**SISTEM PENDUKUNG KEPUTUSAN UNTUK MEMPREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES**

**201851009 Muhammad Arifin**

**201851013 Sandi Prayogo**

**201851032 Teguh Aris Wicaksono**

**201851044 Andika Eka Saputra**

**201851050 Moh Lukman Hakim**

***Abstrak***

*Sistem Pendukung Keputusan untuk menentukan tingkat prediksi kelulusan mahasiswa digunakan untuk menentukan tingkat kelulusan mahasiswa. Prediksi tersebut nantinya digunakan sebagai sumber informasi untuk menghasilkan sebuah keputusan. Penelitian ini menghasilkan sistem pendukung keputusan dengan menggunakan metode Naïve Bayes dengan menggunakan beberapa parameter, yaitu jenis kelamin, status mahasiswa, status pernikahan mahasiswa, IPK. Pengolahan data mining mahasiswa dengan menggunakan metode naïve Bayes dimulai dari proses Data Gathering, Data Preprocessing, Proposed Model/Method, Method Test and Experiment, Result Evaluation and Validation Dalam penelitian ini hasil yang dicapai memiliki akurasi untuk tepat waktu sebesar 53% dan akurasi untuk terlambat sebesar 47%. dengan menggunakan metode Naïve Bayes yang semakin optimal dengan menentukan mahasiswa lulus tepat waktu atau terlambat*

***Kata kunci****: akurasi, data mining, naïve bayes, tepat waktu, terlambat*

1. **PENDAHULUAN**

Memprediksi prestasi akademik siswa sangat penting bagi suatu lembaga pendidikan karena program strategis tersebut dapat direncanakan dalam meningkatkan atau mempertahankan kinerja siswa selama masa studi di lembaga (Ibrohim, 2010). Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing (Kovačić, 2006). Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Berdasarkan penelitian yang dibuktikan bahwa untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tingkat atau tahun ajaran pertama. Ada beberapa faktor yang mempangaruhi prediksi kelulusan mahasiswa yang sesuai dengan waktu studi, diantaranya : NEM SMA, status mahasiswa, jenis kelamin, umur mahasiswa, gaji orang tua, pekerjaan orang tua. Dalam menentukan Prediksi kinerja mahasiswa dengan akurasi yang tinggi bermanfaat untuk mengidentifikasi siswa dengan prestasi akademik (Suhartinah,2010).

Lembaga pendidikan perguruan tinggi menghadapi masalah siswa yang terkait dengan tingkat kelulusan, perguruan tinggi dengan mahasiswa lebih tinggi tingkat retensinya dan cenderung memiliki tingkat kelulusan yang lebih tinggi dalam waktu empat tahun. Rata – rata tingkat retensi nasional mendekati 55% dan di beberapa perguruan tinggi kurang dari 20% dari mahasiswa pascasarjana(Nandeshwar, 2009).

Tujuan dari Penelitian ini untuk menghasilkan sistem pendukung keputusan untuk memprediksi kelulusan mahasiswa di tingkat perguruan tinggi.

1. **METODOLOGI**
   1. Metode Penelitian

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan penelitian seperti berikut (Santoso, 2007):

1. Pengumpulan Data (Data Gathering) Penelitian ini menggunakan data mahasiswa untuk memprediksi kelulusan mahasiswa.
2. Pengolahan Awal Data (Data Preprocessing) Data perolehan di transformasi untuk mendapatkan atribut yang relevan dan sesuai dengan format input algoritma naïve bayes. Atribut yang digunakan meliputi jenis kelamin, status mahasiswa, status pernikahan mahasiswa, IPK ( Indeks Prestasi Kumulatif.
3. Model/Metode Yang Diusulkan (Proposed Model/Method) Metode yang diusulkan adalah metode perbandingan tingkat akurasi dari algoritma naïve bayes yang dapat digunakan untuk memprediksi kelulusan mahasiswa.
4. Eksperimen dan Pengujian Metode (Method Test and Experiment) Pengujian dilakukan dengan menggunakan sebagian data mahasiswa untuk training dan sebagian lagi sebagai data testing. Perhitungan dengan masing-masing algoritma akan diulang beberapa kali untuk mendapatkan besaran parameter terbaik.
5. Evaluasi dan Validasi Hasil (Result Evaluation and Validation) Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan Algoritma Naïve Bayes. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal. Pengukuran kinerja dilakukan dengan membandingkan nilai error hasil prediksi masingmasing algoritma sehingga dapat diketahui algoritma yang lebih akurat.
   1. Penerapan Metode Naïve Bayes

Tabel 1. Data training

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Jenkel | Pekerjaan | Status | IPK | Kelulusan |
| 1 | Andi | L | Mahasiswa | Belum | Baik | Tepat |
| 2 | Budi | L | Bekerja | Belum | Baik | Tepat |
| 3 | Cinta | P | Mahasiswa | Belum | Baik | Tepat |
| 4 | Devita | P | Mahasiswa | Menikah | Baik | Tepat |
| 5 | Erlangga | L | Bekerja | Menikah | Baik | Tepat |
| 6 | Farid | L | Bekerja | Menikah | Kurang | Terlambat |
| 7 | Hima | P | Bekerja | Menikah | Baik | Terlambat |
| 8 | Ika | P | Bekerja | Belum | Cukup | Terlambat |
| 9 | Joko | L | Bekerja | Belum | Kurang | Terlambat |
| 10 | Karina | P | Mahasiswa | Menikah | Kurang | Terlambat |
| 11 | Lala | P | Mahasiswa | Belum | Kurang | Terlambat |
| 12 | Mila | P | Mahasiswa | Belum | Baik | Tepat |
| 13 | Narji | L | Bekerja | Menikah | Baik | Tepat |
| 14 | Okky | L | Mahasiswa | Menikah | Baik | Tepat |
| 15 | Parto | L | Mahasiswa | Belum | Kurang | Terlambat |

Tabel 2. Perhitungan prediksi kelulusan mahasiswa

|  |  |  |  |
| --- | --- | --- | --- |
| Variabel | Sub Variabel | Perhitungan Prediksi Tepat Waktu | Perhitungan Prediksi Terlambat |
| Jumlah | | 8 | 7 |
| 15 | |
| Jenkel | Laki-laki | 5 | 3 |
| Perempuan | 3 | 4 |
|  | | 8 | 7 |
| Status Mahasiswa | Mahasiswa | 5 | 3 |
| Bekerja | 3 | 4 |
|  | | 8 | 7 |
| Status Pernikahan | Menikah | 4 | 3 |
| Belum | 4 | 4 |
|  | | 8 | 7 |
| IPK | Baik | 8 | 1 |
| Cukup | 0 | 1 |
| Kurang | 0 | 5 |
|  |  | 8 | 7 |

1. **HASIL DAN PEMBAHASAN**
   1. Tahapan – tahapan dalam preprocessing data mining
2. 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 performasi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
3. Integrasi data (penggabungan data dari beberapa sumber) Integrasi data dilakukan pada atribut-aribut yang mengidentifikasikan 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.
4. 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 bebrapa karakteristik dari teknik – teknik data mining tertentu yang tergantung pada tahapan ini.
5. 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.
6. 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.
7. 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 (Vercellis,2009).
   1. Metode Naïve Bayes

Teorema keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recoginition). 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 selanjutnya( Santoso, 2007).

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 probabilistik. 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 masing menyatakan jumlah objek 1 dan objek 2, selanjutnya

P(h1)≈ danP(h2 )≈ ………………………………………………………(1)

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

*P(H | X) =*………………………………………………….. (2)

Prediksi kelulusan mahasiswa 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 4 parameter telah mendapatkan akurasi yang cukup tinggi

* 1. Hasil Prediksi

1. Perhitungan tepat waktu:
2. P ( mahasiswa tepat waktu / total mahasiswa data training ) \* P ( Laki – Laki ׀ Tepat waktu ) \*P ( Bekerja ׀ Tepat waktu ) \* P (Menikah ׀ Tepat waktu ) \* P ( Baik ׀ Tepat waktu )
3. Jadi, 8/15 \* 5/8 \* 3/8 \* 4/8 \* 8/8

= 0.063441

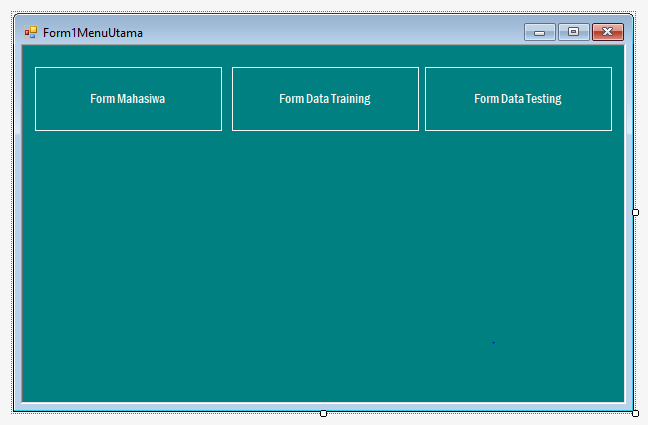
1. Perhitungan terlambat:
2. P ( mahasiswa terlambat / total mahasiswa data training ) \* P ( Laki – Laki ׀ Terlambat ) \*P ( Bekerja ׀ Terlambat ) \* P (Menikah ׀ Terlambat ) \* P ( Baik ׀ Terlambat )
3. Jadi, 7/15 \* 3/7 \* 4/7 \* 3/7 \* 1/7

= 0.0069

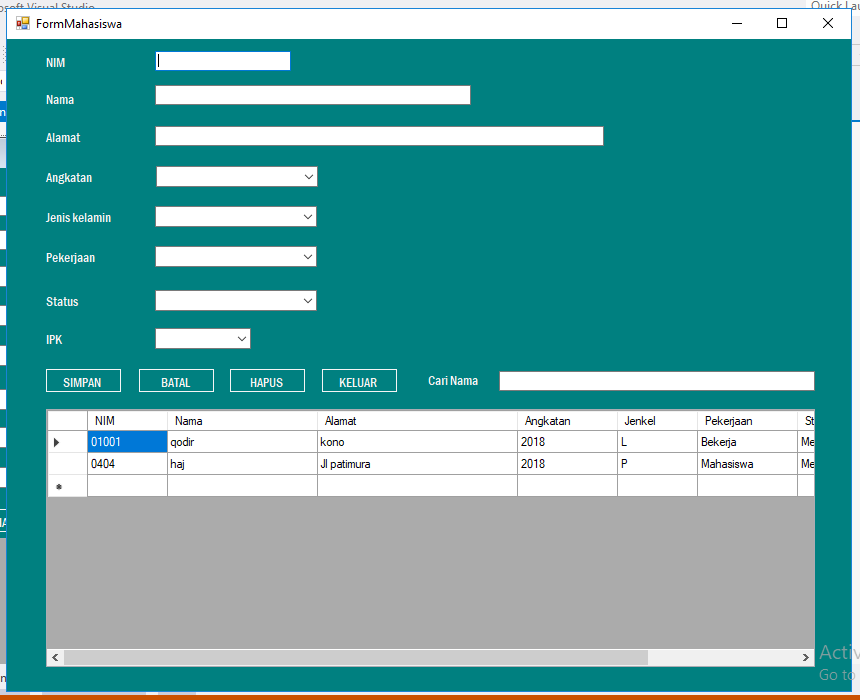
1. Prediksi tepat waktu = 0.063441
2. Prediksi terlambat = 0.0069
3. Penentuan prediksi 0.063441> 0.0069

Jadi, prediksi mahasiswa tersebut adalah **Tepat waktu**

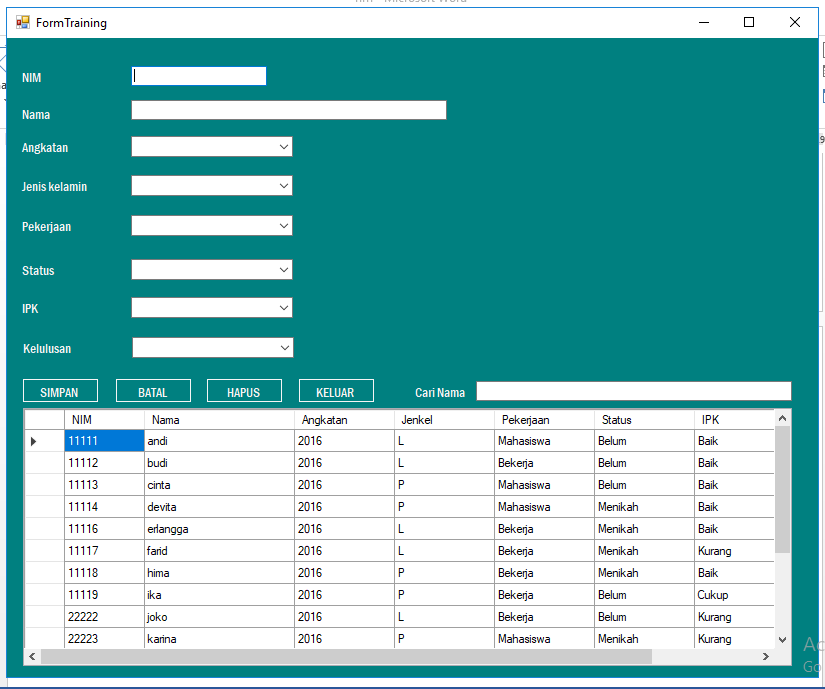
* 1. Tampilan Aplikasi



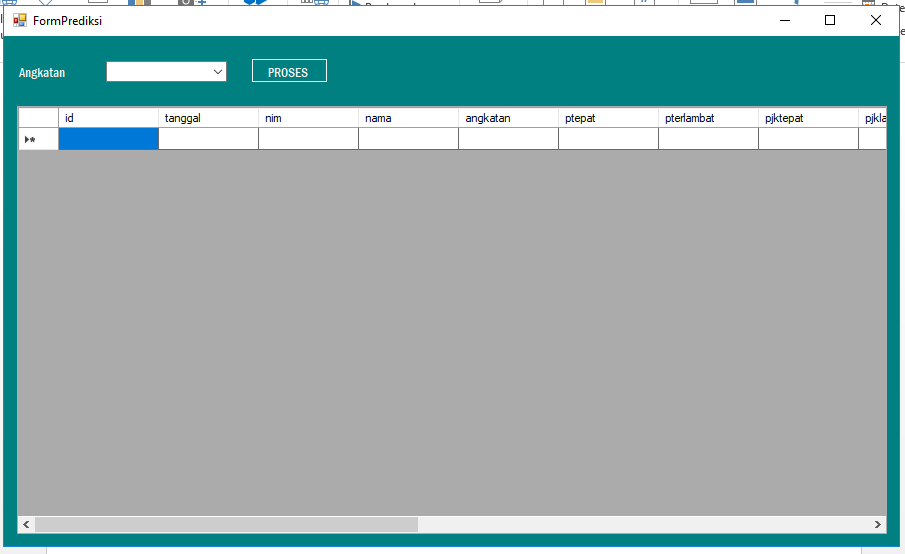
Gambar 1. Halaman Menu Utama



Gambar 2. Form Input Mahasiswa berdasarkan pada tahun angkatan



Gambar 3. Form Data Training



Gambar 4. Form Data Testing berdasarkan tahun angkatan yang di inputkan

1. **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 4 (empat) parameter yaitu jenis kelamin, status mahasiswa, status pernikahan mahasiswa, IPK dengan tingkat akurasi 100%.
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.
4. **DAFTAR PUSTAKA**

Daliela Rusli Zaidah Ibrohim. 2006. *PREDICTING STUDENTS’ ACADEMIC PERFORMANCE: COMPARING ARTIFICIAL.* Annual SAS Malaysia Forum.

Ernastuti Marselina Silvia Suhartinah. 2010. *GRADUATION PREDICTION OFGUNADARMA UNIVERSITY STUDENTS USING ALGORITHM AND NAIVE BAYES C4.5 ALGORITHM.*

Subodh Chaudhari Ashutosh Nandeshwar. 2009. *Enrollment Prediction Models Using Data Mining*.

Zlatko J. Kovačić. 2010. *Early Prediction of Student Success;Mining Students Enrolment Data. Proceedings of Informing Science & IT Education Conference (InSITE).*