KOMPARASI 5 METODE ALGORITMA KLASIFIKASI *DATA MINING* PADA PREDIKSI KEBERHASILAN PEMASARAN PRODUK LAYANAN PERBANKAN

Sari Dewi
Manjemen Informatika AMIK BSI Pontianak
Akademi Manajemen dan Ilmu Komputer Bina Sarana Informatika
Jl.Abdurahman Saleh No 18, Pontianak
sari.sre@bsi.ac.id

Abstract- Utilization data mining in banking marketing strategy is very effective. Prospective customer seg-mentation is one of the processes carried out in the banking marketing strategy. To support the results of the success rate of telemarketing personnel to market the product in its role of banking services that the process requires a prospective customer data, then data mining support is very important in the classification of the prospective customers of the bank so that it can predict the degree of success in product marketing such services. Based on mapping studies of support data mining on prospective customers to come is no classification algorithms are often used for the classification of a borrower among others Neural Network, Naive Bayes, Dececion Tree, K-NN and Logistic Regreesion, of this algorithm can result from the evaluation process by using Cross Validation, confusion matrix, ROC Curve and T-Test to determine the classification of data mining algo-rithms are the most accurate in predicting success in product mar-keting telemarketing services from the bank to do trials in the Neural Network algorithm was more accurate with an accuracy of 89.71% the AUC value of 0872, this may be a comparison of data mining classi-fication Seeing AUC values of the five methods, then five groups of classification algorithms including both because of its AUC value between 0.80-1.00.

Keywords: Comparison of data mining, decision tree, naive Bayes, neural network, k-nn, logistic regression

Intisari- Pemanfaatan data mining dalam strategi pemasaran perbankan sangat efektif. Segmentasi calon nasabah merupakan salah satu proses yang dilakukan dalam strategi pemasaran perbankan. Untuk mendu-kung hasil dari tingkat keberhasilan tenaga telemarketing dalam peran-nya untuk memasarkan produk layanan perbankan yang prosesnya membutuhkan data-data calon nasabah ini, maka dukungan data mining sangat berperan penting dalam klasifikasi calon nasabah bank sehinga dapat memprediksi tingkat keberhasilan dalam pemasaran produk layanan tersebut. Berdasar-kan pemetaan penelitian mengenai dukungan data mining pada calon nasabah didapat ada algoritma klasifikasi yang sering digunakan untuk klasifikasi calon nasabah antara lain Neural Network, Naive Bayes, Decision Tree, K-NN dan Logistic Regreesion, dari algoritma ini di dapat hasil dari proses evaluasi dengan menggunakan Cross Valida-tion, confusion matrix, ROC Curve dan T-Test untuk mengetahui algoritma klasifikasi data mining yang paling akurat dalam prediksi keberhasilan telemarketing dalam pemasaran produk layanan bank dari uji coba yang di lakukan maka algoritma Neural Network lah yang lebih akurat dengan akurasi 89,71% dengan nilai AUC 0.872,hal ini dapat menjadi perbandingan data mining klasifikasi Melihat nilai AUC dari kelima metode tersebut yaitu NN, DC, Naive Bayes , K-NN dan LR, maka lima algoritma tersebut termasuk kelompok klasi-fikasi baik karena nilai AUC-nya antara 0.80-1.00.

kata kunci: decesion tree, komparasi data mining, naive bayes, neural network, k-n, logistic regreesion.

PENDAHULUAN

Pemasaran adalah suatu proses tentang pengembangan produk, periklanan, distribusi dan penjualan (Zhang, 2008) .proses pemasalan sangan erat kaitan nya dengan peran telemarketing,Telemarketing merupakan sebuah cara baru dalam bidang pemasaran yang menggunakan teknologi telekomunikasi sebagai bagian dari pemasaran yang teratur dan terstuktur. Telemarketing (pemasaran jarak jauh) adalah penggunaan telepon dan pusat panggilan untuk menarik prospek, menjual kepada pelanggan yang telah ada dan menyediakan layanan dengan mengambil pesanan dan menjawab pertanyan melalui telepon. Telemarketing membantu perusahaan

dalam meningkatkan pendapatan, mengurangi biaya penjualan, meningkatkan kepuasan pelanggan, Penawaran melalui jalur Telemarketing memberikan solusi bagi nasabah yang memiliki keterbatasan jarak serta waktu untuk tetap dapat melakukan transaksi atas program perlindungan yang dibutuhkan baik perlindungan bagi nasabah sendiri ataupun anggota keluarga

Dukungan data mining pada pemasaran adalah pada *marketing research dan Business Intelligence*. Dalam mengoptimasisasi proses pemasaran diperlukan suatu strategi sehingga dapat digunakan untuk meningkatkan keunggulan kompetetif, *Data mining* dalam strategi pemasaran menggunakan salah satunya menggunakan database marketing untuk melakukan proses pencarian pengetahuan baru guna mendukung pengambilan keputusan,

Oleh karena itu, penelitian inifokus pada pemanfaatan *data mining* untuk memprediksi tingkat keberhasilan telemarketing bank dalam mencari calon nasabah bank dari berbagai produk layanan perbankan, sehingga dapat diketahui apakah calon nasabah yang bersangkutan merupakan nasabah yang berpotensi menjadi nasabah kredit yang produktif atau tidak di liat dari penelitian sebelumnya algoritma yang di pakai adalah *Decision Tree* di gunakan untuk memecahkan masalah tersebut , oleh karna itu penulis ingin menguji algoritma klasifikasi lain apakah tingkat akurasinya lebih baik atau di bawah nilai dari decesion tree.

Untuk menangani permasalahan tersebut, maka akan dibandingkan beberapa algoritma yaitu pohon keputusan C4.5, naive bayes, *neural network, Logistic Regreesion* dan K-NN untuk mengetahui algoritma mana yang lebih akurat dalam memprediksi Tingkat keberhasilan telemarketing dalam layanan produk perbankan

BAHAN DAN METODE

Pengusaha di bidang jasa perbankan sangat menikmati fasilitas atau kemudahan yang diberikan oleh Pemerintah melalui kebijakan deregulasi tersebut. Bank-bank swasta baru bermunculan, bank-bank yang sudah ada menambah kantor cabang, kantor cabang pembantu maupun kantor kasnya. Ekspansi dan pembukaan kantor-kantor bank ini disamping memerlukan sejumlah tenaga kerja yang tidak sedikit, juga membutuhkan strategi pemasaran yang jitu dalam upaya menarik dana pihak ketiga untuk menyimpan uangnya di bank tersebut dan di pihak lain berusaha menyalurkan kredit yang disediakan ke pihak ketiga baik perorangan maupun perusahaan.

Dalam penulisan penelitian ini, penulis menggunakan buku, prosiding, dan jurnal sebagai referensi untuk menjelaskan model algoritma *Decesion tree, Neural Network*, Logistic Reegresion, K-NN, Naive Bayes.

A. Neural Network

Neural Network (Jaringan Saraf Tiruan) adalah prosesor tersebar paralale yang sangat besar dan memiliki kecenderungan untuk menyimpan pengetahuan yang bersifat pengalaman dan membuatnya siap untuk digunakan (puspitaningrum, 2006). NN ini merupakan sistem adaptif yang dapat merubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut. Secara sederhana NN adalah sebuah alat pemodelan data statistik non-linear. NN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola pada data. Neuron juga terdiri dari satu output. Outputnya adalah terbentuk dari pengolahan dari berbagai input oleh neuron-neuron (shukla, 2010).

B. Decision Tree

Decision tree sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak di minati (Wu, 2009) . Dalam decision tree ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit . Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numeric.

C. Naive Bayes

Naïve Bayes merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Naïve Bayes didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Teknik Naïve Bayes (NB) adalah salah satu bentuk sederhana dari Bayesian yang jaringan untuk klasifikasi. Sebuah jaringan Bayes dapat dilihat sebagai diarahkan sebagai tabel dengan distribusi probabilitas gabungan lebih dari satu set diskrit dan variabel stokastik (Liao, 2007) Metode ini penting karena beberapa alasan, termasuk berikut. Hal ini sangat mudah untuk membangun, tidak perlu ada yang rumit Parameter estimasi skema berulang. Ini berarti dapat segera diterapkan untuk besar Data set. Sangat mudah untuk menafsirkan, sehingga

pengguna tidak terampil dalam teknologi classifier dapat memahami mengapa itu adalah membuat klasifikasi itu membuat. Dan, sangat penting, hal itu sering sangat baik: Ini mungkin bukan classifier terbaik dalam setiap diberikan aplikasi, tetapi biasanya dapat diandalkan untuk menjadi kuat dan melakukan dengan sangat baik (Wu, 2009).

D. K-Nearest Neighbor

Algoritma k-*nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik,menurut (Wu, 2009) KNN juga merupakan contoh teknik *lazy learning*, yaitu teknik yang menunggu sampai pertanyaan (*quer*y) datang agar sama dengan data training.

E. Logistic Regreesion.

Regresi logistik (*Logistic regression*) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (respon) merupakan variabel dikotomi. Variabel dikotomi biasanya hanya terdiri atas dua nilai (Santosa, 2007) yang mewakili kemunculan atau tidak adanya suatu kejadian yang biasanya diberi angka 0 atau 1. Tidak seperti regresi linier biasa, regresi logistik tidak mengasumsikan hubungan antara variabel independen dan dependen secara linier.

Ada beberapa penelitian yang menggunakan komparasi algoritma klasifikasi untuk mengukur akurasi terhadap dataset marketing bank:

1. *Could Decision trees Improve the Classification Accuracy and Interpretability of Loan Granting Decision*? penelitian yang dilakukan (Zurada,2010). Yang melakukan komparasi dari beberapa metode diantaranya adalah regresi logistik(LR), jaringan saraf (NN), dasar fungsi jaringan saraf radial (RBFNN), SVM, CBR, dan pohon keputusan (DTs). Dari semua model ternyata tingkat klasifikasi akurasi yang mengungguli adalah Decision trees, DTs tidak hanya mengklasifikasikan lebih baik dari model-model yang lain tapi juga memiliki pengetahuan dalam membentuk aturan yang mudah ditafsirkan, masuk akal dalam menjelaskan tentang alasan penolakan pinjaman.
2 Comparing decision trees with logistic regression for credit risk analysis (Satchidananda & Simha, 2006). Penelitian ini membandingkan dua model algoritma untuk analisa resiko kredit, yaitu Pohon Keputusan dan Regresi Logistik. Data diambil dari dua bank yang berbeda, kemudian untuk mengelompokkan kasus positif dan negatif maka dilakukan klustering data dengan menggunakan k-means. Hasil analisa dari masing-masing model dikomparasi dan kemudian diukur,kemudian didapatkan bahwa algoritma pohon keputusan mempunyai tingkat akurasi yang tinggi dibandingkan algoritma regresi logistik.

Pengumpulan Data

Penulis Memilih metode yang akan digunakan pada saat pengujian data. Metode yang dipilih, berdasarkan penelitian yang terdahulu. Penulis menggunakan Metode Algoritma Decesion Tree, Neural Network.

Evaluasi dan Validasi Hasil

1. ***Cross Validation***

Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. *Validation* juga merupakan pengujian standar yang dilakukan untuk memprediksi *error rate*.

2. Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: *recall, precision*, *acuraccy* dan error rate.Evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah (Wu, 2009)

3. ROC Curve

ROC curves merupakan salah satu cara melakukan analisa terhadap model classifier yang telah dibuat. Penggunaan ROC curves adalah untuk menentukan parameter model yang dinginkan sesuai dengan karakteristik dari model classifier yang

Metode klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi,

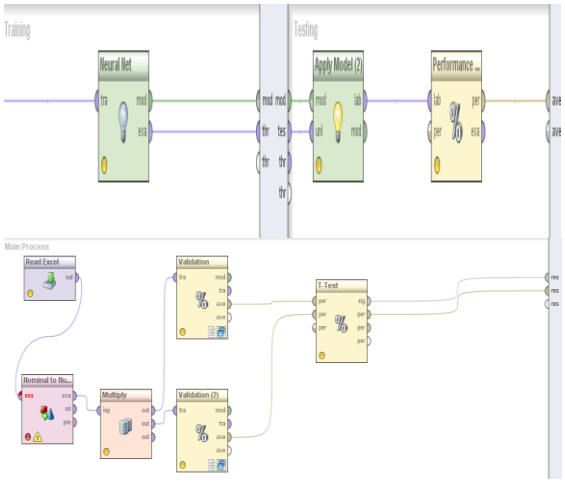
kecepatan, kehandalan, skabilitas dan interpretabilitas (Vecellis, 2009).

4. Validasi

Menurut (Gurenescu,2011) Diperlukan cara yang sistematis untuk mengevaluasi kinerja suatu metoda. Evaluasi klasifikasi didasarkan pada pengujian pada obyek benar dan salah ,menurut (Ian.H,2011) Validasi data digunakan untuk menentukan jenis terbaik dari skema belajar yang digunakan, berdasarkan data pelatihan untuk melatih skema pembelajaran untuk memaksimalkan penggunaan data .

HASIL DAN PEMBAHASAN

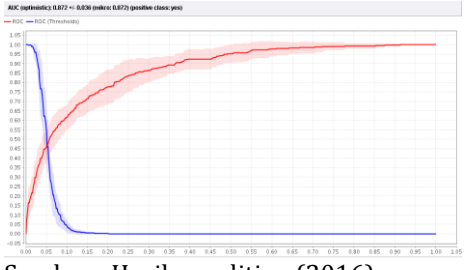
Hasil dari pengujian model yang dilakukan adalah membandingkan algoritma mana yang lebih akurat dan memperbesar akurasi dengan menggunakan T-Test pada Algoritma pada *framework* RapidMiner dengan desain model berikut ini:



Sumber : Hasil penelitian(2016)

Gambar 1.Skema Pengujian Validation dan T-test pada

a. Hasil AUC Algoritma *Neural Network*

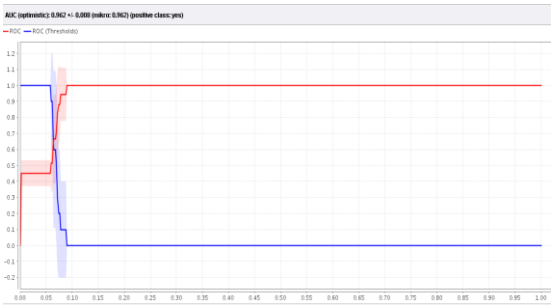


Sumber : Hasil penelitian (2016)

Gambar 2.Grafik AUC *Neural Network*

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukan bahwa ada peningkatan pada akurasi menggunakan *Neural Network* sebesar **89.71%** dan AUC sebesar **0.872**

b. Hasil AUC K-NN

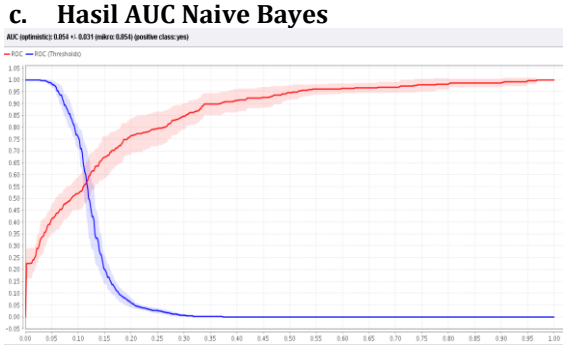


Sumber : Hasil penelitian (2016)

Gambar 3.Grafik AUC K-NN

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukan bahwa ada peningkatan pada akurasi menggunakan *K-NN* sebesar **84.70%** dan AUC sebesar **0.962.**

c. Hasil AUC Naive Bayes



Sumber : Hasil penelitian (2016)

Gambar 4. Grafik AUC Naïve Bayes

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukan bahwa ada peningkatan pada akurasi menggunakan *Naïve Bayes* sebesar **87.79%** dan AUC sebesar **0.854**

d. Hasil AUC *Logistic Regresion*

Sumber : Hasil penelitian (2016)

Gambar 5.Grafik AUC *Logistic Regreesion*

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukan bahwa ada peningkatan pada akurasi menggunakan *Logistic Regreesion* sebesar **89.32%** dan AUC sebesar **0.992.**

e. Hasil AUC Decision Tree

Sumber : Hasil penelitian (2016)

Gambar 6.Grafik AUC *Decision Tree*

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukan bahwa ada peningkatan pada akurasi menggunakan *Dececion Tree* sebesar **89.10%** dan AUC sebesar **0.9.**

Tabel 1.Perbandingan *Performace* Algoritma

	Accuracy	AUC
NN	89.71%	0.872
NB	84.70%	0.854
DC	89.10%	0.959
LR	89.32%	0.993
K-NN	87.79%	0.962

Sumber: Hasil Analisa (2016)

A. Pengujian dengan T-Test

Pengujian T-Test ini akan menguji algoritma Klasifikasi ini agar mendapatkan nilai yang terbaik,dimana dalam pengujian tersebut sampai mendapatkan nilai terkecil ≤ 0,05 dinyatakan sebagai hasil uji yang terbaik (Santoso. S:2010).

1. Hasil T-test antara algoritma Decision Tree dengan Neural Network

T-Test Significance

	0.893 +/- 0.004	0.897 +/- 0.012
0.893 +/- 0.004		0.391
0.897 +/- 0.012		

Analisis Hasil Komparasi

Berdasarkan dari analisi pengujian masing-masing metode diatas maka dapat dirangkumkan hasilnya seperti Tabel berikut

Tabel 2. Perbandingan *Performace* Algoritma Dengan T-Tes

ALGORITMA	DECISION TREE	NB	LR	NN	K-NN
DECISION TREE	-	0.000	0.934	0.391	0.001
NB	0.000	-	0.000	0.000	0.000
LR	0.934	0.000	-	0.392	0.003
NN	0.391	0.000	0.392	-	0.009
K-NN	0.001	0.000	0.003	0.009	-

Sumber: Hasil Analisa (2015)

Melihat hasil perhitungan yang terangkum pada Tabel diatas dengan menerapkan klasifikasi performance keakurasian AUC maka diperoleh hasilpenelitian yaitu, terdapat dua metode yang merupakan kategori *Good Clasification* yaitu untuk metode LR dengan nilai AUC 0.993, K-NN dengan nilai AUC 0.962 dan ,metode Dececion Tree dengan UAC 0.959 dan metode algoritma NN dengan AUC 0.872 dan metode naïve bayes yang termasuk kategori *Fair Clasification* dengan nilai AUC 0.854.

Berdasarkan Tabel di atas juga dapat di lihat bahwa nilai akurasi untuk metode aloritma klasifikasi yang terbaik adalah Algoritma Neural Network memiliki Akurasi yang lebih tinggi dengan nilai 89.71% dibandingan dengan 4 algoritma lain nya sedangkan di urutan ke dua yaitu logistic Regreesion dengan akurasi 89.32 % lalu *Dececion Tree* dengan nilai 89.10% lalu disusul dengan algoritma K-NN dengan nilai 87.79% dan yang terakhir algoritma Naive Bayes dengan Nilai 84.70%.

KESIMPULAN

Dalam penelitian ini dilakukan pembuatan model menggunakan algoritma Klasifikasi yaitu *Neural Network*, Naive Bayes, *Decesion Tree*, K-NN, dan Logistic Regreesion menggunakan data pemasararan pada Bank. Algoritma Neural Network memiliki Akurasi yang lebih tinggi dengan nilai 89.71% dibandingan dengan 4 algoritma lain nya sedangkan di urutan ke dua yaitu logistic Regreesion dengan akurasi 89.32 % lalu Dececion Tree dengan nilai 89.10% lalu disusul dengan algoritma K-NN dengan nilai 87.79% dan yang terakhir algoritma Naive Bayes dengan Nilai 84.70%. Dengan demikian algoritma Neural network dapat memberikan pemecahan untuk permasalahan dalam mengidentifikasi

Tingkat keberhasilan Telemarketing pada pemasaran Bank .

Pada kasus Prediksi Tingkat Keberhasilan Telemarketing Bank menggunakan Algoritma Klasifikasi *data Mining* dapat diterapkan pada data calon nasabah yang dihubungi untuk memprediksi keberhasilan pemasaran pada bank. Berdasarkan data set yang penelitian gunakan ini terbukti bahwa algoritma *Neural Network* ternyata lebih akurat bila dibandingkan dengan algoritma klasifikasi lain nya. Hal ini terlihat dari hasil evaluasi yang telah dilakukan. Dengan hasil ini, menunjukkan bahwa Neural Network merupakan metode yang cukup baik dalam prediksi data sehingga dapat memberikan hasil untuk permasalahan identifikasi calon nasabah.

Untuk keperluan penelitian lebih lanjut mengenai komparasi metode klasifikasi data *mining* dengan menggunakan data *additiona*l bank Ini maka disarankan untuk melakukan penyeleksian atribut, dikarenakan Atribut pday dalam metode algoritma tidak berpengaruh (hal ini di karenakan nilai nya sama) sehingga bisa dianalisa lebih lanjut apakah atribut tersbut diperlukan atau tidak. Penelitian semacam ini dapat dikembangkan pada unit bisnis serupa atau yang lainnya. Penelitian ini dapat dikembangkan dengan algoritma yang lain misalkan saja dengan metode statistik lainnya seperti Support Vector Machine.

DAFTAR PUSTAKA

Diyah, Puspitaningrum. 2006. Pengantar Jaringan Syaraf Tiruan, Penerbit Andi, Yogyakarta.

Gurenescu, 2011, Data maning : *Concept and Techniques*. Verlag berlin Heidelberg: Springer.

Ian H. Witten, frank Eibe, and Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Asma Stephan and Burlington, Eds. United States

Liao. 2007. *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application* . Singapore: World Scientific Publishing

Santosa, B. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Jogyakarta: Graha Ilmu.

Shukla, A. Tiwari, R., & Kala, R. 2010. *Real Life Application of Soft Computing*. Taylor and Francis Groups, LLC.

Vercellis, C. 2009. *Business Intelligent: Data Mining and Optimization for Decision Making. Southern Gate*: John Willey & Sons Inc.

Wu, Xindong & Kumar, Vipin. 2009. *The Top Ten Algorithms in Data Mining. Boca Raton*: CRC Press

Zhang, Guazhen, Zhou, faming, etl., 2008, *Knowledge creation in marketing based on data mining, Intelligent Computation Technology and Automation* (ICICTA), 2008 International *Conference on Page*(s): 782 – 786

BIODATA PENULIS

Sari Dewi, M.Kom Lahir di Cirebon 6 Juli 1989. Lulus S1 dari STMIK Nusa Mandiri Jakarta Program Studi Teknik Informatika 2013 dan S2 STMIK Nusa Mandiri Program Studi Ilmu Komputer 2015. Menjadi pengajar di AMIK BSI dan Pernah menulis pada jurnal paradigma