

Pembangunan Aplikasi Klasifikasi Mahasiswa Baru untuk Prediksi Hasil Studi Menggunakan *Naïve Bayes Classifier*

Mega Kartika Sari¹, Ernawati², Irya Wisnubhadra³

Program Studi Teknik Informatika, Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta
Jl. Babarsari 43, Yogyakarta 55281

E-mail: ¹mega.kartika@mail.uajy.ac.id, ²ernawati@mail.uajy.ac.id, ³irya@mail.uajy.ac.id

Masuk: 23 Oktober 2015; Direvisi: 10 November 2015; Diterima: 11 November 2015

Abstract. Every university hopes to create the best potential graduates. Some of the efforts can be achieved by utilizing university data to be processed into information. The resulting information can help the university to determine the decisions to take in improving the students' academic performance. One of the methods to process the data is by using Naive Bayes Classifier. This method requires some information such as GPA, average credits and attendance percentage. The prediction results are taken based on the data of the students at the university. Because there are large number of data to process it will require an information system that can classify data using Naive Bayes Classifier. This application is built by using C# programming language with an average of 81,67% accuracy results (the accuracy depends on the data used). This application is expected to help the university to increase the achievements of its graduates.

Keywords: Classification, Naive Bayes, Machine learning, Data mining

Abstrak. Setiap universitas tentu berharap dapat meluluskan mahasiswa dengan prestasi terbaik. Usaha meningkatkan prestasi dapat diwujudkan dengan memanfaatkan data universitas untuk diolah menjadi informasi. Informasi yang dihasilkan dapat berguna untuk menentukan keputusan yang harus diambil pihak universitas dalam meningkatkan prestasi. Salah satu metode yang digunakan untuk memproses data ialah metode klasifikasi Naive Bayes. Metode ini menggunakan beberapa informasi seperti IPK, rata-rata sks dan persentase kehadiran mahasiswa sebagai data pelatihan. Data pelatihan tersebut digunakan untuk memprediksi IPK, rata-rata sks dan presentase kehadiran mahasiswa baru sebagai data uji. Adanya data yang cukup besar untuk diolah, maka dibutuhkan aplikasi klasifikasi mahasiswa baru menggunakan metode Naive Bayes Classifier. Aplikasi ini dibangun menggunakan bahasa pemrograman C# dengan rata-rata hasil akurasi 81,67%. Aplikasi ini diharapkan dapat membantu universitas dalam meningkatkan hasil prestasi akademik mahasiswa.

Kata Kunci: Naive Bayes, Pembelajaran mesin, Penambangan data

1. Pendahuluan

Prestasi akademik mahasiswa pada setiap universitas diharapkan selalu menunjukkan peningkatan pada setiap tahunnya. Peningkatan prestasi akademik mahasiswa dapat ditinjau melalui rata-rata IPK dan lama studi mahasiswa. Setiap universitas tentu mengharapakan rata-rata IPK mahasiswa yang menempuh studi pada institusinya terus meningkat, dan rata-rata lama studi semakin singkat pada setiap tahunnya. Pada penelitian ini mengambil beberapa contoh data mahasiswa pada suatu universitas yaitu, data pada tahun 2010, 2011 dan 2012. Ditinjau pada Tabel 1, data rata-rata lama studi pada tahun 2010, 2011 dan 2012 masih lebih besar dari empat tahun dan persentase mahasiswa yang menempuh lama studi lebih dari tujuh tahun masih cukup tinggi. Sedangkan dilihat dari rata-rata IPK mahasiswa yang lulus pada tahun 2010, 2011, dan 2012 sudah cukup baik, namun jika ditinjau dari rata-rata IPK seluruh mahasiswa pada ketiga tahun tersebut rata-rata IPK mahasiswa masih perlu ditingkatkan.

Dari permasalahan tersebut, maka dibutuhkan aplikasi untuk memprediksi IPK, rata-rata SKS dan persentase kehadiran mahasiswa baru. Aplikasi ini dibangun supaya mahasiswa yang diprediksi mendapatkan IPK atau rata-rata SKS dibawah rata-rata dapat diberikan

bimbingan studi untuk meningkatkan IPK atau rata-rata SKS mahasiswa tersebut. Sedangkan mahasiswa yang diprediksi menempuh lama studi lebih dari empat tahun dapat diupayakan untuk dapat menempuh studi kurang dari atau sama dengan empat tahun. Hasil prediksi presentase kehadiran mahasiswa digunakan untuk mengamati keaktifan mahasiswa dalam mengikuti kegiatan belajar mengajar. Aplikasi klasifikasi mahasiswa baru ini dibangun menggunakan bahasa pemrograman C# dengan metode klasifikasi *Naïve Bayes* untuk memprediksikan IPK, rata-rata SKS dan persentase kehadiran data mahasiswa baru.

Tabel 1. Data Mahasiswa Universitas XYZ

Lama Studi	Tahun		
	2010	2011	2012
≥ 7	24,79%	14,14%	9,56%
6,5-6,9	6,83%	5,10%	5,11%
6-6,4	6,05%	5,67%	4,38%
5,5-5,9	9,43%	10,57%	8,91%
5-5,4	11,32%	12,36%	10,70%
4,5-4,9	13,92%	16,94%	20,66%
4-4,4	15,09%	18,79%	18,80%
<4	12,56%	16,43%	21,88%
Rata-rata lama studi(Tahun)	5,98	5,27	4,89
Rata-rata IPK lulusan	3,06	3,11	3,17
Rata-rata IPK mahasiswa	2,41	2,55	2,60

Naïve Bayes Classifier merupakan salah satu metode yang populer digunakan untuk menyelesaikan masalah klasifikasi dan mudah diimplementasikan. Metode tersebut berpotensi baik untuk mengklasifikasikan data karena kesederhanaannya (Ting, dkk., 2011). Algoritma *Naïve Bayes* menghitung perbandingan peluang antara jumlah dari masing-masing kriteria nilai variabel terhadap nilai hasil prediksi sesungguhnya (Jananto, 2013).

Aplikasi klasifikasi mahasiswa baru menggunakan teknologi penambangan data algoritma *Naïve Bayes Classifier*. Tujuan aplikasi ini dibangun adalah untuk membantu universitas dalam memprediksi IPK, rata-rata sks dan persentase kehadiran dari mahasiswa baru berdasarkan program studi, jalur pendaftaran, jenis kelamin, kabupaten/kota dan nilai masuk. Dengan adanya aplikasi klasifikasi mahasiswa baru diharapkan dapat membantu meningkatkan prestasi mahasiswa di bidang akademik pada suatu universitas.

2. Tinjauan Pustaka

2.1. Penelitian Terdahulu

Penggunaan penambangan data dalam meningkatkan performa studi telah dilakukan oleh beberapa peneliti dengan kasus atau metode yang berbeda. Baradwaj & Pal (2011a) mengimplementasikan metode penambangan data yaitu ID3 untuk meningkatkan performa mahasiswa pada setiap semester. Prediksi hasil akhir ujian semester digunakan untuk meningkatkan performa mahasiswa pada setiap semester. Variabel yang digunakan untuk menentukan kelas seperti hasil pada semester sebelumnya, kehadiran, hasil ujian dalam kelas, seminar dan hasil tugas. Pada penelitian ini menghasilkan *rule* yang digunakan untuk klasifikasi dari hasil metode ID3. Hastuti (2012) mengimplementasikan metode klasifikasi data mining untuk prediksi mahasiswa non-aktif. Metode yang digunakan antara lain *logistic regression*, *decision tree*, *naïve bayes*, dan *neural network*. Hasil perbandingan dari beberapa teknik klasifikasi tersebut, *decision tree* merupakan algoritma yang paling akurat namun tidak dominan terhadap algoritma yang lain dengan akurasi 95,29%.

Pandey & Pal (2011) mengimplementasikan metode klasifikasi *bayes* untuk memprediksi preforma mahasiswa. Hasil prediksi ini digunakan untuk meminimalkan jumlah mahasiswa yang dikeluarkan dari universitas karena performa yang kurang baik. Pandey & Pal memilih 600 siswa dari perguruan tinggi yang berbeda di Universitas Dr. RML Awadh, Faizad, India. Data tersebut yang digunakan sebagai data pelatihan untuk memprediksi performa mahasiswa. Kotsiantis, dkk. (2004) mengimplementasikan beberapa metode pembelajaran mesin untuk prediksi performa mahasiswa baru melalui pembelajaran jarak jauh. Beberapa

metode yang dibandingkan untuk prediksi data seperti *Naïve Bayes*, *C4.5*, *Back Propagation*, *Sequential Minimal Optimization*, *3-NN*, dan *Logistic Regression*. Pada penelitian tersebut dinyatakan *Naïve Bayes* ialah metode yang memiliki nilai akurasi tinggi dan mudah untuk diterapkan. Abdul, dkk. (2013) menggunakan metode *Naïve Bayes* untuk memprediksi performa mahasiswa di tahun pertama jurusan Ilmu Komputer. Prediksi performa mahasiswa Ilmu Komputer di tahun pertama menggunakan alat bantu WEKA dengan menggunakan tiga parameter data pelatihan. Hasil dari penelitian tersebut berupa pola yang dapat digunakan untuk memprediksi performa mahasiswa tingkat pertama.

2.2. Penambangan Data

Penambangan data yang juga disebut *Knowledge Discovery from Databases* (KDD) merupakan proses penemuan pengetahuan. Pengetahuan ini diperoleh secara otomatis dari informasi pada dunia nyata serta set data yang besar dan kompleks (Han & Kamber, 2006). KDD merupakan proses pengolahan data yang memiliki beberapa tahapan. Tahapan inti pada proses KDD yaitu penambangan data (Alcala, dkk., 2010). Fungsi utama dari penambangan data ialah mengekstrak pola data yang tersimpan dengan menerapkan berbagai metode dan algoritma (Baradwaj & Pal, 2011a).

2.3. Definisi Klasifikasi

Klasifikasi merupakan salah satu teknik yang sering digunakan untuk penambangan data (Baradwaj & Pal, 2011b). Teknik klasifikasi dilakukan dengan memprediksi label kelas yang belum diketahui menggunakan data berbeda yang memiliki label kelas yang sudah diketahui (Romero & Ventura, 2007, AI-Radaideh, dkk., 2006).

2.4. Definisi Naïve Bayes

Naïve Bayes merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh Naïve Bayes adalah teorema Bayes (Trisedya & Jaiz, 2009). Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal (Shadiq, 2009).

Algoritma Naïve Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Dalam algoritma ini semua atribut yang terdapat pada dataset dianggap tidak mempunyai hubungan, jadi antara atribut dianggap tidak ada keterhubungannya dengan atribut yang lain. Algoritma ini menghitung probabilitas masing-masing atribut untuk dihitung kembali rata-rata probabilitas atribut-atribut tersebut. Metode Naïve Bayes cukup mudah di implementasikan, dan hasil dari metode ini pun baik (Ardianto, 2010). Dasar dari Naïve Bayes digunakan dalam pemrograman dengan Persamaan 1, dimana $P(C_i | x)$ merupakan probabilitas $P(C_i)$ yang terjadi jika x sudah terjadi. $P(C_i)$ adalah kemungkinan C_i pada data yang bersifat independent terhadap x . Sedangkan x adalah kumpulan atribut, dan $P(C_i | x)$ adalah probabilitas x yang terjadi jika C_i benar atau sudah terjadi berdasarkan data pelatihan. Selama nilai $P(x)$ konstan, maka dapat disederhanakan menjadi Persamaan 2. Untuk mencari probabilitas data yang bersifat numerik dapat dilakukan dengan Persamaan 3, dimana $P(A_i | C_j)$ merupakan probabilitas A_i yang terjadi jika C_j sudah terjadi, σ merupakan *sample variance*, dan μ merupakan *mean*.

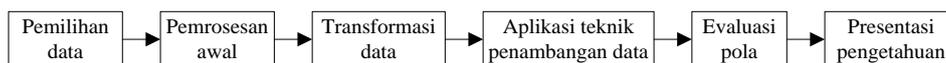
$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)} \quad (1)$$

$$P(C_i | X) = P(X | C_i)P(C_i) \quad (2)$$

$$P(A_i | C_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (3)$$

3. Metodologi Penelitian

Metode yang digunakan dalam penelitian ini mengacu pada tahapan proses KDD, antara lain: (1) Pemilihan data yaitu memilih data yang dapat digunakan untuk penelitian. (2) Pemrosesan awal yaitu membersihkan data yang tidak lengkap atau mengandung duplikasi dan memeriksa data yang tidak konsisten. (3) Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk penambangan data. (4) Aplikasi teknik penambangan data yaitu proses inti dimana teknik penambangan data diterapkan untuk mengekstrak pola-pola tertentu pada data. (5) Evaluasi pola yang ditemukan yaitu hasil dari teknik penambangan data berupa pola-pola yang khas maupun prediksi evaluasi untuk menilai hipotesa yang ada. (6) Presentasi pengetahuan yaitu menggunakan teknik visualisasi untuk menampilkan hasil penambangan data kepada pengguna. Tahapan ini dapat dilihat pada Gambar 1.

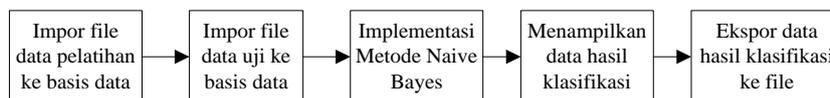


Gambar 1. Tahap Knowledge Discovery from Databases

4. Pembahasan

4.1. Implementasi Sistem

Aplikasi klasifikasi mahasiswa baru dibangun menggunakan bahasa pemrograman C# dengan *tools* Microsoft Visual Studio 2005 dan basis data Microsoft SQL Server 2005. Hasil dari implementasi sistem ini merupakan aplikasi *desktop* yang dapat digunakan untuk mengklasifikasikan data mahasiswa. Klasifikasi data tersebut digunakan untuk memprediksi IPK, lama studi dan persentase kehadiran mahasiswa baru, sehingga jika terdapat mahasiswa yang dimungkinkan memiliki IPK atau rata-rata sks rendah, dengan persentase kehadiran tinggi, maka mahasiswa tersebut dapat dipantau dan dibantu dalam kegiatan belajar mengajar. Sebaliknya, jika terdapat mahasiswa yang diprediksikan memiliki persentase kehadiran rendah, maka dapat diupayakan supaya mahasiswa tersebut menghadiri kegiatan belajar mengajar dan memahami materi perkuliahan. Tahap implementasi pada sistem klasifikasi mahasiswa baru seperti pada Gambar 2.



Gambar 2 Tahap implementasi pada sistem

Tahap pertama pada implementasi penambangan data ialah mengimpor *file* yang berisi data pelatihan untuk memprediksi IPK, rata-rata sks dan persentase kehadiran. Data tersebut disimpan pada basis data sistem pada tabel data pelatihan. Jika sistem sudah memiliki data pelatihan pada basis data, maka tahap pertama ini dapat dilewati menuju tahap kedua. Pada tahap kedua proses implementasi yaitu, mengimpor *file* data uji yang berisi data untuk melakukan pengujian terhadap hasil prediksi yang dihasilkan sistem. Data tersebut juga disimpan pada basis data sistem, tabel data uji. Setelah proses tahap kedua selesai, dilakukan perhitungan metode klasifikasi Naive Bayes untuk memperoleh hasil prediksi IPK, rata-rata SKS dan persentase kehadiran dari data uji, berdasarkan data pelatihan. Tahap keempat dan kelima, menampilkan hasil prediksi data uji yang dapat disimpan menjadi sebuah file dengan mengeksport data hasil klasifikasi.

Sistem klasifikasi mahasiswa baru menggunakan data yang diperoleh dari sistem informasi pada suatu universitas sebagai sampel. Data yang digunakan antara lain program studi, jalur pendaftaran, jenis kelamin, kabupaten/kota dan nilai masuk sebagai variabel serta IPK, rerata SKS dan persentase kehadiran sebagai kelas pada pengujian sistem. Data mahasiswa untuk kelas IPK diklasifikasikan dalam empat bagian yaitu: $IPK \geq 3,00$; IPK antara 2,50 hingga 2,99; IPK antara 2,00 hingga 2,49; dan $IPK < 2,00$. Sedangkan kelas rata-rata SKS dibagi menjadi empat bagian yaitu $SKS \geq 22$; SKS antara 19 hingga 21; SKS antara 16 hingga 18; dan

SKS<16. Pada kelas persentase kehadiran, dibagi menjadi tiga bagian yaitu: kehadiran \geq 75%; Kehadiran antara 51% hingga 74,99%, dan kehadiran<51%. Variabel penentu yang digunakan dalam klasifikasi ialah program studi, jenis kelamin, jalur pendaftaran, kabupaten/kota, dan nilai masuk berdasarkan jenis jalur pendaftaran.

Tabel 2 adalah contoh data yang digunakan sebagai data pelatihan pada aplikasi klasifikasi mahasiswa baru, serta dilanjutkan dengan contoh sederhana untuk perhitungan manual implementasi metode Naïve Bayes dalam aplikasi klasifikasi mahasiswa baru.

Tabel 2. Data Pelatihan

No	Program Studi	Jenis Kelamin	Jalur Pendaftaran	Kab/Kota	Nilai Masuk	IPK
1	Arsitektur	P	Reguler I	Maluku Tenggara Barat	40.25	1.85
2	Arsitektur	L	Reguler II	Makassar	54	2.62
3	Arsitektur	L	Unggulan NEM	Cilacap	8.14	2.09
4	Arsitektur	L	Reguler I	Sleman	65.75	2.63
5	Arsitektur	L	Reguler I	Maluku Barat Daya	30.75	2.46
6	Arsitektur	P	Reguler I	Purworejo	52.5	2.52
7	Arsitektur	P	Reguler II	Sleman	82.25	3.36
8	Arsitektur	L	Reguler II	Jambi	60	2.81
9	Arsitektur	P	Reguler II	Tarakan	67	2.78
10	Arsitektur	P	Reguler II	Dili	46.5	2.37
11	Arsitektur	P	Unggulan NEM	Boyolali	8.1	2.66
12	Arsitektur	L	Unggulan NEM	Tangerang	8.31	3.12
13	Arsitektur	P	Unggulan NEM	Mataram	8.37	2.64
14	Arsitektur	P	Unggulan NEM	Ambon	8.99	2.54
15	Arsitektur	P	Unggulan Raport	Yogyakarta	98	3.56
16	Arsitektur	L	Unggulan Raport	Sleman	87	3.58
17	Arsitektur	L	Unggulan Raport	Yogyakarta	88	3.3
18	Arsitektur	L	Unggulan Raport	Pegunungan Bintang	92	3.17

Langkah-langkah perhitungannya adalah:

1. Nama kelas: IPK (C_i)
2. Label kelas atribut: IPK[IPK \geq 3.00 | 2.50 \leq IPK<3.00 | 2.00 \leq IPK<2.50 | IPK<2.00]
3. Atribut: (Jalur pendaftaran, Kab/Kota, Nilai masuk, jenis kelamin)
4. Data yang akan diklasifikasi: x =(Jalur Pendaftaran: Unggulan Raport; Kab/Kota: Sleman; Nilai Masuk: 85; Jenis Kelamin: P)
5. Menghitung $P(C_i)$:

$$P(\text{IPK} \geq 3.00) = \frac{6}{18}; P(2.50 \leq \text{IPK} < 3.00) = \frac{8}{18}; P(2.00 \leq \text{IPK} < 2.50) = \frac{3}{18}; P(\text{IPK} < 2.00) = \frac{1}{18}$$

6. Menghitung $P(x|C_i)$ untuk masing-masing kelas:

$$P(\text{Jalur Pendaftaran: Unggulan Raport} | \text{IPK} \geq 3.00) = \frac{4}{6}$$

$$P(\text{Jalur Pendaftaran: Unggulan Raport} | 2.50 \leq \text{IPK} < 3.00) = \frac{0}{8} = 0$$

$$P(\text{Jalur Pendaftaran: Unggulan Raport} | 2.00 \leq \text{IPK} < 2.50) = \frac{0}{3} = 0$$

$$P(\text{Jalur Pendaftaran: Unggulan Raport} | \text{IPK} < 2.00) = \frac{0}{1} = 0$$

$$P(\text{Kab/Kota: Sleman} | \text{IPK} \geq 3.00) = \frac{2}{6}; P(\text{Kab/Kota: Sleman} | 2.50 \leq \text{IPK} < 3.00) = \frac{1}{8}$$

$$P(\text{Kab/Kota: Sleman} | 2.00 \leq \text{IPK} < 2.50) = \frac{0}{3} = 0; P(\text{Kab/Kota: Sleman} | \text{IPK} < 2.00) = \frac{0}{1} = 0$$

$$\text{Normalisasi Nilai Masuk} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}$$

$$\text{Unggulan Raport: } x_{\min} = 87; x_{\max} = 98; x = 85$$

$$\text{Normalisasi } x = \frac{(85 - 87)}{(98 - 97)} = -0.1818$$

Jika IPK \geq 3.00: Hasil Normalisasi rata-rata nilai masuk=0.4636; Variance=0.1962

Jika 2.50 \leq IPK < 3.00: Hasil Normalisasi rata-rata nilai masuk=0.5107; Variance=0.1296

Jika 2.00 \leq IPK < 2.50: Hasil Normalisasi rata-rata nilai masuk=0.0150; Variance=0.0007

Jika IPK < 2.00: Hasil Normalisasi rata-rata nilai masuk=0.2714; Variance=0

$$P(\text{Nilai Masuk: } -0.1818 | \text{IPK} \geq 3.00) = 0.3116$$

$$P(\text{Nilai Masuk: } -0.1818 | 2.50 \leq \text{IPK} < 3.00) = 0.1742$$

$$P(\text{Nilai Masuk: } -0.1818 \mid 2.00 \leq \text{IPK} < 2.50) = 4.96 \cdot 10^{-12}$$

$$P(\text{Nilai Masuk: } -0.1818 \mid \text{IPK} < 2.00) = 0$$

$$P(\text{Jenis Kelamin: } P \mid \text{IPK} \geq 3.00) = \frac{2}{6}; P(\text{Jenis Kelamin: } P \mid 2.50 \leq \text{IPK} < 3.00) = \frac{5}{8}$$

$$P(\text{Jenis Kelamin: } P \mid 2.00 \leq \text{IPK} < 2.50) = \frac{1}{3}; P(\text{Jenis Kelamin: } P \mid \text{IPK} < 2.00) = \frac{1}{1} = 1$$

7. Menghitung $P(x|C_i) \cdot P(C_i)$:

$$P(x|C_i) \cdot P(C_i) \mid \text{IPK} \geq 3.00 = P(x \mid \text{IPK} \geq 3.00) \cdot P(\text{IPK} \geq 3.00) = (4/6 \cdot 2/6 \cdot 0.3116 \cdot 2/6) \cdot 6/18 = 0.0076938$$

$$P(x|C_i) \cdot P(C_i) \mid 2.50 \leq \text{IPK} < 3.00 = P(x \mid 2.50 \leq \text{IPK} < 3.00) \cdot P(2.50 \leq \text{IPK} < 3.00) = (0 \cdot 1/8 \cdot 0.1742 \cdot 5/8) \cdot 8/18 = 0$$

$$P(x|C_i) \cdot P(C_i) \mid 2.00 \leq \text{IPK} < 2.50 = P(x \mid 2.00 \leq \text{IPK} < 2.50) \cdot P(2.00 \leq \text{IPK} < 2.50) = (0 \cdot 0 \cdot 4.96 \cdot 10^{-12} \cdot 1/3) \cdot 3/18 = 0$$

$$P(x|C_i) \cdot P(C_i) \mid \text{IPK} < 2.00 = P(x \mid \text{IPK} < 2.00) \cdot P(\text{IPK} < 2.00) = (0 \cdot 0 \cdot 0 \cdot 1) \cdot 1/18 = 0$$

8. Mencari probabilitas terbesar

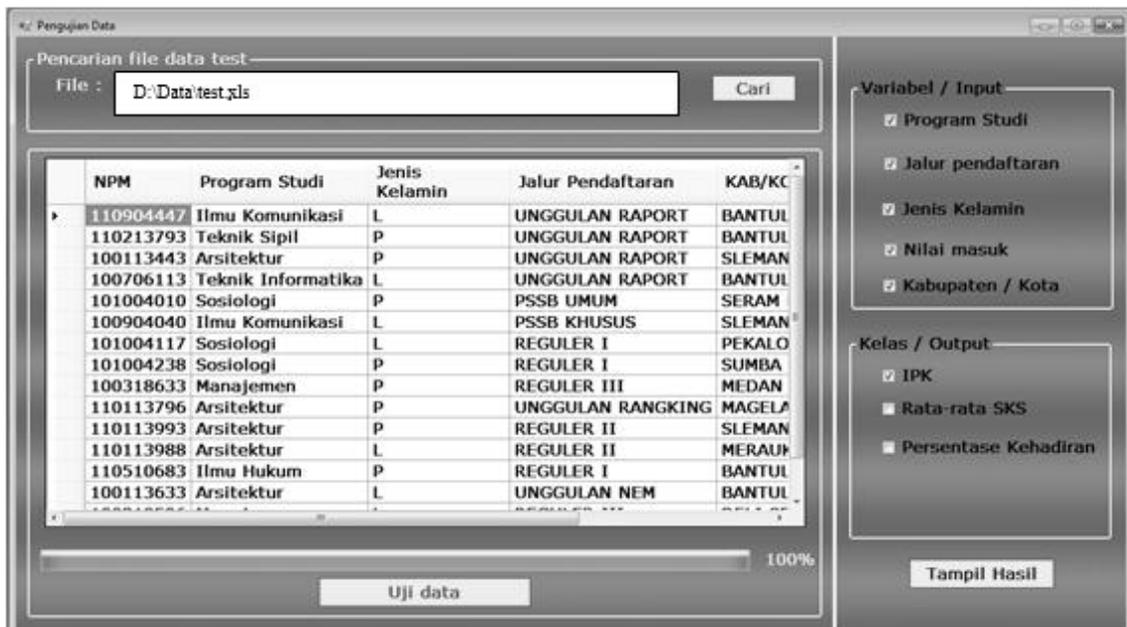
Probabilitas terbesar ialah $P(X|\text{IPK} \geq 3.00) \cdot P(\text{IPK} \geq 3.00)$ dengan nilai probabilitas 0.0076938, maka atribut x diprediksikan pada kelas $\text{IPK} \geq 3.00$.

Beberapa pengujian dilakukan untuk memperoleh nilai akurasi dari sistem klasifikasi mahasiswa baru. Pengujian ini menggunakan 200 data pelatihan dan 20 data uji yang berbeda pada setiap pengujianya. Nilai akurasi pada pengujian pertama sebesar 90%, akurasi pengujian kedua sebesar 60% dan akurasi pengujian ketiga, sebesar 95%. Dari ketiga percobaan tersebut, maka diambil nilai rata-rata akurasi untuk sistem klasifikasi mahasiswa baru sebesar 81,67%.

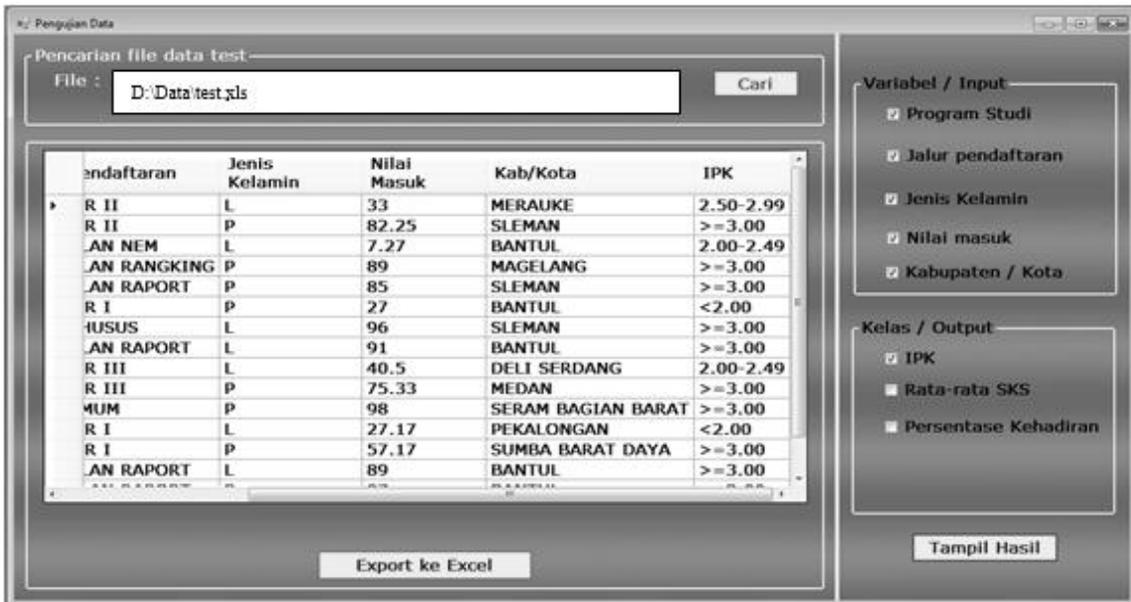
4.2 Hasil Penelitian

4.2.1 Implementasi Antarmuka

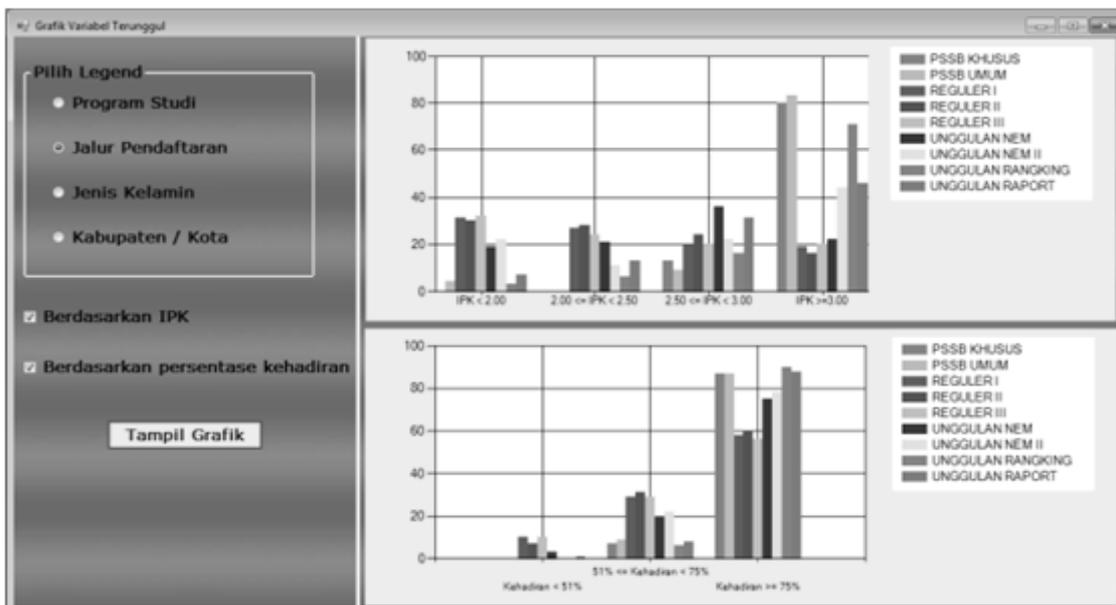
Gambar 3 merupakan antarmuka untuk menguji data yang diimpor dari *file excel* ke dalam basis data. Pada bagian kanan, pengguna dapat memilih atribut dan kelas yang dapat digunakan sebagai parameter pengujian data. Gambar 4 merupakan hasil pengujian data yang sudah diproses oleh sistem. Hasil pengujian tersebut ditampilkan dalam bentuk tabel dan dapat diekspor menjadi sebuah *file excel*. Gambar 5 tampilan grafik untuk mengetahui persentase IPK tertinggi dan persentase kehadiran mahasiswa berdasarkan variabel program studi, jalur pendaftaran, jenis kelamin, atau kab/kota.



Gambar 3. Implementasi antarmuka halaman pengujian data



Gambar 3. Implementasi antarmuka hasil pengujian data



Gambar 4. Implementasi antarmuka grafik variabel terunggul

7. Kesimpulan

Aplikasi klasifikasi mahasiswa baru menghasilkan prediksi IPK, rata-rata SKS, dan persentase kehadiran mahasiswa baru dengan hasil rata-rata keakuratan 81,67% (bergantung pada data yang diujikan). Dengan demikian aplikasi ini dapat digunakan untuk memprediksi mahasiswa yang mungkin memperoleh IPK dibawah rata-rata untuk dibimbing secara lebih detail oleh universitas dalam kegiatan belajar mengajar.

Referensi

Abdul Aziz, Azwa, Ismail, Nur Hafieza, & Ahmad, F. 2013. Mining Students' Academic Performance. *Journal of Theoretical and Applied Information Technology*, 53(3):485.

- AI-Radaideh, Q.A., AI-Shawakfa, E.M. & AI-Najjar, M.L. 2006, December. Mining Student Data using *Decision Trees*. In *International Arab Conference on Information Technology (ACIT'2006)*, Yarmouk University, Jordan, November.
- Alcala, J., Fernandez, A., Luengo, J., Derrac, J., Gracia, S., Sanchez, L. & Herrera, F. 2010. Keel Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17:255-287.
- Ardianto, T. 2010. *Perbandingan Decision Tree C4.5 dan Naïve Bayes dalam Implementasi Klasifikasi Konsumen Asuransi Kesehatan*. Skripsi. Universitas Pendidikan Indonesia. Bandung.
- Baradwaj, B.K. & Pal, S. 2011a. Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications*, 2(6): 63-69.
- Baradwaj, B.K. & Pal, S. 2011b. Data Mining: A Prediction for Performance Improvement using Classification. *International Journal of Computer Science and Information Security*, 9(4): 136-140.
- Han, J. & Kamber, M. 2006. *Data mining: Concepts and Techniques*, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 2nd edition.
- Hastuti, K. 2012. *Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif*. Seminar Nasional Teknologi Informasi & Komunikasi Terapan, Semarang, 23 Juni.
- Jananto, A. 2013. Algoritma Naïve Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Jurnal Teknologi Informasi DINAMIK*, 18 (1): 9-16.
- Kotsiantis, S., Pierrakeas, C. & Pintelas, P. 2004. Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence: An International Journal*, 18(5): 411-426.
- Pandey, U.K. & Pal, S. 2011. Data Mining: A Prediction of Performer or Underperformer using Classification. *International Journal of Computer Science and Information Technology (IJCSIT)*, 2(2): 686-690.
- Romero, C. & Ventura, S. 2007. Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135-146.
- Shadiq, M.A. 2009. *Keoptimalan Naïve Bayes Dalam Klasifikasi*, Program Ilmu Komputer FPMIPA Universitas Pendidikan Indonesia.
- Ting, S.L., Ip, W.H., & Tsang, A.H.C. 2011. Is Naïve Bayes a Good Classifier for Document Classification? *International Journal of Software Engineering and Its Applications*, 5 (3):37-46.
- Trisedya, B.D. , & Jaiz, Hardinal. 2009. *Klasifikasi Dokumen Menggunakan Algoritma Naïve Bayes dengan Penambahan Parameter Probabilitas Parent Category*, Fakultas Ilmu Komputer Universitas Indonesia.