

PEMBUATAN MODEL PROFIL MAHASISWA FAKULTAS TEKNIK UNIVERSITAS PANCASILA

Paryudi

Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pancasila
Jl. Srengseng Sawah, Jagakarsa, Jakarta Selatan, 12640
E-mail: paryudi@rediffmail.com

ABSTRAK: Promosi merupakan suatu keharusan bagi sebuah universitas untuk mendapatkan mahasiswa. Dengan masih awamnya tim promosi terhadap profil mahasiswa yang sudah ada, maka tim promosi melakukan promosi tanpa melihat segmen pasar yang harus dituju. Konsekuensinya adalah biaya promosi menjadi lebih mahal. Untuk melakukan promosi dengan lebih baik, dapat menggunakan metode *direct marketing*. Pada metode ini, model profil dari mahasiswa yang sudah ada harus dibuat terlebih dulu. Dengan menggunakan model profil ini, tim promosi dapat memfokuskan promosi hanya pada calon-calon mahasiswa yang sesuai dengan model. Keuntungan dari metode promosi ini adalah biaya promosi dapat dikurangi, tingkat respon meningkat, dan keuntungan juga meningkat. Untuk membuat model pada *direct marketing* dibutuhkan data dari promosi sebelumnya. Karena data promosi sebelumnya tidak tersedia, dua metode diusulkan untuk membuat model awal. Selanjutnya model awal yang dibuat dites menggunakan perangkat lunak data mining yang tersedia di pasaran. Model dengan akurasi minimal 75% dipilih. Jika ada lebih dari satu model dengan akurasi minimal 75%, maka model dengan akurasi tertinggi dipilih.

Kata kunci: data mining, pemasaran langsung, klasifikasi

ABSTRACT: Promotion is a must for a university to get students. With the innocence of the promotion team about the existing student profile, it will cause the team does not know which segment should be the promotion target. The consequence is that the promotion cost will be higher. In order to have a better promotion, we can use direct marketing method. In this method, a profile model of the existing students must be first created. With this profile model, promotion team can focus the promotion only to candidate student match with the model. The advantages of this method are: promotion cost can be reduced, response rate increase, and profit also increase. In order to create a model in direct marketing, we need previous promotion data. Since previous promotion data is not available, two methods in creating preliminary models are proposed. Next, the preliminary models will be tested using data mining software available in the market. Model with minimal accuracy of 75% will be chosen. If there are more than one model with minimal accuracy of 75%, then model with the highest accuracy will be chosen.

Keywords: data mining, direct marketing, classification

PENDAHULUAN

Krisis ekonomi berpengaruh di segala bidang, tidak terkecuali dunia pendidikan. Krisis ini mengakibatkan terjadinya penurunan jumlah mahasiswa yang mendaftar. Hal ini menyebabkan tim promosi perguruan tinggi harus bekerja keras untuk bisa mendapatkan mahasiswa sesuai dengan target yang telah ditetapkan oleh perguruan tinggi. Dengan masih awamnya tim promosi terhadap profil mahasiswa yang sudah ada, maka tim promosi melakukan promosi secara membabi buta, tanpa melihat segmen pasar yang harus dituju. Konsekuensinya adalah biaya promosi menjadi lebih mahal.

Penulis mengusulkan penggunaan data mining di dunia pendidikan terutama pada saat promosi untuk menjangkau mahasiswa dengan menggunakan metode *direct marketing* [1]. Dengan bantuan data mining, promosi secara *direct marketing* akan lebih terfokus

hanya pada calon-calon mahasiswa yang berpotensi untuk mendaftar. Pemikiran ini merupakan modifikasi apa yang telah dinyatakan oleh Tan, Ling dan Li, dan Javaheri [2, 3, 4]. Tan menyatakan bahwa dengan menggunakan data mining, biaya pengiriman surat penawaran bisa dikurangi [4]. Ling dan Li mengatakan bahwa promosi ke pelanggan bisa dilakukan dengan *direct marketing* [3]. Dengan metode ini, perusahaan terlebih dahulu mempelajari ciri-ciri pelanggannya dan memilih pelanggan tertentu sebagai sasaran promosi. Lebih lanjut mereka juga mengatakan bahwa dengan menggunakan *direct marketing* akan menurunkan biaya promosi, menaikkan tingkat respon, serta menaikkan keuntungan. Sedangkan Javaheri mengatakan bahwa memilih pelanggan bisa dilakukan dengan membuat *response modeling* [2].

Karena belum ada penelitian tentang aplikasi data mining dalam promosi secara *direct marketing* di dunia pendidikan, maka penulis mencoba mengapli-

kasikan data mining, khususnya teknik klasifikasi, untuk permasalahan yang dihadapi. Teknik klasifikasi yang digunakan antara lain: *Decision tree*, *Rule based*, *K-Nearest Neighbor (KNN)*, dan *Naïve Bayes*. Dalam tulisan ini peneliti mengusulkan sebuah metode untuk membuat model. Bagian lain dari tulisan disusun sebagai berikut. Bagian 2 menyajikan tentang *direct marketing*. Bagian 3 membahas tentang permasalahan. Bagian 4 membahas metode penelitian yang diusulkan. Bagian 5 menyajikan kesimpulan.

DIRECT MARKETING

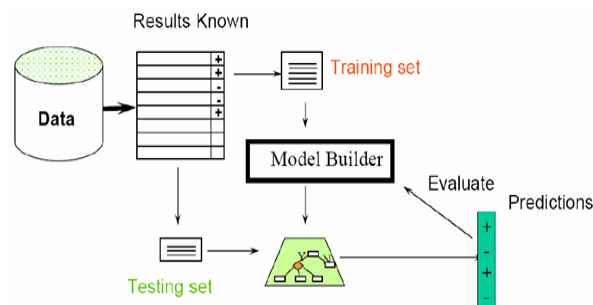
Direct Marketing adalah metode promosi dimana promosi tidak disebarkan secara masal melainkan dengan cara memilih pelanggan tertentu yang sesuai dengan model yang telah dibuat sebelumnya. Saat ini promosi secara *direct marketing* dinilai lebih efektif dari pada promosi secara masal. Promosi secara masal hanya bisa efektif apabila barang yang ditawarkan memang sedang sangat dibutuhkan oleh masyarakat. Contohnya adalah TV pada masa setelah perang dunia ke II. Karena kebutuhan TV saat itu sedang memuncak, maka iklan secara masal tetap efektif. Namun, saat pilihan barang yang ditawarkan pada konsumen sangat banyak serta kualitas antara satu barang dengan barang lain tidak begitu berbeda, maka promosi secara masal tidak efektif lagi. Promosi semacam ini hanya memperbesar biaya promosi. Oleh karena itu pelaku bisnis mulai melirik ke metode promosi secara *direct marketing*. Dengan *direct marketing* biaya promosi bisa ditekan, tingkat respon naik, serta keuntungan juga meningkat [3].

Tabel 1. Perbandingan antara promosi secara *direct marketing* dan *mass marketing* [3]

	<i>Mass Marketing</i>	<i>Direct Marketing</i>
Jumlah pelanggan	600.000	120.000 (20%)
Biaya cetak	\$ 426.000	\$ 85.200
Biaya data mining	\$ 0	\$ 40.000
Total biaya promosi	\$ 424.000	\$ 125.200
Tingkat respon	1 %	3 %
Jumlah penjualan	6.000	3.600
Laba dari penjualan	\$ 420.000	\$ 252.000
Laba bersih dari promosi	-\$ 6.000	\$ 126.800

Langkah-langkah pembuatan model dalam *direct marketing* menurut Javaheri dapat dilihat pada Gambar 1. Data hasil promosi sebelumnya dibagi menjadi dua: *training set* dan *test set*. *Training set* adalah data yang digunakan untuk membuat model. Masing-masing *record* pada *training set* mempunyai

sekumpulan atribut dan salah satu atribut merupakan kelas (*class*). Model dibuat berdasarkan nilai dari masing-masing atribut dan nilai dari *class attribute*. Untuk menguji akurasi dari model, *test set* dimasukkan dalam model. Model dipakai apabila akurasinya mencapai nilai tertentu.



Gambar 1. Langkah-langkah Pembuatan Model [2]

PERMASALAHAN

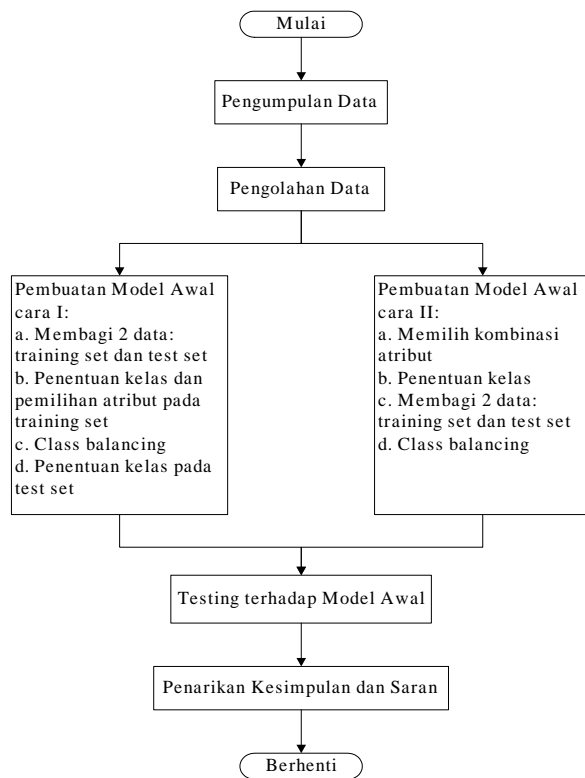
Selama ini tim promosi Fakultas Teknik, Universitas Pancasila melakukan promosi secara masal. Pada metode promosi semacam ini, tim promosi melakukan promosi tanpa memilih segmen pasar tertentu. Konsekuensinya adalah metode ini membutuhkan biaya yang besar sedang keuntungan yang didapat kecil. Untuk mengatasi masalah ini, metode promosi harus diganti dengan metode *direct marketing*. Pada promosi secara *direct marketing*, dibutuhkan model untuk menentukan calon mahasiswa yang berpotensi untuk mendaftar. Dengan melakukan promosi hanya pada calon mahasiswa yang sesuai dengan model, maka *direct marketing* bisa menekan biaya promosi, menaikkan tingkat respon, serta meningkatkan keuntungan. Permasalahannya adalah bagaimana membuat model profil calon mahasiswa yang berpotensi untuk mendaftar.

Ling dan Li menyatakan bahwa ada 2 skenario dalam pembuatan model [3]. Pada skenario pertama seluruh database pelanggan digunakan, dimana sebagian di antaranya adalah pembeli. Database dibagi menjadi dua: *training set* dan *test set*. *Training set* digunakan untuk membuat model berdasarkan data pelanggan yang sudah menjadi pembeli. Sedangkan *test set* digunakan untuk mengetes akurasi dari model yang dihasilkan. Pada skenario kedua, produk baru dipromosikan ke seluruh pelanggan dalam database dimana belum ada pelanggan yang menjadi pembeli. Sebuah survei dilakukan dimana sebagian kecil (misal 5%) pelanggan dipilih secara acak sebagai target promosi. Sebagian pelanggan merespon promosi ini. Dari data pelanggan yang merespon promosi, maka bisa dibuat model pelanggan yang kemungkinan besar membeli produk tersebut.

Perbedaan pada penelitian ini adalah bahwa cara pembuatan modelnya tidak masuk dalam kedua skenario. Dalam penelitian ini tidak ada data promosi sebelumnya yang bisa digunakan untuk membuat model. Oleh karena itu model dibuat hanya dari data mahasiswa yang masih aktif. Sehingga masalah dari penelitian ini adalah bagaimana membuat model profil mahasiswa hanya dari data mahasiswa yang masih aktif. Minimal akurasi model yang bisa diterima adalah 75%.

USULAN METODE PENELITIAN

Kegiatan penelitian yang diusulkan terlihat pada Gambar 2. Penelitian dimulai dengan melakukan pengumpulan data. Pengumpulan data dilakukan dengan cara menyebar kuesioner ke seluruh mahasiswa aktif di Fakultas Teknik, Universitas Pancasila. Informasi yang hendak dikumpulkan pada kuesioner adalah data pribadi mahasiswa namun sedikit lebih detail. Contoh pertanyaan adalah: apa pekerjaan orang tua, berapa penghasilan orang tua per bulan, bagaimanakah status rumah tinggal, apakah punya mobil, dan sebagainya.



Gambar 2. Langkah-langkah penelitian

Pengolahan data dilakukan dengan memasukkan data-data yang sudah dikumpulkan kedalam berkas MS Excel. Pada tahap ini dilakukan sub kegiatan seperti: *data cleaning* dan normalisasi data. *Data cleaning* dilakukan untuk memperbaiki kesalahan

yang ada, menghapus data yang tidak berguna atau data ganda, serta untuk mengisi data yang kosong. Sedang normalisasi data dilakukan agar data yang ada mempunyai skala yang sama sehingga data yang besar tidak menutupi data yang kecil. Ada tiga macam teknik normalisasi yaitu: *Z-score Normalization*, *Min-Max Normalization*, dan *Normalization by decimal scaling*.

Pada pembuatan model awal, peneliti mengusulkan dua cara:

1. Cara I:
 - a. Membagi data menjadi dua: *training set* dan *test set*. Dalam hal ini perbandingan jumlah data adalah 2/3 untuk *training set* dan sisanya untuk *test set*.
 - b. Menentukan kelas dari masing-masing *record*. Misal kita menganggap data mahasiswa yang ada adalah model dari hasil promosi terdahulu. Dengan mengambil asumsi bahwa model tersebut mempunyai akurasi kurang dari 100%, maka harus ada data mahasiswa yang tidak cocok dengan model. Maka tugas penulis untuk menentukan *record* mana yang mempunyai kelas “mahasiswa” dan mana yang “bukan mahasiswa”. Penentuan ini dilakukan berdasarkan nilai atributnya. Sedang pemilihan atribut yang digunakan dilakukan dengan menghitung *F-Score* dari masing-masing atribut. *F-Score* merupakan teknik yang menghitung perbedaan antara dua set bilangan riil. Misalkan ada vektor x_k , dimana $k = 1, 2, \dots, m$. Jika jumlah data positif (mahasiswa) dan data negatif (bukan mahasiswa) adalah n^+ dan n^- , maka *F-Score* dari atribut ke i adalah [2]:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n^+ - 1} \sum_{k=1}^{n^+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n^- - 1} \sum_{k=1}^{n^-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2}$$

Di mana $\bar{x}_i, \bar{x}_i^{(+)}, \bar{x}_i^{(-)}$ adalah rata-rata seluruh data, rata-rata data positif, dan rata-rata data negatif pada atribut ke i . Sedang $x_{k,i}^{(+)}$ adalah data ke k pada atribut ke i dari data positif serta $x_{k,i}^{(-)}$ adalah data ke k pada atribut ke i dari data negatif. Atribut dengan nilai *F-Score* tinggi dipilih dan diikutsertakan dalam pemodelan.

- c. Apabila dari langkah (b) terjadi *class imbalance*, yaitu jumlah kelas “mahasiswa” jauh lebih kecil dari kelas “bukan mahasiswa”, maka *class imbalance* ini harus dihilangkan. Hal ini karena kelas yang jumlahnya lebih banyak menutupi kelas yang jumlahnya sedikit. Kalau menjumpai kasus seperti ini

dimana kelas negatif (bukan mahasiswa) jauh melebihi kelas positif (mahasiswa), maka teknik klasifikasi cenderung menurun kinerjanya. Ada dua macam teknik untuk menghilangkan *class imbalance* yaitu: *under-sample the majority class* dan *over-sample the minority class* [2].

- d. Kalau masing-masing *record* pada *training set* sudah ditentukan kelasnya, maka *record-record* pada *test set* juga ditentukan kelasnya dengan cara yang sama seperti ketika menentukan kelas pada *training set*.

Apabila ingin membuat lebih dari satu model awal, maka proses harus diulang dari langkah (a).

2. Cara II:

- a. Memilih kombinasi atribut. Pemilihan atribut dilakukan untuk mengurangi jumlah data dengan menghilangkan atribut-atribut yang tidak relevan atau *redundant*. Hal ini dimaksudkan untuk mencari jumlah minimum atribut yang bisa mewakili seluruh atribut. Penggalan (*mining*) pada jumlah atribut yang lebih kecil mempunyai keuntungan yaitu mengurangi jumlah atribut yang muncul pada pola yang dihasilkan dan ini menyebabkan pola tersebut lebih mudah dimengerti. Ada 3 macam metode dalam pemilihan atribut: *Stepwise forward selection*, *Stepwise backward elimination*, *Combination of forward selection and backward elimination*, *Decision tree induction* [1].
- b. Penentuan kelas dari masing-masing *record* dilakukan berdasarkan nilai atributnya.
- c. Membagi data menjadi dua: *training set* dan *test set* dengan perbandingan 2/3 : 1/3.
- d. *Class balancing*. Apabila setelah pembagian terjadi *class imbalance*, maka harus dilakukan *class balancing* dengan cara seperti yang sudah dijelaskan sebelumnya.

Apabila mau membuat lebih dari satu model awal, maka proses harus diulang dari langkah (a).

Testing terhadap model awal dilakukan dengan menggunakan empat teknik klasifikasi yaitu *Decision Tree*, *Rule Based*, *K-Nearest Neighbor*, dan *Naïve Bayes*. Penulis mengusulkan untuk melakukan testing menggunakan tiga buah perangkat lunak supaya dapat membandingkan hasil dari satu perangkat lunak dengan yang lain. Hal ini lebih memastikan bahwa model yang didapat memang akurat. Cara testing

adalah sebagai berikut: *training set* dimasukkan kedalam perangkat lunak dan diproses. Selanjutnya *test set* digabungkan dengan *training set* dan diproses untuk mengetahui akurasi dari model awal. Kalau akurasi mencapai minimal 75%, maka model dianggap akurat. Apabila terdapat lebih dari satu model dengan akurasi minimal 75%, maka dipilih model dengan akurasi paling tinggi.

KESIMPULAN

Direct marketing merupakan sebuah metode promosi dimana promosi tidak disebarluaskan secara masal melainkan dengan cara memilih pelanggan tertentu yang sesuai dengan model yang telah dibuat sebelumnya. Model promosi semacam ini efektif apabila produk yang ditawarkan sudah banyak tersedia dimasyarakat dengan kualitas tidak jauh berbeda antara satu dengan yang lainnya. Keuntungan dari *direct marketing* adalah biaya promosi bisa ditekan, tingkat respon naik, dan keuntungan juga meningkat.

Metode penelitian yang diusulkan dimulai dari pengumpulan data, pengolahan data, pembuatan model awal, testing terhadap model awal, dan penarikan kesimpulan. Pembuatan model awal dimasukkan dalam langkah penelitian karena tidak tersedianya data promosi sebelumnya yang bisa digunakan untuk membuat model awal. Oleh karena itu pembuatan model awal harus dilakukan dengan menggunakan data mahasiswa yang masih aktif. Penulis mengusulkan dua cara untuk membuat model awal.

DAFTAR PUSTAKA

1. Paryudi, I., 2008, *Data Mining*, Jurnal Teknologi Informasi & Komputer, Vol. 2, No. 1, 1–18.
2. Javaheri, H. S., 2007, *Response Modeling in Direct Marketing—A Data Mining Based Approach for Target Selection*, Master Thesis, Lulea University of Technology.
3. Ling, C. X. dan Li, C., 1998, *Data Mining for Direct Marketing: Problems and Solutions*, In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, 73 – 79
4. Tan, P. N., Steinbach, M., dan Kumar, V., 2006, *Introduction to Data Mining*, Addison-Wesley.
5. Han, J. dan Kamber, M., 2006, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2nd ed.