

ANALISA NASABAH POTENSIAL TABUNGAN DEPOSITO BERJANGKA MENGGUNAKAN TEKNIK KLASIFIKASI DATA MINING

Candra Agustina

Universitas Bina Sarana Informatika

Sistem Informasi

Jl. Ringroad Barat, Ambarketawang, Gamping, Sleman, Yogyakarta

candra.caa@bsi.ac.id

Abstract

Time deposits are a product of a financial institution, which is currently increasing. The main target of this time deposit marketing is the old customers of the Bank. To increase the effectiveness of marketing customers are grouped into potential and non-potential customers. This means that potential customers have a greater chance to open a time deposit account. Customer data is taken from the UCI repository, originating from Banks in Portugal. Data is processed with rapidminer software using the Decision Tree method with Particle Swarm Optimization, Naïve Bayes with Particle Swarm Optimization and finally processed using Neural Network with Particle Swarm Optimization. Data processing results were compared and showed that the Naïve Bayes Algorithm with Particle Swarm Optimization had the highest accuracy of 97.04%. Therefore an application designed based on Naive Bayes with Particle Swarm Optimization. From the original attribute consisting of 20, only 9 attributes can be used so that the level of accuracy is high. Attributes used have values more than 0.500, while those that have these values are omitted. The design was created using the Unified Modeling Language (UML) and Visual Basic 6.0 to create an User Interface.

Keywords— Decision Tree, Naïve Bayes, Neural Network, Particle Swarm Optimization, Time Deposit

PENDAHULUAN

Perkembangan lembaga keuangan semakin meningkat, hal ini ditandai dengan munculnya lembaga keuangan. Sebagai contoh dengan berkembangnya BMT di Yogyakarta dilihat dari web indukbmt.co.id jumlah BMT tahun 2018 ini di Yogyakarta terdapat 28. Selain BMT masih banyak lagi lembaga keuangan lain. Tentu saja setiap lembaga keuangan mempunyai banyak produk untuk menarik pelanggan, seperti simpanan suka rela dan tabungan berjangka. Salah satu produk dari Bank adalah Deposito Berjangka Deposito berjangka adalah deposito yang mempunyai masa atau jangka waktu yang telah ditentukan dan dalam bilyet dapat dituliskan nama (Kasmir,2013). Untuk memasarkan produknya tentunya Bank harus terus melakukan pemasaran atau marketing. Target produk dari sebuah Bank adalah nasabah Bank itu sendiri, artinya sebelum memasarkan produk untuk non

nasabah sebuah Bank akan menawarkan produknya kepada masyarakat yang sudah menjadi nasabah.

Dalam menawarkan produk Deposito ini Bank tidak melakukannya ke semua nasabah, akan tetapi kepada nasabah yang dianggap potensial. Untuk menentukan nasabah dianggap potensial atau tidak ini sebuah Bank memerlukan sebuah system agar tepat sasaran. Salah satu yang dapat diterapkan adalah dengan menerapkan metode Data Mining. Dengan teknik ini proses menentukan nasabah potensial atau tidak dibantu dengan aplikasi komputer yang dibuat berdasarkan algoritma yang mempunyai akurasi tertinggi. Dengan strategi marketing yang terarah diharapkan biaya yang dikeluarkan semakin kecil sedangkan laba yang didapat semakin besar. Selain itu dengan algoritma yang tepat dapat ditentukan atribut yang dibutuhkan oleh aplikasi sehingga hasil yang didapat mempunyai akurasi yang tinggi.

TINJAUAN PUSTAKA

1. Data Mining

Definisi data mining adalah proses untuk menemukan dan memberikan gambaran berupa pola struktural dalam data sehingga dapat digunakan sebagai alat bantu untuk membantu memberikan penjelasan terhadap sekumpulan data, selain itu juga membuat prediksi dari data tersebut (Witten, Frank & Hall, 2011). Bidang data mining ini adalah bagian dari Knowledge Discovery Data (KDD) yaitu proses untuk meringkas sebuah informasi yang bermanfaat yang sebelumnya belum diketahui dan tidak terlihat (Brammer, 2007). Data mining juga digunakan untuk mengembangkan model yang dapat digunakan untuk memahami fenomena dari analisis data dan prediksi atau perkiraan (Maimon & Rokach, 2010).

Data mining berupa ringkasan pengetahuan dari keseluruhan data yang berjumlah besar (Han & Kamber, 2006) yang tersimpan dalam sebuah komputer, proses pada data mining lebih banyak berdasarkan pada teknologi informasi (Olson & Delen, 2008).

Knowledge Discovery Data terdiri dari 5 tahapan dalam prosesnya (Kusrini & Luthfi, 2009):

1. Langkah pertama adalah pemilihan data atau data selection.
Pada proses ini data secara keseluruhan perlu diseleksi yang akan digunakan dalam proses data mining. File tersebut disimpan dalam sebuah berkas terpisah dari data sebelumnya.
2. Pembersihan data
Dilakukan dengan membuang data yang tidak konsisten, membuang data yang berulang atau sama, memperbaiki data yang kurang lengkap.
3. Transformasi pada data
Data perlu ditransformasikan agar sesuai dengan algoritma data mining yang digunakan.
4. Data Mining
Dilakukan pemrosesan data sesuai dengan algoritma yang di pilih.
5. Evaluasi
Hasil pengolahan dari data mining kemudian di sajikan kedalam laporan atau bentuk yang mudah dimengerti oleh orang lain.

2. Naïve Bayes

Dalam klasifikasi atau pengelompokan komputer memiliki kemampuan hampir sama dengan pohon keputusan atau *decision tree* dan *neural network* (Kusrini & Luthfi 2009).

Aturan Bayes dapat digunakan untuk menghitung besarnya peluang sebuah kelas. Algoritma I Bayes dapat memberikan sebuah cara untuk menggabungkan peluang sebelumnya dengan ketentuan yang ditetapkan menjadi sebuah rumus yang dapat digunakan untuk menghitung peluang dari semua kemungkinan yang bisa terjadi.

Teorema Bayes dapat dirumuskan seperti berikut ini:

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

Keterangan :

- X : Data dengan kelas yang belum diketahui nilainya
- H : Hipotesis data X merupakan sebuah kelas khusus
- P(H|X) : Peluang hipotesis H berdasarkan pada kondisi dari nilai X
- P(H) : Peluang hipotesis dari H
- P(X|H) : Peluang terjadinya X berdasarkan kondisi pada H
- P(X) : Peluang dari X

Kemudian teorema tersebut disederhanakan menjadi Naïve Bayes dengan rumus sebagai berikut :

$$P(H|X) = P(X|H) P(X)$$

Keterangan :

Peraturan Bayes digunakan untuk menghitung hasil posterior dan probabilitas dari data yang sebelumnya. Dalam proses analisis dari komputer, pengelompokan paling akhir dihasilkan dengan cara menggabungkan prior dan posterior untuk menghasilkan probabilitas dengan menggunakan aturan bayes.

3. Neural Network

Algoritma Neural Network diawal kemunculannya dikembangkan pada *machine learning* untuk meniru syaraf pada otak manusia, yang dilakukan dengan mengkombinasikan elemen komputasi atau neurn dalam sebuah komputer yang saling terkait (Guidici & Figini, 2009).

Proses pengolahan data dalam algoritma *neural network* di lakukan dengan cara menyesuaikan masing-masing bobot *neuron* dengan arah mundur didasarkan pada nilai error dalam pemorsesan data Pada setiap proses yang dilakukan, bobot relasi dalam sebuah jaringan di rubah dan dimodifikasi yang berfungsi

untuk meminimalkan nilai dai MSE atau *Mean Squared Error* antaran nilai perkiraan dari sebuah jaringan dengan nilai *real* atau sebenarnya. Langkah memodifikasi hubungan antar jaringan tersebut dilakukan secara mundur ((Kusrini & Luthfi, 2009).

4. Algoritma C.4.5

Algoritma C4.5 pertama kali dikenalkan oleh J. Ross Quinlan, algoritma ini merupakan pengembangan dari algoritma ID3. Algoritma ID3 digunakan untuk membentuk pohon keputusan atau *decision tree*. Pohon keputusan atau *decision tree* dianggap sebagai algoritma paling akurat. Dalam pengelompokan pohon keputusan erdapat sebuah node yang membentuk akar pohon. Node akar ini tidak mempunyai inputan sama sekali tapi menghasilkan output. Adapun node lain yang bukan sebagai akar akan tetapi mempunya satu inputan disebut dengan *test node* atau node internal, sedangkan node lain adalah node daun yang merupakan nilai target yang paling tepat dari salah satu kelas (Maimon & Rokach, 2010).

Decision Tree atau pohon keputusan dihasilkan dengan cara membagi setiap nilai dari atribut menjadi cabang dalam setiap kemungkinan. Cara kerjanya adalah dengan menelusuri mulai dari akar sampai ke cabang-cabang hingga mencapai kelas suatu objek ditemukan. *Instance* di kelompokkan dengan mengarahkan mulai dari akar dampai pada daunnya sesuai dengan hasil tes yang dilakukan melalui node internal. Kemudian mengubah data pada pohon keputusan menjadi serangkaian aturan untuk menghasilkan sebuah keputusan. Algoritma ini sesuai diterapkan untuk mengatasi masalah pengelompokan dan data mining. C4.5 bisa memetakan seluruh nilai atribut menjadi kelas yang dapat di implementasikan untuk pengelompokan baru (Wu & Kumar,2009).

5. Penelitian Terkait

Penelitian ini merupakan penelitan lanjutan dari beberpaa penelitian yang sudah dilakukan para peneliti lain. Adapun penelitian yang pernah dilakukan menggunakan data set ini adalah:

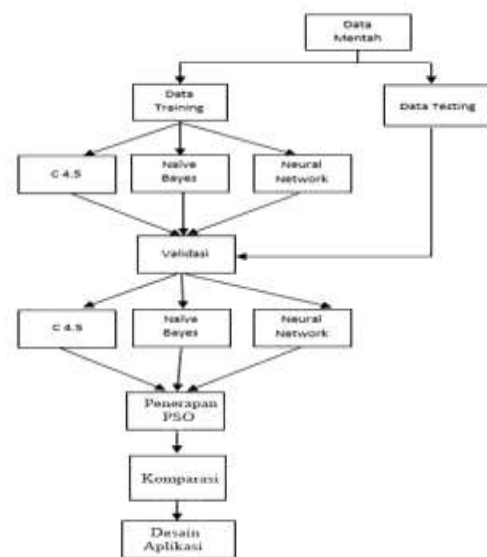
1. Moro et al, 2014 membuat Sistem Pendukung Keputusan untuk telemarketing Bank. Berhasil mengumpulkan data sebanyak 4521. Atribut yang digunakan sebanyak 16 ditambah 1 atribut hasil.
2. Dalam penelitian Ruhmi Suhaelani, 2016 melakukan penelitian menggunakan data set ini diolah dengan algoritma Naïve Bayes kemudian dilakukan seleksi fitur dengan metode *Backward Elimination*. Hasil yang diperoleh data set yang diolah menggunakan algoritma Naïve Bayes

mempunyai akurasi sebesar 89,08% setelah diseleksi fitus menggunakan *Backward Elimination* akurasi meningkat menjadi 90,69%.

3. Dalam penelitian berjudul Pencarian Model Terbaik Antara Algoritma C 4.5 Dan C 4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Promosi Deposito, saeful mujab melakukan penelitian dengan hasil data set yang diolah menggunakan Algoritma C 4.5 memperoleh akurasi sebesar 88.83% setelah diterapkan algoritma PSO menjadi 89,26%. (Mujab,2013)
 4. Dalam penelitian lain, data set diolah menggunakan Naïve Bayes dan Naïve Bayes Particle Swarm Optimization, akurasi yang didapatkan 82,19% setelah di tambahkan algoritma PSO meningkat menjadi 89,70% (Prabowo, 2014)
- Penelitan yang akan dilakukan mencari algoritma yang lebih tinggi dari penelitian yang telah dilakukan, yaitu menerapkan algortima Neural Network dan kemudian dioptimasi.

METODE PENELITIAN

3.1 Kerangka Penelitian



Gambar 1 Kerangka Penelitian

3.2 Metode Pengumpulan Data

University of California Irvine Machine Learning Repository adalah kumpulan database, teori dan data generator yang digunakan oleh komunitas machine learning untuk analisis empiris dari algoritma machine learning. Arsip yang ada telah buat pada tahun 1987 oleh David Aha dan mahasiswa pascasarjana universitas California Irvine. Arsip-arsip tersebut telah banyak digunakan sebagai sumber data

set primer dan telah dijadikan sebagai rujukan lebih dari 1000 kali, sehingga menjadikannya sebagai salah satu dari 100 paper yang paling banyak dikutip dalam bidang ilmu komputer.

Data yang digunakan dalam penelitian ini diambil dari [link https://archive.ics.uci.edu/ml/datasets/bank+marketing](https://archive.ics.uci.edu/ml/datasets/bank+marketing) yang terdiri dari 17 atribut. (Moro, et al, 2014)

HASIL DAN PEMBAHASAN

a) Hasil Klasifikasi Dengan Decision Tree Dan Particle Swarm Optimization



Gambar 2. Desain Proses Menggunakan Decision Tree – PSO

Pemodelan dilakukan dengan memasukkan data training dihubungkan dengan Optimize Weight. Setelah itu double klik optimize Weight sampai muncul jendela baru.



Gambar 3. Desain Proses Menggunakan Decision Tree – PSO

Dalam jendela Optimize Weigh, dimasukkan validation. Yang jika di klik ganda akan memunculkan 2 jendela baru seajar.



Gambar 4. Desain Validation Menggunakan Decision Tree – PSO

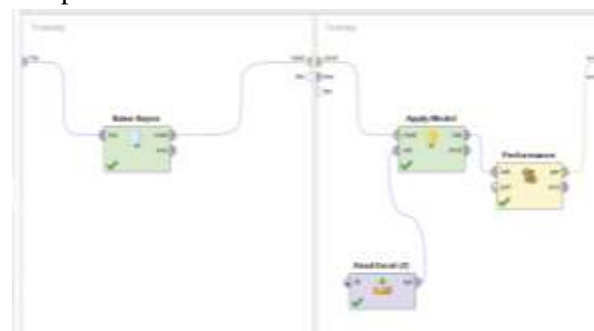
Dalam jendela tersebut dimasukkan Decision Tree, dan untuk jendela sebelah kanan dimasukkan data training, apply model dan performace. Setelah proses dijalankan akan mendapatkan hasil yang bisa dilihat pada menu result.

Tabel akurasi pengolahan data dengan Decision Tree dan Particle Swarm Optimization

accuracy: 89.72%			
	True no	True yes	class precision
pred no	2716	1963	83.20%
pred yes	1424	2418	82.94%
class recall	95.82%	88.19%	

Gambar 5. Gambar Tabel Akurasi Tabel hasil pengolahan data menunjukkan bahwa akurasi yang diperoleh sebesar 89,72%.

b) Hasil Klasifikasi Naïve Bayes Dan Particle Swarm Optimization



Gambar 6. Desain Naïve Bayes – PSO

Tabel 2. Pembobotan Atribut

Model	Bayes	NB-PSO	NN-PSO
pred. nb	78,00	82,00	87,29%
pred. nb	22	17	82,77%
data. real	89,72%	84,29%	

Gambar 7. Tabel Akurasi

c) Hasil Klasifikasi Neural Network Dan Particle Swarm Optimization



Gambar 8. Desain Neural Network - PSO

Model	Bayes	NB-PSO	NN-PSO
pred. nb	78,00	82,00	87,29%
pred. nb	22	17	82,77%
data. real	89,72%	84,29%	

Gambar 9. Akurasi NN - PSO

Hasil pengolahan keseluruhan dapat dilihat pada tabel berikut :

Tabel 1. Perbandingan Akurasi

Algoritma	Akurasi
C4.5 – PSO	89,72
NB - PSO	97,04
NN - PSO	96,96

Dengan demikian Naïve Bayes dengan Particle Swarm Optimization memiliki akurasi paling tinggi yaitu 97,04. Adapun Pembobotan masing-masing atribut sebagai berikut :

Atribut	Nilai
Usia	0.0
Pekerjaan	0.0
Status Pernikahan	0.42
Pendidikan	0.0
Kredi Macet	1.0
KPR	1.0
Pinjaman	0.0
Dihubungi Dengan	0.72
Bulan kontak	1.0
Hari kontak	1.0
Lama kontak	1.0
Frekuensi dihubungi Setelah Promo	1.0
Selang Waktu	0.0
Frekuensi dihubungi Setelah Promo	0.0
Hasil Promo Sebelumnya	0.0
Variasi Tingkat Pekerjaan	0.88
Indeks Harga Konsumen	0.0
Indeks Kepercayaan Konsumen	1.0
Indeks Euribor	0.47473428
Jumlah Karyawan	0.0

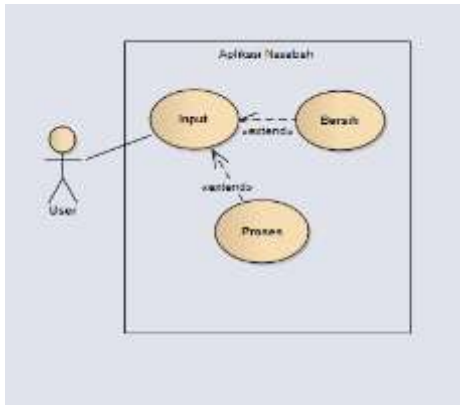
Dari tabel diatas atribut yang mempunyai bobot kurang dari 0,5 dapat dihapus untuk meningkatkan akurasi. Hasil akhirnya sebagai berikut :

Tabel 3. Atribut Yang Digunakan

Atribut	Nilai
Kredit Macet	1.0
KPR	1.0
Dihubungi Dengan	0.72
Bulan kontak	1.0
Hari kontak	1.0
Lama kontak	1.0
Frekuensi dihubungi Setelah Promo	1.0
Variasi Tingkat Pekerjaan	0.88
Indeks Kepercayaan Konsumen	1.0

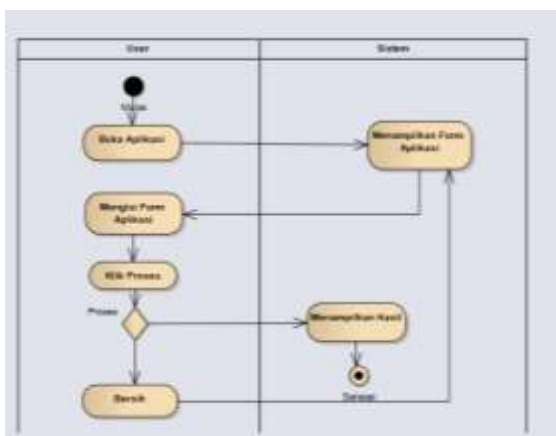
Desain Aplikasi untuk menentukan nasabah potensial :

Use Case



Gambar 10. Use Case

Activity Diagram



Gambar 11. Activity Diagram

Activity Diagram menggambarkan langkah-langkah *system* dalam menyelesaikan pekerjaannya.

Desain User Interface



Gambar IV.11 Desain GUI

User Interface dibuat

KESIMPULAN

Algoritma Naïve Bayes di optimasi dengan Particle Swarm Optimization mempunyai akurasi tertinggi untuk klasifikasi nasabah potensial. Setelah pengolahan data didapatkan hasil bahwa beberapa atribut perlu di hilangkan karena mempunyai bobot kurang dari 0,500. Dari attribute yang tersisa dibuat aplikasi menggunakan software Visual Basic 6.0.

Dalam penelitian berikutnya bisa menggunakan data nasabah Bank dari Indonesia sehingga lebih akurat dan sesuai jika diterapkan di Indonesia. Penelitian juga dapat menggunakan metode optimasi lainnya untuk mendapatkan tingkat akurasi yang lebih tinggi.

UCAPAN TERIMA KASIH

Terima Kasih kepada Universitas Bina Sarana Informatika Jakarta tempat penulis mengabdikan diri selama kurang lebih 12 tahun. Selain saya juga ingin mengucapkan terima kasih kepada Kopertis 5 yang sekarang berubah menjadi LLDIKTI wilayah 5 yang telah membiayai seluruh pengeluaran yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

S. M. Metev & V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, seri Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

S. Zhang, C. Zhu, J. K. O. Sin, &P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

M. Wegmuller, J. P. von der Weid, P. Oberson, & N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," *Prosiding ECOC'00*, 2000, paper 11.3.4, p. 109.

R. E. Sorace, V. S. Reinhardt, &S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

(2002) The IEEE website. [Online].Tersedia: <http://www.ieee.org/>

M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Tersedia: [http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/FLEXChip Signal Processor \(MC68175/D\)](http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/FLEXChip Signal Processor (MC68175/D), Motorola, 1996.), Motorola, 1996.

"PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng.

thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

J. Padhye, V. Firoiu, &D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of

