

ALGORITMA NAÏVE BAYES UNTUK PREDIKSI JUMLAH PENDAFTAR ULANG PADA PENERIMAAN MAHASISWA BARU

Ariesta Damayanti^a

^a Teknik Informatika STM IK AKAKOM, Yogyakarta

INFORMASI ARTIKEL

Diterbitkan Online: 31 Juli 2019

KATA KUNCI

naïve bayes, penerimaan, prediksi, registrasi

KORESPONDENSI

Email : iest_ayanthi@akakom.ac.id

ABSTRACT

STM IK AKAKOM Yogyakarta setiap tahunnya melakukan penerimaan mahasiswa baru yang dilakukan oleh bagian marketing dan admisi, penerimaan mahasiswa baru sangat penting untuk STM IK AKAKOM Yogyakarta karena operasional kampus dibiayai oleh pemasukan yang berasal dari SPP mahasiswa. Sehingga diperlukan suatu sistem untuk bisa melakukan prediksi jumlah mahasiswa baru setiap tahunnya, sebagai informasi bagi manajemen sebagai dasar pengelolaan kegiatan kampus.

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naïve Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi seperti decision tree dan neural network. Naïve Bayes digunakan untuk memprediksi jumlah mahasiswa baru dengan menggunakan data pendaftar ulang di tahun sebelumnya yang memiliki atribut yaitu asal kota, gelombang, program studi, penghasilan orang tua, nilai U N dan status registrasi, sehingga pihak marketing dan admisi STM IK AKAKOM Yogyakarta mendapat gambaran jumlah mahasiswa baru ditahun depan.

Hasil dari penelitian ini adalah sistem yang dapat memprediksi data dengan kelas yaitu registrasi dan tidak registrasi. Dari 1704 data testing yang di proses menggunakan sistem didapatkan hasil prediksi registrasi sebanyak 1226 data dan tidak registrasi 478 data. Untuk pengujian dari 731 data didapatkan hasil prediksi 679 data terprediksi benar dan 52 data salah prediksi. Tingkat akurasi probabilitas yang didapatkan sebesar 92,88%.

1. PENDAHULUAN

Sekolah Tinggi Manajemen Informatika dan Komputer (STM IK) AKAKOM Yogyakarta berdiri 8 Juni 1992, berdasarkan Surat Keputusan Direktorat Jendral Pendidikan Tinggi Departemen Pendidikan dan Kebudayaan RI Nomor 262/DIKTI/Kep/1992, AMIK AKAKOM diubah bentuknya menjadi Sekolah Tinggi Manajemen Informatika dan Komputer (STM IK) AKAKOM yang diselenggarakan oleh yayasan Widya Bakti. STM IK AKAKOM Yogyakarta mempunyai 5 program studi yaitu Teknik Informatika, Sistem Informasi, Manajemen Informatika, Komputer Akutansi dan Teknik Komputer.

Setiap tahunnya STM IK AKAKOM Yogyakarta melakukan penerimaan mahasiswa baru yang dilakukan oleh bagian marketing dan admisi, penerimaan mahasiswa baru sangat penting untuk STM IK AKAKOM Yogyakarta karena operasional kampus dibiayai oleh pemasukan yang berasal dari SPP mahasiswa. STM IK AKAKOM Yogyakarta mengharapkan jumlah mahasiswa baru setiap tahunnya sesuai target, sehingga bagian marketing dan admisi STM IK AKAKOM Yogyakarta memilki jumlah target penerimaan mahasiswa baru setiap tahunnya, perlunya target penerimaan mahasiswa baru digunakan untuk memberikan gambaran jumlah mahasiswa baru ditahun selanjutnya, sehingga diperlukan suatu sistem untuk melakukan prediksi target mahasiswa baru di tahun selanjutnya agar bagian marketing dan admisi mendapatkan gambaran jumlah mahasiswa baru untuk tahun selanjutnya dan dapat melakukan evaluasi

dalam penerimaan mahasiswa baru sehingga dapat mengambil langkah yang tepat untuk meningkatkan penerimaan mahasiswa baru setiap tahunnya.

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naïve Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi seperti decision tree dan neural network (Kusrini. 2009). Selvy Meigra, Kusrini dan Emha Taufiq Luthfi (2018) mengimplementasikan metode Naïve Bayes yaitu tentang prediksi heregistrasi calon mahasiswa baru menggunakan algoritma Naïve Bayes dengan atribut pendapatan ayah, nilai UN, minat studi, gelombang pendaftaran dan status di STMIK AMIKOM.

Penerapan data mining untuk menganalisa kemungkinan pengunduran diri calon mahasiswa baru. Dengan menggunakan data calon mahasiswa 2007-2009 yang digunakan sebagai data sample untuk pembentukan pohon keputusan, didapatkan hasil kecocokan prediksi dengan 600 data calon mahasiswa baru 2010 yang di ujikan terdapat tingkat kecocokan hasil prediksi dengan hasil nyata sebesar 61,89 %.

Pada penelitian ini, sistem akan melakukan klasifikasi terhadap data calon mahasiswa baru yang didapat dari bagian humas dan marketing STMIK AKAKOM ke dalam dua katagori yaitu mahasiswa yang melakukan pendaftaran ulang dan mahasiswa yang tidak melakukan pendaftaran ulang selama lima tahun ke belakang. Setelah data diklasifikasi kemudian dicari prediksi dari setiap kelas berdasarkan prodi.

2. TINJAUAN PUSTAKA

2.1. Penelitian sebelumnya

Berdasarkan penelitian yang telah dilakukan sebelumnya Naïve Bayes memiliki kemampuan yangm unggul dalam melakukan prediksi, sehingga Naïve Bayes digunakan pada sistem prediksi untuk menentukan jumlah pendaftar ulang di STMIK AKAKOM pada semua jurusan untk 3 (tiga) tahun mendatang. Atribut yang digunakan dalam sistem prediksi ini yaitu gelombang pendaftaran, program studi, penghasilan orang tua, nilai UN, biaya kuliah dan status registrasi.

Beberapa penelitian yang pernah dilakukan terkait dengan calon mahasiswa baru dengan metode Naive Bayes diantaranya adalah yang dilakukan oleh Devi Sugianti (2012), tentang prediksi calon mahasiswa baru di STMIK WIDYA PRATAMA. Kemudian penelitian oleh Eko Nur Wahyudi (2013). Teknik klasifikasi untuk melihat kecenderungan calon mahasiswa baru dalam memilih jenjang pendidikan program studi perguruan tinggi. Dan penelitian yang dilakukan oleh Tacbir Hendro Pudjiantoro, Faiza Renaldi, dan Age Teogunadi (2011).

3. KONSEP PERANCANGAN

Kebutuhan data pada sistem yang akan dibuat berupa :

1. Data penerimaan mahasiswa baru tahun 2013- 2019 STMIK Akakom.

2. Data diperoleh dari bagian admin dan marketing STMIK Akakom.
3. Atribut yang digunakan dalam penelitian ini yaitu gelombang pendaftaran, program studi, biaya kuliah, penghasilan orang tua, nilai UN dan status registrasi.
4. Data dibagi menjadi 2 yaitu data latih dan data tes dengan pembagian untuk data latih sebesar 70% dan data tes 30%.

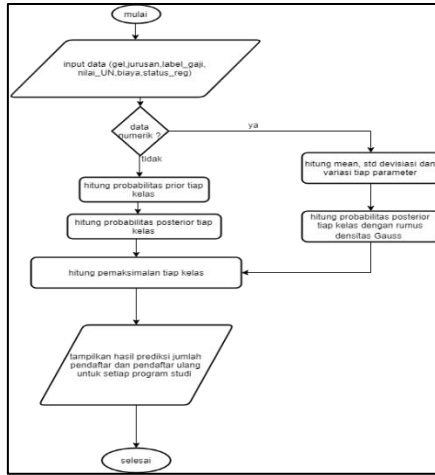
Terdapat tiga proses utama yang dilakukan dalam pengembangan sistem yang dibuat, yaitu

1. Proses input
Data yang digunakan pada proses input ini merupakan data penerimaan mahasiswa baru tahun 2013-2019 STMIK AKAKOM. Atribut yang digunakan yaitu gelombang pendaftaran, program studi, biaya kuliah, penghasilan orang tua dan nilai UN.
2. Proses Pengolahan
Pada subbab proses pengolahan dilakukan proses perhitungan untuk menentukan prediksi jumlah mahasiswa baru di STMIK AKAKOM menggunakan metode Naïve Bayes. Langkah- langkah perhitungan Naïve Bayes yang dilakukan adalah :
 - a) Masukan data penerimaan mahasiswa baru yang ada di STMIK AKAKOM tahun 2013-2019.
 - b) Setelah data dimasukkan, lakukan tahap preprocessing yang terdiri dari:
 - i. Membersihkan nilai kosong. Contoh variabel biaya pada data terdapat nilai “\N” kita dapat mengisi nilai kosong dengan mengisi atribut tersebut secara manual.
 - ii. Menghapus record yang isinya sama. Contoh pada data ada kesamaan nilai antar antar satu baris maka diperlukan menghapus data yang sama tersebut.
 - iii. Menghilangkan missing value, dikatakan missing value jika pada salah satu atribut nilai record tersebut hilang maka record tersebut akan dihapus, karena record tersebut dinilai kehilangan data.
 - iv. Pelabelan untuk data kategori. Contohnya untuk data jurusan diperlukan pelabelan yaitu kode 561 diberikan label “TI” dan sebagainya.
 - c) Setelah data dibersihkan maka tentukan kelas target data, pada penelitian ini menggunakan dua target yaitu pendaftar dan pendaftar ulang. Target disini diambil berdasarkan status registrasi, dengan ketentuan 1 adalah registrasi dan 0 adalah tidak registrasi.
 - d) Selanjutnya membagi data menjadi dua yaitu data latih dan data uji.
 - e) Tahapan selanjutnya mengambil data latih.
 - f) Kemudian menghitung prior probabilitas untuk melakukan prediksi herregistrasi maka akan menghitung prior probabilitas terlebih dulu.
 - g) Selanjutnya hitung probabilitas atribut target prediksi herregistrasi terhadap masing- masing class ($P(X|C_i)$).
 - h) Tahapan berikutnya yaitu hitung perkalian prior probability dengan probability atribut target prediksi herregistrasi ($P(C_i) \times (P(X|C_i))$).
 - i) Selanjutnya mencari nilai maksimal dari $P(C_i) \times (P(X|C_i))$ untuk menentukan hasil prediksi herregistrasi.

3. Proses output.

Menampilkan hasil prediksi jumlah pendaftar dan pendaftar ulang untuk setiap program studi.

Flowchart metode Naive Bayes dapat dilihat pada gambar 1 sebagai berikut:



Gambar 1 . Flowchart Naive Bayes

4. HASIL DAN PEMBAHASAN

Data yang digunakan adalah data pada bagian Penerimaan Mahasiswa Baru sejak tahun 2013 hingga tahun 2019. Contoh data seperti terlihat pada Gambar 2.

GEL	JURUSAN	RANGE GAJI	RATA	BIAYA	STAT_REG
K	561	1.000.000 - 2.500.000	0.00	Orang tua	1
K	561	1.000.000 - 2.500.000	7.97	Orang tua	0
K	541	2.500.000 - 5.000.000	0.00	IN	1
I	561	500.000 - 1.000.000	6.50	IN	1
K	541	2.500.000 - 5.000.000	0.00	Orang tua	1
K	541	5.000.000 - 7.500.000	0.00	Orang tua	0
K	541	7.500.000 - 10.000.000	0.00	Orang tua	0
K	541	1.000.000 - 2.500.000	8.19	Orang tua	1
K	541	1.000.000 - 2.500.000	0.00	Orang tua	1
K	541	1.000.000 - 2.500.000	0.00	Orang tua	1
K	541	1.000.000 - 2.500.000	0.00	Orang tua	1
K	541	1.000.000 - 2.500.000	0.00	Orang tua	1
K	541	1.000.000 - 2.500.000	0.00	Orang tua	0
K	541	1.000.000 - 2.500.000	0.00	IN	1
K	541	> 10.000.000	0.00	Sendiri	1
K	331	1.000.000 - 2.500.000	7.23	Orang tua	1
K	541	2.500.000 - 5.000.000	0.00	Orang tua	1
K	541	1.000.000 - 2.500.000	8.25	Orang tua	1
K	561	500.000 - 1.000.000	0.00	Orang tua	1
K	311	1.000.000 - 2.500.000	8.23	Orang tua	1
K	541	1.000.000 - 2.500.000	0.00	Orang tua	0
K	541	1.000.000 - 2.500.000	0.00	Orang tua	1

Gambar 2. Data Penerimaan Mahasiswa Baru

Gambar 2 merupakan contoh data mahasiswa STMIK AKAKOM Yogyakarta yang diambil, data dikelompokkan sesuai atribut yang dibutuhkan yaitu gelombang, jurusan, range gaji, rata-rata nilai un, biaya dan status registrasi. Data yang diambil sebanyak 2799 data. Data yang telah diambil dibagi menjadi dua yaitu data latih dan data uji. Untuk data latih sebanyak 70% dari data yang sudah diambil dan untuk data uji sebanyak 30% dari data yang diambil. Untuk proses pelabelan data sesuai dengan tampilan pada Tabel 1.

Tabel 1. Pelabelan Data

Atribut	Nilai	Tipe
Biaya	sendiri = 1 , orang tua = 2, instansi = 3, beasiswa pemerintah = 4	float
Gelombang	K = 5,1,2,3,4 1= 0-2.500.000 2=2.500.001-5.000.000 3=5.000.001-7.500.000 4=7.500.001-10.000.000 5=>10.000.000	float
Range Gaji	311, 321, 331, 541, 561	int
Jurusan	0 dan 1	int
Status_registrasi		int

Tahap selanjutnya adalah tahap pemrosesan data awal. Pada tahap ini merupakan tahapan pengolahan dan pembersihan data sebelum diproses oleh sistem. Proses pengolahan data mulai dari menghapus data yang sama atau duplikat dan menghapus data kosong dan membersihkan nilai Nan pada data.

1. Menghapus Data Duplikat

Dalam proses menghapus data duplikat dilakukan dengan menggunakan perintah pada gambar 4.3. Dalam bahasa pemrograman python banyak cara untuk menghapus data duplikat yaitu salah satunya dengan menggunakan library python yaitu Pandas.

2. Menghapus Data kosong dan Nan

Proses menghapus data kosong dan nan yaitu membersihkan data yang kosong pada baris dan kolom, sedangkan untuk data Nan yaitu data yang berisi nilai nol.

Data dari hasil preprocessing dapat dilihat pada gambar 3.

```

1  GEL; JURUSAN; RATA_UN; gaji; BIAYA
2  2; 541; 5.66; 1; 2
3  3; 561; 8.08; 1; 2
4  2; 541; 6.46; 2; 1
5  2; 331; 7.68; 1; 2
6  2; 311; 7.9; 2; 2
7  1; 541; 7.33; 1; 2
8  2; 541; 6.78; 1; 2
9  1; 541; 8; 2; 2
10 3; 541; 6.09; 2; 1
11 2; 541; 8.65; 1; 2
12 1; 541; 8.33; 1; 3
13 3; 541; 5.5; 1; 2
14 1; 331; 6.76; 2; 2
15 3; 561; 6.54; 4; 1
16 5; 561; 5.18; 1; 2
17 1; 541; 9.09; 3; 2
18 2; 541; 7.05; 2; 2
19 1; 541; 8.26; 1; 2
20 2; 541; 7.08; 2; 2
21 3; 561; 6.54; 3; 3
22 1; 541; 7.96; 1; 2
23 5; 541; 9.32; 1; 1
24 2; 331; 6.87; 1; 2
25 3; 561; 7.43; 1; 2
26 3; 311; 6.54; 1; 3
27 1; 541; 4.78; 3; 2
28 2; 561; 6.54; 3; 1
29 2; 561; 3.08; 4; 1
30 3; 561; 5; 1; 2
31 1; 541; 7.2; 1; 2
32 1; 331; 6.54; 3; 1
    
```

Gambar 3. Data Hasil Pemrosesan Awal

Implementasi program Naive Bayes Classifier dilakukan setelah proses preprocessing. Dari 731 data yang digunakan sebagai data uji didapatkan hasil 363 data terprediksi registrasi dengan kode jurusan TI, 211 data terprediksi registrasi dengan kode jurusan SI, 32 data terprediksi registrasi dengan kode jurusan TK, 58 data terprediksi registrasi dengan kode jurusan MI dan 15 data terprediksi registrasi dengan jurusan KA. Untuk rincian hasil prediksi bisa dilihat pada gambar 4. Gambar 4

menunjukkan hasil prediksi jumlah mahasiswa yang melakukan registrasi pada tahun pertama.

	Registrasi	Tidak Registrasi
TI	363	37
SI	211	12
TK	32	2
MI	58	0
KA	15	1

Gambar 4. Hasil prediksi jumlah registrasi pada tahun 1

Untuk hasil prediksi tahun ke 2 dapat dilihat pada gambar 5. Dengan rincian hasil prediksi jumlah mahasiswa registrasi dengan jumlah data uji sebanyak 853 data. 796 data terprediksi ke registrasi dan 57 data terprediksi tidak registrasi.

	Registrasi	Tidak Registrasi
TI	432	39
SI	241	15
TK	44	2
MI	61	0
KA	18	1

Gambar 5. Hasil prediksi jumlah registrasi pada tahun 2

Hasil prediksi pada tahun ke 3 menunjukkan jumlah mahasiswa registrasi dengan jumlah data uji sebanyak 974 data. 911 data terprediksi ke registrasi dan 63 data terprediksi tidak registrasi. Dengan rincian seperti ditunjukkan pada gambar 6.

	Registrasi	Tidak Registrasi
TI	491	41
SI	278	19
TK	51	2
MI	72	0
KA	19	1

Gambar 6. Hasil prediksi jumlah registrasi pada tahun 3

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dari serangkaian proses mulai dari perancangan sistem sampai dengan implementasi sistem pada penelitian ini. Dapat diambil kesimpulan antara lain :

1. Sistem ini dapat melakukan prediksi jumlah pendaftar ulang dengan banyak data dalam format csv dan per data.
2. Dari 731 data uji yang terprediksi benar sebanyak 679 data dan yang terprediksi salah sebanyak 52 data.

3. Dari 731 data yang digunakan sebagai data uji didapatkan hasil 363 data terprediksi registrasi dengan kode jurusan TI, 211 data terprediksi registrasi dengan kode jurusan SI, 32 data terprediksi registrasi dengan kode jurusan TK, 58 data terprediksi registrasi dengan kode jurusan MI dan 15 data terklasifikasi registrasi dengan jurusan KA.
4. Hasil pengujian sistem dengan data sebanyak 731 didapatkan akurasi sebesar 92,88%.

Dalam pengembangan sistem ini masih banyak kekurangan yang perlu diperbaiki. Maka dari itu penulis memberikan beberapa saran untuk dijadikan acuan dalam meningkatkan kualitas dari sistem yang dibuat antara lain :

1. Menambahkan program yang bisa melakukan prediksi dengan banyak atribut.
2. Menambahkan metode prediksi lain yang bisa di pilih seperti Neural Network, K-Nearest Neighbor, Decision Tree dan lain sebagainya.
3. Melakukan pelabelan data secara otomatis atau dengan sistem

DAFTAR PUSTAKA

- Alfa Saleh (2015). *Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. Citec Journal, Vol. 2, No. 3, Mei 2015 – Juli 2015
- Devi Sugianti (2012). *Algoritma Bayesian Classification untuk memprediksi Mahasiswa Baru Di STMIK WIDYA PRATAMA*. Jurnal Ilmiah ICTTech Vol.x No.2.
- Eko Nur Wahyudi (2013). *Teknik Klasifikasi untuk Melihat Kecenderungan Calon Mahasiswa Baru dalam Memilih Jenjang Pendidikan Program Studi di Perguruan Tinggi*. Jurnal Teknologi Informasi DINAMIK Volume 18, No.1.
- Kusrini (2009). *Bayesian Clasification adalah Pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class*.
- Mujib Ridwan, Hadi Suyono, dan M. Sarosa (2013). *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Jurnal EECCIS Vol.7, No. 1, Juni 2013.
- Prasetyo Anugroho, Idris Winarno, Nur Rosyid M (2010). *Klasifikasi Email Spam Dengan Metode Naive Bayes Classifier*. Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember Kampus ITS Keputih Sukolilo Surabaya 60111, Indonesia.
- Tacbir Hendro Pudjiantoro, Faiza Renaldi, dan Age Teogunadi (2011). *Penerapan Data Maining Untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru*. Konferensi Nasional Sistem dan Informatika 2011.
- Turban (2005). *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 database besar.

Selvy Megira, Kusrini, dan Emha Taufiq Luthfi (2018).
Prediksi Herregistrasi Calon Mahasiswa Baru Menggunakan Algoritma Naïve Bayes.

BIODATA PENULIS



Ariesta Damayanti
Lecturer of Informatics, STMIK AKAKOM
Yogyakarta