

**APLIKASI DATA MINING UNTUK ANALISIS MEMBANGUN STRATEGI
MARKETING DENGAN METODE KLASIFIKASI
MENGUNAKAN TEOREMA BAYES
(Studi Kasus: STMIK Bina Patria Magelang)**

Sugeng Wahyudiono

Program Studi Sistem Informasi

STMIK Bina Patria Magelang

Telepon (0293) 362 993

Email : swahyudiono@yahoo.com

Abstrak

Dalam penelitian berjudul “Aplikasi *Data Mining* untuk Analisis Membangun Strategi Marketing dengan Metode Klasifikasi Menggunakan Teorema Bayes (Studi Kasus: STMIK Bina Patria Magelang)” ini dikembangkan sebuah aplikasi *data mining* yang bertujuan untuk memudahkan tim marketing STMIK Bina Patria Magelang dalam menganalisa media promosi yang dipergunakan dan hasil analisa tersebut dapat dipergunakan untuk perencanaan marketing ke depan. Pengembangan aplikasi dilakukan dengan pendekatan berorientasi objek menggunakan UML (*Unified Modelling Language*), sedangkan proses *data mining* menggunakan CRISP-DM (*Cross Industry Standard Process for Data Mining*).

Berdasarkan hasil pengujian terhadap fungsional sistem menggunakan metode *black box*, diperoleh hasil bahwa semua fungsionalitas sistem sudah berjalan sebagaimana yang diharapkan. Pada pengujian terhadap hasil analisa *data mining*, dilakukan pengujian menggunakan data frekuensi media promosi “*Acara di Sekolah*” pada kabupaten Magelang. Hasil pengujian menggunakan aplikasi dan perhitungan secara manual menunjukkan hasil yang sama.

Dapat disimpulkan bahwa aplikasi *data mining* yang dikembangkan sudah memenuhi seluruh fungsionalitas yang diharapkan dan hasil analisa yang diperoleh adalah valid. Hasil analisa berupa prediksi probabilitas frekuensi media promosi untuk kabupaten/ kecamatan tertentu serta grafik frekuensi media promosi akan sangat membantu dalam menyusun strategi marketing, agar lebih terfokus dan tepat sasaran, yang pada akhirnya dapat mengurangi biaya marketing.

Kata Kunci: CRISP-DM, Bayes, *Black Box*

Pendahuluan

1.1 Latar Belakang

STMIK Bina Patria Magelang, sebagai sebuah Perguruan Tinggi Swasta yang juga membuka jurusan teknologi informasi mengalami dinamika dalam hal pertumbuhan

jumlah mahasiswa dari tahun ke tahun. Dinamika ini dipengaruhi

oleh banyak aspek. Salah satu aspek yang cukup memberikan dampak signifikan terhadap hal ini adalah strategi promosi untuk menarik minat calon mahasiswa. Dari hasil

pengumpulan data diperoleh data mengenai jumlah mahasiswa yang masuk, yaitu: tahun 2008 jumlah mahasiswa 145 orang, tahun 2009 231 mahasiswa, tahun 2010 279 mahasiswa, tahun 2011 279 mahasiswa dan tahun 2012 297 mahasiswa (Sumber Bagian Akademik STMIK Bina Patria Magelang). Melihat data di atas dapat diambil kesimpulan bahwa terdapat tren kenaikan dalam hal jumlah mahasiswa.

Pada saat ini proses penetapan rencana dalam strategi marketing di STMIK Bina Patria Magelang masih dilakukan secara tradisional, yaitu didasarkan pengamatan terhadap rekapitulasi jumlah mahasiswa per daerah dan memfokuskan kegiatan-kegiatan promosi pada daerah-daerah yang dipandang potensial serta tetap menjaga relasi dengan sekolah-sekolah yang setiap tahun menjadi sumber masuknya mahasiswa. Untuk saat ini, strategi ini masih dipandang aman untuk dilakukan. Terbukti dengan tren jumlah mahasiswa yang terus meningkat.

Dalam penelitian yang akan dilakukan ini, akan diteliti mengenai penggunaan *data mining* untuk analisis membangun strategi marketing yang digunakan oleh STMIK Bina Patria Magelang menggunakan metode klasifikasi teorema Bayes. Pemilihan teorema Bayes untuk dipergunakan dalam metode klasifikasi adalah karena teorema Bayes mudah dipahami, hanya memerlukan pengkodean yang sederhana dan lebih cepat dalam proses perhitungan. Teorema Bayes merupakan sebuah pendekatan

statistik yang fundamental dalam pengenalan pola (*Pattern Recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan *cost* yang ditimbulkan dalam keputusan-keputusan tersebut.

Alasan yang melatar belakangi penulis mengangkat topik ini, selain hal-hal yang sudah dikemukakan di atas, juga karena penulis sendiri terlibat langsung dalam proses promosi/ marketing STMIK Bina Patria Magelang. Dalam hal ini penulis melihat bahwa diperlukan adanya sebuah alat bantu berupa aplikasi yang dapat membantu tim marketing menganalisa media-media promosi yang dipergunakan saat ini untuk setiap daerah yang menjadi sasaran kegiatan marketing dan bagaimana peluang-peluangnya di tahun mendatang.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- 1) Membangun aplikasi *data mining* untuk membangun strategi marketing menggunakan metode klasifikasi dengan teorema Bayes pada STMIK Bina Patria Magelang.
- 2) Dengan menggunakan hasil analisa yang telah dilakukan menggunakan aplikasi *data mining*, tim marketing STMIK Bina Patria dapat menyusun rencana marketing yang efektif dan efisien.

1.3 Penelitian Pendukung

Hamzah (2012) dalam penelitian berjudul “*Klasifikasi Teks Dengan*

Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis”, diteliti mengenai kinerja klasifikasi Bayes untuk kategorisasi teks berita dan teks akademis. Yang melatar belakangi penelitian ini adalah bahwa perkembangan yang pesat dalam informasi digital telah menyebabkan semakin meningkat pula volume informasi yang berbentuk teks. Di antara berbagai bentuk informasi digital, diperkirakan 80% dokumen digital adalah dalam bentuk teks. Tingginya volume dokumen teks ini misalnya dengan aktivitas yang terus meningkat dari berbagai sumber berita dan aktivitas penulisan dokumen akademis dari kegiatan riset, konferensi dan pertemuan-pertemuan ilmiah. Kondisi “kebanjiran informasi” ini telah menimbulkan kesulitan manusia dalam mencerna informasi. Hal yang lebih menyulitkan dalam analisis adalah bahwa sekitar 80% sampai 85% bentuk informasi tersebut dalam format tidak terstruktur (unstructured data). Melimpahnya informasi teks tidak terstruktur telah mendorongnya munculnya disiplin baru dalam analisis teks, yaitu *text mining* yang mencoba menemukan pola-pola informasi yang dapat digali dari suatu teks yang tidak terstruktur tersebut.

Penelitian ini bertujuan untuk meneliti sejauh mana kinerja klasifikasi Bayes dalam kategorisasi teks yang berupa teks berita dan teks akademis berupa abstrak akademis dari berbagai disiplin ilmu. Fokus kajian dilakukan dalam

hal pemilihan *feature* kata yang akan dijadikan sebagai representasi dokumen dan sejauh mana efek banyaknya kata dalam *vocabulary* mempengaruhi kinerja algoritma klasifikasinya. Hasil penelitian menunjukkan pada dokumen berita akurasi maksimal dicapai 91% sedangkan pada dokumen akademik 82%. Seleksi kata dengan minimal muncul pada 4 atau 5 dokumen memberikan akurasi yang paling tinggi.

Penelitian selanjutnya adalah penelitian berjudul “*Data Mining Untuk Application Scoring Menggunakan Pemodelan Naive Bayes pada PT. Bank Rakyat Indonesia (PERSERO) TBK*” oleh Aprialdo dkk (2012). Sebagaimana diketahui bahwa Bank Rakyat Indonesia bergerak dalam bidang perbankan menyediakan fasilitas kartu kredit bagi para nasabahnya. Pengajuan kartu kredit bisa diajukan oleh para pemohon dengan memenuhi syarat-syarat yang ditentukan oleh Bank Rakyat Indonesia. Pemberian kartu kredit menjadi masalah yang krusial pada tahap evaluasi, karena Bank Rakyat Indonesia masih sulit menentukan pemohon yang dapat menggunakan fasilitas kartu kredit dan yang tidak beresiko menyebabkan kredit macet.

Dalam penelitian ini akan dikembangkan *data mining* untuk pembuatan *scoring credit* pada tahap evaluasi (*Application Scoring*). *Application Scoring* dapat digunakan sebagai alat penyaringan untuk memisahkan pemohon yang baik dan yang buruk. Dengan

menggunakan data mining diharapkan membantu proses evaluasi kredit yang dilakukan oleh Bank Rakyat Indonesia. Dari penelitian ini kemudian didapatkan hasil sebagai berikut:

- 1) Dengan menggunakan Data Mining dapat mempercepat waktu evaluasi kredit yang dilakukan oleh *credit analyst* dalam menentukan *application status* calon pemohon kartu kredit, sehingga *credit analyst* dapat menerima lebih banyak jumlah aplikasi calon pemohon kartu kredit. Hal ini dapat dibuktikan dengan rata-rata waktu yang dibutuhkan sistem untuk menampilkan *Application Status* pemohon adalah kurang dari 1 detik.
- 2) Dengan menggunakan Data Mining, Bank Rakyat Indonesia dapat membantu menentukan dengan lebih akurat calon pemohon yang layak *diapprove* aplikasi kartu kreditnya dengan bantuan pertimbangan sistem. Tingkat akurasi sistem dari pemodelan yang dihasilkan termasuk baik yaitu sebesar 73%. Dengan bantuan penilaian dari sistem, diharapkan *Credit Analyst* dapat mencegah *diapprovenya* calon pemohon yang dapat berpotensi melakukan kredit macet.

Penelitian lain yang juga menggunakan klasifikasi Bayes adalah penelitian yang dilakukan oleh Manjusha dkk (2014) dengan judul “*Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification*”. Pada

penelitian ini diteliti mengenai ketidakpastian data yang bisa disebabkan oleh berbagai faktor, misalnya latensi jaringan, pengukuran presisi dengan batasan, dan kesalahan *sampling*. Pada penelitian ini dikembangkan sebuah metode baru untuk menghitung probabilitas bersyarat teorema Bayes dengan mengaplikasikan teori probabilitas dan statistik pada data yang tidak pasti. Hasil penelitian menunjukkan metode baru mengklasifikasi data dengan tingkat akurasi yang lebih tinggi.

Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode pengembangan *data mining*, yaitu CRISP-DM (*Cross Industry Standar Process for Data Mining*). CRISP-DM merupakan sebuah proses standar untuk *data mining* yang dipergunakan secara luas dalam dunia industri (Olson & Delen, 2008). Crisp-DM terdiri dari enam fase yang merupakan sebuah siklus proses, yaitu:

- 1) *Bussiness understanding*
Pada tahap ini akan dilakukan pendalaman mengenai bisnis yang berjalan sehingga bisa ditetapkan tujuan atau sasaran bisnis, memperkirakan situasi terkini, menentukan sasaran dari *data mining* serta menyusun rencana proyek
- 2) *Data understanding* dan pengumpulan data
Pada tahapan ini, dilakukan proses pengumpulan data melalui pembagian kuisioner kepada seluruh mahasiswa STMIK Bina Patria Magelang. Setelah data yang diperoleh dieksplorasi dan diverifikasi,

kemudian akan dihasilkan sumber data bagi tahapan selanjutnya.

3) *Data preparation*

Data sumber selanjutnya pada fase ini akan dibersihkan dan ditransformasi ke dalam bentuk yang sesuai dengan model yang akan dipergunakan.

4) Perancangan dan pemodelan

Proses perancangan dilakukan dengan pendekatan berorientasi objek menggunakan *Unified Modelling Language* (UML). Dalam tahapan perancangan akan dibuat diagram-diagram UML, yaitu *use-case diagram*, *activity diagram*, *sequence diagram* dan *class diagram*. Selain itu pada tahapan ini juga akan dilakukan proses pemodelan menggunakan teorema Bayes.

5) Evaluasi

Proses untuk mengevaluasi model yang dibangun agar bisa beradaptasi terhadap perubahan-perubahan yang ada.

6) Pengembangan

Berdasarkan perancangan, kemudian akan dikembangkan sebuah aplikasi menggunakan bahasa pemrograman Java dengan editor *NetBeans 7.0.1* dan *MySQL server* sebagai *database server*.

Hasil dan Pembahasan

1.4 Analisis Kebutuhan Sistem

Pada tahapan ini akan diuraikan mengenai kebutuhan data yang akan dipergunakan dalam

proses pembuatan aplikasi. Sebelum dilakukan identifikasi kebutuhan data, terlebih dahulu dilakukan identifikasi terhadap aktor-aktor atau pelaku-pelaku yang nantinya akan berinteraksi dengan sistem. Pada tabel 3.1 ditampilkan aktor-aktor yang berperan di dalam sistem.

Tabel 1 Daftar aktor dan deskripsi

Aktor	Deskripsi
Administrator	Merupakan aktor yang berperan dalam proses administrasi sistem. Proses administrasi sistem meliputi seluruh aktivitas yang berlangsung di dalam sistem. Administrator memiliki hak akses penuh terhadap sistem.
Operator	Merupakan aktor yang berperan dalam proses input data. Hak akses yang dimiliki operator hanya sebatas pada proses input data ke dalam sistem.
Marketing	Merupakan aktor yang paling memiliki kepentingan terhadap sistem. Marketing memiliki hak akses yang lebih tinggi jika dibandingkan dengan operator, namun tidak memiliki hak akses penuh seperti administrator.

Selanjutnya untuk lebih memperjelas peranan setiap aktor, maka pada tabel 3.2 ditampilkan daftar kebutuhan fungsional terkait dengan peranan masing-masing aktor dalam interaksinya dengan sistem.

Tabel 2 Daftar kebutuhan fungsional

Aktor	Kebutuhan
Administrator	• Login ke dalam sistem. • Logout keluar sistem.

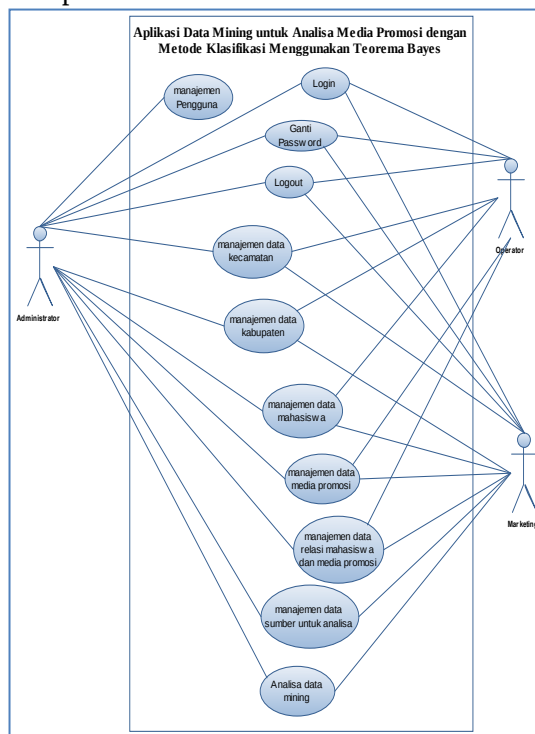
	• Ganti password • Manajemen pengguna (tambah, edit dan menampilkan). • Manajemen data mahasiswa (import dan menampilkan). • Manajemen data kecamatan (tambah, edit dan menampilkan). • Manajemen data kabupaten (tambah, edit dan menampilkan). • Manajemen data media promosi (tambah, edit dan menampilkan). • Manajemen data relasi mahasiswa dan media promosi (tambah, edit dan menampilkan). • Manajemen data sumber untuk analisa. • Analisa <i>data mining</i>.
Operator	• Login ke dalam sistem. • Logout keluar sistem. • Ganti password • Manajemen data mahasiswa (import dan menampilkan). • Manajemen data kecamatan (import dan menampilkan). • Manajemen data kabupaten (tambah, edit dan menampilkan). • Manajemen data media promosi (tambah, edit dan menampilkan). • Manajemen data relasi mahasiswa dan media promosi (tambah, edit dan menampilkan).
Marketing	• Login ke dalam sistem. • Logout keluar sistem. • Ganti password • Manajemen data mahasiswa (import dan menampilkan). • Manajemen data kecamatan (import dan menampilkan). • Manajemen data kabupaten (tambah, edit dan menampilkan). • Manajemen data media promosi (tambah, edit dan menampilkan). • Manajemen data relasi mahasiswa dan media promosi (tambah, edit dan menampilkan). • Manajemen data sumber untuk analisa. • Analisa <i>data mining</i>.

1.5 Rancangan Sistem

Use case diagram aplikasi yang akan dikembangkan dapat dilihat pada gambar 1. Dari diagram

tersebut dapat dilihat setiap interaksi yang terjadi antara aktor dengan sistem. *Use case diagram*

pada gambar 1 menggambarkan fungsional sistem yang sudah dirinci pada tabel 2.



Gambar 1 Use case diagram aplikasi data mining

1.6 Implementasi

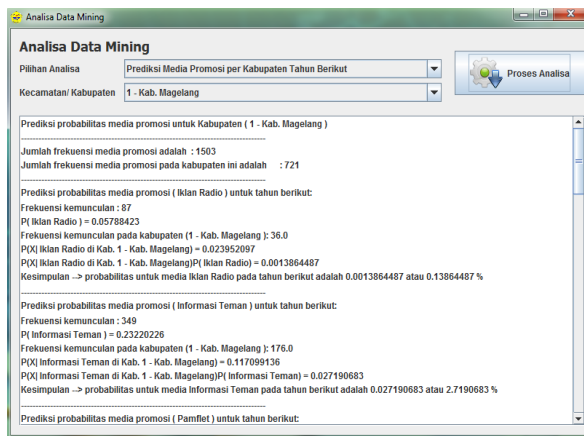
Class AnalisaDataMining merupakan *class* yang paling pokok dalam tahapan implementasi. *Class* ini dipergunakan untuk melakukan proses analisa *data mining*. Implementasi *class AnalisaDataMining* dapat dilihat pada gambar 2. Analisa *data mining* diawali dengan pemilihan jenis analisa dan berdasarkan pilihan analisa ini kemudian pengguna harus memilih kecamatan atau kabupaten tertentu. Kemudian klik tombol “*Proses Analisa*”, dan sistem akan menampilkan hasil

analisa serta menampilkan sebuah grafik.

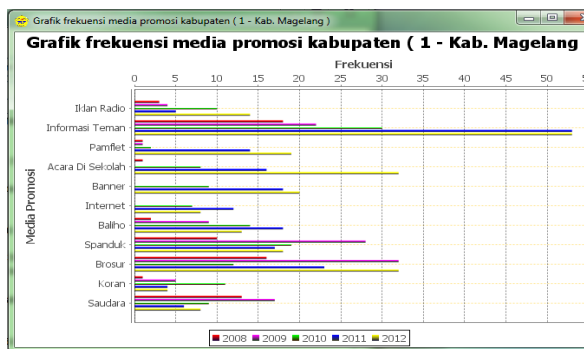
Pada gambar 2, dilakukan analisa “prediksi media promosi per kabupaten untuk tahun berikut” untuk kabupaten “Kab. Magelang”. Untuk pembahasan mengenai analisa data akan dilakukan pada bagian selanjutnya dalam penulisan ini, pada bagian ini hanya akan dikupas mengenai teknis implementasi sistem saja. Hasil dari analisa ini selanjutnya akan ditampilkan ke dalam sebuah *JList*.

Setelah hasil analisa ditampilkan, sistem akan menampilkan sebuah grafik yang memvisualisasikan frekuensi kemunculan setiap media promosi pada kabupaten atau kecamatan tertentu pada setiap tahun. Implementasi grafik ini dapat dilihat pada gambar 3.

Grafik pada gambar 3 diimplementasikan dengan memanfaatkan *library JFreeChart*. Sumber data yang dipergunakan dalam menampilkan disimpan ke dalam sebuah *ResultSet*. Grafis pada gambar 3.25 diimplementasikan pada *class BuatGrafik* mewarisi sifat dari *class ApplicationFrame*. *ApplicationFrame* merupakan sebuah *JFrame Class*.



Gambar 2 Implementasi class AnalisaDataMining



Gambar 3 Grafik frekuensi kemunculan media promosi pada kabupaten/kecamatan tertentu setiap tahun

1.7 Pembahasan

Gambar 3 memperlihatkan grafik frekuensi media promosi di Kab. Magelang untuk setiap tahun. Dari grafik tersebut dapat dilihat tren atau kecenderungan yang terjadi. Misalnya kecenderungan peningkatan frekuensi maupun kecenderungan penurunan frekuensi. Misalnya untuk media promosi “Acara Di Sekolah” diketahui bahwa terjadi kecenderungan

peningkatan setiap tahunnya. Kecenderungan ini dapat dipergunakan untuk menarik beberapa kesimpulan yang bisa bermanfaat dalam menentukan rencana marketing ke depan, misalnya:

- Kegiatan promosi ke sekolah-sekolah melalui acara-acara yang dilakukan antara STMIK Bina Patria Magelang yang melibatkan sekolah-sekolah di Kab. Magelang terbukti mendapatkan respon positif, dan ke depan media promosi ini dapat menjadi salah satu media promosi andalan dalam memperkenalkan STMIK Bina Patria Magelang dalam menjaring calon-calon mahasiswa.
- Jika dilihat dari sisi frekuensi pada grafik terlihat bahwa media promosi “Acara Di Sekolah” memiliki kecenderungan kenaikan yang stabil, jika kegiatan promosi dengan media ini terus dijaga dan ditingkatkan maka diyakini bahwa peningkatan jumlah peminat pun akan mengalami peningkatan yang stabil. Baik hasil perhitungan menggunakan analisa *data mining* maupun analisa menggunakan grafik, keduanya sangat bermanfaat dalam melakukan evaluasi serta penyusunan rencana strategis untuk pemasaran dan memperkenalkan kampus kepada calon-calon mahasiswa. Hasil analisa memberikan prediksi peluang masing-masing media promosi dan jika

dikombinasikan dengan tren atau kecenderungan yang bisa dilihat pada grafik, maka rencana strategis bidang pemasaran akan dapat dibuat secara lebih terukur dan tepat sasaran.

Kesimpulan dan Saran

Dalam penelitian ini dikembangkan sebuah aplikasi *data mining* yang dipergunakan untuk menganalisa media promosi yang dipergunakan STMIK Bina Patria untuk memperkenalkan diri kepada calon-calon mahasiswanya menggunakan metode klasifikasi dengan Teorema Bayes. Pengembangan *data mining* dilakukan menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) dan menggunakan pendekatan berorientasi objek (*Unified Modelling Language*) dalam proses pengembangan aplikasinya. Menggunakan fase-fase pengembangan *data mining* dengan metode CRISP-DM, data yang diperoleh kemudian dianalisa dan diolah sehingga bisa dihasilkan sumber data untuk selanjutnya dianalisa menggunakan teorema Bayes. Hasil analisa ini pun yang digunakan menjadi dasar dalam perancangan dan pengembangan sistem menggunakan pendekatan berorientasi objek dengan UML.

Hasil pengujian sistem menggunakan metode *black box* menunjukkan bahwa:

- 1) Secara fungsionalitas, aplikasi yang dibangun telah memenuhi semua persyaratan fungsionalitas yang ditetapkan.
- 2) Semua fitur yang diharapkan mampu bekerja dengan baik sesuai harapan.

Sedangkan dari pengujian hasil analisa baik perhitungan secara manual

maupun menggunakan hasil perhitungan oleh aplikasi, menunjukkan hasil yang sama. Sebagai contoh pada pengujian secara manual dan secara aplikasi untuk probabilitas media promosi “Acara Di Sekolah” di Kabupaten Magelang, didapatkan kesamaan hasil, yaitu 0.037924152 atau 3,79%. Jika kesamaan ini dipandang sebagai tolok ukur validitas hasil perhitungan sistem, maka dapat disimpulkan bahwa hasil analisa aplikasi yang dikembangkan adalah valid.

Bagaimana hasil dari analisa ini dipergunakan untuk memperbaiki keputusan dalam menyusun rencana marketing? Hasil analisa yang dihasilkan oleh aplikasi *data mining* menampilkan prediksi probabilitas atau peluang kemunculan suatu media promosi tertentu di lokasi tertentu. Dengan prediksi ini maka tim marketing dapat merencanakan strategi yang efektif dan efisien. Efektif sebab dengan adanya prediksi ini, tim marketing bisa memfokuskan proses marketing menggunakan media promosi yang memiliki probabilitas tinggi, dan selanjutnya hal ini akan mendorong terjadinya efisiensi dalam hal pembiayaan jika dibandingkan dengan promosi menggunakan seluruh atau sembarang media tanpa perhitungan atau pertimbangan yang jelas. Grafik yang ditampilkan juga sangat membantu. Tren atau kecenderungan frekuensi media promosi bisa dijadikan dasar dalam memfokuskan promosi menggunakan media promosi tertentu.

Berikut ini adalah saran-saran yang dapat penulis sampaikan agar ke depan penelitian yang dikembangkan ini bisa menghasilkan aplikasi yang handal:

- 1) Jika analisa yang dilakukan juga melibatkan aspek-aspek lain selain frekuensi media promosi, misalnya aspek biaya, aspek akreditasi program studi dan lain-lain, maka dapat dilakukan analisa yang lebih kompleks yang hasilnya tidak hanya bisa dipergunakan untuk menyusun rencana strategi pemasaran saja, namun sekaligus sebagai sarana untuk evaluasi setiap program studi.
- 2) Agar bisa diperoleh data yang *real time* atau terkini, sebaiknya aplikasi yang dihasilkan bisa diintegrasikan dengan sistem akademik yang berjalan

Olson, David L & Deren, Dursun, 2008, “*Advanced Data Mining Techniques*”, Springer, Berlin.

Daftar Pustaka

- Aprialdo, R, Reinhard, M & Ramadhani, S, 2012, “*Data Mining Untuk Application Scoring Menggunakan Pemodelan Naive Bayes pada PT. Bank Rakyat Indonesia (PERSERO) TBK*”, Universitas Bina Nusantara, Jakarta
- Hamzah, A, 2012, “*Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis*”, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, Yogyakarta. 2010
- Manjusha, K, Sankaranarayanan, K, Seen, P, 2014, “*Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification*”, International Journal of Advanced Research in Computer Science and Software Engineering Volume 4 1 Januari 2014.