

# Prediksi Nasabah yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization

Alvino Dwi Rachman Prabowo

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang  
Jl. Nakula 1 No. 5-11, Semarang, 50131, Telp: (024) 3517261, Fax : (024) 325 0165  
E-mail : alvinodwi@gmail.com

---

## Abstrak

Deposito masih merupakan pilihan utama masyarakat untuk berinvestasi dan hal ini merupakan kesempatan bagi bank untuk menentukan strategi pemasaran dan promosi yang lebih efisien dan efektif. Atas dasar permasalahan tersebut, maka dilakukan penelitian untuk memprediksi nasabah yang berpotensi membuka deposito dengan menggunakan teknik data mining khususnya algoritma Naive Bayes berbasis PSO. PSO pada penelitian ini akan digunakan untuk seleksi fitur. Hasil dari penelitian ini, akurasi algoritma Naive Bayes adalah 82,19% dan akurasi algoritma Naive Bayes berbasis PSO adalah 89,70%. Penggunaan algoritma PSO ternyata meningkatkan akurasi sebesar 7,51% dan algoritma tersebut dapat digunakan untuk sistem pendukung keputusan pada penelitian ini.

**Kata kunci :** prediksi nasabah yang berpotensi membuka deposito, data mining, naive bayes, particle swarm optimization, seleksi fitur

## Abstract

Deposits are still the main choice for people to invest this time and it was an opportunity for the banks to determine marketing strategies and promotion of more efficient and effective. From the basis of these problems, the research conducted to predict the potential customers who will make deposits by using data mining techniques, especially Naive Bayes algorithm based PSO. PSO in this study will be used for feature selection. The results of this study, the accuracy of the Naive Bayes algorithm is 82,19% and the accuracy of the PSO based Naive Bayes algorithm is 89,70%. The use of PSO algorithm has increased by 7,51% and the accuracy of the algorithm can be used for decision support systems in this study.

**Keyword :** prediction of the potential customers who will open deposits , data mining, naive bayes, particle swarm optimization, feature selection

## I. PENDAHULUAN

Deposito masih merupakan primadona bagi masyarakat yang ingin berinvestasi namun dengan sedikit resiko. Ini bisa dilihat dari meningkatnya jumlah rekening deposito dari bulan Juni 2013 ke bulan Juli 2013 dari 2.940.178 pemilik rekening menjadi 2.955.539 pemilik rekening [1]. Menurut Undang – Undang Nomor 10 tahun 1998, deposito sendiri memiliki definisi yaitu simpanan yang penarikannya hanya dapat dilakukan pada waktu tertentu berdasarkan perjanjian nasabah penyimpan dengan bank [2]. Hal itu dikarenakan deposito memiliki tingkat suku bunga per tahun yang diterima lebih tinggi apabila dibandingkan dengan suku bunga giro atau simpanan tabungan biasa [3].

Untuk itu, suatu bank harus mengambil kesempatan ini dengan melakukan promosi dan strategi pemasaran yang efisien salah satunya dengan melakukan pemasaran langsung dan salah satu cara yang dapat digunakan yaitu memprediksi nasabah yang berpotensi membuka simpanan deposito pada bank tersebut. Prediksi tersebut dapat digunakan dengan menggunakan data-data nasabah yang sudah ada lalu diproses sehingga menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti statistik dan matematika. Proses tersebut dinamakan Data Mining [4]. Proses prediksi tersebut dapat menggunakan teknik klasifikasi.

Teknik klasifikasi adalah salah satu dari teknik data mining yang termasuk *supervised learning*. Supervised learning artinya proses pembentukan sebuah korespondensi (fungsi) menggunakan sebuah *training* dataset, dilihat sebagai sebuah "pengalaman masa lalu" dari sebuah model. Tujuannya adalah untuk memprediksi dari sebuah nilai (output) dari sebuah fungsi untuk setiap objek baru (input) setelah menyelesaikan proses *training* [5].

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan yaitu pembangunan model sebagai prototype untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada unsur objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya. [6].

Ada beberapa algoritma klasifikasi data mining yang dapat digunakan untuk strategi pemasaran dan promosi, seperti yang ditulis oleh Sergio Moro dan Raul M.S. Laureano diantaranya Naive Bayes (NB), Decision Trees (DT) dan Support Vector Machines (SVM) [7].

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama [6].

Namun Naive Bayes memiliki kelemahan yaitu atribut atau fitur independen sering salah dan hasil estimasi probabilitas tidak dapat berjalan optimal. Untuk mengatasi kelemahan tersebut salah satu caranya dengan metode pembobotan atribut untuk meningkatkan akurasi dari Naive Bayes tersebut [8] dan nantinya hasil bobot atribut tersebut akan digunakan untuk menyeleksi fitur dan atribut yang ada. Untuk melakukan metode pembobotan atribut tersebut, salah satunya menggunakan Particle Swarm Optimization (PSO).

Particle Swarm Optimization (PSO) adalah metode yang terinspirasi dari perilaku sosial

sekelompok burung dan ikan. Metode PSO mudah diterapkan dan dapat memilih parameter yang sesuai. Metode optimasi ini terbukti efektif dan berhasil digunakan untuk memecahkan masalah optimasi multidimensi dan multiparameter pada pembelajaran machine learning seperti Neural Network, SVM, dan algoritma teknik klasifikasi yang lain [9].

Pada penelitian sebelumnya dengan menggunakan data yang sama, namun menggunakan algoritma yang berbeda yaitu menggunakan algoritma C4.5 dan algoritma C4.5 berbasis PSO [10] dan algoritma Naive Baiyes [11] dengan hasil akurasi algoritma C4.5 sebesar 84,13% dengan AUC 0,837. Untuk hasil algoritma C4.5-PSO sebesar 86,09% dan AUC 0,883. Sedangkan untuk algoritma Naive Baiyes menghasilkan akurasi sebesar 87,03% dan AUC 0,813.

Pada penelitian ini, PSO akan digunakan untuk *feature selection* yaitu dengan memilih atribut terbaik dengan memilih atribut yang sudah diberikan bobot untuk meningkatkan hasil akurasi dari Naive Bayes. Hasil dari penelitian ini adalah meningkatkan akurasi prediksi nasabah yang berpotensi membuka simpanan deposito dan menunjukkan bahwa metode tersebut lebih baik daripada metode-metode yang sudah digunakan dalam penelitian sebelumnya.

## II. TINJAUAN PUSTAKA

### A. Related Research

Beberapa penelitian berkaitan tentang klasifikasi data mining yang digunakan untuk prediksi khususnya untuk memprediksi suatu hasil dan menggunakan algoritma optimasi telah banyak dilakukan beberapa diantaranya : Jie Lin dan Jiankun Yu dari Yunnan University of Finance and Economics telah melakukan penelitian mengenai penggunaan Weighted Naive Bayes berbasis algoritma Particle Swarm Optimization (PSO) yang digunakan untuk pembobotan atribut [12]. PSO digunakan untuk fungsi pencarian otomatis ketika mempertahankan integritas dari setiap atribut data. Metode PSO tersebut ternyata dapat meningkatkan akurasi dari algoritma Naive Bayes. Metode tersebut memberikan hasil yang diharapkan setelah menggunakan dataset UCI dan dapat digunakan secara layak dan efektif. Salah satu hasilnya adalah pada data set *transfusion*, jika menggunakan Naive Bayes

mendapatkan akurasi sebesar 0.751336898 atau 75.1336898%. Jika menggunakan PSO Weighted Naive Bayes Classifier (PSOWNBC) maka mendapatkan akurasi sebesar 0.771390374 atau 77.1390374%.

### B. Data Mining

Menurut Tan (2006) pada [6] mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery*.

### C. Naive Bayes

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [13].

Prediksi Bayes didasarkan pada formula teorema Bayes dengan formula umum sebagai berikut [6] :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Sedangkan Formulasi dari *Naive Bayes* untuk klasifikasi adalah [6]

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (2)$$

$P(Y|X)$  adalah probabilitas data dengan vektor  $X$  pada kelas  $Y$ .  $P(Y)$  adalah probabilitas awal kelas  $Y$ .  $\prod_{i=1}^q P(X_i|Y)$  adalah probabilitas independen kelas  $Y$  dari semua fitur dalam vektor  $X$ . Nilai  $P(X)$  selalu tetap sehingga dalam perhitungan prediksi nantinya kita tinggal menghitung bagian  $\frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)}$  dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen  $\prod_{i=1}^q P(X_i|Y)$  tersebut

merupakan pengaruh semua fitur dari data terhadap setiap kelas  $Y$ , yang dinotasikan dengan [6]:

$$P(X|Y = y) = \prod_{i=1}^q P(X_i|Y = y) \quad (3)$$

Setiap set fitur  $X = \{X_1, X_2, X_3, \dots, X_q\}$  terdiri atas  $q$  atribut.

Untuk atribut yang mempunyai tipe data kontinu, maka dinyatakan dalam formulasi berikut [6]:

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp \left\{ -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right\} \quad (4)$$

Dimana parameter  $\mu_{ij}$  bisa didapat dari mean atau rata-rata sampel  $X_i$  ( $\bar{x}$ ) dari semua data latih yang menjadi milik kelas  $y_j$  sedangkan  $\sigma_{ij}^2$  didapat dari varian sampel ( $s^2$ ) dari data latih.

### D. Particle Swarm Optimization (PSO)

PSO adalah metode pencarian populasi, yang berasal dari penelitian untuk pergerakan berkelompok burung dan ikan dalam mencari makan [9]. Seperti algoritma genetika, PSO melakukan pencarian menggunakan populasi (disebut *swarm*) dari individu-individu (disebut partikel) yang diperbarui dari setiap iterasi yang dilakukan. Untuk menemukan solusi yang optimal, setiap partikel bergerak ke arah posisi terbaik yang sebelumnya (*pbest*) dan posisi global terbaik (*gbest*). Kecepatan dan posisi partikel dapat terus diperbarui atau melakukan iterasi dengan persamaan berikut :

$$\begin{aligned} v_i^t &= v_i^{t-1} + c_1 \cdot r_1 \cdot (pbest_i^t - x_i^{t-1}) + c_2 \cdot r_2 \cdot (gbest^t - x_i^{t-1}) \\ x_i^t &= x_i^{t-1} + v_i^t \end{aligned} \quad (5)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (6)$$

$t = t + 1$

Pada penelitian ini, PSO digunakan untuk menyeleksi fitur (*feature selection*). Untuk menyeleksi fitur atau atribut tersebut, maka PSO menggunakan bobot atribut yang telah dihitung dan atribut yang telah diseleksi akan diprediksi menggunakan algoritma *Naive Bayes*. Untuk mendapatkan bobot tersebut, maka fungsi yang digunakan adalah [12] :

Dimana  $P(X|\sim C_i)$  merupakan jumlah probabilitas dari data X yang tidak termasuk dalam kelas  $C_i$ . Langkah-langkah untuk mendapatkan bobot tersebut adalah sebagai berikut [12].

1. Menyiapkan sampel dataset, siap untuk menghitung bobot dari record pertama,
2. Inisialisasi populasi (*swarm*)  $P(t)$  sehingga ketika  $t = 0$ , lokasi  $x_i(t)$  dari setiap partikel  $P_i$   $P(t)$  di suatu ruang yang luas menjadi acak.
3. Melalui setiap posisi partikel saat itu,  $x_i(t)$  mengevaluasi performa dari F.
4. Membandingkan kinerja masing-masing individu saat ini dengan individu yang mempunyai kinerja terbaik yang sejauh ini dimiliki jika  $F x_i t > pbest_i$  maka

$$pbest_i = F x_i t$$

$$X_{pbest_i} = x_i t$$

5. Membandingkan kinerja masing-masing partikel dengan kinerja partikel terbaik secara global, jika  $F x_i t > gbest_i$  maka

$$gbest_i = F x_i t$$

$$X_{gbest_i} = x_i t$$

Ubah kecepatan vector dari particle dengan menggunakan formula 5. Setelah mendapatkan kecepatannya lalu, tempatkan setiap partikel ke lokasi yang baru menggunakan formula 6. Kembali ke langkah nomor 2, ulangi perulangan hingga mendapatkan nilai yang konvergen dan mendapatkan bobot.

6. Jika perhitungan bobot dari record terakhir di data sampel selesai, maka akhiri. Jika tidak, siap untuk menghitung bobot dari record berikutnya dan kembali mengulangi dari langkah nomor 1.

Setelah itu, lalu hitung rata-rata bobot yang telah dihitung dari data sampel tersebut.

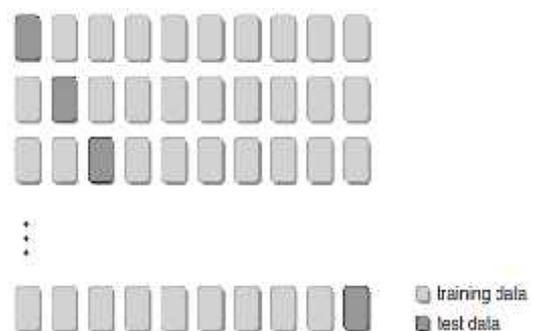
#### E. Cross Validation

*Cross validation* adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *training* dibagi secara *random* ke dalam beberapa bagian dengan perbandingan yang

$$F = P X|C_i - P(X|\sim C_i) \quad (7)$$

sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapat *error rate* secara keseluruhan [14]. *Cross validation* mempunyai banyak model, diantaranya *k-fold cross validation* dan *leave-one-out cross validation*.

Biasanya, 10-fold cross validation direkomendasikan untuk akurasi prediksi karena relatif rendah nilai bias dan varian. pada 10-fold cross validation, data dibagi menjadi 10 bagian terlebih dahulu secara acak dengan perbandingan yang sama, lalu hitung error rate setiap bagian, setelah itu diperoleh error rate secara keseluruhan dari menghitung rata-rata error rate dari semua bagian data [15].



Gambar 1 Ilustrasi 10-fold cross validation (Vercellis, 2009)

#### F. Confusion Matrix

*Confusion matrix* merupakan alat yang berguna untuk menganalisa seberapa baik klasifikasi tersebut dapat mengenali tuple-tuple dari kelas-kelas yang berbeda.

		Predicted class		Total
		yes	no	
Actual class	yes	TP	FN	P
	no	FP	TN	N
Total		P'	N'	P + N

Gambar 2 Confusion Matrix

Dari gambar diatas, TP dan TN menunjukkan

bahwa klasifikasi tersebut bernilai benar atau cocok dengan nilai asli, sedangkan FP dan FN menunjukkan bahwa klasifikasi tersebut bernilai salah atau tidak cocok dengan nilai asli.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

Rumus tersebut merupakan rumus dari *confusion matrix* yang digunakan mendapatkan hasil akurasi dari hasil klasifikasi yang sudah dilakukan.

### III. METODE YANG DIUSULKAN

#### A. CRISP-DM

##### 1) Fase Pemahaman Bisnis (*Business Understanding Phase*)

Pada tahap awal, penulis menggunakan dataset public berjudul *Bank Marketing* yang pernah digunakan oleh S.Moro, R. Laureano dan P. Cortez, yang berjudul *Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology* dari website University of California, Irvine (UCI) Machine Learning.

##### 2) Fase Pemahaman Data (*Data Understanding Phase*)

Dataset *Bank Marketing* tersebut terdiri dari 16 atribut prediktor dan 1 atribut label. Atribut prediktor adalah atribut yang dijadikan penentu nasabah yang hampir pasti membuka rekening deposito dan atribut label yaitu atribut yang dijadikan hasil *learning*.

Atribut prediktor yang digunakan pada penelitian ini adalah *age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous*, dan *poutcome*. Sedangkan atribut hasil atau label adalah *y*.

Tabel 1 Atribut pada Dataset

No	Atribut	Keterangan dan Nilai
1	Age	Umur nasabah
2	Job	Jenis pekerjaan nasabah (admin./unknown/unemployed/management/housemaid/entrepreneur/student/blue-collar/self

		employed/retired/technician/services)
3	Marital	Status pernikahan (married/divorced/single)
4	Education	Pendidikan Nasabah (unknown/secondary/primary/tertiary)
5	Default	Apakah mempunyai kredit gagal? (no/yes)
6	Balance	Rata-rata penghasilan pertahun dalam mata uang Euro
7	Housing	Apakah mempunyai kredit peminjaman rumah? (no/yes)
8	Loan	Apakah mempunyai pinjaman pribadi? (no/yes)
9	Contact	Jenis komunikasi yang digunakan (unknown/telephone/cellular)
10	Day	Tanggal terakhir menghubungi
11	Month	Bulan terakhir menghubungi (Januari/Februari/Maret/April/Mei/Juni/Juli/Agustus/September/Oktober/November/Desember)
12	Duration	Durasi terakhir menghubungi dalam detik
13	Campaign	Jumlah kontak yang dilakukan selama promosi ini dan untuk klien ini
14	Pdays	Jumlah hari yang berlalu setelah nasabah terakhir

		dihubungi dari promosi sebelumnya.
15	Previous	Jumlah kontak dilakukan sebelum promosi ini dan untuk klien ini.
16	Poutcome	Hasil dari promosi pemasaran sebelumnya ( <i>Unknown/Other/Failure/Success</i> )
17	Y	Apakah nasabah berlangganan deposito yang ditawarkan? (no/yes)

### 3) Fase Pengolahan Data (Data Preparation Phase)

Jumlah data yang telah diperoleh adalah 45211 record dengan 16 atribut prediktor dan satu atribut hasil seperti yang telah diterangkan pada tabel 3.1. Namun tidak semua data dan atribut dapat digunakan sehingga harus melewati tahap persiapan data (*data preparation*). Untuk mendapatkan data yang berkualitas maka dapat dilakukan teknik-teknik sebagai berikut (Vercellis,2009).

1. Data Validation, kualitas dari data mungkin tidak memuaskan dikarenakan ada data yang ganjil (outlier/noise), data yang tidak konsisten dan ada data yang tidak lengkap (missing value). Jumlah data awal adalah 45211 record dengan 16 atribut prediktor dan satu atribut hasil dan dari keseluruhan data tidak terdapat data yang *missing*, tidak konsisten ataupun noise.

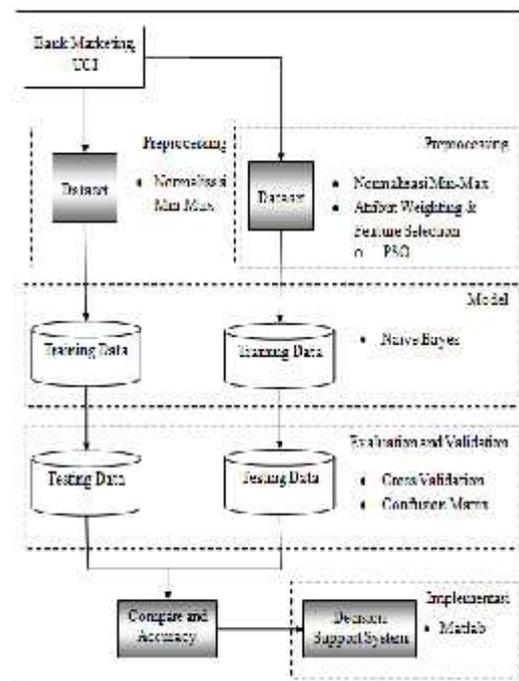
2. Data *integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma dapat dilakukan dengan melakukan beberapa transformasi. Dari dataset yang digunakan, karena untuk mempermudah implementasi kedalam Matlab, maka data nominal yang ada harus ditransformasi ke tipe data numerik terlebih dahulu dan data yang continue atau berlanjut seperti atribut *balance* dan *pdays* menggunakan normalisasi min-max dengan menggunakan formula,

$$V_i = \frac{v_i - \min_A}{\max_A - \min_A} \cdot \text{new max}_A - \text{new min}_A + \text{new min}_A \quad (9)$$

3. Data *size reduction and discretization*, mengurangi atribut atau variabel untuk meningkatkan akurasi dan memproses algoritma dengan lebih efisien tanpa mengorbankan kualitas dari hasil yang akan diperoleh. Dalam penelitian ini, untuk memproses algoritma dengan lebih efisien dan lebih cepat serta untuk meningkatkan akurasi maka digunakan *Feature Selection* dengan pembobotan atribut terlebih dahulu menggunakan algoritma PSO.

### 4) Fase Pemodelan (Modelling Phase)

Penelitian ini menggunakan dua metode yaitu algoritma *Naive Bayes Classifier* (NBC) dan algoritma NBC yang dikombinasikan dengan algoritma PSO yang digunakan sebagai pembobotan atribut untuk *feature selection*. Proses membandingkan dalam penelitian ini akan menggunakan *framework* RapidMiner versi 5.3.008 untuk menemukan algoritma yang akurat dan pengimplementasiannya dari hasil NBC PSO menggunakan Matlab R2010a.



Gambar 3 Model yang diusulkan dalam penelitian

Dimulai dari *preprocessing* yaitu dataset dinormalisasi dan ditransformasi terlebih dahulu, lalu akan dilakukan dua proses yaitu yang hanya dinormalisasi menggunakan normalisasi min-max dan proses yang menggunakan normalisasi min-max serta mencari atribut terbaik untuk meningkatkan akurasi dan efisiensi menggunakan PSO. PSO

berfungsi untuk attribute weighting yang digunakan untuk *feature selection*. Pada penelitian ini atribut yang berbobot 1 akan digunakan karena sangat berpengaruh terhadap akurasi dan efisiensi, sedangkan atribut yang berbobot 0 yang tidak mempengaruhi akurasi, akan dihilangkan untuk efisiensi pada algoritma yang digunakan untuk model penelitian.

Setelah melalui tahap *preprocessing*, dataset tersebut akan dimodel menggunakan algoritma *Naive Bayes*, dimana dataset yang menggunakan PSO dan yang tidak menggunakan PSO akan dibandingkan akurasi dan efisiensinya menggunakan Confusion matrix dan Curva ROC. Akurasi yang terbaik akan digunakan sebagai *Decision Support System* (DSS). DSS tersebut akan menggunakan program Matlab R2010a sebagai implementasinya.

#### 5) Fase Evaluasi dan Validasi (Evaluation Phase)

Dalam tahap ini dilakukan validasi dengan dan keakuratan hasil akurasi serta efisiensi algoritma pada model menggunakan beberapa teknik yang terdapat pada framework rapidminer versi 5.3.008 dan Matlab, antara lain menggunakan *Confusion Matrix*, untuk mengukur akurasi pada model yang digunakan, serta *10-fold Cross Validation* untuk validasi model.

## IV. PEMBAHASAN

Pada percobaan pertama menggunakan algoritma Naive Bayes terlebih dahulu tanpa menggunakan PSO untuk seleksi fitur. Percobaan tersebut menggunakan 45211 data dan 16 atribut prediktor yang sudah dinormalisasi menjadi data numerik terlebih dahulu, yaitu *age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, dan poutcome* serta digunakan juga *10-fold Cross Validation* untuk validasi dan confusion matrix untuk mengetahui tingkat akurasi.

Percobaan tersebut dilakukan selama 10 kali karena setiap percobaan mendapatkan nilai akurasi dan tingkat kesalahan yang berbeda yang disebabkan oleh data yang diambil secara acak pada setiap percobaan walaupun tetap dibagi kedalam 10 bagian yang hampir sama jumlahnya dan diambil nilai akurasi yang tertinggi.

Tabel 2 Hasil Percobaan Naive Bayes

Percobaan Ke -	Tingkat Akurasi	Tingkat Kesalahan	Waktu Proses
1	0,8215	0,1785	2 s
2	0,8216	0,1784	2 s
3	0,8211	0,1789	2 s
4	0,8212	0,1788	2 s
5	0,8217	0,1783	2 s
6	0,8211	0,1789	2 s
7	0,8214	0,1786	2 s
<b>8</b>	<b>0,8219</b>	<b>0,1781</b>	<b>2 s</b>
<b>9</b>	0,8218	0,1782	2 s
10	0,8213	0,1787	2 s

Berdasarkan tabel diatas yang merupakan hasil dari percobaan yang dilakukan di Matlab R2010a, maka nilai akurasi tertingginya adalah 0,8219 atau 82,19% dan tingkat kesalahannya adalah 0,1781 atau 17,81%. Nilai akurasi tersebut didapat dari *confusion matrix* data uji yang didapat dari 45211 record data yang sudah dibagi 10 bagian yang hampir sama. Hasil dari confusion matrix sebagai berikut.

Tabel 3 Hasil Confusion Matrix

Class/Hasil Prediksi	Prediksi 0	Prediksi 1	Total
Class 0	34309	2440	36749
Class 1	5613	2849	8462
Total	39922	5289	45211

Sedangkan hasil percobaan yang menggunakan algoritma Naive Bayes berbasis PSO menggunakan rapidminer dan dimulai dengan memberi nilai parameter *population size* 5-50 dengan kelipatan nilai 5 dan *maximum number of generation* 30-50 dengan kelipatan 10. *Population size* adalah jumlah individu tiap generasi, sedangkan *maximum number of generation* adalah jumlah generasi maksimum untuk menghentikan jalannya algoritma. Terdapat 30 hasil dari percobaan tersebut yang bisa dilihat seperti pada tabel 4 dibawah ini.

Tabel 4 Hasil Percobaan Naive Bayes berbasis PSO dengan maximum number of generation 30-50 dan population size 5-50 menggunakan Rapidminer

Pop size	Max num of gener ation	Akurasi	Waktu Proses	Jumlah Atribut yang terpilih
5	30	88,64%	5 m 46 s	8
5	40	88,41%	8 m 2 s	8
5	50	88,41%	9 m 57 s	8
10	30	88,68%	10 m 50 s	9
10	40	88,69%	12 m 47 s	8
10	50	88,24%	16 m 29 s	9
15	30	88,63%	13 m 46 s	8
15	40	88,68%	18 m 52 s	8
15	50	88,68%	23 m 50 s	8
20	30	89,31%	17 m 16 s	7
20	40	89,31%	23 m 1 s	7
20	50	89,31%	28 m 49 s	7
25	30	88,98%	23 m 9 s	8
25	40	88,98%	32 m 31 s	8
25	50	89,50%	38 m 18 s	6
30	30	89,69%	25 m 14 s	6
30	40	89,45%	35 m 32 s	6
<b>30</b>	<b>50</b>	<b>89,70%</b>	<b>43 m 38 s</b>	<b>5</b>
35	30	89,61%	33 m 8 s	8
35	40	89,56%	43 m 2 s	7
35	50	89,56%	53 m	7
40	30	88,93%	41 m 49 s	9
40	40	88,92%	54 m 7 s	8
40	50	88,85%	1 h 16 m 49 s	9
45	30	88,87%	47 m 11 s	9
45	40	88,88%	1 h 2 m 27 s	9
45	50	88,88%	1 h 28 m 2	9

			s	
50	30	89,65%	39 m	4
50	40	89,60%	1 h 44 s	7
50	50	89,12%	1 h 9 m 14 s	6

Dari tabel 4 diatas, diperoleh *setting* PSO terbaik sebagai pembobot adalah *population size* 30 dan *maximum number of generation* nya 50 dengan akurasi 89,70% dan atribut yang berbobot lebih dari 0 adalah 5 atribut. Jika lebih dari itu maka akan terjadi iterasi yang berlebihan yang berdampak pada bertambahnya waktu eksekusi dan hasil akurasi tidak meningkat. Pernyataan tersebut telah dibuktikan dengan contoh pada percobaan yang dilakukan dengan *population size* 30 dan *maximum number of generation* 100, hasil akurasi adalah 89,57% dengan waktu 1 jam 16 menit 34 detik.

Hasil akurasi tersebut didapatkan dari confusion matrix yang ditunjukkan oleh gambar 5 berikut.

	Actual	Predicted	Confusion
pred 00	3026	305	0,03%
pred 10	103	701	0,14%
Jumlah	3129	1006	

Gambar 4 Confusion Matrix Naive Bayes Berbasis PSO

Hasil atribut dari pembobotan atribut menggunakan PSO adalah seperti yang ditunjukkan pada gambar 6.

atribute	weight
age	0
job	0
marital	1
education	1
defcat	0
balance	0
housing	1
loan	0
contact	0
day	0
month	0
duration	0,856
campaign	0
pdays	0
previous	0
poutcome	1

Gambar 5 Bobot Atribut dari proses Algoritma Naive Bayes berbasis PSO

Berdasarkan hasil diatas, maka atribut yang digunakan ada 5 yaitu *marital*, *education*, *housing*, dan *poutcome* yang bernilai 1 dan *duration* yang bernilai 0,856. Atribut yang bernilai 0 tidak mendapat pembobotan sehingga secara tidak langsung dapat dihilangkan atau diseleksi karena atribut tersebut tidak optimal sehingga dapat meningkatkan hasil akurasi setelah dihilangkan.

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Pada penelitian ini, algoritma Particle Swarm Optimization (PSO) ternyata dapat mengoptimalkan algoritma Naive Baiyes dengan pembobotan atribut terlebih dahulu sehingga akurasi dapat meningkat. Dapat dibuktikan dengan akurasi algoritma Naive Baiyes adalah 82,19% dan untuk algoritma Naive Bayes berbasis Particle Swarm Optimization akurasi meningkat menjadi 89,70%, yang berarti akurasinya meningkat 7,51%.

Sedangkan perbandingan dengan penelitian sebelumnya yaitu penelitian [10] yang menggunakan dataset yang sama namun menggunakan algoritma yang berbeda yaitu algoritma C4.5 mempunyai akurasi sebesar 88,83% dan C4.5-PSO memperoleh akurasi sebesar 89,26% dan untuk Naive baiyes memperoleh akurasi sebesar 82,19% sedangkan algoritma Naive Baiyes-PSO memperoleh akurasi sebesar 89,70%, sehingga dapat disimpulkan pada penelitian ini bahwa algoritma Naive Baiyes-PSO merupakan algoritma terbaik yang digunakan untuk memprediksi dataset bank marketing yang digunakan untuk memprediksi nasabah yang

berpotensi membuka simpanan deposito pada bank.

### B. Saran

Penelitian ini memiliki banyak kekurangan dan masih dapat dikembangkan lagi pada penelitian berikutnya. Beberapa saran untuk pengembangan penelitian selanjutnya adalah

1. Penelitian ini membandingkan antara algoritma Naive Baiyes dengan Naive Baiyes berbasis PSO, untuk penelitian selanjutnya dapat dikembangkan dengan membandingkan algoritma klasifikasi lainnya seperti algoritma *Support Vector Machine*, *Decision Tree*, *Neural Network*, yang dapat dikombinasikan dengan algoritma optimasi lain seperti Genetic Algorithm (GA), Ant Colony Optimization (ACO), Adaboost dan algoritma optimasi yang lainnya.
2. Penelitian ini menggunakan dataset public bank marketing dari UCI Machine Learning Repository, untuk penelitian selanjutnya dapat menggunakan dataset yang lainnya, dari dataset private atau dataset public untuk menguji model algoritma tersebut.
3. Penelitian ini diharapkan dapat digunakan oleh bank yang dapat digunakan untuk menentukan kebijakan dalam melakukan promosi produk deposito terhadap nasabah sehingga dapat melaksanakan promosi yang seefisien mungkin.

## VI. DAFTAR PUSTAKA

- [1] Lembaga Penjamin Simpanan, "Distribusi Simpanan Bank Umum," Lembaga Penjamin Simpanan, 2013.
- [2] Dewan Perwakilan Rakyat Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 10 Tahun 1998," Dewan Perwakilan Rakyat Republik Indonesia, 1998.
- [3] S. Y. Baswara, "Analisa Nilai Hasil Investasi Deposito Rupiah, Deposito Dolar Amerika, dan Dinar Emas Dengan Emas Sebagai Alat Ukur," Skripsi Akuntansi Universitas Diponegoro, 2012.
- [4] D. T. Larose, *Discovering Knowledge In Data*. United States of America: John Wiley & Sons, Inc., 2005.
- [5] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Chennai, India: Springer, 2011.
- [6] E. Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB*, 1st ed. Yogyakarta, Indonesia: Andi, 2012.
- [7] S. Moro and R. M. S. Laureano, "Using Data

- Mining for Bank Direct Marketing: An Application of The CRISP-DM Methodology," Instituto Universitário de Lisboa, 2011.
- [8] N. A. Zaidi, J. Cerquides, M. J. Carman, and G. I. Webb, "Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting," *Journal of Machine Learning Research*, no. 14, pp. 1947-1988, 2013.
- [9] S.-W. Fei, Y.-B. Miao, and C.-L. Liu, "Chinese Grain Production Forecasting Method Based on Particle Swarm Optimization-based Support Vector Machine," in *Recent Patents on Engineering*, Shanghai, 2009, pp. 8-12.
- [10] S. Mujab, "PENCARIAN MODEL TERBAIK ANTARA ALGORITMA C4.5 DAN C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI PROMOSI DEPOSITO," Universitas Dian Nuswantoro Skripsi Teknik Informatika, 2013.
- [11] A. D. Pradibyo, "Analisa Perbandingan Kinerja Algoritma Klasifikasi Data Mining Untuk Prediksi Nasabah Bank yang Berpotensi Membuka Simpanan Deposito Berjangka," Universitas Dian Nuswantoro Skripsi Teknik Informatika, 2013.
- [12] J. Lin and J. Yu, "Weighted Naive Bayes Classification Algorithm Based on Particle Swarm Optimization," *IEEE*, 2011.
- [13] Kusri and E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta, Indonesia: Andi, 2009.
- [14] H. Leidyana, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PENENTUAN RESIKO KREDIT KEPEMILIKAN KENDARAAN BEMOTOR," *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, vol. 1, pp. 65-76, 2013.
- [15] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Waltham, United States of America: Morgan Kaufmann, 2012.