

Application of Naive Bayes Classifier Algorithm in Determining New Student Admission Promotion Strategies

Penerapan Algoritma Naive Bayes Classifier Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru

Ahmad Haidar Mirza¹

¹Informatics Department, Universitas Bina Darma, Palembang, Indonesia
Email: haidarmirza@binadarma.ac.id

Abstract

Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence, machine learning to extract and identify useful information and related knowledge from large databases. Data mining is the process of finding new patterns in data by filtering large amounts of data. Data mining uses pattern recognition technology that is similar to statistical techniques and mathematical techniques. The patterns found can provide useful information for generating economic benefits, effectiveness and efficiency. Algorithm Naive Bayes Classifier is one method of data mining that can be used to support effective and efficient promotion strategies. The Naive Bayes Classifier algorithm is used to predict the interest of the study based on the calculations performed. The data used are new student registration data from 2014 until 2016 at Bina Darma University. The results of this study are new models that are expected to provide important information can be used to assist the Marketing Team of Bina Darma University Palembang in policy making and implementation of appropriate marketing strategy. The results obtained are expected to help to support the promotion strategies that impact on the effectiveness and efficiency of promotion and increase the number of new students who will register.

Keywords: Data Mining, Naive Bayes, Classification

1. PENDAHULUAN

The introduction must contain a general background (shortly), a literature review (state of the art) in order to record the existing method/solutions, to show which is the best of previous researches and to show the main limitation of the previous researches. It has be contain with at least 5 literature in order to justify novelty this paper. The introduction should be clearly contain the gap analysis (why this research needs to be done? What is the uniqueness of this paper Seiring dengan



perkembangan di era globalisasi dan kemajuan dibidang teknologi informasi yang cepat memberikan pengaruh yang cukup besar baik dalam bidang industri maupun dibidang pendidikan. Hal ini juga membawa suatu perubahan besar dalam tingkat persaingan antara instansi dan perusahaan. Banyak perusahaan dan instansi pemerintah yang menggunakan data mining untuk menggali informasi. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar [7]. Selain perusahaan dan instansi lainnya, 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.

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. Naive Bayes Classifier adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition) Pendekatan ini didasarkan pada kuantifikasi trade-off antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang ditimbulkan dalam keputusan-keputusan tersebut [3]. Informasi ini dapat digunakan untuk mendukung strategi promosi agar lebih efektif dan efisien. Sehingga perguruan tinggi dapat mengetahui serta menentukan target atau sasaran pasar dengan lebih rinci.

Ketersediaan data yang banyak dan kebutuhan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis dan dukungan infrastruktur di bidang teknik informatika merupakan cikal- bakal dari lahirnya teknologi data mining. Penggunaan teknik data mining diharapkan dapat membantu mempercepat proses pengambilan keputusan, memungkinkan perusahaan untuk mengelola informasi yang terkandung didalam data transaksi menjadi sebuah pengetahuan (knowledge) yang baru. Dalam hal perencanaan kegiatan promosi guna peningkatan minat calon mahasiswa untuk mendaftar, Unit Pemasaran Universitas Bina Darma membutuhkan masukan – masukan guna meningkatkan minat lulusan Sekolah Menengah Atas (SMA) atau setara yang ada di kota Palembang pada khususnya dan di Sumatera Selatan pada umumnya untuk mejadi Mahasiswa Universitas Bina Darma, hal ini dikarenakan kegiatan promosi yang terdiri dari kegiatan pemasangan iklan, penyebaran brosur dan promosi dengan mengundang sekolah masih belum menyebar luas kedaerah-daerah yang tingkat peminatnya masih rendah sehingga masyarakat yang ada didaerah atau dipedesaan masih belum banyak yang mengetahui tentang Universitas Bina Darma. Dengan demikian, Unit Pemasaran Universitas Bina Darma Palembang perlu untuk menggali atau mencari informasi dari data pendaftaran tahun-tahun sebelumnya sebagai bahan evaluasi dan analisa untuk mendukung strategi promosi tahun berikutnya. Konsep bayesian classification ditujukan untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

2. METODE

1. Metode Penelitian

Metode penelitian yang digunakan adalah metode analisis deskriptif dengan pendekatan kuantitatif Menurut Sugiyono, dapat diartikan sebagai metode penelitian yang berlandaskan pada filsafat positivisme, digunakan untuk meneliti pada populasi atau sampel tertentu. Teknik pengambilan sampel pada umumnya dilakukan secara random, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif / statistik dengan tujuan untuk menguji hipotesis yang telah ditetapkan [6]. Menurut Sugiyono [5] “Metode deskriptif adalah metode yang digunakan untuk menggambarkan analisis suatu hasil penelitian tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas”. Penelitian deskriptif merupakan suatu kegiatan yang dilakukan untuk menemukan pengetahuan mengenai objek dan subjek secara detail dalam kurun waktu tertentu. Selain itu penelitian dengan metode deskriptif juga merupakan penelitian yang bertujuan untuk menjelaskan, menggambarkan atau mendeskripsikan suatu keadaan objek, serta segala sesuatu terhadap variable – variable yang dapat dijelaskan baik secara hitungan angka maupun kata [5].

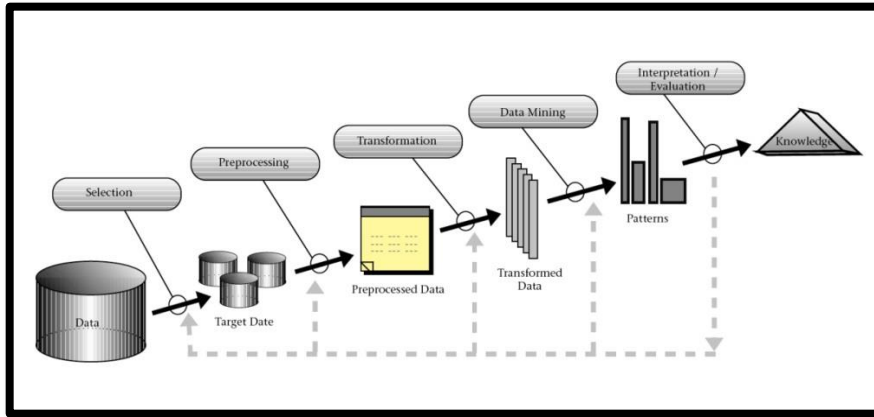
2. Metode Pengumpulan Data

Salah satu tahapan terpenting pada suatu penelitian yaitu pengumpulan data. Pengumpulan data merupakan kegiatan atau cara yang dapat diterapkan oleh peneliti dalam usahanya untuk mengumpulkan data yang dibutuhkan[1]. Selanjutnya dijelaskan oleh [2] bahwa pengumpulan data dapat dilakukan dengan cara pengamatan/observasi, wawancara, dokumentasi, serta gabungan/triangulasi. Pada penelitian ini menggunakan teknik pengumpulan data dengan melakukan pengamatan/observasi, pencarian dokumen, dan wawancara. Melakukan studi pustaka / studi literatur. Studi pustaka/studi literatur adalah suatu cara pengumpulan data dengan mempelajari teori yang berkaitan dengan cara membaca buku- buku/literatur, makalah maupun referensi- referensi.

3. Metode Knowledge Discovery in Database (KDD)

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi serta mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar [1]. Dalam meningkatkan strategi promosi sangat memerlukan adanya proses *data mining* karena proses *data mining* sangat bermanfaat dalam mengambil informasi dari data base yang sudah ada. Istilah data

mining dan *Knowledge Discovery in Databases* (KDD) menurut Kusriani [4], sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD (*Knowledge discovery in databases*) secara garis besar dapat dijelaskan sebagai berikut:



Gambar 1. Tahapan *Knowledge Discovery in Databases* [4]

a. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *knowledge data discovery* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. *Preprocessing* atau *Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *knowledge data discovery*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang *inkonsisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam *knowledge data discovery*

merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation* atau *evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada pada sebelumnya.

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada pada sebelumnya

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data Mining

Data yang akan digunakan pada proses *data mining* menggunakan Algoritma *Naïve Bayes* merupakan data mahasiswa Penerimaan Mahasiswa Baru (PMB) untuk tahun 2014, 2015 dan 2016 yang merupakan gambaran secara nyata mengenai keadaan penerimaan mahasiswa baru untuk seluruh yang ada di lingkungan Universitas Bina Darma. Setelah data di dapat dari Unit Pelaksana Teknis – Sistem Informasi Manajemen (UPT-SIM) Universitas Bina Darma, maka data akan di standarisasi mengikuti proses tahapan *data mining* agar data tersebut layak dan dapat diolah menggunakan *software* pengolahan *data mining RapidMiner* . dengan menggunakan algoritma *Naïve Bayes*.

3.2 Proses Data Mining Knowledge Discovery in Database (KDD)

a. Data Selection

Pada tahap ini data yang digunakan akan diseleksi dengan cara melihat kecenderungan data / kesesuaian data dengan topik / judul penelitian yang akan diteliti oleh penulis, dalam hal ini data yang di peroleh oleh penulis dari UPT-SIM sudah memiliki kesesuaian format data yang terdiri dari atribut Tahun PMB, Nama,

Alamat, Asal_Sekolah, ProgdI dan Status. Adapun atribut yang ada pada masing – masing data sebagai berikut:

Tabel 1. Atribut Database PMB

No	Atribut	Keterangan
1	Tahun PMB	Tahun dari Penerimaan Mahasiswa
2	Nama	Nama Mahasiswa
3	Alamat	Alamat Mahasiswa
4	Asal_Sekolah	Asal Sekolah Mahasiswa
5	ProgdI	Program Studi yang di pilih oleh Mahasiswa
6	Status	Status PMB Mahasiswa

b. *Data Pre Processing* atau *Data Cleaning*

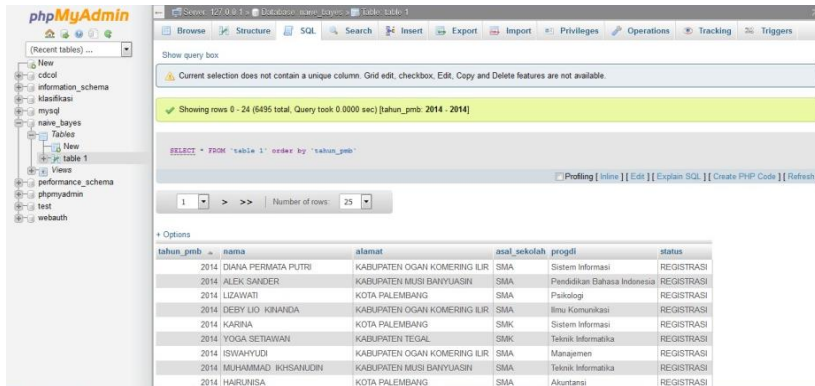
Pada tahap ini akan dilakukan perbersihan data, yakni membuang data yang tidak konsisten dan noise/redudancy data. Pada data yang didapat, terdapat beberapa data yang tidak konsisten dan noise. Pada data tersebut diatas dapat dilihat adanya data yang kosong sehingga data tersebut harus di buang / hapus / delete agar pada saat proses klasifikasi dapat dilakukan perhitungan untuk proses data mining

c. *Data Integration*

Tahap integrasi data adalah tahap penggabungan data dari berbagai sumber. Dataset mahasiswa Penerimaan Mahasiswa Baru (PMB) Universitas Bina Darma berasal dari satu sumber yaitu Unit Pelaksana Teknis – Sistem Informasi Manajemen (UPT-SIM), dan dari data yang ada tersebut dapat digunakan untuk proses pengolahan data mining dengan menggunakan metode Naïve Bayes. Data ini terdiri dari data PMB untuk tahun 2014, 2015 dan 2016 yang kemudian digabungkan kedalam satu dataset untuk di ubah ke format CSV dengan menggunakan Ms. Excel yang nantinya akan di Import ke dalam Database Mysql untuk proses Data Manipulation

d. *Data Transformation*

Ada pun pada tahap ini data akan diubah menjadi bentuk yang sesuai untuk proses data mining. Karena dalam penelitian ini akan dilakukan uji coba secara teoritis dan menggunakan software data mining yaitu RapidMiner, maka data yang telah melalui proses sebelumnya akan di transformasi agar dapat sesuai dengan algoritma yang dipakai yaitu algoritma Naïve Bayes. Pada tahapan ini atribut yang di pakai akan diberi label mengikuti kondisi data - data pada atribut tersebut. Berikut tampilan hasil klasifikasi pada masing – masing atribut pada Mysql:



Gambar 2. Hasil dari Proses Klasifikasi Pada Masing – Masing Atribut

Setelah hasil proses klasifikasi tersebut, maka langkah selanjutnya yaitu mengubah format data hasil klasifikasi menjadi format CSV agar dapat di olah dengan menggunakan software data mining yaitu RapidMiner. Data yang sudah di dilakukan pada proses tahapan diatas ini adalah data yang sudah siap atau sudah bersih dari data yang redundancy dan data yang loss (Kosong) sehingga data tersebut bisa dilakukan proses data mining dengan naive bayes seperti pada proses tahap hasil selanjutnya. Berikut gambar dibawah yang siap untuk dilakukan proses data mining dengan naive bayes.

e. Data Mining dengan Naive Bayes

Berdasarkan hasil dari data set yang telah di klasifikasikan dan siap untuk di olah adapun hasil dari distribusi frekuensi data untuk perhitungan teoritis untuk algoritma Naive Bayes dapat dilihat pada Tabel dibawah ini.

Tabel 2. Distribusi Frekuensi Data PMB

Independent Variabel	Value	count			
		Tidak Registrasi		Tidak Registrasi +1	
		Registrasi	asi	Registrasi +1	asi+1
Alamat	KABUPATEN AGAM	3	0	4	1
	KABUPATEN ASAHAN	1	0	2	1
	KABUPATEN BANGKA	13	0	14	1
	KABUPATEN BANGKA BARAT	19	2	20	3
	KABUPATEN BANGKA SELATAN	9	0	10	1
	KABUPATEN BANGKA TENGAH	1	0	2	1
	KABUPATEN BANGKALAN	1	0	2	1
	KABUPATEN BANTUL	1	0	2	1

KABUPATEN BANYUASIN	367	16	368	17
KABUPATEN BANYUMAS	1	0	2	1
KABUPATEN BATANGHARI	1	1	2	2
KABUPATEN BELITUNG	3	1	4	2
KABUPATEN BELITUNG TIMUR	3	0	4	1
KABUPATEN BENGKULU	9	1	10	2
SRI LANTAN KABUPATEN BLORA	1	0	2	1
KABUPATEN BOGOR	6	1	7	2
KABUPATEN BUNGO	1	1	2	2
KABUPATEN CIANJUR	1	0	2	1
KABUPATEN CILACAP	3	0	4	1
KABUPATEN CIREBON	1	0	2	1
KABUPATEN EMPAT LAWANG	54	2	55	3
KABUPATEN GARUT	2	0	3	1
KABUPATEN HUMBANG	1	0	2	1
HASUNDUTAN KABUPATEN INDRAGIRI HILIR	3	0	4	1
KABUPATEN INDRAGIRI HULU	3	0	4	1
KABUPATEN INDRAMAYU	0	1	1	2
KABUPATEN JENEPONTO	1	0	2	1
KABUPATEN JOMBANG	1	0	2	1
KABUPATEN KAMPAR	1	0	2	1
KABUPATEN KARANGANYAR	1	1	2	2
KABUPATEN KARO	5	0	6	1
KABUPATEN KAUR	1	0	2	1
KABUPATEN KEBUMEN	1	0	2	1
KABUPATEN KEPAHANG	1	1	2	2
KABUPATEN KEPULAUAN	1	1	2	2
MUNTAWAI KABUPATEN KERINCI	3	0	4	1
KABUPATEN KLATEN	1	1	2	2
KABUPATEN KOLAKA	1	0	2	1
KABUPATEN KUANTAN	3	0	4	1
SINGINGI KABUPATEN KULONPROGO	1	0	2	1
KABUPATEN KUTAI	1	0	2	1
KARTANEGARA KABUPATEN LABUHAN BATU	1	1	2	2

KABUPATEN LAHAT	170	14	171	15
KABUPATEN LAMPUNG BARAT	3	1	4	2
KABUPATEN LAMPUNG SELATAN	6	1	7	2
KABUPATEN LAMPUNG TENGAH	3	0	4	1
KABUPATEN LAMPUNG TIMUR	2	1	3	2
KABUPATEN LAMPUNG UTARA	5	0	6	1
KABUPATEN LEBONG	1	0	2	1
KABUPATEN LEMBATA	1	0	2	1
KABUPATEN MADIUN	1	0	2	1
KABUPATEN MALUKU TENGGARA	1	0	2	1
KABUPATEN MANDAILING BARAT	2	0	3	1
KABUPATEN MANGGARAI BARAT	1	0	2	1
KABUPATEN MEGELANG	3	0	4	1
KABUPATEN MERANGIN	1	1	2	2
KABUPATEN MUARA ENIM	280	20	281	21
KABUPATEN MUKO-MUKO	0	1	1	2
KABUPATEN MUSI BANYUASIN	267	20	268	21
KABUPATEN MUSI RAWAS	50	2	51	3
KABUPATEN NGANJUK	1	0	2	1
KABUPATEN NIAS	1	0	2	1
KABUPATEN NIAS SELATAN	1	0	2	1
KABUPATEN OGAN ILIR	252	17	253	18
KABUPATEN OGAN KOMERING ULU	330	22	331	23
KABUPATEN OGAN KOMERING DUNDA	149	7	150	8
KABUPATEN OKU SELATAN	71	0	72	1
KABUPATEN OKU TIMUR	123	9	124	10
KABUPATEN PADANG LAWAS BARAT	1	0	2	1
KABUPATEN PADANG DARUSSALAM	2	0	3	1
KABUPATEN PALI	76	2	77	3
KABUPATEN PATI	1	1	2	2
KABUPATEN PROBOLINGGO	1	0	2	1
KABUPATEN REJANG LEBONG	12	0	13	1
KABUPATEN ROKAN HULU	1	0	2	1
KABUPATEN SAROLANGUN	7	0	8	1

KABUPATEN SELUMA	1	0	2	1
KABUPATEN SEMARANG	3	0	4	1
KABUPATEN SERANG	2	0	3	1
KABUPATEN SIAK	1	0	2	1
KABUPATEN SIDOARJO	1	0	2	1
KABUPATEN SIMALUNGUN	1	0	2	1
KABUPATEN SIMEULUE	1	0	2	1
KABUPATEN SLEMAN	3	0	4	1
KABUPATEN SOLOK	4	1	5	2
KABUPATEN SRAGEN	2	0	3	1
KABUPATEN SUBANG	1	0	2	1
KABUPATEN SUMEDANG	2	0	3	1
KABUPATEN TANGGAMUS	2	0	3	1
KABUPATEN TANJUNG JABUNG	4	0	5	1
<small>RARAT</small> KABUPATEN TANJUNG JABUNG	1	0	2	1
<small>TTIMUR</small> KABUPATEN TAPANULI	1	0	2	1
<small>SFI ATAN</small> KABUPATEN TASIKMALAYA	3	0	4	1
KABUPATEN TEGAL	1	0	2	1
KABUPATEN TEMANGGUNG	1	0	2	1
KABUPATEN TOBA SAMOSIR	1	0	2	1
KABUPATEN TULANG BAWANG	3	1	4	2
KABUPATEN WAJO	1	1	2	2
KABUPATEN WAY KANAN	7	0	8	1
KOTA BANDA ACEH	1	0	2	1
KOTA BANDAR LAMPUNG	11	4	12	5
KOTA BANDUNG	4	3	5	4
KOTA BANJAR	1	0	2	1
KOTA BATAM	5	0	6	1
KOTA BEKASI	2	0	3	1
KOTA BENGKULU	7	1	8	2
KOTA BUKITINGGI	2	1	3	2
KOTA CIMAHI	5	0	6	1
KOTA DEPOK	1	1	2	2
KOTA DUMAI	2	1	3	2

	KOTA JAKARTA BARAT	3	0	4	1
	KOTA JAKARTA PUSAT	1	0	2	1
	KOTA JAKARTA SELATAN	2	0	3	1
	KOTA JAKARTA TIMUR	3	2	4	3
	KOTA JAKARTA UTARA	3	0	4	1
	KOTA JAMBI	23	3	24	4
	KOTA KENDARI	1	0	2	1
	KOTA LUBUK LINGGAU	63	7	64	8
	KOTA MADIUN	1	0	2	1
	KOTA MANADO	1	0	2	1
	KOTA MEDAN	2	0	3	1
	KOTA METRO	2	0	3	1
	KOTA PADANG	9	0	10	1
	KOTA PADANG SIDEMPUAN	1	0	2	1
	KOTA PAGAR ALAM	88	5	89	6
	KOTA PALEMBANG	3273	178	3274	179
	KOTA PANGKAL PINANG	12	0	13	1
	KOTA PEMATANG SIANTAR	2	0	3	1
	KOTA PRABUMULIH	167	13	168	14
	KOTA SALATIGA	1	0	2	1
	KOTA SIBOLGA	1	0	2	1
	KOTA SURABAYA	6	0	7	1
	KOTA SURAKARTA	1	0	2	1
	KOTA TANGERANG	5	2	6	3
	KOTA YOGYAKARTA	2	2	3	3
		6119	376	6120	377
Asal Sekolah	SMA	4393	243	4394	244
	SMK	1472	106	1473	107
	MADRASAH	256	25	257	26
		6121	374	6122	375
Progdi	Administrasi Bisnis	81	6	82	7
	Akuntansi	676	45	677	46

Ilmu Komunikasi	197	15	198	16
Komputerisasi Akuntansi	61	4	62	5
Manajemen	787	64	788	65
Manajemen Informatika	201	10	202	11
Manajemen Perusahaan	67	6	68	7
Pendidikan Bahasa Indonesia	46	6	47	7
Pendidikan Olahraga	185	7	186	8
Psikologi	246	23	247	24
Sastra Inggris	74	5	75	6
Sistem Informasi	1459	85	1460	86
Teknik Elektro	111	8	112	9
Teknik Industri	321	22	322	23
Teknik Informatika	1070	24	1071	25
Teknik Komputer	169	18	170	19
Teknik Sipil	370	26	371	27
	6121	374	6122	375
Status	6121	374	6122	375

Hasil perhitungan *simple distribution* data dari tabel distribusi frekuensi perhitungan teoritis untuk Status Registrasi dan Status Tidak Registrasi sebagai berikut :

$$P(\text{REGISTRASI}) = (6122 / 6495) = 0.942$$

$$P(\text{TIDAK REGISTRASI}) = (375 / 6495) = 0.057$$

f. Proses Pengujian Data

Setelah di dapat hasil *simple distribution* maka dapat dilakukan proses pengujian berdasarkan Persamaan dari *teorema Bayes* adalah:

$$P(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)}$$

Dimana:

X : Data dengan kelas yang belum diketahui

H : Hipotesis data X merupakan suatu kelas spesifik

P(H|X): Probabilitas hipotesis H berdasar kondisi X (*posterior probability*) P(H) :

Probabilitas hipotesis H (*prior probability*)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H $P(X)$: Probabilitas X, Diketahui : X_1 = Alamat, X_2 = Asal Sekolah X_3 = Program Studi Y = Status

Proses Pengujian 1

Jika Diketahui Alamat = Palembang, Asal Sekolah = SMA, ProgdI = Sistem Informasi.

$$P(\text{REGISTRASI}) = (6122 / 6495) = 0.942$$

$$P(\text{TIDAK REGISTRASI}) = (375 / 6495) = 0.057$$

Pengujian

$$P(\text{Palembang} | \text{REGISTRASI}) = 3274 / 6122 = 0.535$$

$$P(\text{Palembang} | \text{TIDAK REGISTRASI}) = 179 / 375 = 0.477$$

$$P(\text{SMA} | \text{REGISTRASI}) = 4394 / 6122 = 0.717$$

$$P(\text{SMA} | \text{TIDAK REGISTRASI}) = 244 / 375 = 0.513$$

$$P(\text{Sistem Informasi} | \text{REGISTRASI}) = 1460 / 6122 = 0.238$$

$$P(\text{Sistem Informasi} | \text{TIDAK REGISTRASI}) = 86 / 375 = 0.229$$

HMAP Kesimpulan :

$$P(\text{REGISTRASI}) = (3274/6122) * (4394/6122) * (1460/6122) * (6122/6495) = 0.0860$$

$$P(\text{TIDAK REGISTRASI}) = (179/375) * (244/375) * (86/375) * (375/6495) = 0.0031$$

Kesimpulan $P(\text{REGISTRASI}) > P(\text{TIDAK REGISTRASI})$ sehingga Calon Mahasiswa dengan kriteria tersebut dapat di prediksi **Registrasi**.

Proses Pengujian 2

Jika Diketahui Alamat = Agam, Asal Sekolah = SMA, ProgdI = Sistem Informasi

$$P(\text{REGISTRASI}) = (6122 / 6495) = 0.942$$

$$P(\text{TIDAK REGISTRASI}) = (375 / 6495) = 0.057$$

Pengujian

$$P(\text{Agam} | \text{REGISTRASI}) = 4 / 6122 = 0.00065$$

$$P(\text{Agam} | \text{TIDAK REGISTRASI}) = 1 / 375 = 0.002666$$

$$P(\text{SMA} | \text{REGISTRASI}) = 4394 / 6122 = 0.717$$

$$P(\text{SMA} \mid \text{TIDAK REGISTRASI}) = 244 / 375 = 0.513$$

$$P(\text{Sistem Informasi} \mid \text{REGISTRASI}) = 1460 / 6122 = 0.238$$

$$P(\text{Sistem Informasi} \mid \text{TIDAK REGISTRASI}) = 86 / 375 = 0.229$$

HMAP Kesimpulan :

$$\begin{aligned} P(\text{REGISTRASI}) &= (4/6122)*(4394/6122)* \\ & (1460/6122)*(6122/6495) = \\ & 0.000104486 \end{aligned}$$

$$\begin{aligned} P(\text{TIDAK REGISTRASI}) &= (1/375)*(244/375)* (86/375)*(375/6495) \\ & = 0.00001785 \end{aligned}$$

Kesimpulan $P(\text{REGISTRASI}) > P(\text{TIDAK REGISTRASI})$ sehingga Calon Mahasiswa dengan kriteria tersebut dapat di prediksi **Registrasi**.

Berdasarkan data hasil dan pembahasan diatas dapat kita jadikan sebagai acuan untuk menentukan strategi promosi pada Universitas Bina Darma Palembang dengan melihat data atribut dari Alamat, Asal Sekolah dan Program Studi. Dimana dari data Penerimaan Mahasiswa Baru (PMB) dari tahun 2014-2016 dengan jumlah data 6495 yang registrasi 6121 dan tidak registrasi 374. Dari atribut alamat data mahasiswa yang mendaftar tertinggi adalah beralamat Kota Palembang dengan jumlah pendaftar 3273, yang mendaftar berdasarkan tingkat sekolah asal adalah SMA dengan jumlah pendaftar 4393 dan pilihan jurusan yang tertinggi adalah Program Studi Sistem Informasi dengan jumlah 1459. Jadi dalam menentukan strategi promosi yang harus kita lihat adalah dari alamat dan asal sekolah yang tingkat pendaftar rendah, dan yang lebih harus ditonjolkan di Universitas Bina Darma adalah Program Studi yang pendaftarnya rendah, Sehingga pendaftar akan tertarik dengan Universitas Bina Darma Palembang.

4. KESIMPULAN

Adapun kesimpulan dari penelitian ini dalam memprediksi proses REGISTRASI dan TIDAK REGISTRASI mahasiswa menggunakan data set yang sudah melalui tahap awal Knowledge Dciscovey in Database (KDD) sebanyak 6495 record yang diambil dari data penerimaan mahasiswa baru tahun 2014, 2015 dan 2016 sebagai berikut:

- 1) Proses pengolahan data mining menggunakan atribut alamat dengan klasifikasi berdasarkan nama kota / nama kabupaten, atribut asal sekolah yang di klasifikasi berdasarkan pendidikan SMA, SMK dan MADRASAH, atribut progdi yang diklasifikasikan berdasakan program studi dari masing – masing fakultas yang ada dilingkungan Universitas Bina Darma dan atribut status sebagai label dalam proses pengolahan data mining dengan klasifikasi REGISTRASI DAN TIDAK REGISTRASI.

- 2) Model data mining di buat menggunakan software pengolahan data mining yaitu Rapidminer, dengan hasil nilai Probalitas untuk REGISTRASI yaitu 0.942 dengan pembulatan nilai menjadi 0.94 dan nilai probabilitas untuk TIDAK REGISTRASI yaitu 0.057 dengan pembulatan nilai menjadi 0.058.
- 3) Model data mining yang di dihasilkan pada Rapidminer dapat memprediksi hasil dengan beberapa kriteria klasifikasi berdasarkan klasifikasi yang dibuat menurut atribut masing – masing mampu menilai seorang calon Mahasiswa Baru.
- 4) Dengan menggunakan metode Naïve Bayes Classifier mampu menunjukkan informasi yang tersembunyi pada data atribut – atribut yang digunakan dalam proses pengolahan data mahasiswa pada proses Penerimaan Mahasiswa Baru (PMB) di Universitas Bina Darma serta memberikan solusi untuk membuat strategi promosi dengan melihat alamat, asal sekolah, dan progdi pilihan dari calon mahasiswa baru.

REFERENSI

- [1]. Arikunto, Suharsimi. 2011. *Prosedur Penelitian: Suatu Pendekatan Praktik*. Edisi Revisi VII. Jakarta: PT. Rineka Cipta.
- [2]. Basuki, Sulisty. 2006. *Metode Penelitian*. Jakarta: Wedatama Widya Sastra.
- [3]. Budi, Santoso, 2007, *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta
- [4]. Kusrini dan Luthfi. E. Taufiq.(2009). *Algoritma Data Mining*. Yogyakarta: Andi. Laksana, Fajar. 2008. *Manajemen Pemasasaran : Pendekatan Praktis*. Edisi Pertama. Cetakan Pertama. Graha Ilmu. Yogyakarta.
- [5]. Setyosari, Punaji. 2010. *Metode Penelitian dan Pengembangan*. Jakarta: Kencana
- [6]. Sugiyono. 2012. *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- [7]. Turban, E., dkk, 2005, *Decicion Support Systems and Intelligent Systems, Andi Offse*
- [8]. www.binadarma.ac.id , *website* Universitas Bina Darma Palembang pada tanggal 29-05-2016