

Optimasi Metode Naïve Bayes Dalam Memprediksi Tingkat Kelulusan Mahasiswa Stekom Semarang

Edy Siswanto

Fakultas Teknologi Informasi, Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia
Email : edy@stekom.ac.id

Abstrak

Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna bagi Perguruan Tinggi jika dimanfaatkan dengan maksimal. Salah satu cara untuk memanfaatkan data tentang mahasiswa yang lulus ini adalah dengan mengolahnya menggunakan data mining. Dengan proses data mining ini dapat ditemukan pola atau aturan yang dapat digunakan untuk menghasilkan suatu informasi seperti prediksi kelulusan mahasiswa. Dari hasil penelitian yang dilakukan dari tahap awal sampai dengan tahap pengujian penerapan metode naïve bayes untuk proses prediksi kelulusan mahasiswa, didapatkan kesimpulan bahwa Penerapan metode algoritma naïve bayes menggunakan 14 (empat belas) parameter yaitu tanggal masuk, kelas, npm, nama, stekom, cabang, tanggal lulus, no. ijazah, sks, IPK, mulai bimbingan, selesai bimbingan, masa studi dan status mahasiswa memperoleh tingkat akurasi sebesar 95,14 %. Kemudian untuk prediksi kelulusan berdasarkan tanggal masuk diperoleh hasil bahwa mahasiswa lulus tepat waktu adalah berdasarkan tanggal masuk antara tanggal 1 sampai 29 november.

Kata Kunci: prediksi, naïve bayes, kelulusan, mahasiswa.

Abstract

Data about graduating students can provide useful information for universities if used optimally. One way to utilize data about students who are passed is by processing it using data mining. With this data mining process, structures or rules can be found that can be used for information such as predictions of student graduation. From the results of the research conducted from the initial to the final stages by using the naïve bayes method for the prediction process, it was concluded that the methods of applying naïve bayes algorithms used 14 (fourteen) parameters, namely date of entry, class, npm, name, stekom, branch, graduation date, no. diploma, credits, GPA, starting guidance, guidance guidance, study period and student status accuracy rate of 95.14%. Then for graduation prediction using the date of entry the result that students graduate on time is the change date between November 1 and 29.

Keywords: prediction, naïve bayes, graduation, students.

1. PENDAHULUAN

STEKOM diawali oleh sebuah kursus komputer dan elektronika PAT yang didirikan pada tahun 1986. Pada awalnya Akademi PAT hanya memberikan perkuliahan dalam taraf Diploma saja, dan berjalannya waktu membuat Akademi PAT melakukan inovasi dan penyesuaian taraf pendidikan, hingga pada tahun 2002, sesuai dengan SK Mendiknas No. 235/D/O/2002, Akademi PAT meningkat menjadi Sekolah Tinggi Komputer dan Elektronika (STEKOM) atau sering pula disebut dengan STEKOM - PAT.

Era digital saat ini, persaingan mahasiswa dalam mendapatkan lapangan pekerjaan menuntut ilmu di perguruan tinggi menghasilkan sarjana yang berkualitas dan memiliki daya saing. Untuk itu, setiap perguruan tinggi selalu melakukan evaluasi performansi mahasiswa. Hasil evaluasi tersebut disimpan dalam basis data akademik. Data tersebut untuk kemudian digunakan sebagai pendukung keputusan oleh pihak manajemen perguruan tinggi. Salah satu variabel indikator efisiensi proses pendidikan adalah informasi mengenai lama masa studi mahasiswa dalam menempuh perkuliahan.

Data yang diperoleh penulis adalah data akademik mahasiswa dari tahun ketahunnya adalah mengambil dari data historis dan biodata kelulusan mahasiswa, dari data-data tersebutlah penulis melakukan prediksi dengan menggunakan algoritma NAÏVE BAYES sebagai metode untuk mendapatkan hasil sesuai dengan keinginan.

Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna bagi Perguruan Tinggi jika dimanfaatkan dengan maksimal. Salah satu cara untuk memanfaatkan data tentang mahasiswa yang lulus ini adalah dengan mengolahnya menggunakan data mining. Dengan proses data mining ini dapat ditemukan pola atau aturan yang dapat digunakan untuk menghasilkan suatu informasi seperti prediksi kelulusan mahasiswa.

2. METODE PENELITIAN

2.1 Perbandingan Penelitian sebelumnya dengan penelitian yang dilakukan

Menurut (Syarli & Muin, 2016) dalam penelitiannya memperoleh hasil bahwa Naïve Bayes dapat melakukan pengklasifikasian dengan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode Naïve Bayes Classification. Implementasi Naïve bayes menggunakan aplikasi WEKA dapat menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus.

Pengelompokkan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata.

Menurut (Testiana, 2018) dalam Penelitiannya mendapatkan hasil observasi terhadap dataset mahasiswa Fakultas Sains dan Teknologi UIN Raden Fatah Palembang dan melalui proses perhitungan menggunakan metode klasifikasi naïve Bayes dengan atribut yang telah dijelaskan di pembahasan sebelumnya, didapatkan sebuah hasil bahwa nilai akurasi terhadap klasifikasi kelulusan sebesar 82.08 %. Dimana 82.08 % bisa juga disebabkan oleh kurang kompleksitas data yang mengakibatkan model dapat memprediksi cukup akurat.

Menurut (Samponu & Kusri, 2017) menyatakan bahwa hasil Penelitian ini untuk melakukan prediksi tingkat kelulusan tepat waktu dengan melakukan analisis menggunakan teknik klasifikasi data mining. Algoritma Naïve Bayes yang digunakan untuk penelitian ini akan dibahas juga sebagai acuan dalam melakukan penelitian. Penulis melakukan serangkaian skenario percobaan yang berbeda / cross validation untuk melakukan perbandingan yang dapat memberikan perbedaan dalam tingkat akurasi yang dihasilkan dari penelitian ini. Hasil penelitian ini menunjukkan dengan penambahan skenario pengujian Cross Validation terdapat peningkatan akurasi sebesar 2% dari pengujian yang dilakukan.

2. 2 Proses Preprocessing

Tujuan dari preprocessing adalah untuk membersihkan data, menghilangkan data yang muncul berulang, dan membakukan kalimat yang tidak baku sehingga mendapatkan data bersih (Pudjajana & Manongga, 2018). Tahap preprocessing dalam penelitian ini sebagai berikut :

1. Seleksi data mahasiswa

Data yang telah dikumpulkan pada tahap pengumpulan data ternyata ada data yang berulang dikarenakan adanya kesalahan input dari pihak kampus cabang. Hasil seleksi yang didapatkan untuk kemudian digunakan dalam pengolahan penelitian.

2. Cleaning

Cleaning merupakan proses pembersihan tiap data mahasiswa dari hal-hal yang tidak dibutuhkan dalam penelitian. Beberapa hal yang akan dilakukan proses cleaning pada penelitian ini yaitu penulisan tahun yang tidak sesuai dengan kelas.

3. Normalisasi nama kelas

Normalisasi nama kelas merupakan proses untuk mengubah data nama kelas yang tidak sesuai menjadi Sesuai standar dari kampus. Hal ini dilakukan karena ditemukan banyak nama kelas dari berbagai cabang yang menggunakan nama yang tidak sesuai dengan standar kampus sehingga akan sulit dilakukan pengujian data.

2. 3 Pembuatan Data Latih

Penelitian ini membagi data-data yang telah dilakukan proses *preprocessing* menjadi dua bagian yaitu data latih dan data uji. Pembuatan data latih ini bertujuan sebagai pedoman dalam menentukan hasil data uji dan pada penelitian ini menggunakan 150 data kelulusan mahasiswa dari tahun 2015 sampai 2017.

2. 4 Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah algoritma yang melakukan klasifikasi dengan cara menghitung peluang berdasarkan rumus Bayesian Rule berikut (Samponu & Kusri, 2017) :

$$P(C | X) = (P(X | C) \times P(C)) / (P(X)) \dots\dots\dots(1)$$

Dimana algoritma ini bisa digunakan untuk memecahkan masalah klasifikasi data dengan feature bernilai nominal ataupun numerik. Jenis Naïve Bayes untuk berbagai tipe data adalah sebagai berikut :

1. Gaussian Naïve Bayes.
2. Bernoulli Naïve Bayes.
3. Multinomial Naïve Bayes.

3. ANALISA DAN PEMBAHASAN

Pada bagian ini berisi analisa, hasil serta pembahasan dari topik penelitian, yang bisa di buat terlebih dahulu metodologi penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, tabel dan lainnya.

1. Cleaning atau Pembersihan data (untuk membuang data yang tidak konsisten dan noise)

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak relevan dengan hipotesa data mining yang kita miliki. Pembersihan data yang tidak relevan akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (penggabungan data dari beberapa sumber) Integrasi data dilakukan pada atribut-atribut yang

- mengidentifikasi entitas - entitas yang unik. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya. Hasil integrasi data sering diwujudkan dalam sebuah data warehouse.
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di mining)
Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Karenanya data berupa angka numeric yang berlanjut perlu dibagi – bagi menjadi beberapa interval. Proses ini sering disebut binning. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining yang nantinya karena ada beberapa karakteristik dari teknik – teknik data mining tertentu yang tergantung pada tahapan ini.
 4. Aplikasi teknik data mining
Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Beberapa teknik data mining sudah umum dipakai. Ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di bidang tertentu atau untuk data tertentu.
 5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai)
Dalam tahap ini hasil dari teknik data mining berupa polapola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.
 6. Presentasi pola yang ditemukan untuk menghasilkan aksi
Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining

3.1 Metode Naïve Bayes

Metode Naïve Bayes yang digunakan untuk memprediksi kelulusan mahasiswa dengan menggunakan beberapa variable, antara lain kelas, npm, nama, jenis kelamin, jurusan, sks, ipk, lama studi dan status mahasiswa. Tabel 1 menunjukkan tabel untuk perhitungan prediksi kelulusan mahasiswa.

Menurut Santoso dalam (Fithri & Darmanto, 2014) menyatakan bahwa Teorema keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition). Pendekatan ini didasarkan pada kuantifikasi tradeoff antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang ditimbulkan dalam keputusan keputusan tersebut. Ide dasar dari bayes adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek. Misalkan terdapat dua jenis objek dengan kemungkinan kemunculan random, selanjutnya ingin diprediksi objek apa yang akan lewat

Objek pertama diwakili oleh h1 dan objek kedua diwakili oleh h2. karena apa yang akan muncul bersifat probablistik maka h adalah suatu variabel yang harus di deskripsikan secara probablistik. Selanjutnya probabilitas a priori, (P h1) dan (P h2) masing-masing melihat peluang munculnya objek 1 dan objek 2. Walaupun probabilitas kemunculan kedua objek tersebut tidak diketahui dengan pasti tapi setidaknya dapat diestimasi dari data yang tersedia. Misalkan N adalah jumlah total kedua objek, kemudian N1 dan N2 masing menyatakan jumlah objek 1 dan objek 2, (Fithri & Darmanto, 2014) selanjutnya seperti dibawah ini:

$$P(h1) \approx \frac{N1}{N} \text{ dan } P(h2) \approx \frac{N2}{N} \dots\dots\dots(2)$$

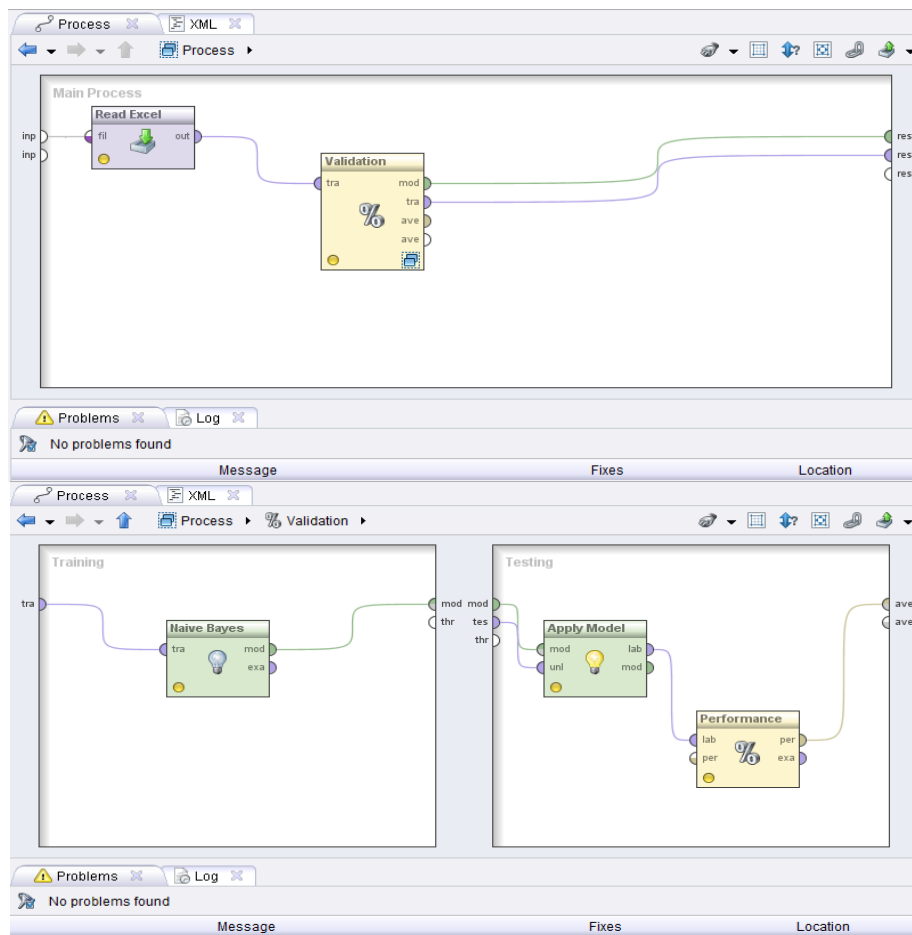
Untuk mengestimasi probabilitas digunakan rumus, yang sering disebut m-estimate :

$$P(a_i/v_i) = \frac{N_{a_i v_i} + m P(a_i)}{N_{v_i} + m} \dots\dots\dots(3)$$

Prediksi kelulusan mahasiswa di stekom digunakan untuk meningkatkan mutu dan kualitas mahasiswa, juga memberikan motivasi dan dukungan kepada mahasiswa agar dapat menempuh gelar sarjana dalam waktu yang telah ditentukan, yaitu sekitar 4 tahun dalam 8 semester. Penelitian yang berjudul Prediksi kelulusan mahasiswa menggunakan metode Naïve Bayes dengan 9 parameter telah mendapatkan akurasi yang cukup tinggi yaitu sebesar 95,14%.

4. IMPLEMENTASI

Dari gambar dibawah ini menunjukkan hasil analisa prediksi kelulusan berdasarkan prediksi variabel utama dengan beberapa variabel pendukung menggunakan metode naïve bayes. Dimulai dari tahap sebagai berikut: yaitu memasukkan data kelulusan mahasiswa tiga tahun terakhir kedalam aplikasi rapidminer, kemudian menambahkan tool validation untuk menentukan algoritma naïve bayes seperti yang terlihat pada gambar 1 dibawah ini :

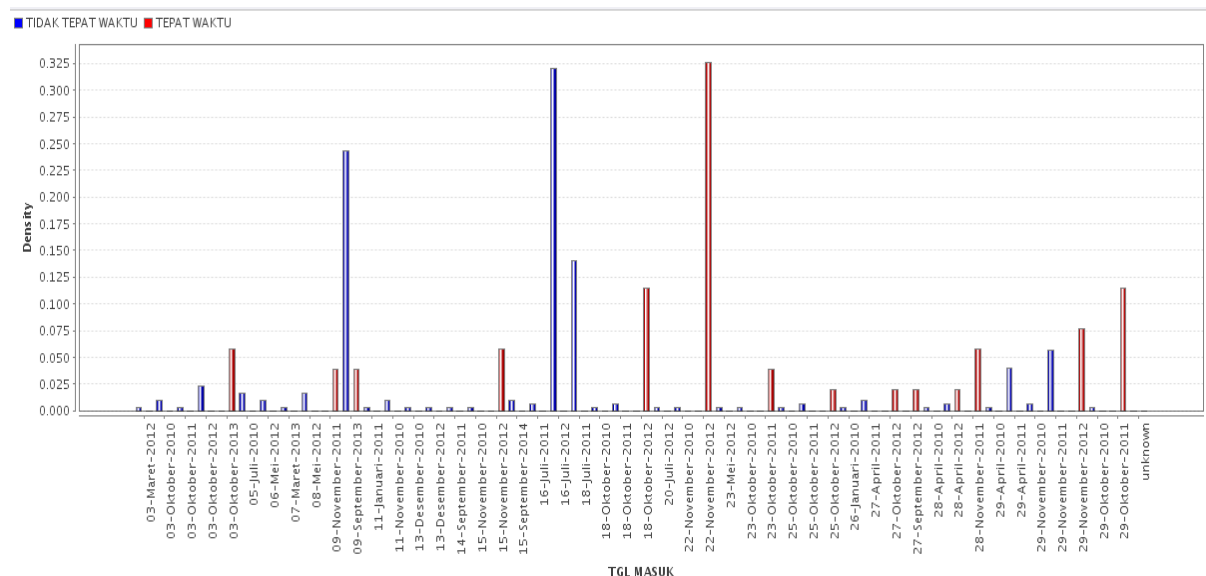


Gambar 1. Tahap awal analisa dengan aplikasi

Dari analisa diatas diperoleh hasil sebagai berikut:

1. Variabel utama dengan variabel tanggal masuk

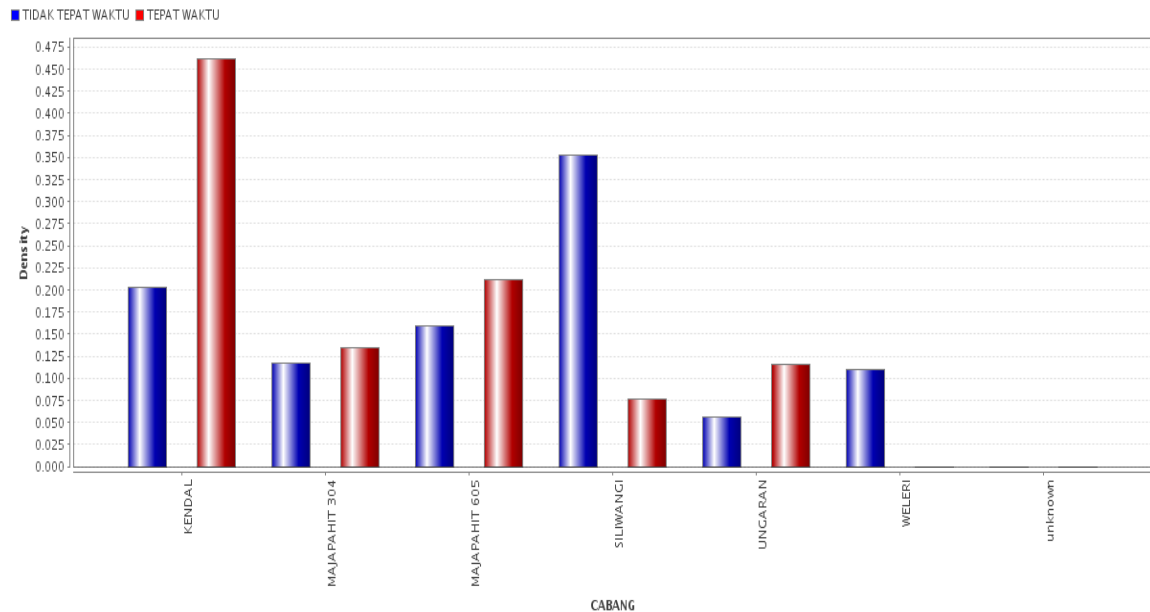
Variabel tanggal masuk menunjukkan bahwa tanggal masuk dari tiga tahun terakhir mahasiswa yang lulus menunjukkan bahwa mahasiswa yang lulus tepat waktu adalah mahasiswa yang pendaftarannya sekitar tanggal 1 sampai tanggal 29 november. Berikut hasil gambar yang diperoleh.



Gambar 2. Grafik perbandingan variabel tgl masuk

2. Variabel utama dengan variabel stekom cabang

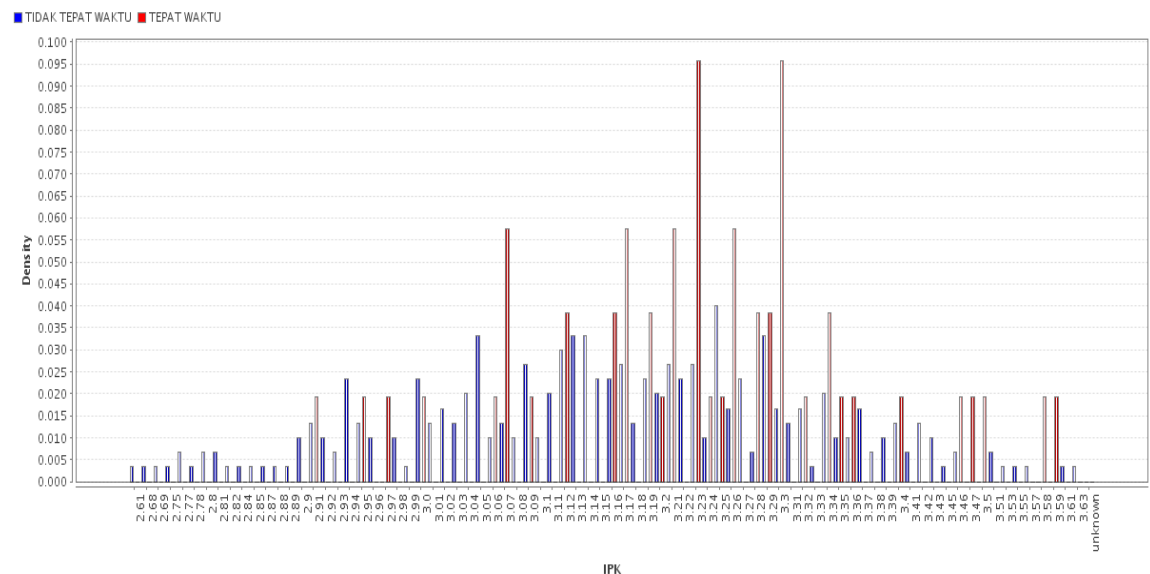
Dari gambar dibawah menunjukkan bahwa variabel utama dengan variabel stekom cabang menunjukkan bahwa stekom cabang yang lulus tepat waktu adalah stekom cabang kendal. Berikut hasil dari prediksi kelulusan berdasarkan stekom cabang tiga tahun terakhir.



Gambar 3. Grafik perbandingan variabel stekom cabang

3. Variabel utama dengan variabel ipk

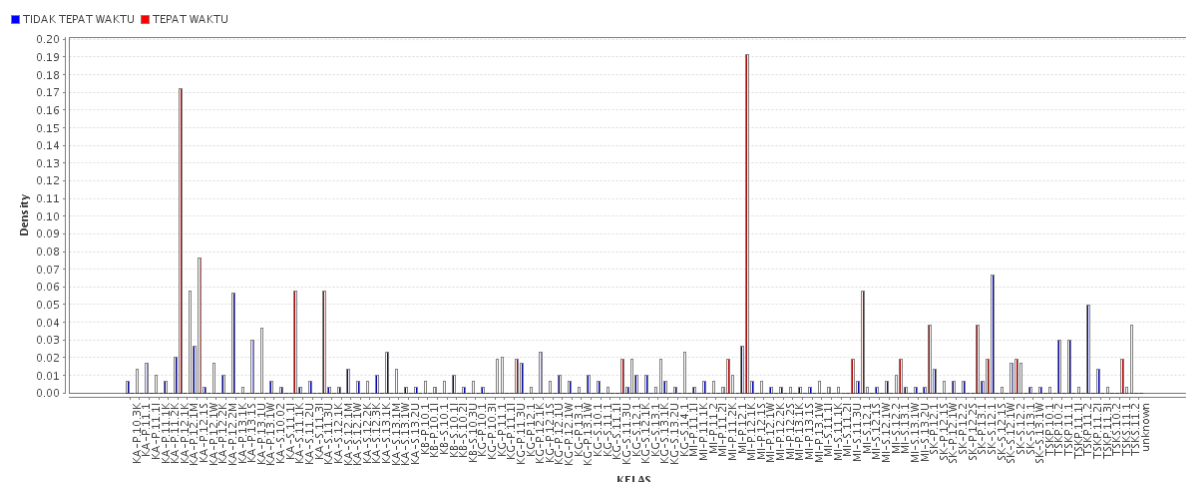
Dari gambar dibawah menunjukkan bahwa variabel utama dengan variabel ipk menunjukkan bahwa ipk yang lulus tepat waktu adalah mahasiswa dengan ipk 3.23. Berikut hasil dari prediksi kelulusan berdasarkan stekom cabang tiga tahun terakhir.



Gambar 4. Grafik perbandingan variabel ipk

4. Variabel utama dengan variabel kelas

Dari gambar dibawah menunjukkan bahwa variabel utama dengan variabel kelas menunjukkan bahwa kelas yang lulus tepat waktu adalah mahasiswa dengan kelas MI.P.12.1, yaitu kelas dengan jurusan Manajemen Informatika angkatan tahun 2012 gelombang yang pertama. Berikut hasil dari prediksi kelulusan berdasarkan kelas tiga tahun terakhir.



Gambar 5. Grafik perbandingan variabel kelas

Dari gambar dibawah ini menunjukkan hasil analisa prediksi kelulusan mahasiswa stekom dari tabel / plot view berdasarkan performancevektor menunjukkan nilai accuracy 95.14%.

<input checked="" type="radio"/> Table / Plot View <input type="radio"/> Text View <input type="radio"/> Annotations			
Criterion Selector <input checked="" type="radio"/> Multiclass Classification Performance <input type="radio"/> Annotations			
accuracy precision recall AUC (optimistic) AUC AUC (pessimistic)			
<input checked="" type="radio"/> Table View <input type="radio"/> Plot View			
accuracy: 95.14% +/- 4.62% (mikro: 95.17%)			
	true TIDAK TEPAT WAKTU	true TEPAT WAKTU	class precision
pred. TIDAK TEPAT WAKTU	286	3	98.96%
pred. TEPAT WAKTU	14	49	77.78%
class recall	95.33%	94.23%	

Gambar 6. Plot view

5. KESIMPULAN

Dari hasil penelitian yang dilakukan dari tahap awal sampai dengan tahap pengujian penerapan metode naïve bayes untuk proses prediksi kelulusan mahasiswa, didapatkan kesimpulan bahwa:

1. Penerapan metode algoritma naïve bayes menggunakan 14 (empat belas) parameter yaitu tanggal masuk, kelas, npm, nama, stekom, cabang, tanggal lulus, no. ijazah, sks, IPK, mulai bimbingan, selesai bimbingan, masa studi dan status mahasiswa dengan tingkat akurasi 95,14 %.
2. Metode Naïve Bayes digunakan untuk menghitung probabilitas dengan kemungkinan tepat waktu atau terlambat dalam menentukan prediksi kelulusan mahasiswa.
3. Dalam pengolahan data mining seharusnya dilakukan tahapan-tahapan dalam data mining, yaitu cleaning data, integrasi data, transformasi data, aplikasi data mining dan evaluasi pola yang ditemukan dalam data mining.

REFERENCES

- [1] Fithri, D. L., & Darmanto, E. (2014). SISTEM PENDUKUNG KEPUTUSAN UNTUK MEMPREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES. Prosiding SNATIF, ISBN: 978-602-1180-04-4.
- [2] Pudjajana, A. M., & Manongga, D. (2018). SENTIMEN ANALISIS TWEET PORNOGRAFI KAUM HOMOSEKSUAL INDONESIA DI TWITTER DENGAN NAIVE BAYES. Jurnal SIMETRIS Vol. 9 No. 1, ISSN: 2252-4983.
- [3] Samponu, Y. B., & Kusrini. (2017). OPTIMASI ALGORITMA NAIVE BAYES MENGGUNAKAN METODE CROSS VALIDATION UNTUK MENINGKATKAN AKURASI PREDIKSI TINGKAT KELULUSAN TEPAT WAKTU. Jurnal ELTIKOM, Vol. 1 No. 2, ISSN 2598-3245 (Print).
- [4] Syarli, & Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). Jurnal Ilmiah Ilmu Komputer, Vol. 2, No. 1 (P) ISSN 2442-4512.
- [5] Testiana, G. (2018). Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah. JUSIFO (Jurnal Sistem Informasi), Vol. 02, No.1 ISSN: 2460-0921.