaan melalui penggunaan layanan secara berkelanjutan [3]. Oleh sebab itu, perusahaan

Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

perlu memahami perilaku pelanggan secara lebih mendalam agar dapat mengidentifikasi pelanggan yang berpotensi menghentikan penggunaan layanan sebelum keputusan tersebut benar-benar terjadi[4].

Salah satu permasalahan utama yang sering dihadapi perusahaan telekomunikasi adalah customer churn, yaitu kondisi ketika pelanggan berhenti menggunakan layanan atau berpindah ke penyedia layanan lain. Fenomena ini menjadi indikator penting yang dapat memengaruhi keberlangsungan bisnis perusahaan. Tingginya tingkat *customer churn* menyebabkan penurunan pendapatan, meningkatnya biaya pemasaran untuk memperoleh pelanggan pengganti, berkurangnya pangsa pasar, serta menurunnya keuntungan perusahaan dalam jangka panjang. Berbagai faktor dapat memengaruhi terjadinya *customer churn*, seperti kualitas layanan yang kurang memuaskan, harga layanan yang tidak kompetitif, kualitas jaringan yang tidak stabil, munculnya penawaran yang lebih menarik dari kompetitor, maupun perubahan kebutuhan pelanggan. Apabila perusahaan terlambat mengidentifikasi pelanggan yang berpotensi melakukan *churn*, maka peluang untuk mempertahankan pelanggan tersebut akan semakin kecil [5].

Identifikasi pelanggan yang berpotensi melakukan *customer churn* secara manual menjadi semakin sulit dilakukan karena volume data pelanggan yang terus bertambah serta karakteristik data yang semakin kompleks. Analisis secara konvensional umumnya hanya mampu memberikan informasi deskriptif tanpa mampu mengungkap pola hubungan antarvariabel yang bersifat kompleks. Kondisi tersebut menyebabkan perusahaan membutuhkan pendekatan yang mampu melakukan analisis secara otomatis terhadap data historis pelanggan sehingga dapat menghasilkan prediksi yang lebih cepat, objektif, dan akurat. Salah satu pendekatan yang banyak digunakan untuk mengatasi permasalahan tersebut adalah penerapan *Machine Learning*[6]. Dengan memanfaatkan data historis pelanggan, algoritma *Machine Learning* mampu mempelajari pola perilaku pelanggan sehingga dapat mengidentifikasi pelanggan yang memiliki kemungkinan tinggi untuk melakukan *customer churn*. Hasil prediksi tersebut dapat digunakan sebagai dasar penyusunan strategi retensi pelanggan, pemberian promosi yang lebih tepat sasaran, maupun peningkatan kualitas layanan[7].

Berbagai algoritma klasifikasi telah dikembangkan untuk menyelesaikan permasalahan prediksi *customer churn*, seperti Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbor, Artificial Neural Network, Gradient Boosting, Extreme Gradient Boosting (XGBoost), Random Forest, dan Support Vector Machine (SVM). Masing-masing algoritma memiliki karakteristik, kelebihan, dan keterbatasan yang berbeda sehingga performanya dapat berubah bergantung pada karakteristik dataset yang digunakan. Oleh karena itu, pemilihan algoritma yang tepat menjadi salah satu faktor penting dalam menghasilkan model prediksi dengan tingkat akurasi yang tinggi[8].

Random Forest merupakan salah satu metode *ensemble learning* yang bekerja dengan membangun sejumlah *decision tree* menggunakan teknik *bootstrap aggregating* (*bagging*). Hasil prediksi akhir diperoleh melalui mekanisme *majority voting* dari seluruh pohon keputusan yang dibangun. Pendekatan tersebut mampu meningkatkan kemampuan generalisasi model, mengurangi risiko *overfitting*, serta menghasilkan performa klasifikasi yang stabil pada berbagai jenis data. Di sisi lain, Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan mencari *hyperplane* optimal yang memiliki margin maksimum untuk memisahkan dua kelas data. Dengan dukungan berbagai fungsi *kernel*, SVM mampu menangani data yang memiliki hubungan linier maupun nonlinier sehingga sering digunakan pada berbagai permasalahan klasifikasi dengan dimensi data yang tinggi[9].

Selain proses pembangunan model, tahapan eksplorasi data (*Exploratory Data Analysis* atau EDA) juga memiliki peranan penting dalam memahami karakteristik dataset sebelum dilakukan proses klasifikasi. Melalui EDA, peneliti dapat mengidentifikasi distribusi data, pola hubungan antaratribut, keberadaan data kosong (*missing value*), data pencilan (*outlier*), maupun distribusi kelas yang tidak seimbang. Informasi tersebut sangat membantu dalam menentukan strategi *preprocessing* yang sesuai sehingga model yang dihasilkan memiliki performa yang lebih optimal. Dalam penelitian ini, Microsoft Power BI dimanfaatkan sebagai perangkat visualisasi data untuk mengeksplorasi karakteristik pelanggan secara interaktif, sedangkan Orange Data Mining digunakan untuk melakukan proses *preprocessing*, pembangunan model klasifikasi, serta evaluasi performa algoritma[2].

Beberapa penelitian terdahulu telah membahas prediksi *customer churn* menggunakan berbagai algoritma *Machine Learning*. Penelitian pertama menunjukkan bahwa algoritma Random Forest mampu menghasilkan tingkat akurasi yang tinggi karena memanfaatkan pendekatan *ensemble* yang dapat mengurangi risiko *overfitting* serta meningkatkan stabilitas model pada data pelanggan telekomunikasi. Penelitian kedua membandingkan beberapa algoritma klasifikasi dan menemukan bahwa Support Vector Machine memiliki kemampuan yang baik dalam memisahkan pelanggan yang berpotensi *churn* dan *non-churn*, khususnya pada data dengan karakteristik yang kompleks. Penelitian ketiga mengimplementasikan metode Decision Tree, Naïve Bayes, dan Logistic Regression untuk memprediksi *customer churn*, di mana hasil penelitian menunjukkan bahwa performa model sangat dipengaruhi oleh kualitas proses *preprocessing*, pemilihan atribut, serta distribusi data yang digunakan. Penelitian keempat memanfaatkan teknik *ensemble learning* dan *feature selection* untuk meningkatkan performa prediksi, sehingga menghasilkan peningkatan nilai akurasi dan *Area Under Curve* (AUC) dibandingkan metode klasifikasi tunggal. Meskipun demikian, hasil dari berbagai penelitian tersebut masih menunjukkan perbedaan performa karena dipengaruhi oleh karakteristik dataset, teknik transformasi data, proses normalisasi, metode validasi, serta parameter algoritma yang digunakan[10].

Berdasarkan hasil telaah terhadap penelitian-penelitian sebelumnya, masih terdapat beberapa kesenjangan penelitian (*research gap*) yang menjadi dasar dilaksanakannya penelitian ini. Pertama, sebagian besar penelitian hanya berfokus pada evaluasi satu algoritma klasifikasi sehingga belum memberikan gambaran yang objektif mengenai perbandingan performa beberapa algoritma pada dataset yang sama. Kedua, beberapa penelitian lebih menitikberatkan

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

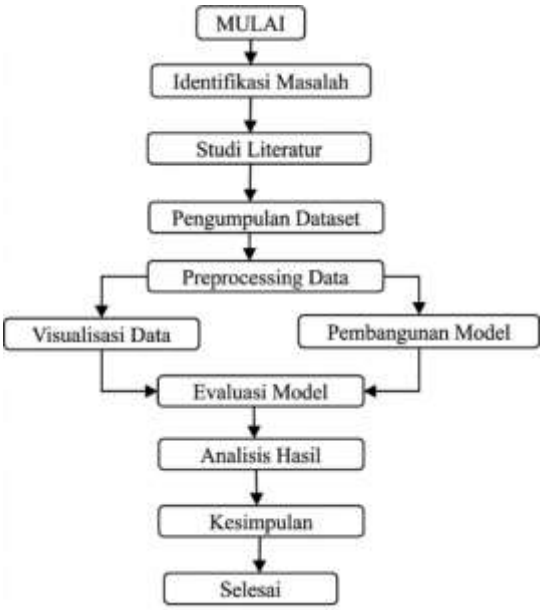
pada peningkatan nilai akurasi tanpa mempertimbangkan metrik evaluasi lain seperti *Precision*, *Recall*, *F1-Score*, dan *Area Under Curve (AUC)* yang penting untuk mengevaluasi kemampuan model secara menyeluruh[1]. Ketiga, masih terbatas penelitian yang mengombinasikan proses eksplorasi data menggunakan Microsoft Power BI dengan proses pembangunan model menggunakan Orange Data Mining dalam satu alur analisis yang terintegrasi. Keempat, sebagian penelitian menggunakan teknik validasi yang berbeda sehingga hasil evaluasi sulit dibandingkan secara langsung. Kelima, belum banyak penelitian yang secara khusus membandingkan Random Forest dan Support Vector Machine pada IBM Telco Customer Churn Dataset dengan menggunakan tahapan *preprocessing* dan metode evaluasi yang identik sehingga hasil perbandingan menjadi lebih objektif[4].

Berdasarkan permasalahan dan kesenjangan penelitian tersebut, penelitian ini bertujuan untuk membandingkan performa algoritma Random Forest dan Support Vector Machine dalam memprediksi customer churn menggunakan IBM Telco Customer Churn Dataset. Penelitian diawali dengan proses eksplorasi data menggunakan Microsoft Power BI untuk memahami karakteristik dataset, kemudian dilanjutkan dengan proses *preprocessing* menggunakan Orange Data Mining yang meliputi pembersihan data, transformasi atribut kategorikal, serta normalisasi data. Selanjutnya dibangun model klasifikasi menggunakan algoritma Random Forest dan Support Vector Machine, kemudian dilakukan evaluasi menggunakan metode 10-Fold Cross Validation dengan metrik Accuracy, Precision, Recall, F1-Score, AUC, dan *Confusion Matrix*. Hasil penelitian diharapkan mampu memberikan rekomendasi mengenai algoritma yang memiliki performa terbaik dalam memprediksi *customer churn* serta menjadi referensi bagi perusahaan telekomunikasi dalam menyusun strategi retensi pelanggan yang lebih efektif dan berbasis data[3].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif untuk membandingkan performa algoritma Random Forest dan Support Vector Machine (SVM) dalam memprediksi customer churn pada perusahaan telekomunikasi. Tahapan penelitian meliputi identifikasi masalah, studi literatur, pengumpulan dataset, preprocessing data, visualisasi data, pembangunan model klasifikasi, evaluasi model, analisis hasil, dan penarikan kesimpulan. Pengolahan data dilakukan menggunakan Orange Data Mining, sedangkan Microsoft Power BI digunakan untuk visualisasi data. Alur tahapan penelitian secara keseluruhan disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, tahapan penelitian dimulai dari identifikasi masalah dan studi literatur untuk memperoleh landasan teori yang relevan. Selanjutnya dilakukan pengumpulan dataset IBM Telco Customer Churn, kemudian dilanjutkan dengan proses preprocessing menggunakan Orange Data Mining. Data yang telah diproses selanjutnya divisualisasikan menggunakan Microsoft Power BI dan digunakan untuk membangun model klasifikasi menggunakan algoritma Random Forest dan Support Vector Machine. Tahap berikutnya adalah evaluasi model menggunakan metode 10-Fold Cross Validation dengan beberapa metrik pengujian, yaitu Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC). Hasil evaluasi kemudian dianalisis untuk menentukan algoritma yang memiliki performa terbaik sehingga dapat ditarik kesimpulan sesuai tujuan penelitian[11].

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

Tahap berikutnya adalah pengumpulan data menggunakan IBM Telco *Customer churn* Dataset yang diperoleh melalui platform Kaggle. Dataset ini mengandung berbagai informasi tentang pelanggan, seperti karakteristik pelanggan, layanan yang digunakan, jenis kontrak, metode pembayaran, biaya layanan, lama berlangganan, dan status churn pelanggan. Dataset ini kemudian disiapkan untuk tahap pengolahan data. Sebelum data diproses, Orange Data Mining digunakan pada tahap berikutnya. Sebelum digunakan dalam pembangunan model, proses ini bertujuan untuk meningkatkan kualitas data. Proses yang dilakukan termasuk memilih atribut yang akan digunakan, menyesuaikan tipe data, menangani data dengan nilai yang hilang, dan mengubah data agar sesuai dengan kebutuhan algoritma klasifikasi. Setelah *preprocessing* selesai, visualisasi data menggunakan Microsoft Power BI dilakukan untuk melihat distribusi dan fitur data pelanggan.

Data yang telah diproses kemudian digunakan pada tahap pembangunan model menggunakan algoritma Random Forest dan Support Vector Machine. Kedua algoritma dilatih menggunakan dataset yang sama sehingga hasil yang diperoleh dapat dibandingkan secara objektif. Selanjutnya dilakukan evaluasi model menggunakan metode 10-Fold Cross Validation dengan beberapa metrik pengujian, yaitu *Accuracy*, Precision, *Recall*, F1-Score, *Area under curve (AUC)*, dan Matthews Correlation Coefficient (MCC)[12]. Pada langkah terakhir, nilai evaluasi digunakan untuk membandingkan kinerja kedua algoritma. Hasil analisis ini digunakan sebagai dasar untuk menentukan algoritma mana yang memiliki kinerja terbaik dalam memprediksi kelangkaan pelanggan. Berdasarkan hasil ini, dibuat kesimpulan yang menjawab tujuan penelitian dan memberikan saran untuk penelitian selanjutnya.

2.2 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah IBM Telco *Customer churn* Dataset, yaitu dataset publik yang banyak dimanfaatkan dalam penelitian klasifikasi *customer churn*[13]. Dataset ini diperoleh melalui platform Kaggle dan berisi data pelanggan perusahaan telekomunikasi yang mencakup informasi mengenai karakteristik pelanggan, layanan yang digunakan, jenis kontrak, metode pembayaran, lama berlangganan (tenure), biaya bulanan (MonthlyCharges), total biaya (TotalCharges), serta status pelanggan (Churn). Data tersebut digunakan sebagai dasar dalam proses pembangunan dan evaluasi model klasifikasi[14].

Dataset yang digunakan terdiri dari 7.043 data pelanggan yang memiliki kombinasi atribut numerik dan kategorikal. Variabel target penelitian ini adalah Churn, yang memiliki dua kelas: Yes dan No. Pelanggan dalam kelas Yes menunjukkan bahwa mereka telah berhenti menggunakan layanan, sedangkan pelanggan dalam kelas No menunjukkan bahwa mereka masih menggunakan layanan. Dataset ini memiliki karakteristik yang membuatnya sesuai untuk digunakan dalam metode klasifikasi yang menggunakan algoritma Random Forest dan Support Vector Machine. Untuk memberikan gambaran umum mengenai dataset yang digunakan pada penelitian ini, karakteristik utama dataset dirangkum pada Tabel 1. Informasi tersebut meliputi nama dataset, sumber data, jumlah data, jumlah atribut, variabel target, jumlah kelas, jenis data, serta perangkat lunak yang digunakan dalam proses analisis dan visualisasi.

Tabel 1. Karakteristik Dataset Penelitian

Karakteristik	Keterangan
Nama Dataset	IBM Telco *Customer churn* Dataset
Sumber Dataset	IBM Sample Data (Kaggle)
Jumlah Data	7.043 data pelanggan
Jumlah Atribut	21 atribut
Variabel Target	Churn
Jumlah Kelas	2 (Yes dan No)
Jenis Data	Numerik dan Kategorikal
Platform Analisis	Orange Data Mining
Platform Visualisasi	Power BI

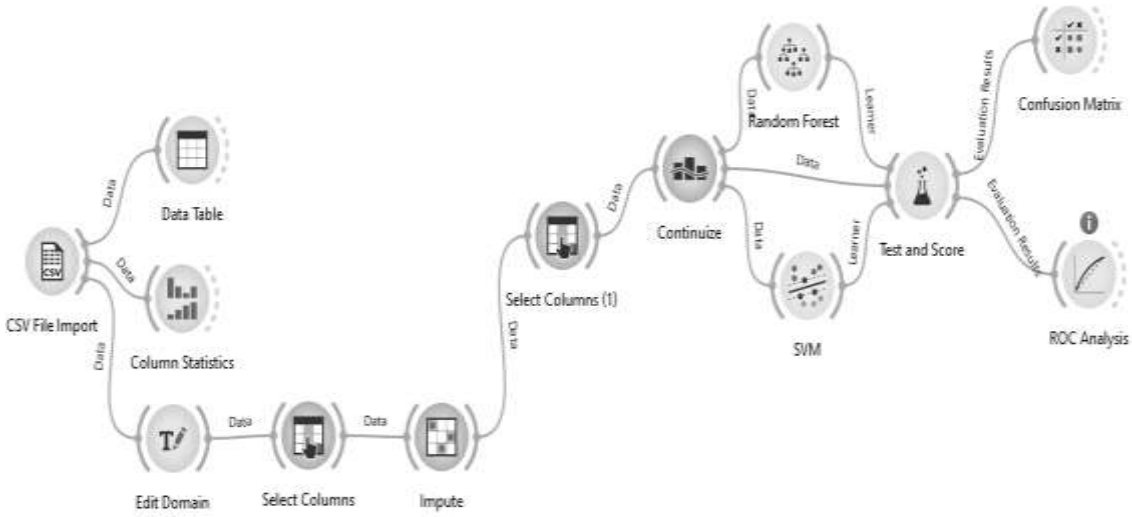
Berdasarkan Tabel 1, dataset penelitian terdiri atas 7.043 data pelanggan dengan 21 atribut yang mencakup karakteristik pelanggan, layanan yang digunakan, jenis kontrak, metode pembayaran, biaya layanan, dan status pelanggan. Variabel target yang digunakan adalah Churn yang terdiri atas dua kelas, yaitu Yes dan No. Kombinasi atribut numerik dan kategorikal memungkinkan algoritma Random Forest dan Support Vector Machine membangun model prediksi yang dapat dibandingkan secara objektif.

2.3 Prepocessing Data

Sebelum klasifikasi dilakukan, tahap *preprocessing* dilakukan. Tujuan dari tahap ini adalah untuk meningkatkan kualitas data sehingga data yang digunakan lebih konsisten, lengkap, dan sesuai dengan kebutuhan algoritma pembelajaran mesin[15]. Perangkat lunak Orange Data Mining digunakan untuk melakukan seluruh proses *preprocessing* dalam penelitian ini. Sebelum klasifikasi dilakukan, proses ini mencakup penyesuaian tipe data, pemilihan atribut, penanganan nilai yang hilang, dan transformasi data. Untuk memastikan bahwa model yang dibangun dapat berfungsi dengan baik, tahapan ini sangat penting[16]. Pada langkah pertama, tipe data disesuaikan dengan widget Edit Domain. Pada tahap ini, setiap atribut diperiksa untuk memastikan tipe data telah sesuai dengan atributnya. Variabel Churn digunakan sebagai variabel target (class), sedangkan atribut lainnya digunakan sebagai variabel prediktor.

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

Setelah itu, widget Select Columns digunakan untuk memilih atribut yang akan digunakan selama proses klasifikasiSelanjutnya, widget Impute digunakan untuk menangani nilai kosong, tahap berikutnya adalah transformasi data menggunakan widget Continuize untuk memastikan bahwa seluruh data dapat digunakan tanpa mengurangi jumlah data yang tersedia. Tujuan dari tahap transformasi ini adalah untuk mengubah atribut kategorikal menjadi bentuk numerik sehingga algoritma Random Forest dan Support Vector machine dapat memprosesnya. Kemudian, hasil dari seluruh tahapan *preprocessing* dimasukkan ke dalam proses pembangunan model klasifikasi. Selanjutnya, data yang telah diproses digunakan untuk melatih dan menguji algoritma Random Forest dan Support Vector Machine. Diharapkan bahwa tahapan *preprocessing* yang dilakukan mampu meningkatkan kualitas data sehingga proses klasifikasi dapat melakukan pekerjaan dengan lebih baik.



Gambar 2. *Preprocessing* Data dan Pembangunan Model Menggunakan Orange Data Mining

Gambar 2 menunjukkan alur pengolahan data yang dilakukan menggunakan Orange Data Mining. Proses dimulai dari pembacaan dataset menggunakan widget CSV File Import, kemudian dilakukan penyesuaian tipe data melalui Edit Domain, pemilihan atribut menggunakan Select Columns, penanganan missing value dengan Impute, serta transformasi data menggunakan Continuize. Data yang telah diproses selanjutnya digunakan untuk membangun model klasifikasi menggunakan algoritma Random Forest dan Support Vector Machine. Kinerja kedua model kemudian dievaluasi menggunakan Test and Score, sedangkan hasil klasifikasi dianalisis lebih lanjut melalui *Confusion matrix* dan ROC Analysis.

2.4 Metode Klasifikasi

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu Random Forest dan Support Vector Machine (SVM).[7] Kedua algoritma dipilih karena memiliki karakteristik yang berbeda dalam membangun model klasifikasi.[4] Random Forest dikenal mampu menangani data berdimensi tinggi serta mengurangi risiko overfitting,[6] sedangkan Support Vector Machine memiliki kemampuan yang baik dalam memisahkan data menggunakan *hyperplane* optimal.[7] Perbandingan kedua algoritma dilakukan untuk mengetahui metode yang memberikan performa terbaik dalam memprediksi *customer churn* pada perusahaan telekomunikasi.

2.4.1 Random Forest

Random Forest merupakan algoritma *ensemble learning* yang bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dari sampel data yang dipilih secara acak[12]. Setiap pohon menghasilkan prediksi, kemudian hasil akhir ditentukan berdasarkan suara mayoritas (*majority voting*). Pendekatan ini mampu meningkatkan akurasi klasifikasi serta mengurangi risiko *overfitting* yang sering terjadi pada penggunaan satu pohon keputusan.

$$mode\left(T_{1(x)}, T_{2(x)}, \dots, T_{n(x)}\right) \quad (1)$$

Model Random Forest menghasilkan prediksi akhir yang dinyatakan dengan $\hat{y}$, yaitu *output prediction* yang diperoleh melalui mekanisme *majority voting*. Setiap *decision tree* yang dibangun dalam model akan menghasilkan prediksi masing-masing, yang dinyatakan sebagai $T_1(x), T_2(x), \dots, T_n(x)$. Jumlah keseluruhan *decision tree* yang digunakan dalam proses pembentukan model dinyatakan dengan n. Selanjutnya, seluruh hasil prediksi dari setiap *decision tree* dikombinasikan menggunakan fungsi mode, yaitu fungsi yang memilih nilai atau kelas yang paling sering muncul (*majority voting*), sehingga diperoleh prediksi akhir $\hat{y}$sebagai hasil klasifikasi model Random Forest. Berdasarkan Persamaan (1), mekanisme *majority voting* digunakan oleh algoritma Random Forest untuk membuat keputusan akhir.

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

Metode ini dapat meningkatkan akurasi klasifikasi dan mengurangi kemungkinan overfitting jika dibandingkan dengan pendekatan yang menggunakan satu *decision tree*.

2.4.2 Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu algoritma *Machine Learning* berbasis *supervised learning* yang banyak digunakan untuk menyelesaikan permasalahan klasifikasi[7]. Algoritma ini bekerja dengan membentuk *hyperplane* optimal yang berfungsi sebagai batas pemisah antar kelas sehingga mampu memaksimalkan jarak (margin) antara data dari kelas yang berbeda. Pendekatan tersebut memungkinkan SVM memiliki kemampuan generalisasi yang baik, terutama pada data berdimensi tinggi dan data dengan karakteristik yang kompleks. Karena karakteristik tersebut, algoritma SVM banyak diterapkan dalam berbagai penelitian klasifikasi, termasuk prediksi *customer churn* pada perusahaan telekomunikasi.

$$f(x) = W^T x + b \tag{2}$$

Pada metode Support Vector Machine (SVM), fungsi keputusan dinyatakan sebagai $f(x)$, yaitu fungsi yang digunakan untuk menentukan kelas dari suatu data berdasarkan posisi data terhadap *hyperplane*. Fungsi tersebut dibentuk oleh vektor bobot w, yang berperan dalam menentukan arah atau orientasi *hyperplane*, serta vektor data masukan x, yang merepresentasikan atribut atau fitur dari setiap data yang akan diklasifikasikan. Selain itu, terdapat nilai bias $b$yang berfungsi untuk menggeser posisi *hyperplane* sehingga pemisahan antar kelas dapat dilakukan secara optimal. Kombinasi antara w, x, dan $b$menghasilkan nilai fungsi keputusan $f(x)$, yang selanjutnya digunakan untuk menentukan kelas prediksi suatu data. Berdasarkan Persamaan (2), algoritma Support Vector Machine menentukan kelas suatu data berdasarkan posisi data terhadap *hyperplane* yang dibentuk. Tujuan utama algoritma ini adalah memperoleh *hyperplane* dengan margin terbesar sehingga kemampuan klasifikasi terhadap data baru menjadi lebih baik. Karakteristik tersebut SVM mampu menghasilkan model klasifikasi yang memiliki tingkat akurasi dan kemampuan generalisasi yang baik pada berbagai permasalahan klasifikasi.[7]

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengujian Model

Bab ini menyajikan hasil pengujian dan pembahasan terhadap performa algoritma Random Forest dan Support Vector Machine (SVM) dalam memprediksi customer churn menggunakan IBM Telco Customer Churn Dataset. Dataset yang telah melalui tahap preprocessing menggunakan Orange Data Mining selanjutnya digunakan untuk membangun model klasifikasi. Evaluasi performa dilakukan menggunakan metrik Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC). Selain itu, hasil klasifikasi dianalisis melalui Confusion Matrix, kurva Receiver Operating Characteristic (ROC), serta didukung dengan visualisasi data menggunakan Microsoft Power BI untuk memperkuat interpretasi hasil penelitian.

Sebelum dilakukan pembahasan lebih lanjut, dilakukan pengujian terhadap model klasifikasi yang dibangun menggunakan algoritma Random Forest dan Support Vector Machine (SVM). Pengujian bertujuan untuk membandingkan kemampuan kedua algoritma dalam memprediksi pelanggan yang berpotensi melakukan *customer churn* berdasarkan data yang telah melalui proses *preprocessing*. Evaluasi dilakukan menggunakan beberapa metrik kinerja, yaitu Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC). Hasil evaluasi tersebut disajikan pada Tabel 2 sebagai dasar untuk menganalisis tingkat akurasi, kemampuan klasifikasi, serta efektivitas masing-masing algoritma dalam mengidentifikasi pelanggan yang berpotensi berhenti menggunakan layanan.

Tabel 2. Hasil Evaluasi Model

Algoritma	AUC	*Accuracy*	F1-Score	Precision	*Recall*	MCC
Random Forest	0.832	0.796	0.787	0.785	0.796	0.442
SVM	0.575	0.661	0.659	0.657	0.661	0.121

Berdasarkan Tabel 2, algoritma Random Forest menunjukkan performa yang lebih baik dibandingkan Support Vector Machine (SVM) pada hampir seluruh metrik evaluasi. Nilai Accuracy yang lebih tinggi menunjukkan bahwa Random Forest memiliki tingkat ketepatan klasifikasi yang lebih baik dalam membedakan pelanggan yang berpotensi melakukan customer churn dan pelanggan yang tetap menggunakan layanan. Selain itu, nilai Precision, Recall, dan F1-Score yang lebih tinggi mengindikasikan bahwa algoritma Random Forest lebih efektif dalam mengenali pelanggan yang benar-benar berpotensi melakukan customer churn dengan tingkat kesalahan klasifikasi yang lebih rendah dibandingkan SVM. Hasil tersebut menunjukkan bahwa pendekatan ensemble learning yang digunakan oleh Random Forest mampu meningkatkan kemampuan model dalam mengenali pola yang kompleks pada data pelanggan. Sebaliknya, performa Support Vector Machine dipengaruhi oleh karakteristik data dan parameter model sehingga menghasilkan nilai evaluasi yang sedikit lebih rendah. Temuan ini menunjukkan bahwa Random Forest lebih sesuai diterapkan pada IBM Telco Customer Churn Dataset yang digunakan dalam penelitian ini.

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

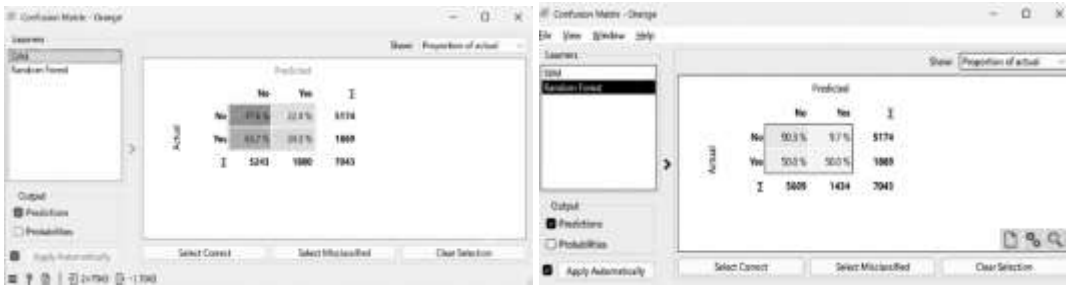
Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0.575	0.661	0.659	0.657	0.661	0.121
Random Forest	0.832	0.796	0.787	0.785	0.796	0.442

Gambar 3. Hasil Evaluasi Model Menggunakan Test and Score

Berdasarkan Gambar 3, algoritma Random Forest mampu mengklasifikasikan sebagian besar data pelanggan dengan benar, baik pada kelas customer churn maupun non-customer churn. Jumlah prediksi yang benar lebih besar dibandingkan prediksi yang salah, sehingga menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik. Kesalahan klasifikasi masih ditemukan pada sebagian kecil data, namun jumlahnya relatif rendah sehingga tidak memberikan pengaruh yang signifikan terhadap performa model secara keseluruhan. Hasil tersebut sejalan dengan nilai Accuracy, Precision, Recall, dan F1-Score yang diperoleh pada Tabel 2.

3.2 Analisis *Confusion Matrix*

Confusion matrix digunakan untuk mengevaluasi kemampuan model klasifikasi dalam mengidentifikasi kelas aktual dan kelas hasil prediksi sehingga tingkat kesalahan klasifikasi dapat dianalisis secara lebih rinci. Jumlah prediksi yang benar dan salah untuk masing-masing kelas dapat dihitung dengan menggunakan *Confusion matrix*. Informasi ini menunjukkan tingkat ketepatan model dalam membedakan pelanggan yang terus menggunakan layanan dari pelanggan yang berpotensi melakukan churn.



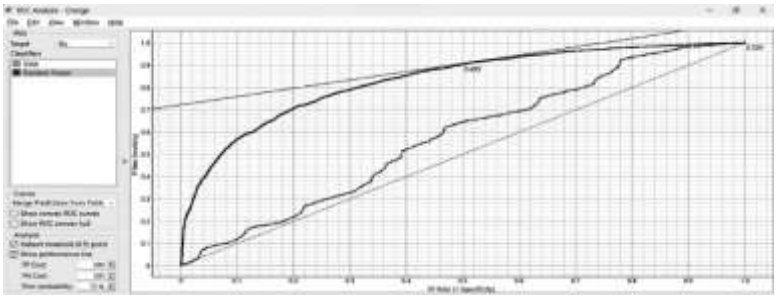
(a) Support Vector Machine (b) Random Forest

Gambar 4. Hasil *Confusion Matrix* Algoritma Random Forest dan Support Vector Machine

Berdasarkan Gambar 4, algoritma Random Forest menghasilkan jumlah prediksi yang benar lebih tinggi dibandingkan Support Vector Machine pada kedua kelas data. Sebaliknya, Support Vector Machine masih menghasilkan jumlah kesalahan klasifikasi yang lebih besar, terutama pada data pelanggan yang melakukan *customer churn*. Hasil tersebut menunjukkan bahwa Random Forest memiliki kemampuan yang lebih tinggi dalam mengenali pola data pelanggan pada dataset penelitian ini. Berdasarkan hasil *Confusion matrix*, kedua algoritma menunjukkan kemampuan yang berbeda dalam mengklasifikasikan data *customer churn*. Random Forest memiliki proporsi prediksi yang lebih baik dibandingkan Support Vector Machine pada kelas No. Namun, kedua algoritma mengalami kesalahan klasifikasi pada kelas Yes, tetapi Random Forest mampu memberikan hasil yang lebih seimbang dibandingkan Support Vector Machine. Hal ini menunjukkan bahwa Random Forest memiliki kemampuan yang lebih baik dalam mengenali pola data dalam penelitian ini.

3.3 Analisis Receiver Operating Characteristic (ROC)

Kurva *Receiver operating characteristic (ROC)* digunakan untuk mengevaluasi kemampuan model klasifikasi dalam membedakan kelas positif dan kelas negatif. Hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) digambarkan pada berbagai nilai ambang. Semakin dekat dengan sudut kiri atas dan nilai *Area under curve (AUC)*, semakin baik kemampuan model untuk klasifikasi.



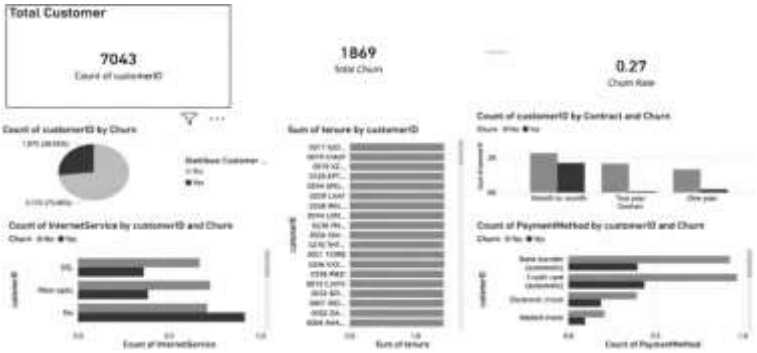
Gambar 5. Kurva ROC *Algoritma* Support Vector Machine dan Random Forest

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

Berdasarkan Gambar 5, algoritma Random Forest menghasilkan kurva ROC yang berada lebih dekat ke sudut kiri atas dibandingkan Support Vector Machine. Hal tersebut menunjukkan bahwa Random Forest memiliki kemampuan yang lebih baik dalam membedakan pelanggan yang berpotensi melakukan customer churn dan pelanggan yang tidak melakukan customer churn. Nilai Area Under Curve (AUC) sebesar 0,832 mengindikasikan bahwa model memiliki kemampuan klasifikasi dalam kategori baik. Sebaliknya, kurva ROC yang dihasilkan oleh Support Vector Machine berada di bawah Random Forest, sehingga menunjukkan kemampuan diskriminasi kelas yang lebih rendah. Hasil analisis ROC ini konsisten dengan hasil evaluasi menggunakan Accuracy, Precision, Recall, F1-Score, dan Matthews Correlation Coefficient (MCC), yang menunjukkan bahwa algoritma Random Forest memberikan performa terbaik pada dataset penelitian ini.

3.4 Visualisasi Data Menggunakan Power BI

Visualisasi data menggunakan Microsoft Power BI dilakukan untuk menyajikan informasi hasil analisis dalam bentuk grafik dan diagram yang lebih mudah dipahami. Penyajian visual membantu menggambarkan karakteristik pelanggan, distribusi data, serta pola yang berkaitan dengan *customer churn*. Dengan demikian, hasil penelitian tidak hanya ditampilkan dalam bentuk angka, tetapi juga didukung oleh visualisasi yang lebih informatif.



Gambar 6. Visualisasi Data *Customer churn* Menggunakan Microsoft Power BI

Gambar 6 menunjukkan dashboard visualisasi data yang dibangun menggunakan Microsoft Power BI. Dashboard tersebut menyajikan informasi mengenai distribusi pelanggan berdasarkan status customer churn, jenis kontrak, metode pembayaran, masa berlangganan (tenure), layanan internet, serta biaya layanan yang digunakan pelanggan. Penyajian data dalam bentuk grafik dan diagram memudahkan proses eksplorasi data dibandingkan penyajian dalam bentuk tabel karena hubungan antarvariabel dapat diamati secara lebih jelas.

Hasil visualisasi menunjukkan bahwa beberapa karakteristik pelanggan memiliki hubungan dengan status customer churn. Misalnya, pelanggan dengan masa berlangganan yang lebih singkat dan jenis kontrak bulanan cenderung memiliki tingkat customer churn yang lebih tinggi dibandingkan pelanggan dengan kontrak jangka panjang. Informasi tersebut mendukung hasil klasifikasi yang diperoleh dari algoritma Random Forest dan Support Vector Machine sehingga proses interpretasi hasil penelitian menjadi lebih komprehensif. Dengan demikian, Microsoft Power BI tidak hanya berfungsi sebagai media visualisasi, tetapi juga membantu proses analisis pola data sebelum dan sesudah pembangunan model klasifikasi.

3.5 Pembahasan

Berdasarkan hasil evaluasi model, algoritma Random Forest menunjukkan performa yang lebih unggul dibandingkan Support Vector Machine (SVM) dalam memprediksi customer churn pada IBM Telco Customer Churn Dataset. Keunggulan tersebut terlihat secara konsisten pada seluruh metrik evaluasi, yaitu Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC). Hasil tersebut menunjukkan bahwa Random Forest memiliki kemampuan yang lebih baik dalam mengenali pola hubungan antaratribut pelanggan sehingga mampu menghasilkan prediksi yang lebih akurat dibandingkan SVM.

Keunggulan Random Forest tidak terlepas dari mekanisme kerja algoritma yang memanfaatkan pendekatan *ensemble learning*. Algoritma ini membangun sejumlah *decision tree* dari sampel data yang berbeda, kemudian menggabungkan hasil prediksi setiap pohon menggunakan mekanisme *majority voting*. Pendekatan tersebut memungkinkan model menghasilkan prediksi yang lebih stabil serta mengurangi risiko overfitting. Pada penelitian ini, karakteristik dataset yang terdiri atas berbagai atribut pelanggan, seperti masa berlangganan, jenis kontrak, metode pembayaran, layanan internet, dan biaya bulanan, dapat dipelajari dengan baik oleh Random Forest. Oleh karena itu, algoritma ini mampu menghasilkan prediksi yang lebih akurat dibandingkan SVM. Karakteristik tersebut membuat Random Forest lebih adaptif terhadap data dengan kombinasi atribut numerik dan kategorikal seperti yang terdapat pada IBM Telco Customer Churn Dataset. Sebaliknya, algoritma Support Vector Machine juga mampu melakukan klasifikasi dengan baik, namun performanya masih berada di bawah Random Forest. Berdasarkan hasil pengujian, SVM masih menghasilkan kesalahan klasifikasi pada beberapa data pelanggan. Kondisi tersebut menunjukkan bahwa proses

Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

pembentukan *hyperplane* belum sepenuhnya mampu memisahkan data pelanggan yang melakukan *customer churn* dengan pelanggan yang tetap menggunakan layanan. Selain itu, performa SVM sangat dipengaruhi oleh pemilihan fungsi kernel dan parameter model. Apabila parameter yang digunakan belum optimal, kemampuan model dalam mengklasifikasikan data juga akan menurun.

Hasil analisis Confusion Matrix semakin memperkuat temuan tersebut. Algoritma Random Forest menghasilkan jumlah prediksi benar (True Positive dan True Negative) yang lebih tinggi dibandingkan Support Vector Machine. Sebaliknya, jumlah kesalahan klasifikasi (False Positive dan False Negative) pada Random Forest lebih rendah dibandingkan SVM. Kondisi ini menunjukkan bahwa Random Forest lebih mampu mengenali karakteristik pelanggan yang berpotensi melakukan customer churn maupun pelanggan yang tetap menggunakan layanan, sehingga tingkat kesalahan prediksi dapat diminimalkan. Hasil evaluasi menggunakan kurva Receiver Operating Characteristic (ROC) juga mendukung temuan tersebut. Nilai Area Under Curve (AUC) yang lebih tinggi pada algoritma Random Forest menunjukkan kemampuan model yang lebih baik dalam membedakan kelas pelanggan customer churn dan non-customer churn. Hal ini mengindikasikan bahwa Random Forest memiliki kemampuan diskriminasi yang lebih tinggi dibandingkan SVM sehingga lebih andal digunakan untuk proses prediksi pada dataset penelitian ini.

Selain pengujian menggunakan Orange Data Mining, penelitian ini juga memanfaatkan Microsoft Power BI untuk menyajikan hasil visualisasi data. Dashboard yang telah dibuat memberikan gambaran mengenai karakteristik pelanggan berdasarkan beberapa atribut, seperti status *customer churn*, jenis kontrak, metode pembayaran, masa berlangganan, serta layanan yang digunakan pelanggan. Penyajian informasi dalam bentuk grafik mempermudah proses interpretasi data dan membantu mengidentifikasi pola-pola yang sulit diamati apabila hanya disajikan dalam bentuk tabel. Dengan demikian, visualisasi menggunakan Power BI tidak hanya berfungsi sebagai pelengkap hasil penelitian, tetapi juga mendukung proses analisis secara keseluruhan.

Hasil penelitian ini sejalan dengan berbagai penelitian terdahulu yang menyatakan bahwa Random Forest memiliki kemampuan klasifikasi yang baik pada data dengan jumlah atribut yang banyak dan karakteristik yang beragam.[4], [7] Kemampuan algoritma dalam mengombinasikan banyak *decision tree* menjadikannya lebih stabil dan mampu menghasilkan prediksi yang akurat pada berbagai kasus klasifikasi, termasuk prediksi *customer churn*. Meskipun demikian, SVM tetap memiliki kelebihan, terutama pada dataset dengan dimensi tinggi dan pola pemisahan kelas yang jelas. Oleh karena itu, pemilihan algoritma tetap perlu disesuaikan dengan karakteristik data yang digunakan agar diperoleh hasil klasifikasi yang optimal.

Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma *Random Forest* memberikan performa yang lebih baik dibandingkan *Support Vector Machine* dalam memprediksi *customer churn* pada perusahaan telekomunikasi. Keunggulan tersebut terlihat secara konsisten pada seluruh tahapan evaluasi, mulai dari *Test and Score*, *Confusion Matrix*, hingga kurva ROC. Selain menghasilkan model klasifikasi dengan performa yang lebih baik, penelitian ini juga menunjukkan bahwa pemanfaatan Orange Data Mining dan Microsoft Power BI dapat mendukung proses analisis data secara lebih sistematis dan mudah diinterpretasikan. Oleh karena itu, hasil penelitian ini diharapkan dapat menjadi referensi bagi perusahaan telekomunikasi dalam memilih metode klasifikasi yang tepat untuk mendukung penyusunan strategi retensi pelanggan secara lebih efektif.

4. KESIMPULAN

Penelitian ini berhasil membandingkan performa algoritma Random Forest dan Support Vector Machine (SVM) dalam memprediksi customer churn pada perusahaan telekomunikasi menggunakan IBM Telco Customer Churn Dataset. Seluruh tahapan penelitian telah dilakukan secara sistematis, mulai dari proses preprocessing data menggunakan Orange Data Mining, pembangunan model klasifikasi, evaluasi performa menggunakan metode 10-Fold Cross Validation, hingga visualisasi data menggunakan Microsoft Power BI. Berdasarkan hasil evaluasi terhadap metrik Accuracy, Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC), algoritma Random Forest menunjukkan performa yang lebih baik dibandingkan Support Vector Machine. Hasil tersebut menunjukkan bahwa pendekatan ensemble learning pada Random Forest mampu mengenali pola hubungan antaratribut pelanggan secara lebih efektif sehingga menghasilkan kemampuan klasifikasi yang lebih baik pada dataset yang digunakan. Temuan ini juga diperkuat oleh hasil analisis Confusion Matrix dan kurva Receiver Operating Characteristic (ROC) yang memperlihatkan kemampuan diskriminasi kelas yang lebih tinggi pada algoritma Random Forest. Selain itu, pemanfaatan Microsoft Power BI memberikan dukungan terhadap proses analisis melalui penyajian visualisasi yang mempermudah identifikasi karakteristik pelanggan berdasarkan atribut yang berkaitan dengan customer churn. Dengan demikian, penelitian ini menunjukkan bahwa algoritma Random Forest merupakan metode yang lebih efektif dibandingkan Support Vector Machine untuk memprediksi customer churn pada perusahaan telekomunikasi serta dapat dijadikan sebagai alternatif dalam mendukung pengambilan keputusan dan penyusunan strategi retensi pelanggan berdasarkan hasil analisis data.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas HKBP Nommensen Pematangsiantar atas dukungan yang diberikan selama proses penelitian hingga

Vol. 4, No. 2, Juli 2026, Hal 143 - 152
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.377
https://journal.grahamitra.id/index.php/jurikti

penyusunan artikel ini. Penulis juga menyampaikan apresiasi kepada semua pihak yang telah memberikan bantuan dan masukan sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENCES

[1] N. Januari, I. K. M. Bili, I. W. Sudiarta, M. Yuditia, N. K. A. Rosdiana, and P. Rafiana, "Analisis dan Prediksi Customer Churn pada Platform Streaming Berbasis Langganan Menggunakan Metode Random Forest membicarakan suatu layanan setelah jangka waktu tertentu . Dalam dunia bisnis , churn layanan . Data menunjukkan durasi akses , frekuensi penggunaan , tingkat konsumsi , dan layanan . Mengumpulkan data pengguna yang memengaruhi penggunaan layanan . Data ini," no. November 2025, 2026.

[2] D. M. Putri, R. D. Mulyani, and F. Mawarni, "Pemanfaatan Data Mining untuk Klasifikasi Customer Churn Menggunakan Algoritma Random Forest dalam Mendukung Strategi Bisnis," *J. Data Sci. Informatics Eng.*, vol. 1, no. 2, pp. 63–68, 2026, doi: 10.64803/jodsie.v1i2.32.

[3] H. D. Syaputra, M. S. Wisnubroto, F. Farid, H. S. Ramadhan, and D. R. Luthfi, "Penerapan Metode Ensemble Learning Untuk Prediksi Churn Customer Pada Layanan Telekomunikasi," *J. RESTIKOM Ris. Tek. Inform. dan Komput.*, vol. 8, no. 1, pp. 515–525, 2026.

[4] M. I. Fauzi, N. Fitriyah, and M. Karimah, "Prediksi Customer Churn Menggunakan Decision Tree dan Random Forest dengan Pendekatan SMOTE untuk Mendukung Customer Intelligence pada Industri Telekomunikasi," *J. Inf. Syst. Bus. Technol.*, vol. 2, no. 3, pp. 785–794, 2026.

[5] A. Sikri, R. Jameel, S. M. Idrees, and H. Kaur, "Enhancing customer retention in telecom industry with machine learning driven churn prediction," *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-63750-0.

[6] M. Z. Alotaibi and M. A. Haq, "Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 3, pp. 14572–14578, 2024, doi: 10.48084/etasr.7480.

[7] Oktaviana Putri Agung, Fairuza Mayla Faizal, and Irenia Mascharenhas, "Penerapan Algoritma Random Forest Dalam Prediksi Customer Churn Untuk Mendukung Strategi Retensi Pelanggan," *J. Ris. Tek. Komput.*, vol. 3, no. 2, pp. 90–96, 2026, doi: 10.69714/84q65g10.

[8] L. N. Wakhidah, A. K. Zyen, and B. B. Wahono, "Evaluation of Telecommunication Customer Churn Classification with SMOTE Using Random Forest and XGBoost Algorithms," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 89–95, 2025, doi: 10.30871/jaic.v9i1.8740.

[9] J. I. Komputer *et al.*, "Perbandingan Algoritma Random Forest dan Support Vector Machine dalam Memprediksi Customer Churn pada Perusahaan Telekomunikasi," vol. 9, no. 99, pp. 1–10.

[10] M. Basri, "A Comparative Study : Predicting Customer Churn in Banking Using Logistic Regression & Random Forest," *Ultim. J. Tek. Inform.*, vol. 17, no. 1, pp. 72–81, 2025, doi: 10.31937/ti.v17i1.4075.

[11] N. Y. Nhu, T. Van Ly, and D. V. Truong Son, "Churn prediction in telecommunication industry using kernel Support Vector Machines," *PLoS One*, vol. 17, no. 5 May, 2022, doi: 10.1371/journal.pone.0267935.

[12] D. A. Kusuma, A. R. Dewi, and A. R. Wijaya, "Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 10, no. 2, pp. 186–194, 2025, doi: 10.14421/jiska.2025.10.2.186-194.

[13] Y. Setiawan, A. I. Hadiana, and F. R. Umbara, "Customer Churn Prediction Using the Random Forest Algorithm," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 3, pp. 209–216, 2024, doi: 10.33387/jiko.v7i3.8711.

[14] Aqilla Nurul Hasanah Aqilla and Hafiyyan Putra Pratama, "Klasifikasi Spesies Bunga Iris Menggunakan Logistic Regression Dan Support Vector Machine," *J. Komput. Teknol. Inf. Sist. Komput.*, vol. 5, no. 1, pp. 704–713, 2026, doi: 10.62712/juktisi.v5i1.1096.

[15] P. Meilina, "Penerapan Data Mining dengan Metode Klasifikasi," *J. Teknol. Univ. Muhammadiyah Jakarta*, vol. 7, no. 1, pp. 11–20, 2015, [Online]. Available: jurnal.ftumj.ac.id/index.php/jurtek

[16] M. D. N. Alif and N. F. Fahrudin, "Performance Analysis of Oversampling and Undersampling on Telco Churn Data Using Naive Bayes, SVM And Random Forest Methods," *E3S Web Conf.*, vol. 484, 2024, doi: 10.1051/e3sconf/202448402004.