

## Analisis Data Bank Direct Marketing dengan Perbandingan Klasifikasi Data Mining Berbasis Optimize Selection (Evolutionary)

**Ahmad Fauzi**

Teknik Informatika, Universitas Pamulang, Jl, Raya Puspitek No. 46, Buaran, Kecamatan Pamulang, Kota Tangerang Selatan, Banten. 15310  
e-mail : dosen02621@unpam.ac.id

Submitted Date: January 26<sup>th</sup>, 2021

Revised Date: June 26<sup>th</sup>, 2021

Reviewed Date: June 26<sup>th</sup>, 2021

Accepted Date: June 15<sup>th</sup>, 2021

### Abstract

In determining marketing strategies, the bank performs a classification from a customer database, the database will be analyzed by a decision maker and this is not easy for a decision maker, because of the complexity of the vast data and the many attributes of the data owned, so that it becomes an obstacle and obstacle. in decision making. This of course can have a negative effect on the company's business processes because there will be delays in determining marketing strategies. Data mining method is a method that can classify large data to determine the level of accuracy of a database. In overcoming these problems, it is necessary to do a database analysis to determine the accuracy level of the database classification owned by the company. For this reason, in this study a classification process will be carried out with the Bank Direct Marketing dataset taken from the UCI Machine Learning Repository web, using the Naïve Bayes algorithm, K-Nearest Neighbor, Support Vector Machine with Optimize Selection (Evolutionary) optimization, the calculation process using a data mining application. namely Rapidminer 5.3, to find the highest accuracy value from the calculation algorithm. Test method with 10-fold cross validation. In this study, the classification results with the highest level of accuracy were obtained using Optimize Selection (Evolutionary) optimization, namely the Naïve Bayes algorithm 90.18%, then K-Nearest Neighbor 86.66%, and Support Vector Machine 89.40%.

Keyword: Bank Direct Marketing; Data Mining; Naïve Bayes; K- Nearest Neighbor; Support Vector Machine; Optimize selection (Evolutionary).

### Abstrak

Dalam menentukan startegi-startegi pemasaran bank melakukan sebuah klasifikasi dari sebuah database nasabah, database akan di analisa oleh seorang pengambil keputusan dan ini tidak mudah bagi seorang pengambil keputusan karena kompleksitas data yang luas serta banyaknya atribut dari data yang dimiliki, sehingga menjadi kendala dan penghambat dalam pengambilan keputusan. Hal ini tentunya dapat memberikan efek negatif dalam proses bisnis perusahaan karena akan terjadinya keterlambatan dalam penetapan starategi pemasaran. Metode data mining adalah metode yang dapat mengklasifikasikan sebuah data besar untuk mengetahui tingkat akurasi dari sebuah database. Dalam mengatasi permasalahan tersebut perlu dilakukan sebuah analisa database untuk mengetahui tingkat akurasi dari klasifikasi database yang dimiliki oleh perusahaan. Untuk itu dalam penelitian ini akan dilakukan proses klasifikasi dengan dataset Bank Direct Marketing yang diambil dari web UCI Machine Learning Repository, menggunakan algoritma Naïve Bayes, K-Nearest Neighbor, Support Vector Machine dengan optimasi Optimize Selection (Evolutionary), proses penghitungan menggunakan aplikasi data mining yaitu rapidminer 5.3, untuk mencari nilai akurasi tertinggi dari penghitungan algoritma tersebut. Metode test dengan 10-fold cross validation. Pada penelitian ini didapatkan hasil klasifikasi dengan tingkat akurasi tertinggi menggunakan optimasi Optimize Selection (Evolutionary) yaitu algoritma Naïve Bayes 90.18%, lalu K-Nearest Neighbor 86.66%, dan Support Vector Machine 89.40%.

Kata Kunci: Bank Direct Marketing, Data Mining, Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Optimize selection (Evolutionary).

## 1. Pendahuluan

Dalam perkembangan dunia perbankan terjadi begitu banyak perubahan-perubahan yang harus di hadapi dan disesuaikan oleh para pelaku usaha perbankan untuk menjaga eksistensi dalam sektor keuangan. Pada aktifitas keuangan setiap harinya database yang dimiliki oleh bank selalu mengalami penambahan dan disimpan menjadi database perusahaan, dari database tersebutlah bank melakukan analisa untuk melakukan strategi-strategi pemasaran serta penjualan produk. Dengan jumlah data nasabah atau pelanggan yang besar atau banyak bank membutuhkan sebuah metode dan sistem yang mampu memberikan keefektifan dan efisiensi waktu maupun biaya dalam melakukan proses tersebut, sehingga dari data tersebut akan didapati klasifikasi sebagai prediksi tingkat keberhasilan pemasaran produk dari para nasabah untuk menjadi data dalam proses marketing. Metode data mining adalah metode yang dapat mengklasifikasikan sebuah data besar untuk mengetahui tingkat keberhasilan penawaran dari sebuah database yang dimiliki oleh bank.

Dalam data mining terdapat berbagai jenis metode penghitungan yaitu algoritma Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Neural Network dan sebagainya.

Pada prosesnya nilai akurasi yang dihasilkan oleh algoritma data mining tersebut bukanlah hasil yang final atau akhir namun dari hasil tersebut dapat ditingkatkan kembali dengan menambahkan metode optimasi seperti Particle Swarm Optimization, dan Optimize Selection (Evolutionary) serta masih banyak yang metode optimasi lainnya yang dalam beberapa penelitian terbukti mampu meningkatkan hasil akurasi pada algoritma data mining. Pada penelitian ini penulis menggunakan sebuah metode optimasi Optimize Selection (Evolutionary) untuk meningkatkan hasil akurasi pada algoritma Naïve Bayes, K-Nearest Neighbor dan Support Vector Machine dalam dataset Bank Direct Marketing yang diambil dari UCI machine learning repository.

Dari penelitian sebelumnya terdapat beberapa literatur yang melakukan perbandingan metode data mining dalam mengklasifikasikan sebuah data contohnya adalah penelitian yang dilakukan oleh penulis sendiri dalam analisa potensi dan retail pada Nasabah PT. Bank Tabungan Negara (Persero) Tbk. Metode

Decision Tree dan Naïve Bayes berbasis Optimize Selection (Evolutionary) dengan validasi menggunakan 10fold cross-validation, dalam pengujiannya menghasilkan sebuah perbandingan nilai akurasi tertinggi antara metode Decision Tree dan Naïve Bayes dimana nilai akurasi decision tree setelah dioptimasi 94,01%, sementara naïve bayes setelah dioptimasi 94,47%, (Ahmad Fauzi, T. 2019).

Penelitian selanjutnya yaitu Komparasi Metode Data Mining Untuk Prediksi Pemasaran Bank Secara Langsung (Amalia, H. 2016), penelitian dilakukan dengan data set 45.212 record, penelitian menggunakan empat metode yaitu algoritma c4.5, neural network, naïve bayes dan SVM tanpa metode optimasi, setelah dilakukan pengujian didapati hasil akurasi algoritma C4.5 88,48%, neural network 88,41%, naïve bayes 86,71%, dan SVM 89,27%.

Selanjutnya penelitian Data Mining untuk Bank Telemarketing menggunakan Kombinasi Algoritma Naïve Bayes dan Algoritma Genetik (Ahmad Asifudin Aqhom, K. D. 2019), pada penelitian ini dataset diambil dari data public UCI repository. Pengujian dengan algoritma Naïve Bayes memiliki hasil nilai akurasi 86.71%, sementara jika menggunakan cross validation dengan algoritma genetik menghasilkan nilai akurasi tertinggi 90.27%

Penelitian dengan Perbandingan Teknik Klasifikasi dalam Data Mining untuk Bank Direct Marketing (Irvi Oktanisa, A. A. 2018), pada penelitian ini melakukan pengujian menggunakan tools Orange dengan sembilan teknik klasifikasi yaitu SVM, adaboost, cosntant, KNN, Tree, Random Forest, Stochastic Gradient Descent, dan CN2 Rule, untuk mengklasifikasi respon pelanggan pada dataset Bank Direct Marketing, dari hasil pengujian masing-masing metode didapati nilai akurasi tertinggi yaitu pada klasifikasi stochastic gradient descent dengan nilai akurasi 0,972 (97,2%) serta memiliki hasil visualisasi yang lebih jelas dalam mengklasifikasikan target dari dataset Bank Direct Marketing.

Dan penelitian data mining dalam telemarketing bank menggunakan algoritma Naïve Bayes dikombinasikan dengan algoritma Genetik (Ahmad Asifuddin Aqhom, K. D. 2019), pada penelitian ini dataset diambil dari data public UCI repository. Hasil akurasi dari kombinasi algoritma Naïve Bayes 86.71% dan

hasil akurasi dengan cross validation ditambah algoritma Genetik memiliki akurasi tertinggi 90.27%

Pada penelitian yang lain oleh Yunial, A. H (2020). Pengaruh dari optimasi metode Adaboost dan Bagging terhadap hasil algoritma Support Vector Machine, Decision Tree, dan Neural Network. Dari penghitungan tersebut dapat diketahui bahwa metode optimasi Adaboost berhasil meningkatkan nilai akurasi algoritma Support Vector Machine 88.93% menjadi 89.10%, algoritma Decision Tree 90.24% menjadi 90.36%, dan algoritma Neural Network 88.53% menjadi 88.61%. Sementara optimasi metode Bagging hanya berhasil meningkatkan nilai akurasi algoritma Decision Tree dengan nilai akurasi 90.55% dan Neural Network menjadi 90.38%.

Dari beberapa penelitian atau literatur tersebut terbukti bahwa metode data mining dapat mengklasifikasikan dataset dalam jumlah yang besar, serta nilai akurasi yang didapat dari hasil data mining tersebut dapat dioptimasi menggunakan beberapa metode optimasi dengan proses penghitungan menggunakan beberapa aplikasi data mining. Sehingga dari penelitian atau literature tersebut penulis melakukan penelitian untuk mengetahui tingkat akurasi sebagai perbandingan bank dalam mengklasifikasi database yang dimilikinya. Adapun algoritma yang digunakan adalah algoritma Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine dengan metode optimasi menggunakan Optimize Selection (Evolutionary), data yang dijadikan dataset adalah data Bank Direct Marketing yang diambil dari UCI machine learning repository. Sementara aplikasi data mining yang digunakan adalah aplikasi Rapidminer 5.3.

## 2. Metode Penelitian

### 2.1. Analisis Kebutuhan

Dalam melakukan penelitian penulis menggunakan sebuah perangkat komputer baik perangkat keras maupun perangkat lunak. Adapun perangkat keras yang digunakan adalah komputer laptop dengan spesifikasi system Core i3, RAM 4 GB, sedangkan untuk perangkat lunak adalah operating system menggunakan windows 10, dan aplikasi data mining menggunakan Rapidminer 5.3.

Dalam penelitian ini Data yang ada pada penelitian ini adalah data sekunder dimana data sudah tersedia dari situs web uci machine learning sehingga tidak ada pengambilan data

langsung kepada perusahaan. Situs pengambilan data yang digunakan yaitu (<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>), terdapat 17 atribut dan satu atribut yang dijadikan prediksi. Berikut dataset yang telah diambil:

Tabel. 1 Dataset dan atribut Bank Direct Marketing

Atribut	Tipe	Keterangan
<b>y</b>	binominal	Output(Hasil Penawaran)
<b>age</b>	integer	Usia
<b>job</b>	polynomial	Pekerjaan
<b>marital</b>	polynomial	Status Pernikahan
<b>education</b>	polynomial	Pendidikan
<b>default</b>	binominal	Tunggakan Hutang(Kredit Macet)
<b>balance</b>	integer	Saldo Tabungan
<b>housing</b>	binominal	Pinjaman KPR
<b>loan</b>	binominal	Kepemilikan Kredit
<b>contact</b>	polynomial	Cara Menghubungi
<b>day</b>	integer	Hari&Tanggal dihubungi terakhir
<b>month</b>	polynomial	Bulan Terakhir dihubungi
<b>duration</b>	integer	Durasi waktu saat dihubungi
<b>campaign</b>	integer	Jumlah Waktu yang dihubungi
<b>pdays</b>	integer	Jarak hari untuk menghubungi nasabah setelah penawaran (Nasabah yang belum pernah dihubungi maka akan diberi tanda -1)
<b>previous</b>	integer	Jumlah panggilan kepada nasabah sebelum melakukan penawaran.
<b>poutcome</b>	polynomial	Proses penawaran yang telah dilakukan.

Dari data yang diambil penulis memilih dataset yang berjumlah 4.521, dengan 17 atribut di mana dari 17 atribut tersebut terdapat satu atribut prediksi, data yang diambil berupa file csv dan kemudian data diubah ke dalam format excel dengan cara mengconvert file csv ke excel sehingga data dapat tampil berupa tabel.

Selanjutnya data akan di upload ke dalam aplikasi Rapidminer 5.3 untuk dijadikan dataset dan diproses dengan metode algoritma yang telah ditentukan.

## 2.2. Rancangan Penelitian

Dalam melakukan penelitian ini pertama penulis menentukan topik penelitian yang akan dilakukan sehingga proses penelitian dapat berjalan. Kedua setelah menentukan topik penelitian penulis membuat batasan-batasan pada proses penelitian, untuk membuat penelitian tidak keluar dari topik yang dibahas. Ketiga penulis melakukan pemilihan metode klasifikasi untuk menyelesaikan masalah pada topik penelitian.

Pada langkah keempat penulis melakukan pengambilan data serta memeriksa data yang telah diambil yang selanjutnya akan dijadikan dataset penelitian sehingga proses penelitian mendapatkan hasil yang sesuai dengan topik masalah yang ditentukan.

## 2.3. Teknik Analisis

Dalam penelitian ini, penulis melakukan pengambilan data pada dataset yang terdapat pada situs UCI respository setelah dilakukan pengambilan penulis melakukan penyeleksian data dikarenakan data yang diunduh terdapat dua file dataset yaitu data bank dan bank-full dengan jumlah data yang berbeda dimana untuk dataset bank sebanyak 4.521 dan dataset bank-full sebanyak 45.211 sementara atribut dari kedua file tersebut sama berjumlah 17 atribut dengan satu atribut prediksi.

Setelah melakukan analisa data penulis menentukan data yang diambil adalah dataset bank dikarenakan sudah terjadi perbaikan dari data yang tidak sesuai, menurut keterangan publikasi dataset Bank Direct Marketing. Selanjutnya data bank akan diproses pada aplikasi Rapidminer 5.3 untuk dihitung dengan algoritma Naive Bayes, K-Nearest Neighbor (K-NN), Support Vector

Machine (SVM), dengan 10-fold cross validation ini merupakan standar terbaik dalam teknik pengambilan sampel secara random yang terdapat pada aplikasi rapidminer dengan 10

pembagian kemudian error rate akan dihitung berdasarkan pembagian tersebut.

Setelah nilai akurasi dari metode algoritma masing-masing tersebut dihitung, maka selanjutnya metode algoritma Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine, akan dioptimasi menggunakan Optimize Selection (Evolutionary), sehingga dari hasil optimasi tersebut akan dibandingkan untuk mengetahui tingkat akurasi tertinggi untuk dijadikan hasil dalam mengklasifikasikan dataset Bank Direct Marketing.

Tabel. 2 Perbandingan Hasil Akurasi

Metode	Accuracy		
	Tanpa Opimasi	+ Optimize Selection (Evolutionary)	Persentase Kenaikan
Naive Bayes			
KNN			
SVM			

## 3. Hasil

Dalam mencari hasil klasifikasi untuk menentukan tingkat akurasi tertinggi menggunakan aplikasi rapidminer 5.3 dengan 10-fold cross validation ini merupakan standar terbaik dalam teknik pengambilan sampel yang terdapat pada aplikasi rapidminer 5.3 (Gorunescu, 2011). Dengan sistem data secara random akan dihitung dengan pembagian 10 bagian atau partisi yang akan dibandingkan dengan perbandingan yang sama, kemudian error rate akan dihitung berdasarkan pembagian tersebut, dan hasil dari penghitungan rata-rata error rate dari 10 bagian tersebut menjadi nilai error rate secara keseluruhan.

Selanjutnya data akan diproses melalui import data dengan aplikasi rapidminer 5.3, dan dari meta data view yang dihasilkan secara sistem maka akan terlihat ada atau tidak missings dari dataset, serta akan terlihat juga parameter data yang menjadi sampel dari penelitian ini, berikut gambar meta data view pada dataset Bank Direct Marketing:

Role	Name	Type	Statistics	Range	Missings
prediction	y	binominal	mode = no (4000), least = yes (5)	no (4000), yes (521)	0
regular	age	integer	avg = 41.170 +/- 10.576	[19.000 ; 87.000]	0
regular	job	polynomial	mode = management (969), least = unemployed (128), services (417)	unemployed (128), services (417)	0
regular	marital	polynomial	mode = married (2797), least = c	married (2797), single (1196), di	0
regular	education	polynomial	mode = secondary (2306), least =	primary (678), secondary (2306),	0
regular	default	binominal	mode = no (4445), least = yes (7)	no (4445), yes (76)	0
regular	balance	integer	avg = 1422.658 +/- 3009.638	[-3313.000 ; 71188.000]	0
regular	housing	binominal	mode = yes (2559), least = no (1	no (1962), yes (2559)	0
regular	loan	binominal	mode = no (3830), least = yes (6	no (3830), yes (691)	0
regular	contact	polynomial	mode = cellular (2896), least = te	cellular (2896), unknown (1324),	0
regular	day	integer	avg = 15.915 +/- 8.248	[1.000 ; 31.000]	0
regular	month	polynomial	mode = may (1398), least = dec	oct (80), may (1398), apr (293), ju	0
regular	duration	integer	avg = 263.961 +/- 259.857	[4.000 ; 3025.000]	0
regular	campaign	integer	avg = 2.794 +/- 3.110	[1.000 ; 50.000]	0
regular	pdays	integer	avg = 39.767 +/- 100.121	[-1.000 ; 871.000]	0
regular	previous	integer	avg = 0.543 +/- 1.694	[0.000 ; 25.000]	0
regular	poutcome	polynomial	mode = unknown (3705), least =	unknown (3705), failure (490), otl	0

Gambar 1. Meya Data View

### 3.1. Hasil Klasifikasi Algoritma Naive Bayes

Setelah mengimport data ke dalam aplikasi rapidminer 5.3, selanjutnya algoritma Naive Bayes akan dihitung untuk mencari nilai akurasi, tanpa menggunakan metode optimasi ataupun dengan metode optimasi.

#### a. Algoritma Naive Bayes Tanpa Optimasi

Dengan penghitungan tanpa menggunakan metode optimasi nilai akurasi yang didapat algoritma Naive Bayes adalah 86.71%, berikut gambar dari hasil algoritma Naive Bayes pada aplikasi rapidminer 5.3:

```

PerformanceVector
PerformanceVector:
accuracy: 86.71% +/- 0.89% (mikro: 86.71%)
ConfusionMatrix:
True:  no  yes
no:   3660 261
yes:  340  260
precision: 43.48% +/- 2.73% (mikro: 43.33%) (positive class: yes)
ConfusionMatrix:
True:  no  yes
no:   3660 261
yes:  340  260
recall: 49.89% +/- 4.08% (mikro: 49.90%) (positive class: yes)
ConfusionMatrix:
True:  no  yes
no:   3660 261
yes:  340  260
AUC (optimistic): 0.844 +/- 0.019 (mikro: 0.844) (positive class: yes)
AUC: 0.844 +/- 0.019 (mikro: 0.844) (positive class: yes)
AUC (pessimistic): 0.844 +/- 0.019 (mikro: 0.844) (positive class: yes)
    
```

Gambar 2. Hasil Akurasi Algoritma Naive Bayes Tanpa Metode Optimasi

Di bawah ini merupakan table view yang merupakan Confusion Matrix dari algoritma Naive Bayes

Tabel 3. Confusion Matrix Algoritma Naive Bayes

accuracy: 86.71% +/- 0.89% (mikro: 86.71%)			
	true no	true yes	class precision
pred. no	3660	261	93.34%
pred. yes	340	260	43.33%
class recall	91.50%	49.90%	

Dari tabel 3 dapat dilihat persentase kesesuaian prediksi no dengan class precision 93.34% yaitu 3660 yang sesuai dan 261 yang tidak sesuai. Sementara untuk prediksi yes persentasi class precision 43.33% yaitu 340 yang tidak sesuai dan 260 yang sesuai.

Dari hasil tersebut dapat diketahui bahwa:

$$TP = 260$$

$$FP = 261$$

$$TN = 3660$$

$$FN = 340$$

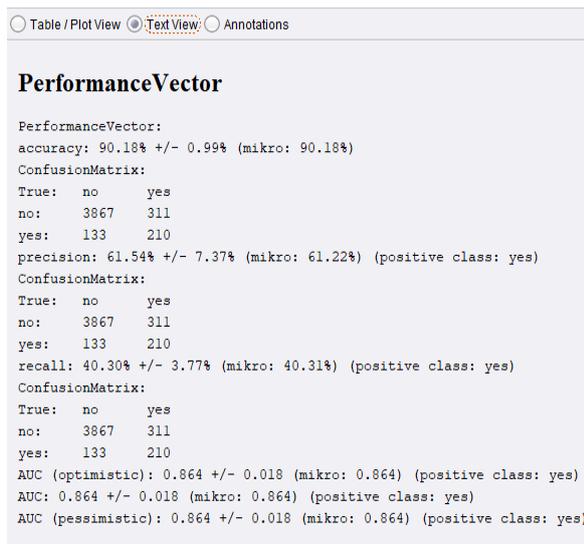
$$\text{Akurasi} = (TP+TN)/(TP+TN+FP+FN)$$

$$= (260+3660) / (260+3660+261+340)$$

$$= 0.8670648086706$$

Nilai akurasi yang didapat adalah 86.71%.

b. Algoritma Naive Bayes Dengan Optimasi.  
 Penghitungan algoritma Naive Bayes dengan metode optimasi Optimize Selection (Evolutionary) berhasil meningkatkan hasil akurasi dengan persentase kenaikan sebesar 3.47% sehingga hasil akurasi klasifikasi menjadi 90.18%. berikut Gambar hasil akurasi dari optimasi algoritma Naive Bayes dengan Optimize Selection (Evolutionary):



Gambar 3. Nilai Akurasi Optimasi Algoritma Naive Bayes Dengan Optimize Selection (Evolutionary)

Di bawah ini merupakan table view yang merupakan Confusion Matrix dari algoritma Naive Bayes yang telah dioptimasi.

Tabel 4. Confusion Matrix Algoritma Naive Bayes dengan Optimize Selection (Evolutionary)

<b>accuracy: 90.18% +/- 0.99% (mikro: 90.18%)</b>			
	true no	true yes	class precision
pred. no	3867	311	92.56%
pred. yes	133	210	61.33%
class recall	96.68%	40.31%	

Dari tabel 4 dapat dilihat persentase kesesuaian prediksi no dengan class precision 92.56% yaitu 3867 yang sesuai dan 311 yang tidak sesuai. Sementara untuk prediksi yes persentase class precision 61.22% yaitu 133 yang tidak sesuai dan 210 yang sesuai.

Dari hasil tersebut dapat diketahui bahwa:  
 TP = 210

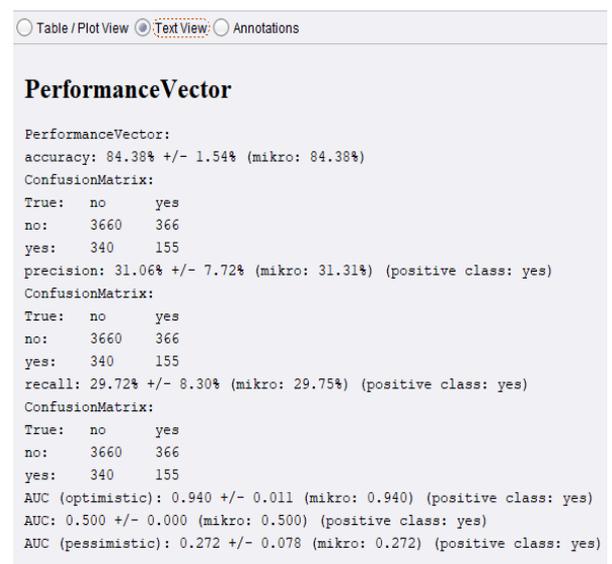
FP = 311  
 TN = 3867  
 FN = 133  
 Akurasi = (TP+TN)/(TP+TN+FP+FN)  
 = (210+3867) / (210+3867+311+133)  
 = 0.9017916390179  
 Nilai akurasi yang didapat adalah 90.18%.

### 3.2. Hasil Klasifikasi Algoritma K-Nearest Neighbor

Pada algoritma K-Nearest Neighbor dilakukan penghitungan akurasi dari data yang sudah dijadikan dataset pada aplikasi rapidminer 5.3, penghitungan dilakukan tanpa metode optimasi dan dengan metode optimasi.

#### a. Algoritma K-Nearest Neighbor (KNN) Tanpa Optimasi

Dengan penghitungan tanpa menggunakan metode optimasi nilai akurasi yang didapat algoritma K-Nearest Neighbor adalah sebesar 84.38%, berikut gambar dari hasil algoritma K-Nearest Neighbor pada aplikasi rapidminer 5.3:



Gambar 4. Hasil Akurasi Optimasi Algoritma K-Nearest Neighbor Tanpa Metode Optimasi

Di bawah ini merupakan table view yang merupakan Confusion matrix Algoritma K-Nearest Neighbor.

Tabel 5. Confusion matrix Algoritma K-Nearest Neighbor Tanpa Metode Optimasi

accuracy: 84.38% +/- 1.54% (mikro: 84.38%)			
	true no	true yes	class precision
pred. no	3660	366	90.91%
pred. yes	340	155	31.31%
class recall	91.50%	29.75%	

Dari tabel 5 dapat dilihat persentase kesesuaian prediksi no dengan class precision 90.91% yaitu 3660 yang sesuai dan 366 yang tidak sesuai. Sementara untuk prediksi yes persentase class precision 31.31% yaitu 340 yang tidak sesuai dan 155 yang sesuai.

Dari hasil tersebut dapat diketahui bahwa:

$$TP = 155$$

$$FP = 366$$

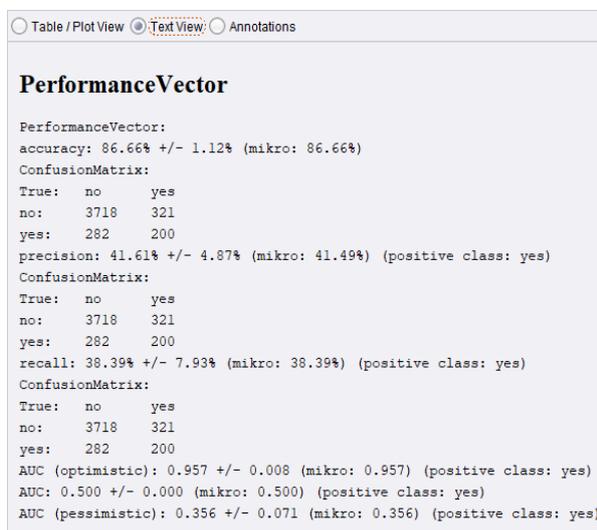
$$TN = 3660$$

$$FN = 340$$

$$\begin{aligned} \text{Akurasi} &= (TP+TN)/(TP+TN+FP+FN) \\ &= (155+3660) / (155+3660+366+340) \\ &= 0.8438398584383 \end{aligned}$$

Nilai akurasi yang didapat adalah 84.38%.

#### b. Algoritma K-Nearest Neighbor (KNN) Dengan Optimasi



Gambar 5. Nilai Akurasi Optimasi Algoritma K-Nearest Neighbor Dengan Metode Optimasi

Penghitungan algoritma K-Nearest Neighbor dengan metode optimasi Optimize Selection (Evolutionary) berhasil meningkatkan

hasil akurasi dengan persentase kenaikan sebesar 2.28% sehingga hasil akurasi klasifikasi menjadi 86.66%. berikut Gambar nilai akurasi dari optimasi algoritma K-Nearest Neighbor dengan Optimize Selection (Evolutionary).

Di bawah ini merupakan table view yang merupakan Confusion matrix Algoritma K-Nearest Neighbor.

Tabel 6. Confusion matrix Algoritma K-Nearest Neighbor Dengan Metode Optimasi

accuracy: 86.66% +/- 1.12% (mikro: 86.66%)			
	true no	true yes	class precision
pred. no	3718	321	92.05%
pred. yes	282	200	41.49%
class recall	92.95%	38.39%	

Dari tabel 6 dapat dilihat persentase kesesuaian prediksi no dengan class precision 92.05% yaitu 3718 yang sesuai dan 321 yang tidak sesuai. Sementara untuk prediksi yes persentase class precision 41.49% yaitu 282 yang tidak sesuai dan 200 yang sesuai.

Dari hasil tersebut dapat diketahui bahwa:

$$TP = 200$$

$$FP = 321$$

$$TN = 3718$$

$$FN = 282$$

$$\begin{aligned} \text{Akurasi} &= (TP+TN)/(TP+TN+FP+FN) \\ &= (200+3718) / (200+3718+321+282) \\ &= 0.8666224286662 \end{aligned}$$

Nilai akurasi yang didapat adalah 86.66%.

### 3.3. Hasil Klasifikasi Support Vector Machine

Pada algoritma Support Vector Machine dilakukan penghitungan nilai akurasi dari data yang sudah dijadikan dataset pada aplikasi rapidminer 5.3, penghitungan dilakukan tanpa metode optimasi dan dengan metode optimasi, yaitu:

#### a. Algoritma Support Vector Machine (SVM) Tanpa Optimasi

Dengan penghitungan tanpa menggunakan metode optimasi nilai akurasi yang didapat algoritma Support Vector Machine adalah sebesar 80.93%, berikut gambar dari hasil algoritma Support Vector Machine pada aplikasi rapidminer 5.3 :

**PerformanceVector**

PerformanceVector:  
 accuracy: 80.94% +/- 17.21% (mikro: 80.93%)  
 ConfusionMatrix:  
 True: no yes  
 no: 3595 457  
 yes: 405 64  
 precision: 13.65% (positive class: yes)  
 ConfusionMatrix:  
 True: no yes  
 no: 3595 457  
 yes: 405 64  
 recall: 12.12% +/- 28.34% (mikro: 12.28%) (positive class: yes)  
 ConfusionMatrix:  
 True: no yes  
 no: 3595 457  
 yes: 405 64  
 AUC (optimistic): 0.363 +/- 0.155 (mikro: 0.363) (positive class: yes)  
 AUC: 0.363 +/- 0.155 (mikro: 0.363) (positive class: yes)  
 AUC (pessimistic): 0.363 +/- 0.155 (mikro: 0.363) (positive class: yes)

Gambar 6. Nilai Akurasi Optimasi Algoritma Support Vector Machine Tanpa Metode Optimasi

Di bawah ini merupakan table view yang merupakan Confusion Matrix Algoritma Support Vector Machine.

Tabel 7. Confusion matrix Algoritma Support Vector Machine Tanpa Metode Optimasi

accuracy: 80.94% +/- 17.21% (mikro: 80.93%)			
	true no	true yes	class precision
<b>pred. no</b>	3595	457	88.72%
<b>pred. yes</b>	405	64	13.65%
<b>class recall</b>	89.88%	12.28%	

Dari tabel 7 dapat dilihat persentase kesesuaian prediksi no dengan class precision 88.72% yaitu 3595 yang sesuai dan 457 yang tidak sesuai. Sementara untuk prediksi yes persentase class precision 13.65% yaitu 405 yang tidak sesuai dan 64 yang sesuai.

Dari hasil tersebut dapat diketahui bahwa:

$$TP = 64$$

$$FP = 457$$

$$TN = 3595$$

$$FN = 405$$

$$\text{Akurasi} = (TP+TN)/(TP+TN+FP+FN)$$

$$= (64+3595) / (64+3595+457+405)$$

$$= 0.8093342180933$$

Nilai akurasi yang didapat adalah 80.93%.

b. Algoritma Support Vector Machine (SVM) Dengan Optimasi

**PerformanceVector**

PerformanceVector:  
 accuracy: 89.41% +/- 0.56% (mikro: 89.40%)  
 ConfusionMatrix:  
 True: no yes  
 no: 3952 431  
 yes: 48 90  
 precision: 66.88% +/- 10.23% (mikro: 65.22%) (positive class: yes)  
 ConfusionMatrix:  
 True: no yes  
 no: 3952 431  
 yes: 48 90  
 recall: 17.28% +/- 4.31% (mikro: 17.27%) (positive class: yes)  
 ConfusionMatrix:  
 True: no yes  
 no: 3952 431  
 yes: 48 90  
 AUC (optimistic): 0.811 +/- 0.114 (mikro: 0.811) (positive class: yes)  
 AUC: 0.811 +/- 0.114 (mikro: 0.811) (positive class: yes)  
 AUC (pessimistic): 0.811 +/- 0.114 (mikro: 0.811) (positive class: yes)

Gambar 7. Hasil Akurasi Optimasi Algoritma Support Vector Machine Dengan Metode Optimasi

Penghitungan algoritma Support Vector Machine dengan metode optimasi Optimize Selection (Evolutionary) berhasil meningkatkan hasil akurasi dengan persentase kenaikan sebesar 8.47% sehingga hasil akurasi klasifikasi menjadi 89.40%. berikut Gambar hasil akurasi dari optimasi algoritma Support Vector Machine dengan Optimize Selection (Evolutionary).

Di bawah ini merupakan table view yang merupakan Confusion Matrix Algoritma Support Vector Machine.

Tabel 8. Confusion Matrix Algoritma Support Vector Machine Dengan Metode Optimasi

accuracy: 89.41% +/- 0.56% (mikro: 89.40%)			
	true no	true yes	class precision
<b>pred. no</b>	3952	431	90.17%
<b>pred. yes</b>	48	90	65.22%
<b>class recall</b>	98.80%	17.27%	

Dari tabel 8 dapat dilihat persentase kesesuaian prediksi no dengan class precision 90.17% yaitu 3952 yang sesuai dan 431 yang tidak sesuai. Sementara untuk prediksi yes persentase class precision 65.22% yaitu 48 yang tidak sesuai dan 90 yang sesuai.

Dari hasil tersebut dapat diketahui bahwa:

$$TP = 90$$

$$FP = 431$$

$$TN = 3952$$

$$FN = 48$$

$$\text{Akurasi} = (TP+TN)/(TP+TN+FP+FN)$$

$$= (90+3952) / (90+3952+431+48)$$

= 0.8940499889404

Nilai akurasi yang didapat adalah 89.40%.

Dari hasil penghitungan masing-masing metode baik tanpa optimasi maupun dengan optimasi dapat dibuatkan perbandingan tingkat akurasi untuk dijadikan hasil atau acuan dalam klasifikasi Bank Direct Marketing, seperti pada tabel 9 di bawah ini:

Tabel 9. Akurasi Klasifikasi Bank Direct Marketing

Metode	Accuracy		
	Tanpa Optimasi	+ Optimize Selection (Evolutionary)	Persentase Kenaikan
Naive Bayes	86.71%	90.18%	3.47%
KNN	84.38%	86.66%	2.28%
SVM	80.93%	89.40%	8.47%

#### 4. Kesimpulan

Pada penelitian ini telah dilakukan penghitungan metode Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine, dengan metode optimasi Optimize Selection (Evolutionary), dari dataset Bank Direct Marketing. Maka dapat diambil kesimpulan berdasarkan tujuan dari penelitian ini antara lain:

- Algoritma Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine, dapat di implementasikan dalam pengklasifikasian dataset Bank Direct Marketing pada aplikasi rapidminer 5.3.
- Dengan penambahan metode optimasi Optimize Selection (Evolutionary) mampu meningkatkan hasil dari nilai akurasi pada penghitungan algoritma Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine. Nilai akurasi tertinggi adalah 90.18% dari hasil penghitungan algoritma Naive Bayes.
- Dalam peningkatan nilai akurasi dari algoritma Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine, dengan metode optimasi menghasilkan nilai akurasi yang berbeda yaitu algoritma Naive Bayes 3.47%, K-Nearest Neighbor 2.28%, dan Support Vector Machine menghasilkan peningkatan nilai akurasi tertinggi yaitu 8.47%.

#### 5. Saran

Dari Penelitian ini dapat diketahui tingkat akurasi algoritma Naive Bayes, K-Nearest

Neighbor, dan Support Vector Machine, baik tanpa optimasi ataupun dengan optimasi menggunakan aplikasi rapidminer 5.3, sehingga dapat dijadikan penelitian lanjutan dalam pengolahan metode data mining.

Saran-saran dalam penelitian yang telah dilakukan ini serta dalam penelitian berikutnya yaitu :

- Sebagai perbandingan untuk mencari nilai akurasi terbaik dapat dilakukan penghitungan dengan aplikasi data mining yang berbeda seperti WEKA, ORANGE, RATLLE dan lain sebagainya.
- Pada penelitian ini metode optimasi yang digunakan adalah Optimize Selection (Evolutionary), untuk penelitian lanjutan dapat digunakan metode optimasi lainnya yang terdapat pada aplikasi rapidminer.
- Dapat melakukan penghitungan dengan dataset Bank Direct Marketing menggunakan algoritma selain algoritma Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine, untuk hasil nilai akurasi yang lebih tinggi.
- Penelitian selanjutnya dapat menggunakan dataset Bank Direct Marketing bank-full, untuk mengetahui hasil tingkat akurasi terbaik.

#### Referensi

- Ahmad Fauzi, T. (2019). Analisis Potensi Dana Retail Pada Nasabah PT. Bank Tabungan Negara (Persero).Tbk, dengan Metode Decision Tree dan Naive Bayes Berbasis Optimize Selection (Evolutionary). *Jurnal Administrasi dan Manajemen Vol 9 No 1* 1-7.
- Amalia, H. (2016). Komparasi Metode Data Mining Untuk Prediksi Pemasaran Bank Secara Langsung. *SNIPTEK2016*, 1-6.
- Arisawati, E. (2017). Penerapan K-Nearest Neighbor Berbasis Genetic Algorithm Untuk Penentuan Pemberian Kredit. *Jurnal Sains Komputer & Informatika (J-Sakti)*, 1-11.
- Irvi Oktanisa, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 567-576.
- Sulaehani, R. (2016). Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination. *Jurnal Ilmiah ILKOM*, 1-8.
- Yunial, A. H. (2020). Analisis Optimasi Algoritma Klasifikasi Support Vector Machine, Decision Tree, dan Neural Network

- Menggunakan Adaboost dan Bagging. *Jurnal Informatika Universitas Pamulang*, 247-260.
- A.Elsalamony, H. (2014). Bank Direct Marketing Analysis of Data Mining Techniques . *International Jurnal of Computer Applications*, 12-22.
- Akhmad, E. P. (2019). Evaluasi Telemarketing Kartu Kredit Bank Menggunakan algoritma Genetika untuk Seleksi Fitur dan Naive Bayes. *Jurnal Aplikasi Palayaran dan Kepelabuhan*, 12-22.
- Bonggo Bawono, R. W. (2019). Perbandingan Metode random Forest dan Naive Bayes Untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit. *Seminar Nasional Edusaintek FMIPA UNIMUS*, 343-348.
- Ispandi. (2015). Penerapan Algoritma Genetika untuk Optimasi Parameter pada Support Vector Machine untuk Meningkatkan Prediksi Pemasaran Langsung. *Jurnal Of Intelligent Systems*, 115-119.
- Pungkas Subarkah, E. P. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing. *Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 139-148.
- Rousyati, a. N. (2021). Komparasi Adaboost dan Bagging dengan Naive Bayes pada DataSet Bank Direct Marketing . *Bianglala Informatika*, 12-16.