ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

PREDIKSI NASABAH CHURN DENGAN ALGORITMA DECISION TREE, RANDOM FOREST DAN SUPPORT VECTOR MACHINE

Namira[1*], Isnandar Slamet[2], Irwan Susanto[3]
[1]Program Studi Statistika, Universitas Sebelas Maret, Surakarta
e-mail: *[1]namiramirra@student.uns.ac.id, [2]isnandarslamet@staff.uns.ac.id,
[3]irwansusanto@staff.uns.ac.id

Abstrak

Persaingan bank tidak luput dari perkembangan zaman. Masing masing bank akan memberikan fasilitas-fasilitas terbaiknya untuk tetap menjaga para nasabahnya dan meminimalisir nasabahnya untuk pindah ke bank kompetitor. Nasabah bank yang sebelumnya telah memiliki akun bank dan rekening yang aktif pada bank, memutuskan untuk berhenti dan beralih ke bank lain disebut dengan nasabah *churn*. Kemampuan bank untuk menjaga para nasabah sangat penting untuk dilakukan agar mempertahankan kestabilan dan profitabilitas bank. Penelitian ini dilakukan untuk mengidentifikasi pola perilaku nasabah guna mengetahui nasabah yang berpotensi *churn* dan tidak, sehingga dapat dilakukan pencegahan lebih awal. Penentuan proses klasifikasi menggunakan proses *data mining* dengan metode *decision tree, random forest* dan *support vector machine*. Objek penelitian menggunakan 10000 *record* data dengan variabel dependen berupa status *churn* nasabah dan variabel independen yang digunakan adalah *customer id, credit score, country, gender, age, tenure, balance, product number, credit card, active member* dan *estimated salary*. Hasil klasifikasi menunjukan bahwa metode *random forest* merupakan metode terbaik yang mampu memprediksi status *churn* nasabah. Diperoleh nilai akurasi dari *random forest* sebesar 83,86%, nilai presisi sebesar 85,82%, nilai *recall* sebesar 81,37% dan F1-*score* sebesar 83,54%.

Kata kunci — Nasabah *churn*; Klasifikasi; *Decision Tree*; *Random Forest*; *Support Vector Machine*.

Abstract

Bank competition has not escaped the times. Each bank will provide its best facilities to keep its customers and minimize their customers from moving to competing banks. Bank customers who previously had a bank account and an active account at the bank, decided to quit and switch to another bank is called customer churn. The bank's ability to keep customers is very important to do in order to maintain the stability and profitability of the bank. This research is conducted to identify patterns of customer behavior to find out which customers have the potential to churn and not, so that early prevention can be done. Determination of the classification process using data mining process with decision tree, random forest and support vector machine methods. The research object uses 10000 data records with the dependent variable being customer churn status and the independent variables used are customer id, credit score, country, gender, age, tenure, balance, product number, credit card, active member and estimated salary. The classification results show that the random forest method is the best method that can predict customer churn status. The accuracy value of random forest is 83.86%, precision value is 85.82%, recall value is 81.37% and F1-score is 83.54%.

Keywords—*Nasabah churn, classification, Decision Tree, Random Forest, Support Vector Machine.*

I. PENDAHULUAN

Perkembangan zaman yang dibersamai oleh perkembangan teknologi saat ini telah berkembang pesat terlebih pada revolusi industri 4.0 yang juga mempengaruhi pola hidup masyarakat dan kebutuhannya yang semakin tinggi. Untuk menghadapi permasalahan tersebut banyak

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

pihak yang memberikan fasilitas terbaiknya untuk tetap menunjang kebutuhan masyarakat, salah satunya adalah pihak perbankan. Persaingan bank juga tidak luput dari perkembangan zaman ini. Masing masing bank akan memberikan fasilitas-fasilitas terbaiknya untuk tetap menjaga para nasabahnya dan meminimalisir nasabahnya untuk pindah ke bank kompetitor. Kemampuan bank untuk menjaga para nasabah sangat penting untuk dilakukan agar mempertahankan kestabilan dan profitabilitas bank.

Perilaku pelanggan bank yang sebelumnya telah memiliki akun bank dan rekening yang aktif, memutuskan untuk berhenti menggunakan layanan tersebut disebut nasabah *churn* (Irmanda dkk., 2019). Irmada juga mengungkapkan bahwa akibat dari *churn* nasabah akan menimbulkan dampak negatif yang cukup signifikan bagi bank, seperti penurunan pendapatan dan penurunan reputasi bank. Salah satu pendekatan untuk memprediksi nasabah *churn* yaitu dengan menggunakan metode klasifikasi *data mining* pada data pelanggan dengan memahami pola pola nasabah untuk mengetahui nasabah yang memiliki potensi *churn* atau tidak. Dengan melakukan analisis tersebut, perusahaan perbankan dapat mengambil langkah lebih awal untuk menjaga nasabahnya terlebih nasabah yang berpotensi untuk *churn.*

Beberapa penelitian terkait pelanggan *churn* sudah pernah dilakukan. Salah satunya oleh Clementine dan Arum (2022) tentang klasifikasi *data mining* untuk memprediksi *churn* pelanggan pada perbankan menggunakan metode *Naïve Bayes* dan ID3. Hasil menunjukan bahwa nilai akurasi tertinggi diperoleh oleh metode *Naïve Bayes* sebesar 85,17%.

Allaam (2023) juga melakukan analisis untuk memprediksi *churn* menggunakan algoritma *Random Forest* dan *Fuzzy C-means* dengan hasil akurasi model *Random Forest* sebesar 99,8% dan akurasi *Fuzzy C-means* sebesar 53% .

Karvana (2019) dalam penelitiannya menganalisis dan memprediksi nasabah *churn* menggunakan model *data mining* pada industri perbankan dengan membandingkan beberapa metode klasifikasi. Dari penelitian ini didapat bahwa metode *Support Vector Machine* merupakan metode terbaik dengan perbandingan sampling kelas data 50:50 yang memiliki nilai akurasi sebesar 99,83%.

Berdasarkan dengan uraian di atas, penelitian berikut akan dibandingkan implementasi tiga pemodelan *machine learning* yaitu *Decision Tree, Random Forest* dan *Support Vector Machine* untuk memprediksi nasabah *churn.*

II. METODOLOGI PENELITIAN

Sebagai langkah awal untuk memahami permasalahan dalam penelitian, dikemukakan sumber data, dasar teori dan langkah-langkah yang digunakan sebagai landasan dalam penelitian ini.

2.1 Sumber Data

Sampel data yang digunakan dalam penelitian ini merupakan data sekunder nasabah Bank ABC (*Arab Banking Corporation*) yang diambil dari website Kaggle dengan jumlah sebanyak 10.000 data. Data ini mencakup informasi mengenai nasabah bank yang terdiri dari 12 atribut termasuk atribut status *churn* nasabah. Variabel independen yang digunakan adalah *customer id, credit score, country, gender, age, tenure, balance, product number, credit card, active member* dan *estimated salary*. Penjelasan mengenai variabel tercantum pada Tabel 1.

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

Atribut	Keterangan
customer_id	Kode unik setiap nasabah
credit_score	Skor kredit nasabah
country	Negara domisili nasabah
gender	Jenis kelamin nasabah
age	Umur nasabah
tenure	Jumlah tahun nasabah menjadi anggota bank
balance	Saldo rekening nasabah
products_number	Jumlah produk yang dimiliki nasabah
credit_card	Kepemilikan kartu kredit pada nasabah, dengan kategori : 0 = tidak aktif 1 = aktif
active_member	Keaktifan nasabah dalam anggota bank, dengan kategori : 0 = tidak aktif 1 = aktif
estimated_salary	Perkiraan gaji nasabah
churn	Status nasabah dalam menggunakan layanan bank, dengan kategori : 0 = tidak *churn* 1 = *churn*

Tabel 1. Daftar Atribut

2.2 *Data Preprocessing*

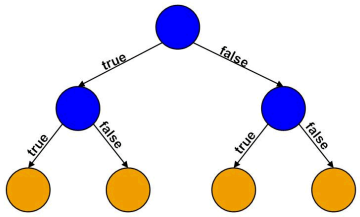
Tahapan *data preprocessing* dapat dikatakan merupakan langkah kunci dalam analisis data, karena *output* yang dihasilkan harus berupa data yang berkualitas dan kesiapan data tersebut dapat mempengaruhi hasil akhir model yang dikembangkan.

Pada Penelitian ini dilakukan beberapa langkah *preprocessing* data yaitu menghitung statistik deskriptif, melakukan *encoding* pada data dan cek *imbalance* data hingga mengatasi *imbalaced* data dengan teknik SMOTE.

2.3 *Decision Tree*

Metode *Decision tree* merupakan salah satu teknik yang dimanfaatkan dalam proses pengambilan keputusan, pengelompokan, dan prediksi. Pendekatan ini mengubah data menjadi bentuk pohon yang memiliki simpul (*node*) dan cabang (*edge*). Tiap simpul dalam pohon mewakili atribut yang telah diuji, sementara setiap cabang menggambarkan hasil dari uji tersebut. Di sisi lain, simpul daun (*leaf*) memperlihatkan kelompok kelas spesifik (Nasrullah, 2021).

Illustrasi metode *decision tree* dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi Metode *Decision Tree*
(sumber : https://help.sap.com)

Dasar dari perhitungan algoritma *decision tree* adalah mencari nilai *entropy* sebagai tahap pertama. Nilai *entropy* diukur untuk menilai sejauh mana sebuah atribut masukan memberikan informasi dalam menciptakan suatu atribut (Nurzahputra dkk., 2017). Rumus dasar *entropy* dapat dilihat pada Persamaan 1.

$$Entropy\ (S) = \sum_{i=1}^{n} -P_i \log_2 P_i \quad (1)$$

dengan *S* adalah himpunan dataset, *n* adalah jumlah kelas klasifikasi dan P_i adalah probabilitas frekuensi kelas ke-i dalam dataset. Dalam pemilihan atribut sebagai akar, dilakukan berdasarkan nilai maksimal dari atribut yang tersedia. Perhitungan nilai

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

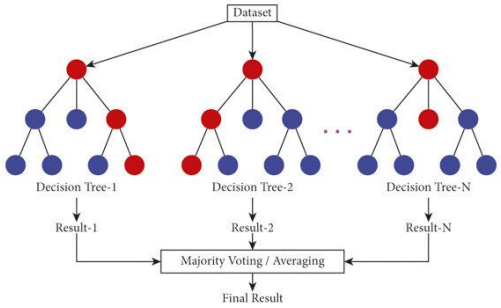
gain menggunakan rumus sesuai dengan Persamaan 2.

$$Gain\ (S, A) = Entropy\ (S) - (\sum_v \frac{|S_v|}{|S|} Entropy(S_v)) \quad (2)$$

dengan *A* merupakan variabel, *v* merupakan nilai yang mungkin untuk variabel A, $|S_i|$ adalah jumlah sampel untuk nilai v, $|S|$ adalah jumlah seluruh sampel data dan $Entropy(S_v)$ merupakan *entropy* untuk sampel yang memiliki nilai v.

2.4 *Random Forest*

Metode *random forest* merupakan pengembangan dari metode *decision tree*, yaitu dengan menerapkan metode *bootstrap aggregating* dan *random feature selection* untuk mengumpulkan banyak-nya *tree*. Dalam metode *random forest*, sejumlah besar pohon ditumbuhkan untuk membentuk sebuah hutan (*forest*) dan setelah itu analisis dilakukan pada sekumpulan pohon tersebut (Dewi dkk., 2011). Pada algoritma permodelannya memiliki ilustrasi seperti pada Gambar 2.



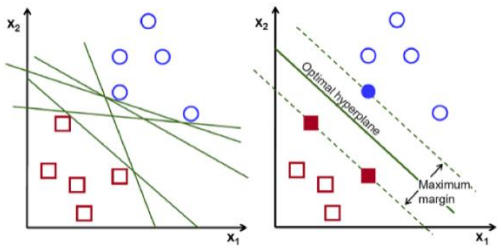
Gambar 2. Ilustrasi Metode *Random Forest* (sumber : https://www.trivusi.web.id)

Hasil prediksi dari *random forest* diperoleh dengan mengambil hasil mayoritas dari setiap pohon keputusan individual (Liparas et al., 2014).

2.5 *Support Vector Machine*

Metode klasifikasi *support vector machine* merupakan metode klasifikasi untuk mengidentifikasi *hyperplane* optimal yang memisahkan dua kelas di ruang input (Wibawa dkk., 2018).

Hyperplane dalam *Support Vector Machines* (SVM) adalah konsep utama yang memfasilitasi pengelompokan titik-titik data dengan membentuk batas antara kelas-kelas yang berbeda. Titik data yang berada di satu sisi *hyperplane* akan diklasifikasikan ke satu kelas, sementara yang berada di sisi yang berlawanan akan diklasifikasikan ke kelas yang berbeda (Kapoor & Kumar, 2022). Illustrasi metode *support vector machine* dapat dilihat pada Gambar 3.



Gambar 2. Ilustrasi Metode *Support Vector Machine* (sumber : https://codingstudio.id)

Bentuk umum persamaan *hyperplane* untuk memisahkan kedua kelas 1 dan 0 dinotasikan pada Persamaan 3.

$$w.x + b = 0 \quad (3)$$

dengan *w* adalah nilai bobot, *x* adalah data ke-x dan *b* merupakan nilai bias.

Support vector machine bekerja untuk menemukan hyperplane optimalnya dengan meminimalkan fungsi tujuan pada Persamaan (4) dan memperhatikan pembatas pada persamaan (5) dan (6).

$$min\frac{1}{2}||w||^2, \quad (4)$$

$$(w.x_i + b) \geq +1 \text{ untuk } y_i = 1, \quad (5)$$

$$(w.x_i + b) \geq -1 \text{ untuk } y_i = 0. \quad (6)$$

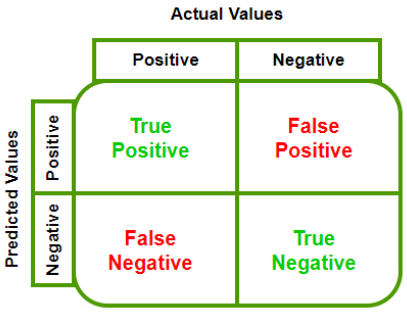
Nilai *w* dan *b* yang diperoleh nantinya akan digunakan untuk proses klasifikasi data uji. Fungsi klasifikasi dinyatakan pada Persamaan 7 (Gunn, 1998).

$$f(x) = sign(w.x + b) \quad (7)$$

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

2.6 *Confussion Matrix*

Confusion matrix merupakan alat yang digunakan untuk mengevaluasi performa suatu model klasifikasi dengan memperlihatkan hasil klasifikasi aktual dan prediksi dalam format tabel. Tabel *confusion matrix* memberikan informasi mengenai klasifikasi *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) (Fibrianda & Bhawiyuga, 2018). Illustrasi tabel *confusion matrix* dapat dilihat pada Gambar 4.

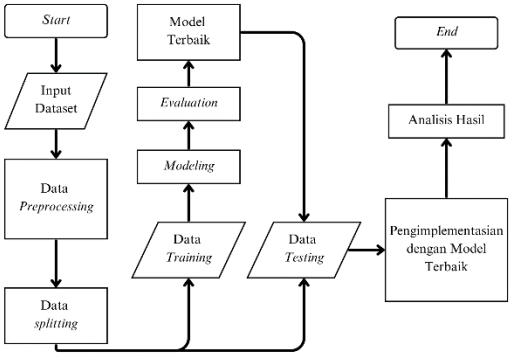


Gambar 4. Ilustrasi *Tabel Confusion Matrix*
(sumber : https://www.geeksforgeeks.org)

Confusion matrix dapat menghasilkan berbagai *performance metrics* seperti nilai akurasi, *precision*, *recall* dan F1-*score*. Akurasi menilai seberapa besar prediksi yang benar secara keseluruhan, presisi memusatkan perhatian pada sejauh mana prediksi positif yang tepat, *recall* memusatkan perhatian pada seberapa baik instance positif diidentifikasi dengan benar, dan F1-*score* adalah pengukuran rata-rata harmonik dari presisi dan *recall* (Han et al., 2012)

2.7 Langkah Penelitian

Tahapan penelitian ini diantaranya adalah melakukan analisis deskriptif pada data, melakukan *preprocessing* pada dataset, kemudian membangun model *decision tree, random forest* dan *support vector machine* dan melatih dengan menggunakan data latih dan menghitung nilai evaluasi untuk dilakukan penentuan model terbaik, model terbaik yang telah didapatkan akan diterapkan pada data uji.



Alur penelitian dapat dilihat pada gambar 5.

Gambar 5. *Flowchart* Penelitian

III. HASIL DAN PEMBAHASAN

Berikut merupakan hasil dan pembahasan dari penelitian berserta penjelasan dan penjabarannya.

3.1 Data *Preprocessing*

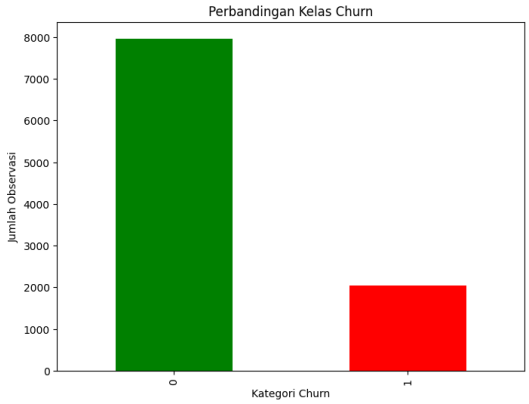
Pada dataset nasabah *churn*, terdapat 12 variabel dengan 2 variabel berbentuk kategorik non numerik. Variabel kategorik tersebut adalah *country* dan *gender*, sementara 10 variabel lainnya sudah berbentuk data numerik. Untuk mempermudah proses pengolahan data diperlukan *encoding* pada data kategorik. Hasil dari proses *encoding* dapat dilihat pada Tabel 2.

Variabel	Hasil *Encoding*
Gender	*0 : Male*
	1 : Female
Country	*0 : France*
	1 : Spain
	2 : Germany

Tabel 2. Hasil *Encoding Data*

Sementara itu, untuk variabel respon akan dibuat diagram proporsi klasifikasi dari status nasabah *churn* yang ditunjukkan pada Gambar 6.

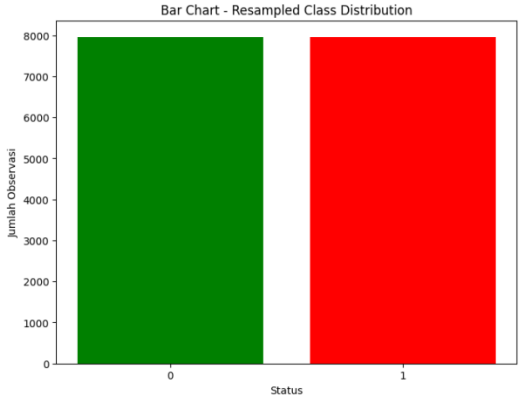
ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594



Gambar 6. Proporsi Status Nasabah *Churn*

Gambar 6 menunjukan bahwa sebanyak 79,6% dari keseluruhan nasabah atau 7963 nasabah memiliki status tidak churn yang dilambangkan dengan warna hijau dan sebanyak 20,4% dari keseluruhan nasabah atau 2037 nasabah memiliki status churn yang dilambangkan oleh warna merah. Terlihat bahwa terdapat ketidakseimbangan kelas pada data sehingga diperlukan adanya penanganan untuk data *imbalace.*

Untuk mengatasi permasalahan tersebut, dilakukan teknik SMOTE supaya proporsi status *churn* seimbang sehingga dapat meningkatkan akurasi prediksinya. Klasifikasi sebelum dilakukan *handle imbalanced* data dapat dilihat pada Gambar 6 dan klasifikasi setelah dilakukannya teknik SMOTE dapat dilihat pada Gambar 7.



Gambar 7. Proporsi Status Nasabah Churn Setelah Melakukan SMOTE

Metode SMOTE dapat menangani masalah ini dengan baik, hal ini ditunjukan pada Gambar 7 yang menunjukan kedua status kelas memiliki proporsi yang seimbang sebesar 50% atau sebanyak 7963 data pada kedua kelas.

Selanjutnya dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Pada analisis ini dilakukan pembagian dataset dengan rasio pembagian sebesar 80:20, yang berartikan bahwa 80% dari dataset akan digunakan sebagai data latih untuk membangun model dan sisanya 20% akan digunakan sebagai data uji untuk menguji kinerja model.

3.2 Metode *Decision Tree*

Proses pembentukan model dilakukan dengan proses *tuning* parameter dengan teknik *grid search* untuk mengoptimalkan hasil prediksi. Beberapa parameter yang digunakan yaitu *max_features*, *max_depth*, *criterion*, *min_samples_split*, *min_samples_leaf* dan *max_leaf_nodes*. Hasil dari proses *tuning* parameter dapat dilihat pada Tabel 3 berikut.

Tabel 3. Hasil Proses *Tuning* Parameter Metode *Decision Tree*

Parameter	Parameter *Values*	Parameter Terbaik
max_features	*auto, sqrt, log2*	*auto*
max_depth	5, 10, 20, 30	10
criterion	*gini, entropy*	*entropy*
min_samples_split	50, 100, 200, 500	200
min_samples_leaf	25, 50, 100, 200	25
max_leaf_nodes	11, 13, 15, 30	30

3.3 Metode *Random Forest*

Proses pembentukan model dilakukan dengan proses *tuning* parameter

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

dengan teknik *grid search* untuk mengoptimalkan hasil prediksi. Beberapa parameter yang digunakan yaitu *n_estimators*, *max_features*, *max_depth*, *criterion*, *min_samples_split*, *min_samples_leaf* dan *max_leaf_nodes*. Hasil dari proses *tuning* parameter dapat dilihat pada Tabel 4 berikut.

Tabel 4. Hasil Proses *Tuning* Parameter Metode *Random Forest*

Parameter	Parameter *Values*	Parameter Terbaik
n_estimators	10, 25, 50, 100	25
max_features	*auto, sqrt, log2*	*auto*
max_depth	5, 10, 20, 30	10
criterion	*gini, entropy*	*gini*
min_samples_split	50, 100, 200, 500	50
min_samples_leaf	25, 50, 100, 200	25
max_leaf_nodes	11, 13, 15, 30	30

3.4 Metode SVM

Proses pembentukan model dilakukan dengan proses *tuning* parameter dengan teknik *grid search* untuk mengoptimalkan hasil prediksi. Beberapa parameter yang digunakan yaitu *kernel* dengan fungsi *linear, polimonial, sigmoid* dan *radial basis function*, lalu nilai *cost* (C) dan nilai gamma (γ) untuk hasil dari proses *tuning* parameter dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Proses *Tuning* Parameter Metode *Support Vector Machine*

Parameter	Parameter *Values*	Parameter Terbaik
kernel	*linear, poly, sigmoid, rbf*	*rbf*
C	0.1, 1, 10, 100, 1000	1000
γ	0.01, 0.005, 1, 1.5	0.01

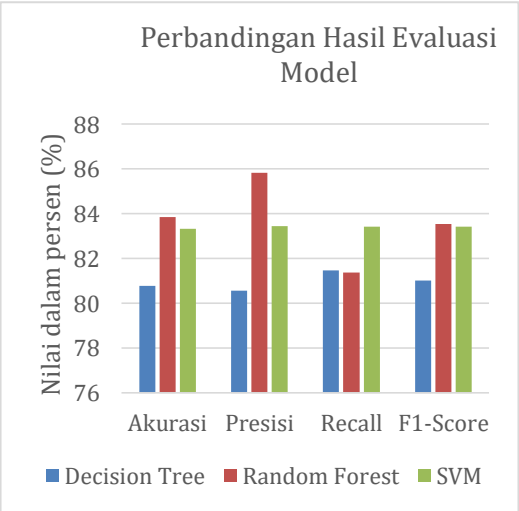
3.5 Evaluasi

Setelah ketiga model diperoleh dan dievaluasi hasil performa model tersebut, pembandingan hasil performa ketiga model perlu diperlukan untuk menentukan model mana yang lebih akurat dalam mengklasifikasikan status *chrun* nasabah. Perbandingan akan dilakukan pada keempat poin metrik evaluasi yaitu akurasi, presisi, *recall* dan F1-*score* dari ketiga model. Berdasarkan perhitungan dari nilai *confusion matrix*, diperoleh performa klasifikasi nasabah *churn* menggunakan *decision tree*, *random forest* dan SVM seperti pada Tabel 6.

Tabel 6. Hasil Evaluasi Performa Model

	Akurasi	Presisi	*Recall*	F1-*score*
DT	80,78%	80,56%	81,5%	81,02%
RF	83,86%	85,82%	81,37%	83,54%
SVM	83,32%	83,44%	83,42%	83,43%

Berdasarkan Tabel 6, perbandingan hasil evaluasi performa ketiga model juga dapat digambarkan pada Gambar 8 berikut.



Gambar 8. Grafik Perbandingan Nilai Evaluasi Performa Model

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

Pada hasil evaluasi performa diatas, disimpulkan bahwa model terbaik ditunjukan oleh model *random forest* yang memiliki nilai akurasi tertinggi sebesar 83,86% dan nilai F1-*score* tertinggi sebesar 83,54%. Model *random forest* dianggap paling akurat untuk mengklasifikasikan status nasabah *churn* dibandingkan dengan model *decision tree* dan SVM

3.6 Hasil Klasifikasi

Berdasarkan hasil evaluasi model pada data *testing*, diperoleh model yang dibentuk memilki nilai akurasi data *testing* sebesar 83,34%, presisi sebesar 85,42%, *recall* sebesar 79,76% dan F1-*score* sebesar 82,49%. Nilai akurasi diperoleh berdasarkan perhitungan dari *confusion matrix* seperti yang tertera pada Tabel 7.

Tabel 7. Hasil *Confusion Matrix Random Forest* pada Data *Training*

		Prediksi	
		Churn	*Non Churn*
Aktual	*Churn*	1876	476
	Non Churn	320	2106

IV. KESIMPULAN

Metode yang paling akurat dalam memprediksi status *churn* pada nasabah adalah metode *Random Forest* yang terbukti unggul dengan memiliki nilai akurasi sebesar 83,86% dan F1-*score* sebesar 83,54%. Hasil klasifikasi menjabarkan bahwa sebanyak 1876 nasabah yang berpotensi *churn* dapat diklasifikasikan secara tepat sebagai nasabah yang berpotensi *churn* dan 2106 nasabah yang tidak berpotensi *churn* juga dapat diklasifikasikan dengan tepat sebagai nasabah yang tidak berpotensi *churn*, sementara itu 476 nasabah yang berpotensi *churn* diklasifikasikan secara tidak tepat sebagai nasabah yang tidak berpotensi *churn* dan 320 nasabah yang tidak berpotensi *churn* diklasifikasikan secara tidak tepat sebagai nasabah yang berpotensi *churn*.

V. SARAN

Berdasarkan proses analisis penelitian yang telah dilakukan dengan berbagai keterbatasan, peneliti menyarankan beberapa pengembangan analisis yaitu dengan dapat melakukan penyesuaian *hyperparameter* untuk meningkatkan nilai akurasi maupun mengurangi kesalahan klasifikasi.

VI. DAFTAR PUSTAKA

Allaam, A. (2023). *Prediksi Churn Konsumen Menggunakan Algoritma Random Forest dengan Fuzzy C-Means untuk Meningkatkan Produktivitas Penjualan Bisnis*. Universitas Islam Negeri Syarif Hidayatullah, Jakarta.

Clementine, M., & Arum. (2022). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naive Bayes dan ID3. *Jurnal Processor*, *17*(1), 9–18.

Dewi, N. K., Mulyadi, S. Y., & Syafitri, U. D. (2011). Penerapan Metode Random Forest Dalam Driver Analysis. *Forum Statistika Dan Komputasi*, *16*(1), 35–43.

Fibrianda, M. F., & Bhawiyuga, A. (2018). Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(9).

Gunn, S. R. (1998). Support Vector Machines for classification and regression. *ISIS Technical Report*, *14*(1), 5–16.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In

ESCAF 3rd 2024
p-ISSN : 2962 - 7710
e-ISSN : 3021-8594

Data Mining: Concepts and Techniques.

Irmanda, H. N., Astriratma, R., & Afrizal, S. (2019). Perbandingan Metode Jaringan Syaraf Tiruan dan Pohon Keputusan untuk Prediksi Churn Universitas Pembangunan Nasional Veteran Jakarta. *JSI : Jurnal Sistem Informasi (E-Journal)*, *11*(2).

Kapoor, K. G., & Kumar, A. (2022). SVM Hyperplane Misclassification Control by Finding Optimum Cost of Misclassification with Boundary Value Analysis Technique. *Journal of Positive School Psychology*, *6*(3), 2812–2816.

Karvana, K. (2019). *Analisis dan Prediksi Nasabah Churn Menggunakan Model Data Mining pada Industri Perbankan: Studi Kasus pada PT BANK XYZ*. Universitas Indonesia, Depok.

Liparas, D., HaCohen-Kerner, Y., Moumtzidou, A., Vrochidis, S., & Kompatsiaris, I. (2014). News articles classification using random forests and weighted multimodal features. 63–75.

Nasrullah, A. H. (2021). Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris. *Jurnal Ilmiah Ilmu Komputer*, *7*(2), 45–51.

Nurzahputra, A., Ratna Safitri, A., & Aziz Muslim, M. (2017). Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree. *PRISMA : Prosiding Seminar Nasional Matematika*, 717–722.

Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, *3*(1).