

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

Prediction of Potential Deposit of Customers Using Naive Bayes-Based Particle Swarm Optimization

Alvino Dwi Rachman Prabowo¹, Muljono²

^{1,2}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.com¹, muljono@dsn.dinus.ac.id²

Abstrak

Deposito masih merupakan pilihan utama bagi masyarakat untuk berinvestasi saat ini dan hal itu merupakan kesempatan bagi bank-bank untuk menentukan strategi pemasaran dan promosi yang lebih efisien dengan tidak terlalu banyak menggunakan biaya sehingga masyarakat tertarik untuk berinvestasi pada produk deposito dari bank tersebut. 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 feature selection yaitu dengan memilih atribut terbaik dengan memilih atribut yang sudah diberikan bobot sehingga dapat meningkatkan hasil akurasi dari prediksi menggunakan algoritma Naive Bayes. Hasil dari prediksi nasabah yang berpotensi membuka deposito dengan menggunakan Naive Baiyes memiliki akurasi sebesar 82,19%. Sedangkan prediksi yang menggunakan Naive Baiyes berbasis PSO memiliki akurasi sebesar 89,70%. Penggunaan algoritma PSO ternyata meningkatkan akurasi sebesar 7,51% dan algoritma Naive Baiyes berbasis PSO tersebut dapat digunakan untuk decision support system nasabah yang berpotensi membuka deposito karena menjadi model algoritma yang terbaik.

Kata kunci--prediksi nasabah yang berpotensi membuka deposito, *data mining*, *naive baiyes*, *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 by not using too much cost so that people interested in investing in the bank's deposit products. 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 is to select the best attributes by selecting the attributes that have been assigned weights so as to improve the accuracy of prediction results using Naive Bayes algorithm. The results of the predictive potential customers make deposits by using Naive Bayes has an accuracy of 82.19%. Other predictions using a Naive Bayes based PSO has an accuracy of 89.70%. The use of PSO algorithm turns out to improve the accuracy until 7.51% and a PSO-based Bayes Naive algorithm can be used for decision support system that could potentially open a deposit customers for being the best algorithm model.

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

1. 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. Ada dua pendekatan dalam melakukan promosi atau pemasaran yaitu dengan pemasaran massal dengan target yang acak atau dengan pemasaran langsung dengan target yang umum [4]. Untuk pemasaran massal masih banyak kelemahannya yaitu respon dari nasabah terhadap pemasaran tersebut masih sangatlah rendah yaitu masih dibawah 1% dari studi yang sama [5]. Sedangkan pemasaran secara langsung dengan target yang sudah spesifik lebih menarik untuk dilakukan karena lebih efisien dari segi biaya dan sumber daya pemasaran [5]. Namun pemasaran secara langsung memiliki beberapa kelemahan. Beberapa diantaranya adalah memicu sikap negatif seseorang terhadap bank karena mereka merasa terganggu karena diulas hal-hal yang merasa privasi bagi mereka [6]. Dengan pemasaran langsung, 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 [7].

Data Mining sendiri memiliki akar yang panjang dari beberapa bidang ilmu seperti kecerdasan buatan, *machine learning*, statistik, database dan juga *information retrieval* [8] serta merupakan dari teknologi *business intelligence*. Data mining dapat menyelesaikan beberapa pekerjaan dan dibagi menjadi empat kelompok yaitu model prediksi (*prediction modelling*), analisis kelompok (*cluster analysis*), analisis asosiasi (*association analysis*), dan deteksi anomaly (*anomaly detection*) [9]. 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* [10]. 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 [9].

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 Naïve Bayes (NB), Decision Trees (DT) dan Support Vector Machines (SVM) [11]. 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 [9]. Namun Naive Bayes memiliki kelemahan yaitu attribut atau fitur independen sering salah dan hasil estimasi probabilitas tidak dapat berjalan optimal. Ada dua cara untuk mengatasi masalah tersebut yaitu menggunakan metode Semi-Naive Bayes. Metode ini bertujuan untuk meningkatkan akurasi Naive Bayes dengan meringankan asumsi independensi bersyarat antara attribut yang diberikan label atau kelas. Cara kedua yaitu dengan

cara metode pembobotan attribut untuk meningkatkan akurasi dari Naive Bayes tersebut [12]. Untuk melakukan metode pembobotan attribut 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 [13].

Pada penelitian sebelumnya dengan menggunakan data yang sama, namun menggunakan algoritma yang berbeda yaitu menggunakan algoritma C4.5 dan algoritma C4.5 berbasis PSO [14] dan algoritma Naïve Baiyes [15] 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 Naïve Baiyes menghasilkan akurasi sebesar 87,03% dan AUC 0,813. Pada penelitian ini, PSO akan digunakan untuk *feature selection* yaitu dengan memilih attribut terbaik dengan memilih attribut 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.

2. METODE PENELITIAN

2.1 CRISP-DM

2.1.1 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.

2.1.2 Fase Pemahaman Data (*Data Understanding Phase*)

Dataset *Bank Marketing* tersebut terdiri dari 16 attribut prediktor dan 1 attribut label. Attribut prediktor adalah attribut yang dijadikan penentu nasabah yang hampir pasti membuka rekening deposito dan attribut label yaitu attribut yang dijadikan hasil *learning*. Attribut prediktor yang digunakan pada penelitian ini adalah *age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, dan outcome*. Sedangkan attribut hasil atau label adalah *y*.

2.1.3 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.

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

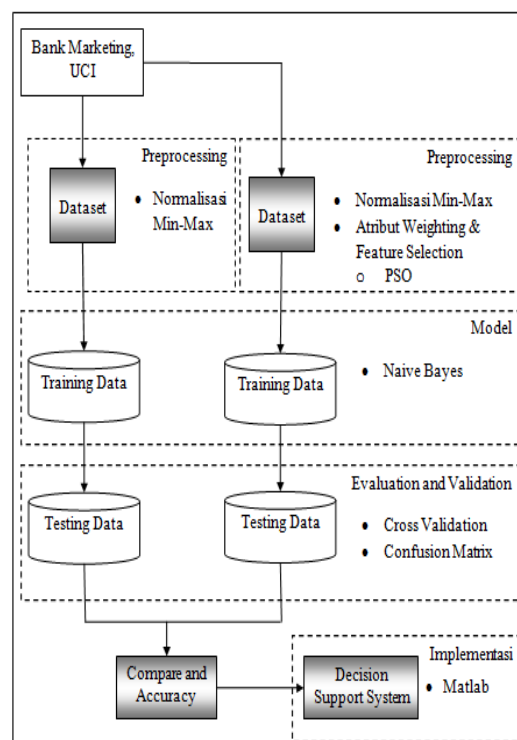
ditransformasi ke tipe data numerik terlebih dahulu dan data yang continue atau berlanjut seperti atribut balance dan pdays menggunakan normalisasi min-max dengan menggunakan formula,

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (\text{new max}_A - \text{new min}_A) + \text{new min}_A \quad (1)$$

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.

2.1.4 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 1 Model yang diusulkan dalam penelitian

Dimulai dari *preprocessing* itu 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 *Confussion matrix* dan *Curva ROC*. Akurasi yang terbaik akan digunakan sebagai *Decision Support System* (DSS). DSS tersebut akan menggunakan program Matlab R2010a sebagai implementasinya.

2.1.5 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 *Confussion Matrix*, untuk mengukur akurasi pada model yang digunakan, serta *10-fold Cross Validation* untuk validasi model.

2.2 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 [8].

Prediksi Bayes didasarkan pada formula teorema Bayes dengan formula umum sebagai berikut [9] :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (2)$$

Kaitan antara Naive Bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti adalah fitur-fitur yang menjadi *input* dalam klasifikasi. Jika X merupakan vektor masukan yang berisi fitur dan Y adalah label kelas. *Naive Bayes* dituliskan dengan $P(Y|X)$. Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini juga disebut probabilitas akhir (*posterior probability*) untuk Y, sedangkan $P(Y)$ disebut probabilitas awal (*prior probability*) Y.

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir $P(Y|X)$ pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari nilai Y' dengan memaksimalkan nilai $P(Y'|X')$ yang didapat.

Formulasi dari *Naive Bayes* untuk klasifikasi adalah [9]

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

2.3 Particle Swarm Optimization (PSO)

PSO adalah metode pencarian populasi, yang berasal dari penelitian untuk pergerakan berkelompok burung dan ikan dalam mencari makan [13]. 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*).

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*. Langkah-langkah untuk mendapatkan bobot tersebut adalah sebagai berikut [16].

1. Menyiapkan sampel dataset, siap untuk menghitung bobot dari record pertama.
2. Inilialisasi populasi (*swarm*) $P(t)$ sehingga ketika $t = 0$, lokasi $x_i(t)$ dari setiap partikel $P_i \in P(t)$ 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

$$\begin{cases} pbest_i = F(x_i(t)) \\ X_{pbest_i} = x_i(t) \end{cases} \quad (4)$$

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

$$\begin{cases} gbest_i = F(x_i(t)) \\ X_{gbest_i} = x_i(t) \end{cases} \quad (5)$$

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

3. HASIL DAN 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 1 Hasil Percobaan *Naive Bayes*

Perco- baan Ke -	Tingkat Akurasi	Tingkat Kesalaha n	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
8	0,8219	0,1781	2 s
9	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 2 Hasil Confusion Matrix

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

Jumlah *True Positive* (TP) pada confusion matrix data uji diatas adalah 34309 data yang diklasifikasikan sebagai 0 atau TIDAK BERPOTENSI, *True Negative* (TN) adalah 2849 data yang diklasifikasikan sebagai 1 atau BERPOTENSI, *False Positive* (FP) adalah 2440 data yang diklasifikasi sebagai TIDAK BERPOTENSI namun diprediksi sebagai BERPOTENSI, dan *False Negative* (FN) adalah 5613 data yang diklasifikasi sebagai BERPOTENSI namun diprediksi sebagai TIDAK BERPOTENSI.

Percobaan 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 3 dibawah ini.

Tabel 3 Hasil Percobaan Naive Bayes berbasis PSO dengan maximum number of generation 30-50 dan population size 5-50 menggunakan Rapidminer

Percobaan ke-	Population size	Max number of generation	Akurasi	Waktu Proses	Jumlah Atribut yang terpilih
1	5	30	88,64%	5 m 46 s	8
2	5	40	88,41%	8 m 2 s	8
3	5	50	88,41%	9 m 57 s	8
4	10	30	88,68%	10 m 50 s	9
5	10	40	88,69%	12 m 47 s	8
6	10	50	88,24%	16 m 29 s	9
7	15	30	88,63%	13 m 46 s	8
8	15	40	88,68%	18 m 52 s	8
9	15	50	88,68%	23 m 50 s	8
10	20	30	89,31%	17 m 16 s	7

11	20	40	89,31%	23 m 1 s	7
12	20	50	89,31%	28 m 49 s	7
13	25	30	88,98%	23 m 9 s	8
14	25	40	88,98%	32 m 31 s	8
15	25	50	89,50%	38 m 18 s	6
16	30	30	89,69%	25 m 14 s	6
17	30	40	89,45%	35 m 32 s	6
18	30	50	89,70%	43 m 38 s	5
19	35	30	89,61%	33 m 8 s	8
20	35	40	89,56%	43 m 2 s	7
21	35	50	89,56%	53 m	7
22	40	30	88,93%	41 m 49 s	9
23	40	40	88,92%	54 m 7 s	8
24	40	50	88,85%	1 h 16 m 49 s	9
25	45	30	88,87%	47 m 11 s	9
26	45	40	88,88%	1 h 2 m 27 s	9
27	45	50	88,88%	1 h 28 m 2 s	9
28	50	30	89,65%	39 m	4
29	50	40	89,60%	1 h 44 s	7
30	50	50	89,12%	1 h 9 m 14 s	6

Dari tabel 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 akurasinya tidak meningkat. Pernyataan tersebut telah dibuktikan dengan contoh pada percobaan yang dilakukan dengan *population size* 30 dan *maximum number of generation* 100, hasil akurasinya adalah 89,57% dengan waktu 1 jam 16 menit 34 detik.

Tabel 4 Confusion Matrix Naive Bayes Berbasis PSO pada Rapidminer

Accuracy: 89.70% +/- 0.44% (mikro: 89.70%)			
	True 0.0	True 1.0	Class precision
Pred 0.0	38293	3026	92.68%
Pred 1.0	1629	2263	58.14%
Class recall	95.92%	42.79%	

Dari 45211 data, Jumlah *True Positive* (TP) pada confusion matrix data uji diatas adalah 38293 data yang diklasifikasikan sebagai 0 atau TIDAK BERPOTENSI, *True Negative* (TN) adalah 2263 data yang diklasifikasikan sebagai 1 atau BERPOTENSI, *False Positive* (FP) adalah 1629 data yang diklasifikasi sebagai TIDAK BERPOTENSI namun diprediksi sebagai BERPOTENSI, dan *False Negative* (FN) adalah 3026 data yang diklasifikasi sebagai BERPOTENSI namun diprediksi sebagai TIDAK BERPOTENSI.

Tabel 5 Bobot Atribut dari proses Algoritma Naive Bayes berbasis PSO

Attribute	Weight
Age	0
Job	0
Marital	1
Education	1
Default	0
Balance	0
Housing	1
Loan	0
Contact	0
Day	0
Month	0
Duration	0.856
Campaign	0
Pday	0
Previous	0
Poutcome	1

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.

Setelah mengetahui atribut yang akan digunakan, maka buat dataset baru yang hanya berisi 5 atribut tersebut yaitu *marital*, *education*, *housing*, *poutcome*, dan *duration*, dengan jumlah data tetap yaitu 45211 data. Dataset baru tersebut akan digunakan untuk diproses di Matlab R2010a agar dapat dibandingkan dengan algoritma Naive Bayes tanpa PSO sebelumnya.

Percobaan pada algoritma Naive Bayes berbasis PSO menggunakan Matlab R2010a sama dengan Naive bayes tanpa PSO yaitu dilakukan 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, berbeda dengan yang dilakukan pada Rapidminer, karena memiliki hasil yang tetap.

Tabel 6 Hasil Percobaan Naive Bayes berbasis PSO

Perco- baan Ke -	Tingkat Akurasi	Tingkat Kesalahan	Wakt u Proses
1	0,8968	0,1032	2 s
2	0,8968	0,1032	2 s
3	0,8969	0,1031	2 s
4	0,8968	0,1032	2 s
5	0,8967	0,1033	2 s
6	0,8968	0,1032	2 s
7	0,8968	0,1032	2 s
8	0,8966	0,1034	2 s
9	0,8970	0,1030	2 s
10	0,8968	0,1032	2 s

Dari hasil percobaan dari tabel diatas yang dilakukan di Matlab R2010a, maka nilai akurasi tertingginya adalah 0,8970 atau 89,7% dan tingkat kesalahannya adalah 0,1030 atau 10,3%. Nilai akurasi tersebut didapat dari *confusion matrix* data uji yang didapat dari 45211 record data yang sudah dibagi 10 bagian yang hampir sama.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Pada penelitian ini, algoritma Particle Swarm Optimization (PSO) ternyata dapat mengoptimalkan algoritma Naive Baiyes dengan pembobotan atribut terlebih dahulu sehingga akurasi dapat meningkat. Terbukti, pada dataset bank marketing, akurasi algoritma Naive Baiyes adalah 82,19% dan setelah menggunakan algoritma Particle Swarm Optimization untuk mengoptimalkan algoritma Naive Baiyes dengan pembobotan atribut, akurasi meningkat menjadi 89,70%, yang berarti akurasi meningkat 7,51%.

Sedangkan perbandingan dengan penelitian sebelumnya yaitu penelitian [14] yang menggunakan dataset yang sama namun menggunakan algoritma yang berbeda yaitu algoritma C4.5 dan C4.5-PSO memperoleh hasil algoritma C4.5 mendapatkan akurasi sebesar 88,83% sedangkan algoritma Naive baiyes memperoleh akurasi sebesar 82,19% sedangkan yang menggunakan algoritma PSO, algoritma C4.5-PSO mendapatkan akurasi sebesar 89,26% sedangkan algoritma Naive Baiyes-PSO memperoleh akurasi sebesar 89,70%, sehingga dapat disimpulkan 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.

4.2 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.

DAFTAR PUSTAKA

- [1] Lembaga Penjamin Simpanan, "Distribusi Simpanan Bank Umum," Lembaga Penjamin Simpanan, Jakarta, 2013.
- [2] Dewan Perwakilan Rakyat Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 10 Tahun 1998," Dewan Perwakilan Rakyat Republik Indonesia, Jakarta, 1998.
- [3] Satsya Yoga Baswara, "Analisa Nilai Hasil Investasi Deposito Rupiah, Deposito Dolar Amerika, dan Dinar Emas Dengan Emas Sebagai Alat Ukur," Skripsi Akuntansi Universitas Diponegoro, Semarang, 2012.
- [4] Charles X. Ling and Chenghui Li, "Data Mining for Direct Marketing: Problems and Solutions," in *Proceedings of the 4th KDD conference*, 1998, pp. 73-79.
- [5] Chuangxin Ou, Chunnian Liu, Jiajing Huang, and Ning Zhong, "On Data Mining for Direct Marketing," in *Proceedings of the 9th RSFDGrC conference*, 2003, pp. 491-498.
- [6] Carole Page and Ye Luding, "Bank Manager's Direct Marketing Dilemmas – Customer's Attitudes and Purchase Intention," *International Journal of Bank Marketing*, vol. 21, no. 3, pp. 147-163, 2003.
- [7] Daniel T. Larose, *Discovering Knowledge In Data*. United States of America: John Wiley & Sons, Inc., 2005.
- [8] Kusrini and Emha Taufiq Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta, Indonesia: Andi, 2009.
- [9] Eko Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB*, 1st ed. Yogyakarta, Indonesia: Andi, 2012.
- [10] Florin Gorunescu, *Data Mining Concepts, Models and Techniques*. Chennai, India: Springer, 2011.
- [11] Sergio Moro and Raul M.S. Laureano, "Using Data Mining for Bank Direct Marketing: An Application of The CRISP-DM Methodology," Instituto Universitário de Lisboa, Lisboa, 2011.
- [12] Nayyar A. Zaidi, Jesus Cerquides, Mark J. Carman, and Geoffrey I. Webb, "Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting," *Journal of Machine Learning Research*, no. 14, pp. 1947-1988, 2013.

- [13] Sheng-Wei Fei, Yu-Bin Miao, and Cheng-Liang 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.
- [14] Syaeful Mujab, "Pencarian Model Terbaik antara Algoritma C4.5 dan C4.5 Berbasis Particle Swarm Optimization untuk Prediksi Promosi Deposito," Universitas Dian Nuswantoro, Semarang, Skripsi Teknik Informatika 2013.
- [15] Abbie Dwi Pradibyo, "Analisa Perbandingan Kinerja Algoritma Klasifikasi Data Mining Untuk Prediksi Nasabah Bank yang Berpotensi Membuka Simpanan Deposito Berjangka," Universitas Dian Nuswantoro, Semarang, Skripsi Teknik Informatika 2013.
- [16] Jie Lin and Jiankun Yu, "Weighted Naive Bayes Classification Algorithm Based on Particle Swarm Optimization," *IEEE*, 2011.