

Analisis Perbandingan Optimasi Berbasis *Evolutionary* pada Algoritma Klasifikasi Penentuan Profile Resiko Nasabah

Comparative Analysis of Evolutionary-based Optimization on the Classification Algorithm for Determining Customer Risk Profile

Asep Arifyan¹, Tjahjanto²

¹Program Magister Ilmu Komputer, Fakultas Teknologi Informasi Universitas Budi Luhur
E-mail: ¹2011600810@student.budiluhur.ac.id, ²cahyanto2000@gmail.com

Abstrak

Penelitian ini mengungkapkan bagaimana pentingnya penanganan deteksi profile nasabah untuk meminimalisir terjadinya penyalahgunaan akun nasabah. Kebutuhan akan nasabah baru dengan berbagai macam produk perbankan membuat layanan *screening* awal begitu penting dilakukan oleh pihak perbankan agar mendapatkan informasi profile resiko nasabah sejak dini. Oleh karena itu, tujuan dari penelitian ini adalah menentukan model terbaik dari klasifikasi data profile nasabah dengan cara membandingkan model klasifikasi data mining *Naïve Bayes*, *Decision Tree*, *Random Forest*, *KNN*, *SVM*. Model dengan akurasi terbaik inilah yang nantinya akan direkomendasikan sebagai model alternatif untuk melakukan prediksi profile resiko nasabah. Pada penelitian ini juga dilakukan proses optimasi dengan menggunakan *Optimize Selection (Evolutionary)* pada setiap teknik klasifikasi yang digunakan. Dimana dihasilkan bahwa model algoritma *Random Forest* mendapatkan hasil total akurasi paling baik yaitu sebesar 82.55% dengan angka kenaikan optimasi sebesar 2.51%. Proses training dan testing pada penelitian ini menggunakan komposisi 80% dataset training dan 20% dataset untuk testing sedangkan metodologi yang digunakan pada penelitian ini adalah dengan menggunakan CRISP-DM.

Kata kunci: Klasifikasi Data Mining, *Random Forest*, *Optimize Selection (Evolutionary)*, CRISPDM, Banking

Abstract

This research reveals how important it is to handle customer profile detection to minimize the occurrence of misuse of customer accounts. The need for new customers with various banking products makes initial screening services so important for banks to get information on customer risk profiles early on. Therefore, the purpose of this study is to determine the best model for classifying customer profile data by comparing the data mining classification models of Naïve Bayes, Decision Tree, Random Forest, KNN, SVM. This best model will later be recommended as an alternative model for predicting customer risk profiles. In this study, an optimization process was also carried out using Optimize Selection (Evolutionary) for each classification technique used. Where it is generated that the Random Forest algorithm model gets the best total accuracy of 82.55% with an optimization increase of 2.51%. The training and testing process in this study uses a composition of 80% training dataset and 20% dataset for testing while the methodology used in this study is CRISP-DM.

Keywords: Klasifikasi Data Mining, *Random Forest*, *Optimize Selection (Evolutionary)*, CRISP DM, Banking

1. PENDAHULUAN

Perkembangan teknologi informasi didunia perbankan dan finance saat ini semakin berkembang dengan pesat. Data Bank umum berdasarkan Badan Pusat Statistik tahun 2021 adalah

sebanyak 107 Bank dan 1.632 untuk Bank Perkreditan / Pembiayaan Rakyat. Hal ini menyurutkan setiap Bank untuk mendapatkan jumlah nasabah sebanyak banyaknya sehingga proses *profiling* nasabah harus secara ketat dilakukan oleh pihak perbankan. Data mining dapat menjadi salah satu alternatif yang dapat dilakukan untuk meningkatkan kinerja bank pada permasalahan identifikasi dan klasifikasi profile nasabah [1]. Dimana pada penelitian sebelumnya metode AI dan data mining dapat digunakan untuk menganalisis efisiensi bank, Management Resiko, Performansi Bank, Regulasi bank, Merger akuisisi, Relasi Nasabah, dan *Fintech* Perbankan [2].

Penelitian sebelumnya yang berelasi dengan penelitian ini telah dirangkum dan dieksplorasi sebagai bahan rujukan dari penelitian yang berjudul *Analisis Perbandingan Optimasi berbasis Evolutionary pada Algoritma Klasifikasi Penentuan Profile Resiko Nasabah*. Penelitian yang menjadi rujukan tersebut berasal dari penelitian skala nasional dan Internasional yang digali dari tahun 2017 sampai dengan 2021.

Tabel 1 Penelitian Metode Klasifikasi Data Mining pada sektor Perbankan

No	Judul	Sektor	Index	Peneliti /Tahun	Hasil
1	Analisis Data Bank Direct Marketing dengan Perbandingan Klasifikasi Data Mining Berbasis Optimize Selection (Evolutionary) [3]	Banking / Finance	Google Scholar	Ahmad Fauzi 2021	optimasi Optimize Selection (Evolutionary) yaitu algoritma Naïve Bayes 90.18%
2	Analisis Data Mining Kelayakan Calon Debitur di PT. Bank BRI dengan Menggunakan Metode Klasifikasi Naïve Bayes [4]	Banking / Finance	Google Scholar	Ahmad Turmudzy, Donny, Sofyan 2018	Akurasi Naïve Bayes sebesar 78.88 %
3	Analisis Optimasi Algoritma Klasifikasi Support Vector Machine, Decision Trees, dan Neural Network Menggunakan Adaboost dan Bagging [5]	Banking / Finance	Google Scholar	Agus Heri Yunial 2020	akurasi svm 89.10%, Decision Tree 90.36%
4	Perbandingan Teknik Klasifikasi Dalam Data Mining untuk Bank Direct Marketing [6]	Banking / Finance	Google Scholar	Irvi Oktanisa, Ahmad Afif Supianto 2017	Akurasi terbaik Stochastic Gradient Descent sebesar 97.2%
5	Penerapan K-Nearest Neighbor Berbasis Genetic Algorithm untuk Penentuan Pemberian Kredit [7]	Banking / Finance	Google Scholar	Ester Arisawati 2017	Akurasi KNN sebesar 68.56%
6	Analisis Potensi Dana Retail pada Nasabah PT. Bank Tabungan Negara (Persero), Tbk. Dengan Metode Decision Tree dan Naive Bayes Berbasis Optimize Selection (Evolutionary) [8]	Banking / Finance	Google Scholar	Ahmad Fauzi, Tukiyat 2019	Akurasi naïve bayes terbaik 94.47%
7	Perbandingan Metode random Forest dan Naïve Bayes untuk Klasifikasi Debitur berdasarkan Kualitas Kredit [9]	Banking / Finance	Google Scholar	Bonggo Bawono, Rochdi Wasono 2019	Akurasi Random Forest 98.16%
8	Evaluation of classification algorithms for banking customer's behavior under Apache Spark Data Processing System [10]	Banking / Finance	ScienceDirect	Wael Ataiwi, Mariam Biltawi, Ghazi Naymat 2017	Precision Recall F-Measure NB 4% 49% 7.3% SVM 0.18% 10% 0.3%
9	Predictive Model for Cutting Customers Migration from banks: Based on machine learning classification algorithms [11]	Banking / Finance	IEEE	Kahlid Alkhatib, Sayel Abualigah 2020	Akurasi Random Forest 86%
10	Predictions of Loan Defaulter - A Data Science Perspective [12]	Banking / Finance	IEEE	P. Maheswari, Narayana 2020	Akurasi Random Forest 79%
11	Analysing Data Mining Techniques on Bank Customers for Credit Score [13]	Banking / Finance	IEEE	Durgesh Kumar Singh, Noopur Goel 2020	Akurasi SVM 90.4%

Pada penelitian yang telah dilakukan oleh Ahmad Fauzi pada Tahun 2021 tentang Analisis Data Bank Direct Marketing dengan Perbandingan Klasifikasi Data Mining Berbasis Optimize Selection (Evolutionary) menghasilkan akurasi klasifikasi terbaik adalah dengan algoritma *Naïve Bayes* yang telah dioptimalisasi dengan *Evolutionary* yakni sebesar 90.18% . Sementara untuk *KNN* sebesar 86.66% dan *SVM* sebesar 89.40% [3]. Penelitian yang dilakukan

Ahmad Turmudzy, Donny, dan Sofyan pada tahun 2018 tentang Analisis Data Mining Kelayakan Calon Debitur di PT. Bank BRI dengan Menggunakan Metode Klasifikasi *Naïve Bayes* mendapatkan Akurasi *Naïve Bayes* sebesar 78.88 % [4]. Kemudian Penelitian yang dilakukan Agus Heri Yunial pada tahun 2020 tentang Analisis Optimasi Algoritma Klasifikasi *Support Vector Machine, Decision Trees, dan Neural Network* Menggunakan *Adaboost* dan *Bagging* mendapatkan Akurasi *SVM* sebesar 89.10 % dan *Decision Tree* 90.36% [5].

Penelitian yang dilakukan Irvi Oktanisa dan Ahmad Afif Supianto 2017 tentang Perbandingan Teknik Klasifikasi Dalam Data Mining untuk Bank Direct Marketing mendapatkan akurasi *Tree, Naïve Bayes, Constant, dan Stochastic Gradient Descent* sebesar 97.2% sementara untuk *SVM* 67.1% kemudian *K-Nearest Neighbor* dan *Random Forest* 97.1% [6]. Penelitian yang dilakukan Ester Arisawati tahun 2017 tentang Penerapan K-Nearest Neighbor Berbasis Genetic Algorithm untuk Penentuan Pemberian Kredit mendapatkan Akurasi *KNN* sebesar 68.56% [7]. Penelitian yang dilakukan Ahmad Fauzi, Tukiyat 2019 tentang Analisis Potensi Dana Retail pada Nasabah PT. Bank Tabungan Negara (Persero), Tbk. Dengan Metode *Decision Tree* dan *Naive Bayes* Berbasis *Optimize Selection (Evolutionary)* mendapatkan Akurasi *Naïve Bayes* sebesar 94.47 % dan *Decision Tree* 94.01% [8]. Kemudian Penelitian yang dilakukan Bonggo Bawono, Rochdi Wasono pada tahun 2019 tentang Analisis Perbandingan Metode *random Forest* dan *Naïve Bayes* untuk Klasifikasi Debitur berdasarkan Kualitas Kredit mendapatkan Akurasi *Random Foest* sebesar 98.16% dan *Naïve Bayes* 95.93% [9].

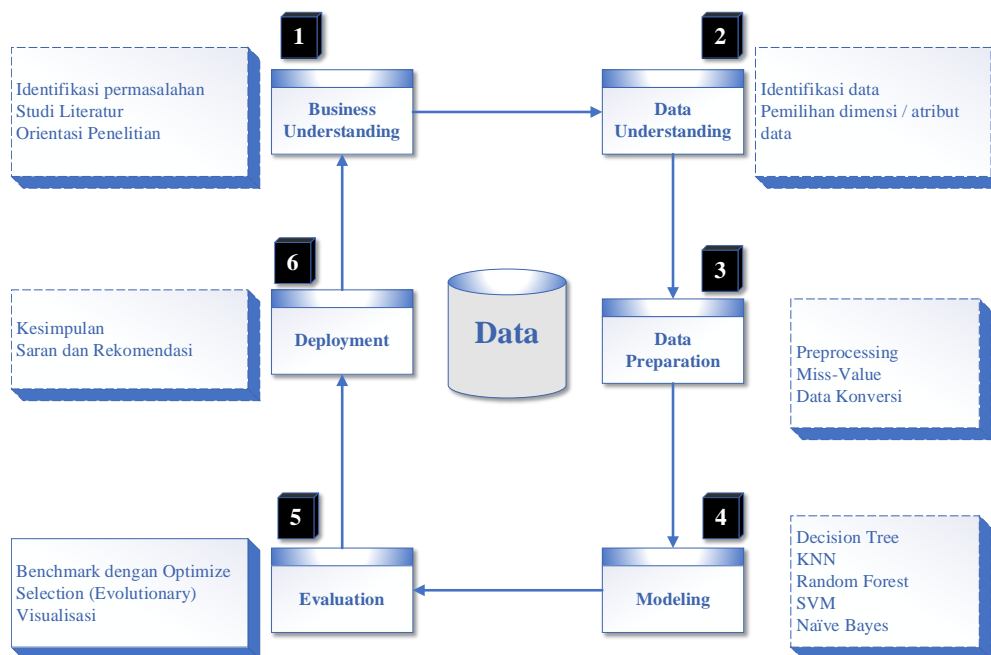
Kemudian Penelitian yang dilakukan Wael Ataiwi, Mariam Biltawi, Ghazi Naymat 2017 tentang *Evaluation of classification algorithms for banking customer's behavior under Apache Spark Data Processing System* mendapatkan *Precision Recall F-Measure NB 4% 49% 7.3% SVM 0.18% 10% 0.3%* [10]. Penelitian yang dilakukan Kahlid Alkhatib, Sayel Abualigah 2020 tentang *Predictive Model for Cutting Customers Migration from banks: Based on machine learning classification algorithms* mendapatkan Akurasi *Decision Tree 80%, Random Forest 86%, Naïve Bayes 84%, K Nearest Neighbor 82%* dan *Logistic Regression 82%* [11]. Selanjutnya Penelitian yang dilakukan P. Maheswari, Narayana 2020 tentang *Predictions of Loan Defaulter - A Data Science Perspective* mendapatkan Akurasi *Random Forest 79%, K Nearest Neighbor 78%* dan *Logistic Regression 80%* [12]. Penelitian yang dilakukan Durgesh Kumar Singh, Noopur Goel 2020 tentang *Analysing Data Mining Techniques on Bank Customers for Credit Score* mendapatkan Akurasi *Decision Tree 85.7% , K Nearest Neighbor 79.4%* dan *SVM 90.4%* [13].

Berdasarkan hasil akurasi yang relatif cukup tinggi pada penelitian diatas, penulis tertarik untuk melakukan penelitian perbandingan model klasifikasi data mining *Naïve Bayes, Decision Tree, Random Forest, KNN, SVM* dengan melakukan optimasi *Optimize Selection (Evolutionary)* pada setiap algoritma klasifikasi. Perbandingan model ini dilakukan dengan tujuan untuk mendapatkan model terbaik dari model klasifikasi data profile nasabah. Sehingga diharapkan menjadi model alternatif untuk melakukan deteksi dari profile resiko nasabah sejak dini. Penelitian ini juga dilakukan dengan melakukan perbandingan antara original algoritma kasifikasi dan algoritma klasifikasi yang sudah dioptimasi dengan *Evolutionary*. Adapun aspek pembeda dari penelitian sebelumnya adalah terletak pada penyelesaian permasalahan yaitu mengatasi permasalahan manajemen resiko dari sisi nasabah dengan cara mendeteksi profile nasabah sejak dini apakah beresiko tinggi (*High*) atau rendah (*Low*). Kemudian aspek tujuan penelitian yang menitikberatkan kepada pencarian model terbaik dari teknik klasifikasi data mining dengan membandingkan model dan menambahkan 2 tambahan model klasifikasi dari jenis *Tree* yaitu *Decision Tree* dan *Random Forest*. Kemudian aspek pembeda lainnya adalah asepek data, dimana dataset yang digunakan pada penelitian ini adalah dataset profile nasabah Bank xyz yang sudah dimasking, sedangkan penelitian sebelumnya adalah Bank Direct Marketing yang diambil dari UCI machine learning repository atau data word [14]. Training dan Pengujian algoritma

klasifikasi dan metode Optimasi *Evolutionary* pada penelitian ini adalah dengan memanfaatkan tools Rapid Miner 9.8.

2. METODE PENELITIAN

Metodologi penelitian yang digunakan penulis dalam penelitian ini adalah dengan menggunakan metode Cross-Industry Standard Process for Data Mining (CRISP-DM) yang terdiri dari enam tahap yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. CRISP-DM merupakan salah satu metodologi terbaik dari 3 metodologi data mining lainnya seperti SEMMA, KDD, dan Data Science Process [15]:



Gambar 1 Proses Metodologi CRISP-DM

Berikut ini merupakan tahapan proses CRISP-DM yang dilakukan pada penelitian ini :

a. *Business Understanding* (Pemahaman Bisnis) adalah:

Pemahaman bisnis meliputi penetapan identifikasi permasalahan, tujuan atau orientasi penelitian, perencanaan pemilihan data, perencanaan penggalian data, perencanaan pengembangan penelitian [16]. Pada fase ini skop penelitian terkait kontribusi umum dan akademik harus ditetapkan sehingga dapat menghasilkan penelitian yang terukur sesuai dengan batasan atau skop yang sudah ditetapkan. Berbagai studi literatur sangat diperlukan pada tahap ini.

b. *Data Understanding* (Pemahaman Data) adalah:

Setelah tujuan penelitian dan rencana penelitian ditetapkan, pemahaman data dalam mempertimbangkan data yang dibutuhkan atau menyesuaikan dengan sumber data yang ada / tersedia. Langkah ini bisa meliputi pemilihan kolom yang akan digunakan sebagai atribut dan label untuk proses klasifikasi data dengan menggunakan *Naïve Bayes*, *Decision Tree*, *Random Forest*, *KNN*, *SVM* [17]. Pada fase ini juga diperlukan verifikasi

kepakaran data dari seorang pakar dibidang perbankan atau yang linier. Hal ini berguna agar data yang dipilih sebagai atribut klasifikasi sesuai dengan disiplin ilmu pada sektor perbankan atau finance. Pada penelitian ini penulis mendapatkan komposisi atribut data dan label sesuai dengan tabel 2.

Tabel 2 Komposisi atribut , tipe data, dan keterangan dari dataset

Atribut	Tipe	Keterangan
branch	integer	Cabang Bank dari nasabah terdaftar
document	integer	Dokumen yang digunakan untuk verifikasi data nasabah. Misalnya : KTP, KTPS, SIM, SIUP, NPWP
sex	binomial	Jenis kelamin nasabah
category	binomial	Kategori nasabah apakah perorangan atau non perorangan
tipe_prod	integer	Tipe produk yang digunakan misalnya tabungan dan non tabungan
produk	integer	Produk spesifik bank misalnya tabungan rencana, tabungan payroll, tabungan anak, tabungan biasa dan lain sebagainya.
status	integer	Status produk
usia	integer	Usia nasabah
kelas	binomial	Label kelas yang digunakan berupa LOW RISK dan HIGH RISK

c. *Data Preparation* (Persiapan Data) adalah:

Setelah sumber data yang tersedia diidentifikasi, sumber data tersebut perlu diseleksi, dibersihkan, dibangun ke dalam wujud yang dikehendaki dan dibentuk. Pembersihan dan transformasi data dalam persiapan model data perlu dilakukan pada tahap ini. Eksplorasi data secara lebih mendalam juga dapat diterapkan dalam tahap ini, dan penggunaan model-model tambahan sekali lagi memberikan peluang untuk melihat berbagai pola berdasarkan pemahaman bisnis. Perlu diketahui bahwa persiapan data dilakukan agar data yang dipersiapkan pada tahap ini dapat diproses pada tahap modeling. Dengan kata lain data available untuk semua jenis algoritma klasifikasi yang digunakan (*Naïve Bayes, Decision Tree, Random Forest, KNN, SVM.*) [18]. Dengan kondisi tersebut penulis melakukan data preprocessing, missing value, dan konversi data dari bentuk data asal menjadi data numerik. Dataset yang digunakan pada penelitian ini sebanyak 2.781 data dengan Prosentase Training 80% dan Testing 20%.

Sebelum dilakukan konversi kedalam bentuk numerik data asli profil nasabah masih banyak menggunakan data dengan tipe karakter.

Tabel 3 Contoh data asli sebelum konversi ke numerik

BRANCH	DOKUMEN	SEX	KATEGORI	TIPE_PROD	PRODUK	STATUS	USIA	KELAS
2	APP	M	U	D	ESCROW	1	14	LOW RISK
1102	KTP	M	A	S	TABHAR	2	43	LOW RISK
1202	KTP	M	A	S	TABHAR	2	47	LOW RISK
1100	KTP	M	A	S	TABHAR	2	50	LOW RISK
3002	KTP	F	A	S	TABHAR	2	70	HIGH RISK
4121	KTP	F	A	S	TABHAR	2	55	HIGH RISK
3003	KTP	M	A	S	TABHJR	2	46	HIGH RISK
1405	KTP	F	A	S	TABHJR	2	53	HIGH RISK

Tabel 4 Contoh data yang sudah dikonversi ke numerik

branch	document	sex	category	tipe_prod	produk	status	usia	kelas
2	1	1	1	1	1	1	14	LOW RISK
13	2	1	2	2	2	2	43	LOW RISK
3	2	1	2	2	2	2	52	LOW RISK
19	2	2	2	2	2	2	43	LOW RISK
3	2	2	2	2	2	2	45	LOW RISK
20	2	1	2	2	2	2	47	LOW RISK
11	2	1	2	2	2	2	50	LOW RISK
61	2	2	2	2	3	2	37	HIGH RISK
20	2	1	2	2	2	9	44	HIGH RISK
18	2	2	2	2	3	2	48	HIGH RISK
11	2	2	2	2	2	2	46	HIGH RISK
56	2	2	2	2	13	2	35	HIGH RISK
60	2	2	2	2	2	2	38	HIGH RISK

d. *Modeling* (Pembuatan Model) adalah:

Pada tahapan ini dijalankan proses klasifikasi algoritma *Naïve Bayes*, *Decision Tree*, *Random Forest*, *KNN*, dan *SVM* dengan menggunakan tools Rapid Miner 9.8 . Tahapan training terlebih dahulu dilakukan sehingga akan menghasilkan model yang nanti akan digunakan pada saat testing data. Sebagaimana kita ketahui pada penelitian ini training menggunakan 80% dari dataset sedangkan 20% digunakan sebagai data testing. Visualisasi terkait statistik , hasil model, dan hasil performasi label harus dimunculkan sebagai alat ukur yang dapat dievaluasi pada tahap selanjutnya. Pada tahap ini juga model algoritma klasifikasi akan dijalankan dengan 2 tipe yaitu dijalankan dengan original algoritma dan dijalankan dengan melakukan optimasi *Optimize Selection (Evolutionary)* untuk setiap algoritma.

e. *Evaluation* (Evaluasi) adalah:

Hasil model sebaiknya dievaluasi dalam konteks tujuan bisnis yang ditetapkan pada tahap awal (pemahaman bisnis). Hal ini dilakukan agar hipotesis awal dari penelitian ini dapat diketahui (sesuai atau tidak sesuai dengan hipotesis awal). Evaluasi juga dilakukan agar mengetahui apakah pengaruh *Optimize Selection (Evolutionary)* akan mempengaruhi akurasi kepada semua algoritma klasifikasi. Evaluasi juga dilakukan untuk mengetahui algoritma klasifikasi manakah yang mendapatkan total akurasi terbaik.

Disamping itu efek peningkatan akurasi akibat penggunaan *Optimize Selection (Evolutionary)* perlu menjadi catatan untuk setiap algoritma sebagai rekomendasi pada penelitian selanjutnya.

f. *Deployment* (Pelaksanaan) adalah:

Penggalian data dapat digunakan baik untuk membuktikan hipotesis sebelumnya, ataupun untuk penemuan hasil pengetahuan yang merupakan kontribusi praktis dan akademik yang dapat dijadikan sebagai laporan hasil penelitian yang dapat dijabarkan pada tahapan ini [19]. Hasil penelitian yang menjadi kesimpulan dari tujuan penelitian dapat menjadi sumber referensi untuk penelitian selanjutnya. Saran dan rekomendasi dari penelitian ini juga hendaknya diperlukan agar penelitian selanjutnya dapat lebih baik lagi.

3. HASIL DAN PEMBAHASAN

Laporan dari hasil dan pembahasan terkait penelitian ini akan dipaparkan kedalam hasil performansi training dan testing untuk setiap algoritma Klasifikasi. Model Klasifikasi yang dijalankan adalah model original dan dengan tambahan Optimasi *Optimize Selection (Evolutionary)* yang dilakukan pada algoritma *Decision Tree*, *KNN*, *Random Forest*, *SVM*, dan *Naïve Bayes*. Kemudian untuk proses evaluasi model Klasifikasi menggunakan *Confusion Matrix*.

a. Hasil Klasifikasi *Decision Tree*

a. Algoritma *Decision Tree* tanpa Optimasi

Berikut merupakan Tabel hasil Confusion Matrix *Decision Tree* tanpa menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 5 Hasil Confusion Matrix *Decision Tree* tanpa optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	297	53	84.86%	79.86%
Pred. HIGH RISK	59	147	71.36%	
Class Recall	83.43%	73.50%		

Dari tabel 5 dapat diketahui bahwa persentase class precision 84.86% sebanyak 297 untuk prediksi LOW RISK yang sesuai dan 53 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 71.36% sebanyak 147 yang sesuai dan 59 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

$$TP = 147, FP = 53, TN = 297, FN = 59$$

$$Akurasi = (TP+TN) * 100% / (TP+TN+FP+FN) = (147+297) * 100% / (147+53+297+59)$$

$$Akurasi = 79.86%$$

b. Algoritma *Decision Tree* dengan Optimasi

Berikut merupakan Tabel hasil Confusion Matrix *Decision Tree* dengan menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 6 Hasil Confusion Matrix *Decision Tree* dengan optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	303	51	85.59%	81.29%
Pred. HIGH RISK	53	149	73.76%	
Class Recall	85.11%	74.50%		

Dari tabel 6 dapat diketahui bahwa persentase class precision 85.57% sebanyak 303 untuk prediksi LOW RISK yang sesuai dan 51 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 73.76% sebanyak 149 yang sesuai dan 53 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

TP = 149 , FP = 51 , TN = 303 , FN = 53

Akurasi=(TP+TN)*100%/(TP+TN+FP+FN) = (149+303)*100%/(149+303+51+53)

Akurasi= 81.29%

b. Hasil Klasifikasi KNN

a. Algoritma KNN tanpa Optimasi

Berikut merupakan Tabel hasil Confusion Matrix KNN tanpa menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 7 Hasil Confusion Matrix KNN tanpa optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	306	71	81.17%	78.24%
Pred. HIGH RISK	50	129	72.07%	
Class Recall	85.96%	64.50%		

Dari tabel 7 dapat diketahui bahwa persentase class precision 81.17% sebanyak 306 untuk prediksi LOW RISK yang sesuai dan 71 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 72.07% sebanyak 129 yang sesuai dan 50 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

TP = 129 , FP = 71 , TN = 306 , FN = 50

Akurasi=(TP+TN)*100%/(TP+TN+FP+FN) = (129+306)*100%/(129+306+71+50)

Akurasi= 78.24%

b. Algoritma KNN dengan Optimasi

Berikut merupakan Tabel hasil Confusion Matrix KNN dengan menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 8 Hasil Confusion Matrix KNN dengan optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	316	81	79.60%	78.24%
Pred. HIGH RISK	40	119	74.84%	
Class Recall	88.76%	59.50%		

Dari tabel 8 dapat diketahui bahwa persentase class precision 79.60% sebanyak 316 untuk prediksi LOW RISK yang sesuai dan 81 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 74.84% sebanyak 119 yang sesuai dan 40 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

TP = 119, FP = 81 , TN = 316 , FN = 40

Akurasi=(TP+TN)*100%/(TP+TN+FP+FN) = (119+316)*100%/(119+316+81+40)

Akurasi= 78.24%

c. Hasil Klasifikasi Random Forest

a. Algoritma Random Forest tanpa Optimasi

Berikut merupakan Tabel hasil Confusion Matrix Random Forest tanpa menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 9 Hasil Confusion Matrix *Random Forest* tanpa optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	311	66	82.49%	80.04%
Pred. HIGH RISK	45	134	74.86%	
Class Recall	87.36%	67.00%		

Dari tabel 9 dapat diketahui bahwa persentase class precision 82.49% sebanyak 311 untuk prediksi LOW RISK yang sesuai dan 66 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 74.86% sebanyak 134 yang sesuai dan 45 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

$$TP = 134, FP = 66, TN = 311, FN = 45$$

$$Akurasi = (TP+TN) * 100\% / (TP+TN+FP+FN) = (134+311) * 100\% / (134+311+66+45)$$

$$Akurasi = 80.04\%$$

b. Algoritma *Random Forest* dengan Optimasi

Berikut merupakan Tabel hasil Confusion Matrix *Random Forest* dengan menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 10 Hasil Confusion Matrix *Random Forest* dengan optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	326	67	82.95%	82.55%
Pred. HIGH RISK	30	133	81.60%	
Class Recall	91.57%	66.50%		

Dari tabel 10 dapat diketahui bahwa persentase class precision 82.95% sebanyak 326 untuk prediksi LOW RISK yang sesuai dan 67 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 81.60% sebanyak 133 yang sesuai dan 30 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

$$TP = 133, FP = 67, TN = 326, FN = 30$$

$$Akurasi = (TP+TN) * 100\% / (TP+TN+FP+FN) = (133+326) * 100\% / (133+326+67+30)$$

$$Akurasi = 82.55\%$$

d. Hasil Klasifikasi SVM

a. Algoritma *SVM* tanpa Optimasi

Berikut merupakan Tabel hasil Confusion Matrix *SVM* tanpa menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 11 Hasil Confusion Matrix *SVM* tanpa optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	315	105	75.00%	73.74%
Pred. HIGH RISK	41	95	69.85%	
Class Recall	88.48%	47.50%		

Dari tabel 11 dapat diketahui bahwa persentase class precision 75.00% sebanyak 315 untuk prediksi LOW RISK yang sesuai dan 105 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 69.85% sebanyak 95 yang sesuai dan 41 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

$$TP = 95, FP = 105, TN = 315, FN = 41$$

$$Akurasi = (TP+TN) * 100\% / (TP+TN+FP+FN) = (95+315) * 100\% / (95+315+105+41)$$

$$Akurasi = 73.74\%$$

b. Algoritma SVM dengan Optimasi

Berikut merupakan Tabel hasil Confusion Matrix SVM dengan menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 12 Hasil Confusion Matrix SVM dengan optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	318	98	76.44%	75.54%
Pred. HIGH RISK	38	102	72.86%	
Class Recall	89.33%	51.00%		

Dari tabel 12 dapat diketahui bahwa persentase class precision 76.44% sebanyak 318 untuk prediksi LOW RISK yang sesuai dan 98 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 72.86% sebanyak 102 yang sesuai dan 38 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

TP = 102 , FP = 98 , TN = 318 , FN = 38

Akurasi=(TP+TN)*100%/(TP+TN+FP+FN) = (102+318)*100%/(102+318+98+38)

Akurasi= 75.54%

e. Hasil Klasifikasi Naïve Bayes

a. Algoritma Naïve Bayes tanpa Optimasi

Berikut merupakan Tabel hasil Confusion Matrix Naïve Bayes tanpa menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 13 Hasil Confusion Matrix Naïve Bayes tanpa optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	26	6	81.25%	39.57%
Pred. HIGH RISK	330	194	37.02%	
Class Recall	7.30%	97.00%		

Dari tabel 13 dapat diketahui bahwa persentase class precision 81.25% sebanyak 26 untuk prediksi LOW RISK yang sesuai dan 6 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 37.02% sebanyak 194 yang sesuai dan 330 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

TP = 194 , FP = 6 , TN = 26 , FN = 330

Akurasi=(TP+TN)*100%/(TP+TN+FP+FN) = (194+26)*100%/(194+26+6+330)

Akurasi= 39.57%

b. Algoritma Naïve Bayes dengan Optimasi

Berikut merupakan Tabel hasil Confusion Matrix Naïve Bayes dengan menggunakan optimasi *Optimize Selection (Evolutionary)*.

Tabel 14 Hasil Confusion Matrix Naïve Bayes dengan optimasi

	True LOW RISK	True HIGH RISK	Class Precision	Akurasi
Pred. LOW RISK	297	127	70.05%	66.55%
Pred. HIGH RISK	59	73	55.30%	
Class Recall	83.43%	36.50%		

Dari tabel 14 dapat diketahui bahwa persentase class precision 70.05% sebanyak 297 untuk prediksi LOW RISK yang sesuai dan 127 yang tidak sesuai . Sementara untuk prediksi HIGH RISK class precision 55.30% sebanyak 73 yang sesuai dan 59 yang tidak sesuai. Sehingga dari hasil tersebut diketahui bahwa :

TP = 73 , FP = 127 , TN = 297 , FN = 59

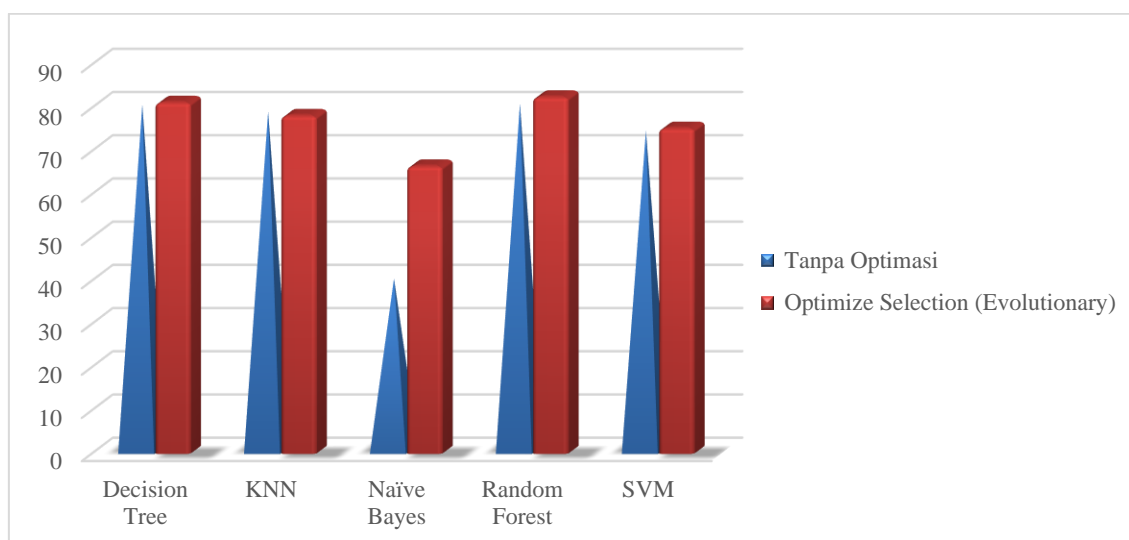
$$\text{Akurasi} = \frac{(TP+TN) \cdot 100\%}{(TP+TN+FP+FN)} = \frac{(73+297) \cdot 100\%}{(73+297+127+59)}$$

Akurasi = 66.55%

Tabel 15 Hasil Akurasi dan Kenaikan Akurasi setelah dilakukan optimasi

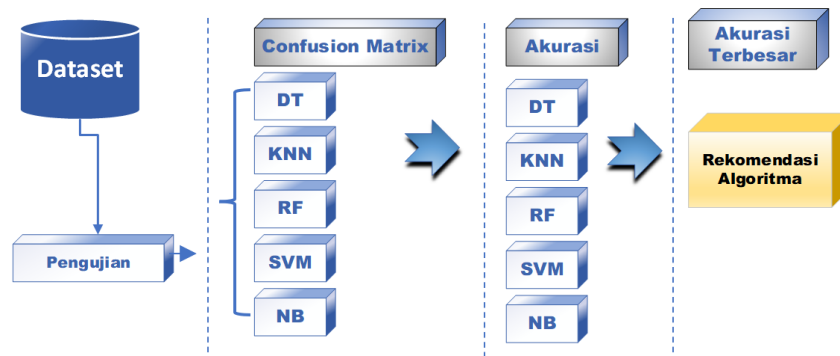
Algoritma	Akurasi		
	Tanpa Optimasi	Dengan Optimize Selection (Evolutionary)	Kenaikan Persentase
Decision Tree	79.86%	81.29%	1.43%
KNN	78.24%	78.24%	0%
Random Forest	80.04%	82.55%	2.51%
SVM	73.74%	75.54%	1.8%
Naïve Bayes	39.57%	66.55%	26.98%

Dari Tabel 15 menunjukkan bahwa algoritma *Decision Tree* mendapatkan akurasi total sebesar 81.29% , *KNN* 78.24% , *Random Frest* 82.55% , *SVM* 75.54%, dan *Naïve Bayes* 66.55% . Sementara untuk kenaikan akurasi setelah dilakukan proses optimasi *Decision Tree* menunjukkan kenaikan sebesar 1.43%, *KNN* 0%, *Random Frest* 2.51% , *SVM* 1.8% , dan *Naïve Bayes* 26.98%.



Gambar 2 Grafik perbandingan akurasi *Decision Tree*, *KNN*, *Naïve Bayes*, *Random Forest* dan *SVM*

Pada Gambar 2 menunjukkan bahwa algoritma *Random Forest* mendapatkan akurasi paling maksimal / tertinggi yaitu sebesar 82.55% namun untuk kenaikan pascaoptimasi algoritma *Naïve Bayes* mendapatkan kenaikan paling signifikan yaitu sebesar 26.98%. Sementara algoritma yang tidak terpengaruh terhadap *Optimize Selection (Evolutionary)* pada studi kasus penelitian ini adalah *KNN* yaitu tetap berada pada akurasi 78.24%. Kemudian Secara keseluruhan analisis evaluasi hasil Klasifikasi dilakukan dengan alur seperti yang ditunjukkan pada Gambar 3 (Perhitungan *Confusion Matrix*, Akurasi, pemilihan akurasi terbesar).



Gambar 3 Cara Evaluasi Perbandingan model klasifikasi

Pada penelitian ini teknik analisis perbandingan antar model Klasifikasi dilakukan dengan membandingkan nilai akurasi dari hasil evaluasi *Confusion Matrix* yang telah didapatkan pada proses pengujian sebelumnya. Dimana nilai terbesar dari akurasi model Klasifikasi akan menjadi model yang direkomendasikan sebagai model pendeteksi data profile resiko nasabah yang akan diterapkan pada proses *screening* awal pendataan nasabah baru. Dari data yang bisa kita lihat pada Tabel 15 menunjukkan bahwa algoritma *Random Forest* adalah algoritma yang direkomendasikan untuk digunakan. Karena *Random Forest* mendapatkan akurasi tertinggi dibandingkan dengan model algoritma yang lain yaitu dengan akurasi sebesar 82.55%.

4. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah sebagai berikut :

- a. Algoritma klasifikasi *Random Forest* menjadi algoritma yang mendapatkan akurasi total paling optimal pada studi kasus penelitian ini yaitu dengan akurasi 82.55%. Algoritma *Random Forest* direkomendasikan menjadi model yang akan digunakan untuk proses deteksi profile nasabah dalam studi kasus ini.
- b. Algoritma Naive Bayes adalah algoritma yang mendapatkan kenaikan paling besar pada studi kasus penelitian ini yaitu setelah dilakukan proses *Optimize Selection (Evolutionary)* kenaikan tersebut sebesar 26.98%. Namun Naive Bayes secara total memiliki akurasi yang terkecil dibandingkan dengan *Decision Tree*, *KNN*, *SVM*, dan *Random Forest*.
- c. Algoritma *KNN* merupakan satu-satunya Algoritma pada studi kasus penelitian ini yang tidak mengalami perubahan akurasi sebelum dan sesudah dilakukan proses optimasi.

Adapun beberapa saran terkait penelitian ini adalah sebagai berikut :

- a. Perlu dilakukan penelitian lebih lanjut terkait optimasi lain yang dapat dijadikan sebagai alat ukur untuk meningkatkan akurasi algoritma klasifikasi seperti penggunaan algoritma genetika dan partikel swarm.
- b. Perlu dilakukan penelitian lebih lanjut tentang proporsi data dan penggunaan optimasi yang tepat untuk masing-masing model. Sehingga dapat berlaku adil bagi masing-masing algoritma.

DAFTAR PUSTAKA

- [1] D. Broby, "The use of predictive analytics in finance," *The Journal of Finance and Data Science*, vol. 8, pp. 145–161, Nov. 2022, doi: 10.1016/j.jfds.2022.05.003.
- [2] M. Doumpos, C. Zopounidis, D. Gounopoulos, E. Platanakis, and W. Zhang, "Operational research and artificial intelligence methods in banking," *European Journal of Operational Research*, Apr. 2022, doi: 10.1016/j.ejor.2022.04.027.
- [3] A. Fauzi, "Analisis Data Bank Direct Marketing dengan Perbandingan Klasifikasi Data Mining Berbasis Optimize Selection (Evolutionary)," *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 1, p. 102, Mar. 2021, doi: 10.32493/informatika.v6i1.9291.
- [4] A. Turmudzy *et al.*, "Analisis Data Mining kelayakan Calon Debitur PT. Bank BRI dengan Menggunakan Metode Klasifikasi Naive Bayes," ISSN : 2407-3903, 2018.
- [5] A. H. Yunial, "Analisis Optimasi Algoritma Klasifikasi Support Vector Machine, Decision Trees, dan Neural Network Menggunakan Adaboost dan Bagging," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 3, p. 247, Sep. 2020, doi: 10.32493/informatika.v5i3.6609.
- [6] I. Oktanisa and A. A. Supianto, "Perbandingan Teknik Klasifikasi dalam Data Mining untuk Bank Direct Marketing," vol. 5, no. 5, pp. 567–576, 2018, doi: 10.25126/jtiik20185958.
- [7] "Penerapan K-Nearest Neighbor Penentuan Pemberian Kredit (Ester Arisawati)|1", [Online]. Available: <http://ejournal.tunasbangsa.ac.id/index.php/jsakti>
- [8] A. Fauzi, S. Eresha, and T. Selatan, "Analisis Potensi Dana Retail pada Nasabah PT. Bank Tabungan Negara (Persero), Tbk. Dengan Metode Decision Tree dan Naive Bayes Berbasis Optimize Selection (Evolutionary) (Study Kasus: PT. Bank Tabungan Negara Kantor Kas Season City)." [Online]. Available: <http://ejournal.urindo.ac.id/index.php/administrasimanajemen>
- [9] B. Bawono and R. Wasono, "Perbandingan Metode Random Forest dan naive bayes untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit," 2019. [Online]. Available: <http://prosiding.unimus.ac.id>
- [10] W. Etaiwi, M. Biltawi, and G. Naymat, "Evaluation of classification algorithms for banking customer's behavior under Apache Spark Data Processing System," in *Procedia Computer Science*, 2017, vol. 113, pp. 559–564. doi: 10.1016/j.procs.2017.08.280.
- [11] K. Alkhatib and S. Abualigah, "Predictive Model for Cutting Customers Migration from banks: Based on machine learning classification algorithms," in *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, Apr. 2020, pp. 303–307. doi: 10.1109/ICICS49469.2020.239544.
- [12] P. Maheswari and C. H. V. Narayana, "Predictions of loan defaulter - A data science perspective," Oct. 2020. doi: 10.1109/ICCS49678.2020.9277458.
- [13] Durgesh Kumar Singh and Noopur Goel, *Analysing Data Mining Techniques on Bank Customers for Credit Score*, ICRITO 2020., vol. June 4-5, 2020. Amity University, Noida, India: IEEE, 2020.
- [14] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 187–192, Feb. 2021, doi: 10.29207/resti.v5i1.2813.
- [15] D. C. Corrales, A. Ledezma, and J. C. Corrales, "From theory to practice: A data quality framework for classification tasks," *Symmetry (Basel)*, vol. 10, no. 7, Jul. 2018, doi: 10.3390/sym10070248.
- [16] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah

- Hujan Berpotensi Banjir,” 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [17] S. M. B. and J. W. B. Charles A. Worrell, “Comparison of data classification methods for predictive ranking of banks exposed to risk of failure”.
- [18] B. Ruyu, H. Mo, and L. Haifeng, “A Comparison of Credit Rating Classification Models Based on Spark- Evidence from Lending-club,” in *Procedia Computer Science*, 2019, vol. 162, pp. 811–818. doi: 10.1016/j.procs.2019.12.054.
- [19] T. R. Rivanthio and M. Ramdhani, “Penerapan Teknik Clustering Data Mining untuk Memprediksi Kesesuaian Jurusan Siswa (Studi Kasus SMA PGRI 1 Subang),” *Faktor Exacta*, vol. 13, no. 2, p. 125, Aug. 2020, doi: 10.30998/faktorexacta.v13i2.6588.