

**MENGGUNAKAN DATA MINING UNTUK SEGMENTASI CUSTOMER PADA
BANK UNTUK MENINGKATKAN CUSTOMER RELATIONSHIP
MANAGEMENT (CRM) DENGAN METODE KLASIFIKASI (AGORITMA J-48,
ZERO-R DAN NAIVE BAYES)**

Maghfirah, Teguh Bharata Adji, and Noor Akhmad Setiawan

Department of Electrical Engineering and Information Technology

Universitas Gadjah Mada

Jalan Grafika No. 2, Yogyakarta, 55281 Indonesia

Email: maghfirah.mti13@mail.ugm.ac.id, adji@mti.ugm.ac.id, noorwewe@ugm.ac.id

Abstrak

Paper ini akan membahas mengenai salah satu model strategi pemasaran yaitu Customer Segmentation yang membantu pihak bank untuk membagi pasar menjadi kelompok nasabah yang terbedakan dengan kebutuhan, karakteristik atau tingkah laku yang berbeda yang mungkin membutuhkan produk atau bauran pemasaran yang terpisah. Customer Segmentation dapat dilakukan dengan bantuan teknik Data Mining. Oleh karena itu, dalam paper ini akan dilakukan analisis dari dataset yang berasal dari data Bank Marketing dari marketing sebuah Bank di Portugis yang berhubungan dengan berlangganan Deposito Bank dengan menggunakan salah satu dari teknik data mining yaitu teknik Classification dengan membandingkan algoritma Naive Bayes, Rules Zero-R, dan Trees J-48. Dan hasil dari penerapan ketiga algoritma tersebut dalam paper ini menunjukkan bahwa dengan algoritma J-48 memberikan hasil terbaik dengan error rate terkecil, yaitu 5.8769%. Sedangkan jika dilihat dari segi efisiensi waktu dan hasil klasifikasi, algoritma Zero-R memberikan hasil terbaik (0,03 detik). Selanjutnya dari hasil yang telah diperoleh tersebut diharapkan dapat dihasilkan Customer Segmentation yang sesuai dengan kebutuhan bank yang dapat meningkatkan kualitas servis dan revenue dari bank tersebut.

Kata Kunci : *Bank Customer Segmentation, Classification, Datamining*

I. PENDAHULUAN

Dalam perekonomian modern, kehidupan masyarakat menjadi tidak asing dengan aktivitas perbankan. Perbankan merupakan bagian dari sistem keuangan yang melayani masyarakat yang surplus dana maupun masyarakat yang defisit dana. Oleh karena itu pengelolaan perbankan harus dilakukan dengan manajemen profesional, karena mengelola bank sangat berbeda dengan mengelola usaha industri. Kunci dari keberhasilan manajemen dana bank adalah bagaimana bank tersebut bisa merebut hati masyarakat sehingga peranannya sebagai financial intermediary dapat berjalan dengan baik.

Apabila suatu bank mempunyai situasi keuangan yang baik maka bank tersebut akan lebih mudah untuk menentukan segmentasi pasar, penguasaan pasar dan posisi dipasar yang mana akan dimasuki oleh bank sesuai dengan arah kebijakan manajemen bank itu sendiri. Bank yang memasarkan produk ke pasar konsumen dan pasar bisnis, dapat dilihat bahwa bank tidak dapat menarik semua nasabah di pasar tersebut. Hal ini disebabkan : nasabah terlalu banyak, nasabah tersebar luas, nasabah mempunyai kebutuhan dan kebiasaan yang bervariasi. Oleh karena itu manajemen bank harus mempunyai suatu perencanaan pemasaran yang baik, hal ini dapat dilakukan dengan mengidentifikasi segmen pasar yang paling menarik yang dapat dilayani secara efektif, daripada harus bersaing di semua segmen.

Ada sumber data yang dapat digunakan sepenuhnya di bank untuk *customer segmentation* berdasarkan data warehouse dan data mining. *Data mining* yang efektif dan analisis di bank terhadap *customer segmentation* dapat membantu bank menemukan pelanggan potensial saat ini dan dapat menciptakan manfaat bagi bank-bank secara akurat, dan membantu bank mengembangkan produk baru untuk memenuhi permintaan konsumen dan menemukan alat pemasaran yang efektif untuk mempertahankan pelanggan.

II. MATERIALS AND METHOD

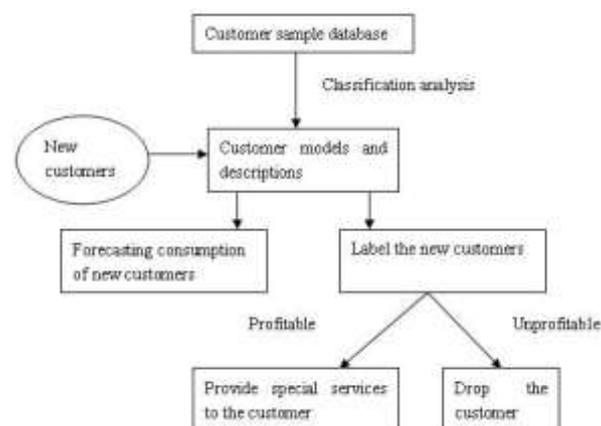
a. Bank Direct Marketing Data

Dataset yang digunakan dalam paper ini adalah berkaitan dengan data (marketing) dari sebuah institusi Bank di Portugis. Dataset ini diambil dari penelitian yang berjudul “Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology” yang bersumber dari Paulo Cortez (Univ. Minho) and Sérgio Moro (ISCTE-IUL) yang dilakukan pada tahun 2011. Dataset memiliki 17 attribute (16 + output attribute) dan 45211 *instances* [S.Moro dkk., 2011].

b. Classification Method

Berbeda dengan analisis cluster, analisis klasifikasi berarti studi di bawah instruksi, dan mengklasifikasikan data yang tidak diketahui ke beberapa data yang dikenal, yang telah ditentukan kategori sesuai dengan konsep dari berbagai kategori, dan menggunakan deskripsi ini untuk membangun kategori Model (biasanya ditandai dengan peraturan atau Modus pohon pengambilan keputusan). Kategori deskripsi dan Model klasifikasi dapat diperoleh dengan memanfaatkan Bank dataset melalui algoritma tertentu. Klasifikasi dapat juga disebut perkiraan atribut, yaitu, mengklasifikasikan data ke dalam kategori yang paling tepat. Algoritma utama adalah: pohon, statistik Bayesian, BP jaringan saraf pengambilan keputusan, Algoritma Genetika, teori himpunan kasar, logika fuzzy dan sebagainya[Wu Kun dkk., 2010].

Metode ini banyak digunakan dalam klasifikasi pelanggan dengan menganalisis data dalam sampel database pelanggan dan membuat deskripsi yang tepat untuk masing-masing kategori untuk mendirikan model pelanggan yang berbeda. Para pelanggan baru di masa depan juga bias diklasifikasikan dalam kategori yang ada pada perusahaan untuk menegaskan kesukaan nasabah, kebiasaan nasabah, kecenderungan perilaku nasabah, kebutuhan nasabah dan sebagainya, membuat "Dipasang" program pemasaran bagi nasabah, membangun sistem layanan pelanggan one-to-on dan melakukan diversifikasi manajemen pelanggan. Misalnya, dalam bank dataset, atribut output yang dihasilkan adalah apakah nasabah akan melakukan deposit pada bank tersebut, sehingga dengan metode klasifikasi pada data mining dapat dianalisis atribut-atribut untuk mengetahui nasabah profit. Classification bisa juga memperkirakan kegiatan konsumsi pelanggan baru dan membentuk model prediksi, menganalisis pelanggan dengan kegiatan yang sama dalam kondisi yang berbeda dan menyaring potensi masa depan pelanggan. Dengan memanfaatkan metode klasifikasi, perusahaan dapat mengklasifikasikan pelanggan potensial dalam kategori yang ada, jika pelanggan tersebut tidak dapat membawa keuntungan atau membuat defisit, mereka akan ditinggalkan, dan jika pelanggan tersebut bisa membawa keuntungan bagi perusahaan, maka perusahaan perlu menggabungkan sumber daya dan layanan pelanggan tersebut untuk tertarik dan memberikan layanan yang positif kepada individu, yaitu mengubah pelanggan potensial menjadi mantan pelanggan [Wu Kun dkk., 2010]. Penerapan metode klasifikasi ditunjukkan pada Gambar. 2 berikut ini.



Gambar 2. Application of Classification[Wu Kun dkk., 2010]

III. HASIL DAN PEMBAHASAN

1. Preprocessing

Dataset yang digunakan dalam paper ini merupakan data mengenai nasabah bank yang berbentuk file microsoft excel. File tersebut perlu diubah dulu menjadi ekstensi .csv baru kemudian menjadi .arff untuk dapat diolah menggunakan Weka. Dataset memiliki 45211 instances dan 17 attributes. Pada masing masing attribute tidak terdeteksi adanya nilai yang kosong (*missing value*) namun jelas terdapat beberapa informasi yang tidak diketahui (*unknown*). Sekiranya *unknown* data termasuk dalam kategori *missing value* untuk menangani missing value yang ada dapat dilakukan treatment untuk memprediksinya dengan metode-metode statistik seperti modus, asosiasi, maupun metode yang lain. Namun pada dataset ini, kebanyakan *unknown value* terjadi pada atribut atribut yang bertipe nominal sehingga penggunaan metode classification dirasa lebih tepat untuk diterapkan.

2. Evaluasi Klasifikasi dengan beberapa metode klasifikasi

Bank Dataset ini diuji dengan menggunakan beberapa metode sebagai berikut :

a. Naive Bayes

=== Evaluation on training set ===

Correctly Classified Instances	39811	88.056 %
Incorrectly Classified Instances	5400	11.944 %
Kappa statistic	0.4409	
Mean absolute error	0.153	
Root mean squared error	0.3078	
Relative absolute error	74.0529 %	
Root relative squared error	95.778 %	
Total Number of Instances	45211	

=== Confusion Matrix ===

a	b	<-- classified as
37014	2908	a = no
2492	2797	b = yes

Dari hasil dengan tools Weka di atas didapatkan bahwa error rate dari metode Naive Bayes sebesar 11.944% dengan total 5400 instances yang salah dalam proses klasifikasi instances.



Gambar 4. Visualisasi dari klasifikasi dengan naive bayes

Gambar 4 menunjukkan klasifikasi error hubungan antara class attribute yes dan no ketika menggunakan naive bayes. Kelas direpresentasikan dengan sumbu x dan klasifikasi dengan sumbu y.

Ditinjau dari segi efisiensi, waktu yang dibutuhkan untuk metode ini adalah 0.21 detik.

b. Rules--Zero-R

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	39922	88.3015 %
Incorrectly Classified Instances	5289	11.6985 %
Kappa statistic	0	
Mean absolute error	0.2066	
Root mean squared error	0.3214	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	45211	

=== Confusion Matrix ===

a	b	<-- classified as
39922	0	a = no
5289	0	b = yes

Dari hasil dengan tools Weka di atas didapatkan bahwa error rate dari metode Rules--Zero-X sebesar 11.6985% dengan total 5289 instances yang salah dalam proses klasifikasi instances.



Gambar 5. Visualisasi dari klasifikasi dengan naive Zero-R

Gambar 5 menunjukkan klasifikasi error pada hubungan antara class attribute yes dan no ketika menggunakan Rules—Zero-R. Kelas direpresentasikan dengan sumbu x dan klasifikasi dengan sumbu y.

Ditinjau dari segi efisiensi, waktu yang dibutuhkan untuk metode ini adalah 0.03 detik.

c. Trees.J48

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	42554	94.1231 %
Incorrectly Classified Instances	2657	5.8769 %
Kappa statistic	0.6876	
Mean absolute error	0.0953	
Root mean squared error	0.2183	
Relative absolute error	46.1174 %	
Root relative squared error	67.9119 %	
Total Number of Instances	45211	

=== Confusion Matrix ===

a	b	<-- classified as
39139	783	a = no
1874	3415	b = yes

Dari hasil dengan tools Weka di atas didapatkan bahwa error rate dari metode Trees-J48 sebesar 5.8769% dengan total 2657 instances yang salah dalam proses klasifikasi instances.



Gambar 6. Visualisasi dari klasifikasi dengan naive J-48

Gambar 6 menunjukkan klasifikasi error pada hubungan antara class attribute yes dan no ketika menggunakan Trees-J48. Kelas direpresentasikan dengan sumbu x dan klasifikasi dengan sumbu y.

Ditinjau dari segi efisiensi, waktu yang dibutuhkan untuk metode ini adalah 4.87 detik.

IV. KESIMPULAN

Aplikasi dari teknik data mining dalam CRM muncul menjadi trend dalam industri, termasuk dalam industri perbankan, dimana *Data mining* dipergunakan oleh perusahaan di berbagai industri untuk melakukan analisis baik secara logika (*logical analysis*) maupun statistik (*statistical analysis*) data transaksi, dan menemukan sebuah pola (*pattern*) yang dapat membantu proses pengambilan keputusan. Termasuk di industri perbankan, seorang analis di sebuah perusahaan industri perbankan melakukan *data mining* dari sistem CRM (*Customer Relationship Management*) khususnya untuk customer segmentation yang bertujuan menemukan pola aktivitas pelayanan pelanggan dan calon pelanggan agar dapat meningkatkan kualitas servis dan revenue dari perusahaan tersebut untuk mendapatkan pelanggan yang profit. Dalam paper ini dilakukan analisis bank dataset dengan menggunakan beberapa algoritma dari metode classification, dan diperoleh kesimpulan hasil seperti yang ditunjukkan pada tabel berikut ini:

Metode	Class Yes	Class No	Error Rate	Waktu (Detik)
Naive Bayes	2908	37014	11.944%	0.21
Rules--Zero -R	0	39922	11.6985%	0.03
Trees-J48	783	39139	5.8769 %	4.87

Tabel 1. Kesimpulan Hasil beberapa metode klasifikasi [K.Wisaeng, 2013]

Untuk dataset yang digunakan yaitu bank dataset, metode classification dengan algoritma J-48 memberikan hasil terbaik dengan error rate terkecil, yaitu 5.8769%. Sedangkan jika dilihat dari segi efisiensi waktu dan hasil klasifikasi, algoritma Zero-R memberikan hasil terbaik (0,03 detik), walaupun tidak terlalu signifikan dibandingkan dengan Naive Bayes. Kemudian instances kelas Yes yang paling banyak mengalami kesalahan klasifikasi adalah ketika menggunakan Zero-R, sementara untuk instances kelas No yang paling banyak mengalami kesalahan klasifikasi adalah ketika menggunakan Naive Bayes.

Di penelitian selanjutnya, dari paper ini dapat dilakukan pengolahan data yang lebih detail lagi tentang bank dataset untuk kemudian dapat menggunakan dataset tersebut dengan mencoba beberapa metode data mining yang lain yang mungkin akan menghasilkan tingkat akurasi data yang lebih besar dari hasil yang diperoleh dalam paper ini.

DAFTAR PUSTAKA

- [1] K. Wisaeng. "A Comparison of Different Classification Techniques for Bank Direct Marketing". *International Journal of Soft Computing and Engineering (IJSCE)*. 2013.
- [2] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), *Proceedings of the European Simulation and Modelling Conference - ESM'2011*, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.
- [3] Wu Kun, Liu Feng-Ying. *Application of Data Mining in Customer Relationship Management*. Tianjin Polytechnic University Tianjin, China. 2010.
- [4] Z.Bosnjak, O.Grljevic. "Credit Users Segmentation for improved Customer Relationship Management in Banking". *IEEE International Symposium on Applied Computational Intelligence and Informatics*.2011.