

# IMPLEMENTASI DATA MINING DENGAN NAIVE BAYES CLASSIFIER UNTUK Mendukung STRATEGI PEMASARAN DI BAGIAN HUMAS STMIK AMIKOM YOGYAKARTA

Erik Hadi Saputra<sup>1)</sup>, Burhan Alfironi Mukhtar<sup>2)</sup>

<sup>1), 2)</sup> Teknik Informatika STMIK AMIKOM Yogyakarta  
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55283  
Email : [erik@amikom.ac.id](mailto:erik@amikom.ac.id)<sup>1)</sup>, [alfironi3697@gmail.com](mailto:alfironi3697@gmail.com)<sup>2)</sup>

## Abstrak

Salah satu metode dalam data mining adalah Naive Bayes Classifier yang merupakan teknik data mining yang memiliki kemampuan klasifikasi. Algoritma naive bayes classifier dapat digunakan untuk memprediksi minat studi berdasarkan bukti-bukti yang diberikan. Algoritma Naive Bayes Classifier merupakan salah satu metode data mining yang dapat digunakan untuk mendukung strategi pemasaran yang efektif dan efisien. Hasil dari penelitian ini adalah aplikasi data mining yang dibangun menggunakan algoritma Naive Bayes Classifier yang dapat memberikan informasi penting berupa hasil prediksi minat studi calon mahasiswa yang bisa digunakan untuk membantu Team Marketing STMIK AMIKOM Yogyakarta. Data yang digunakan adalah data pendaftaran mahasiswa tahun sebelumnya. Dari hasil prediksi yang didapat, diharapkan dapat membantu sebagai pendukung strategi pemasaran sehingga berdampak pada efektifitas dan efisiensi marketing serta peningkatan jumlah mahasiswa baru yang mendaftar.

**Kata Kunci:** data mining, Naive Bayes Classifier, prediksi minat studi

## 1. Pendahuluan

Banyak perusahaan dan instansi pemerintah yang menggunakan data mining untuk menggali informasi. Informasi ini sangat berguna untuk meningkatkan keuntungan ataupun membantu dalam penyusunan strategi pemasaran. Selain perusahaan dan instansi pemerintah, penerapan data mining juga dilakukan dalam dunia pendidikan baik oleh perguruan tinggi negeri maupun swasta. Hal ini karena banyak perguruan tinggi yang berupaya untuk mendapatkan competitive intelligence. STMIK AMIKOM Yogyakarta merupakan salah satu perguruan tinggi swasta yang berkedudukan di Yogyakarta di bawah naungan Yayasan AMIKOM. STMIK AMIKOM Yogyakarta merupakan sebuah

perguruan tinggi hasil pengembangan dari AMIK "AMIKOM Yogyakarta" [2].

Salah satu penerapan data mining dengan naive bayes classifier dalam perguruan tinggi adalah untuk mengetahui minat dan ketertarikan calon mahasiswa baru terhadap program studi yang tersedia dengan memprediksi probabilitas keanggotaan suatu class. Informasi ini dapat digunakan untuk mendukung strategi pemasaran agar lebih efektif dan efisien. Dengan informasi ini suatu perguruan tinggi dapat mengetahui tingkat ketertarikan calon mahasiswa terhadap suatu program studi yang terdapat di perguruan tinggi tersebut. Sehingga perguruan tinggi dapat mengetahui serta menentukan target atau sasaran pasar dengan lebih rinci.

### 1.1 Klasifikasi

Algoritma klasifikasi menggunakan data training untuk membuat sebuah model. Model yang sudah dibangun tersebut kemudian digunakan untuk memprediksi label kelas data baru yang belum diketahui.

Dalam proses klasifikasi terdapat banyak algoritma yang telah dikembangkan oleh para peneliti seperti K-Nearest Neighbor, Artificial Neural Network, Support Vector Machine, Decision Tree, Naive Bayes Classifier dan lain sebagainya. [5]

Salah satu pengukur kinerja klasifikasi adalah tingkat akurasi. Sebuah sistem dalam melakukan klasifikasi diharapkan dapat mengklasifikasi semua set data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% akurat [2].

- Untuk menghitung akurasi digunakan formula :

$$\begin{aligned} \text{Akurasi} &= \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \\ &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \end{aligned}$$

- Untuk menghitung kesalahan prediksi (error) digunakan formula:

$$\text{Error} = \frac{\text{jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}}$$

$$= \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}}$$

**1.2 Konsep Naive Bayes Classifier**

Naive bayes classifier merupakan salah satu metode yang digunakan dalam data mining yang didasarkan pada teori keputusan Bayes. Naive bayes classifier memiliki kemampuan klasifikasi seperti metode decision tree dan neural network. Metode ini dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class[3].

Seperti yang dikatakan Eko Prasetyo (2012:61), formulasi Naive Bayes Classifier adalah sebagai berikut [4];

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

Dimana untuk formula diatas :

- P(Y|X) adalah probabilitas data dengan vector X pada kelas Y.
- P(Y) adalah probabilitas awal kelas Y.

- $\prod_{i=1}^q P(X_i|Y)$  adalah probabilitas independen kelas Y dari semua fitur dalam vector X.

- Nilai P(X) selalu tetap sehingga dalam perhitungan prediksi nantinya tinggal menghitung bagian

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

dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi.

Naive Bayes Classifier (NBC) membutuhkan jumlah record data yang sangat besar untuk mendapatkan hasil yang baik. Jika kategori prediktor tidak ada dalam data training, maka naive bayes classifier mengasumsikan bahwa record baru dengan kategori predictor memiliki probabilitas nol.

**2. Pembahasan**

**2.1 Analisis Kinerja Sistem (Performance Analysis)**

Dari analisa kinerja sistem lama di HUMAS STMIK AMIKOM Yogyakarta didapatkan hasil dengan parameter Throughput bahwa Jumlah total pendaftar pada tahun ajaran 2013/2014 yang dihasilkan dari proses marketing kurang lebih sebesar 10.000 orang. Kemudian yang melakukan verifikasi pendaftaran sebesar 5.168

orang dan diterima sebesar 2080 orang. Berdasarkan Resptime pelayanan marketing dari customer service marketing kurang Resptime untuk lebih 10 menit/orang.

**2.2 Analisis Informasi (Information Analysis)**

Analisa informasi di HUMAS STMIK AMIKOM Yogyakarta didapatkan hasil dari sistem yang lama belum ada informasi yang akurat berhubungan dengan hasil klasifikasi dan prediksi minat studi proses marketing. Untuk mendapatkan perhitungan informasi mengenai hasil analisa yang berhubungan dengan klasifikasi dan prediksi minat studi serta statistik teknik marketing secara manual membutuhkan waktu kurang lebih 2 jam. Informasi secara umum yang berhubungan dengan data pendaftar dalam proses marketing dinilai relevan karena dikelola dengan baik oleh Innovation Center.

**2.3 Analisis Ekonomi**

Analisa ekonomi di HUMAS STMIK AMIKOM Yogyakarta didapatkan hasil bahwa Anggaran biaya yang dikeluarkan untuk proses marketing dialokasikan untuk pembuatan brosur, periklanan dengan berbagai media seperti televisi, koran, dan radio, kunjungan, serta proses marketing lainnya.

**2.4. Analisis Kendali**

Analisa kendali di HUMAS STMIK AMIKOM Yogyakarta didapatkan hasil bahwa kontrol terhadap data dilakukan oleh Innovation Center dengan melakukan back up data setiap malam dan keamanan data dikelola dengan pemberian hak akses dan keamanan database serta didukung dengan keamanan sistem operasi dan jaringan oleh Innovation Center.

**2.5 Analisis Efisiensi (Efficiency Analysis)**

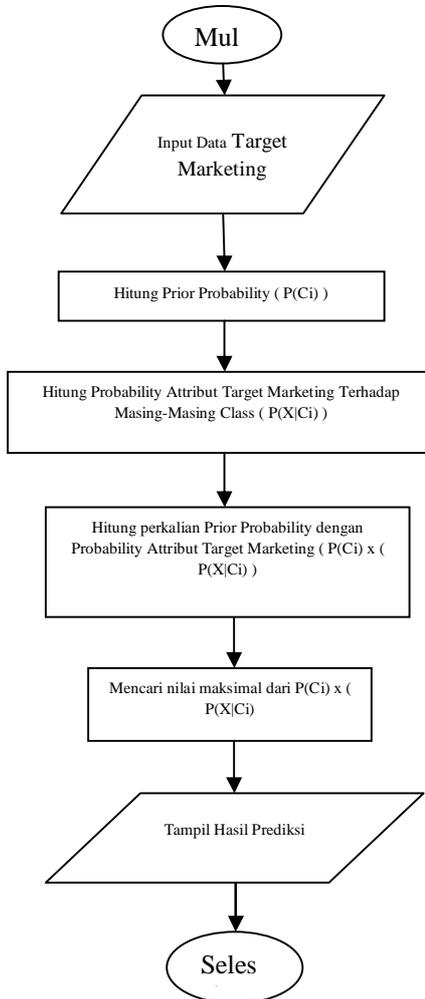
Analisa efisiensi di HUMAS STMIK AMIKOM Yogyakarta didapatkan hasil seperti Dengan proses marketing yang dilakukan dan biaya yang dikeluarkan dihasilkan jumlah pendaftar kurang lebih 10.000 orang, sehingga proses marketing dinilai efisien.

**2.6 Analisis Pelayanan (Service Analysis)**

Analisa pelayanan di HUMAS STMIK AMIKOM Yogyakarta didapatkan hasil seperti Pelayanan pada customer service marketing di STMIK AMIKOM Yogyakarta dilaksanakan dari jam 08.00 – 16.00 WIB pada hari senin sampai dengan jumat dan jam 08.00 – 14.00 WIB pada hari sabtu.

**2.7 Flowchart**

Perancangan flowchart ini bertujuan untuk memberi gambaran bagaimana aliran proses dari aplikasi untuk prediksi minat studi ini berjalan. Mulai dari awal ketika task menu prediksi minat studi diklik hingga tampil hasil prediksi minat studi.



Gambar 1. Alur Aplikasi Program

### 3. Uji Coba Sistem dan Program

#### 3.1 White-box Testing

White-box testing dilakukan dengan menguji logika jalur program yang meliputi :

1. Menggunakan semua modul minimal satu kali untuk memberikan jaminan bahwa semua jalur program berfungsi seperti yang diharapkan.
2. Menggunakan semua keputusan logis pada sisi true dan false.
3. Mengeksekusi semua loop pada batasan operasional mereka.

Hasil yang didapatkan menunjukkan tidak adanya kesalahan logika dalam pemrograman.

Pengujian *syntax* program dilakukan dengan pengecekan pada setiap *listing* program. Pada *tool Microsoft Visual Studio 2010* jika terdapat kesalahan *syntax* maka akan ditunjukkan dengan tanda *error* garis merah seperti yang ditunjukkan pada gambar 2. Setelah dilakukan pengecekan terhadap seluruh *source code* program menunjukkan tidak ada kesalahan *syntax*.

```

namespace Bayes.dao
{
    interface IHitungProbability
    {
        double priorProbability(string prodi);
        double probabilityAnggotaS1TI(DataPendaftar dataPendaftar);
        double probabilityAnggotaD3I(DataPendaftar dataPendaftar);
        double probabilityAnggotaS1MI(DataPendaftar dataPendaftar);
        double probabilityAnggotaD3MI(DataPendaftar dataPendaftar);
        double probabilityS1TI(DataPendaftar dataPendaftar);
        double probabilityD3TI(DataPendaftar dataPendaftar);
        double probabilityS1SI(DataPendaftar dataPendaftar);
        double probabilityD3MI(DataPendaftar dataPendaftar);
    }
}
    
```

Gambar 2. Contoh Syntax Error

#### 3.2.Black-box Testing

Black-box testing dilakukan pada seluruh modul program untuk mengetahui apakah program yang dibuat sudah sesuai dengan kebutuhan fungsionalnya.

Tabel 1. Uji Coba Kebutuhan Fungsional

No	Kebutuhan Fungsional	Keterangan
1	User dapat melakukan penginputan data target marketing yang berupa sekolah asal, kota, jenis kelamin, dan jurusan.	Terpenuhi
2	Sistem dapat melakukan perhitungan probability keanggotaan suatu class.	Terpenuhi
3	User dapat menganalisa hasil output perhitungan probability.	Terpenuhi
4	Sistem dapat memberikan hasil prediksi minat studi yang diinginkan oleh target marketing.	Terpenuhi
5	Sistem dapat memberikan statistik keberhasilan teknik marketing berdasarkan dari informasi yang didapat oleh para pendaftar untuk masuk STMIK AMIKOM Yogyakarta.	Terpenuhi

Hasil dari uji coba dari 1000 data testing dengan aplikasi *Smart Marketing* untuk prediksi minat studi ditampilkan pada daftar hasil uji coba prediksi minat studi (terlampir). Dengan menggunakan rumus (2.1), hasil uji coba prediksi minat studi tersebut mendapatkan nilai akurasi sebagai berikut :

$$Akurasi = \frac{927}{1000} = 0,927 = 92,7 \%$$

Dari perhitungan akurasi untuk kinerja klasifikasi didapatkan akurasi sebesar 92,7%, kemudian dengan hasil uji coba prediksi minat studi menggunakan rumus (2.2) didapatkan kesalahan prediksi (*error*) sebagai berikut :

$$Error = \frac{73}{1000} = 0,073 = 7,3 \%$$

Setelah dilakukan perhitungan kesalahan prediksi (*error*) dari 1000 *data testing* didapatkan bahwa kesalahan prediksi klasifikasi sebesar 7,3%.

**Tabel 2.** Nilai Akurasi dan Kesalahan Klasifikasi (*Error*)

Algoritma	Data Training	Data Testing	Akurasi	Error
Naive Bayes Classifier	Data pendaftar tahun 2009 sampai dengan 2011	1000 data pendaftar tahun 2012 yang diambil secara acak	92,7%	7,3%

**3.3. Pembahasan Interface**

Dalam aplikasi *Smart Marketing* ini terdiri dari beberapa *interface* yaitu Menu Utama, Menu Prediksi Minat Studi, Menu Statistik *Marketing*, Menu *About*, Menu *Credit*, dan Menu *Help*.

**3.3.1. Halaman Menu Utama**

Menu utama adalah halaman yang akan ditampilkan setelah *user* berhasil *login*, pada menu utama ini terdapat beberapa pilihan menu yaitu Prediksi Minat Studi, Statistik *Marketing*, *About*, *Credit*, *Help*, *Log Out*, dan *Exit*.

**3.3.2. Halaman Prediksi Minat Studi**

Prediksi Minat Studi adalah halaman yang akan tampil ketika *user* menekan *task item* Prediksi Minat Studi pada menu utama. Pada halaman Prediksi Minat Studi ini akan menampilkan *combo box* untuk *input* data target *marketing* berupa provinsi, kota, sekolah asal, jenis kelamin, dan jurusan. Pada halaman ini juga terdapat 2 tombol yaitu tombol PROSES dan tombol RESET. Selain itu pada halaman ini juga terdapat *chart* dan *text field* untuk menampilkan hasil dari prediksi minat studi.



**Gambar 3.** Tampilan Prediksi Minat Studi

**3.3.3. Halaman Statistik Marketing**

Statistik *Marketing* adalah halaman yang akan tampil ketika *user* menekan *task item* Statistik *Marketing* pada menu utama. Pada halaman Statistik *Marketing* ini akan menampilkan *combo box* untuk memilih tampilan statistik, apakah ingin ditampilkan statistik marketing dari semua data yang ada atau berdasarkan kota yang dipilih. Pada halaman ini juga terdapat 2 tombol yaitu tombol PROSES dan tombol RESET. Selain itu pada halaman ini juga terdapat *chart* dan *text field* untuk menampilkan hasil dari perhitungan statistik *marketing*.



**Gambar 4.** Tampilan Statistik Marketing

**3.3.4. Halaman Help**

Halaman *Help* adalah halaman yang akan tampil ketika *user* menekan *task item* *Help* pada menu utama. Pada halaman *Help* ini akan menampilkan manual program untuk mempermudah *user* dalam mengoperasikan aplikasi *Smart Marketing*.



**Gambar 5.** Tampilan Help

**4. Kesimpulan**

Kesimpulan yang dapat diambil adalah sebagai berikut:

1. *Smart Marketing* adalah aplikasi *data mining* menggunakan algoritma *naive bayes classifier* yang dapat memprediksi minat studi terhadap program studi S1 Teknik Informatika, S1 Sistem Informatika,

D3 Teknik Informatika, D3 Manajemen Informatika di STMIK AMIKOM Yogyakarta dari target marketing dengan harapan dapat mendukung strategi pemasaran khususnya di bagian HUMAS STMIK AMIKOM Yogyakarta.

2. Hasil dari uji coba 1000 *data testing* yang diambil secara acak dari data induk mahasiswa tahun 2012 dengan *data training* yang diambil dari data induk mahasiswa tahun 2009 sampai dengan 2011 menggunakan aplikasi *Smart Marketing* untuk prediksi minat studi didapatkan akurasi sebesar 92,7% dan error sebesar 7,3%.

## DAFTAR PUSTAKA

- [1] Anonim, *Sejarah Berdirinya Sekolah Tinggi Manajemen Informatika dan Komputer AMIKOM Yogyakarta*, <http://www.amikom.ac.id/index.php/profile/history>, diakses pada 12 Juli 2013 pukul 10.45 WIB.
- [2] Booch, G., James, R., dan Ivar, J., 1998, *The Unified Modeling Language User Guide First Edition*, Addison Wesley Professional.
- [3] Kusriani, dan Emha, T.L., 2009, *Algoritma Data Mining*, Yogyakarta : Andi.
- [4] Prasetyo, E., 2012, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Yogyakarta : Andi.
- [5] Santoso, B., 2007, *Data Mining : Teknik Pemanfaatan Data Untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu.

## Biodata Penulis

**Erik Hadi Saputra**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2003. Memperoleh gelar Master (M.Eng.) Program Pasca Sarjana Magister Teknologi Informasi Jurusan Teknik Elektro Universitas Gajah Mada Yogyakarta, lulus tahun 2011. Saat ini menjadi Dosen dan Kepala bagian Humas di STMIK AMIKOM Yogyakarta.

**Burhan Alfironi Muktamar**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2013, saat ini sebagai Ketua Forum Asisten di STMIK AMIKOM Yogyakarta.



