

**ANALISA KINERJA METODE KLASIFIKASI DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES UNTUK PREDIKSI KETEPATAN WAKTU KELULUSAN MAHASISWA (STUDI KASUS STIKES SYEDZA SAINTIKA)****Nurul Abdillah, Sarjon Defit, Sumijan**

STIKES Syedza Sainatika

(Abdillahadik15@gmail.com, +62 822-6812-0465)

**ABSTRAK**

Tingkat kelulusan merupakan salah satu parameter efektifitas institusi pendidikan. Penurunan tingkat kelulusan mahasiswa berpengaruh terhadap akreditasi perguruan tinggi. Database perguruan tinggi menyimpan data administrasi dan akademik mahasiswa, apabila dieksplorasi dengan tepat menggunakan teknik data mining maka dapat diketahui pola atau pengetahuan untuk mengambil keputusan. Algoritma naive bayes bertujuan untuk mengukur tingkat akurasi untuk diterapkan dalam kasus ketepatan waktu kelulusan mahasiswa. Metode Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik untuk memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya. Penelitian ini menggunakan data mahasiswa program studi Pendidikan Teknik Informatika Universitas Negeri Padang angkatan 2011. Variabel yang digunakan dalam penelitian ini adalah: NIM, nama, jenis kelamin, status masuk, IPK, daerah asal dan status pekerjaan. Berdasarkan hasil pengujian dengan mengukur kinerja metode, diketahui bahwa naive bayes memiliki nilai akurasi yang bagus yaitu sebesar 93,48%. Dari nilai akurasi tersebut dapat ditarik kesimpulan bahwa algoritma naive bayes memiliki kinerja yang baik dalam memprediksi ketepatan waktu kelulusan mahasiswa.

**Kata kunci: DataMining; Klasifikasi; Naïve Bayes; Ketepatan Kelulusan Mahasiswa.**

**ABSTRACT**

*The excellent service is the best way that we give according with the expectation or more than costumers expect. The purpose of this service is to determine the relationship between the independent and dependent variable. Type of this research is analytic research with a sectional research design has been done on 30th desember 2019 until 5th October 2020. The population of the research is an outpatient at RSUD Sultan Thaha Saifuddin Tebo as total 5.593 suspect. The total of sample this research is 107 patient. The technical of sample used is Purposive sampling. Data analysis was done by univariat and bivariat testing used statistical test. Chi-Square with the accuracy 95% $\alpha=0,05$ . Research result got 59,8% Patient feel bad service, 56,1% less ability, 68,2% less attitude, 56,1% less appearance 68,2% less of attention, 55,1% not alert, 54,2% not responsible. Ther is a link between Excellent service ( $p=0,042$ ), Attitude ( $p=0,024$ ), Appearance ( $p=0,028$ ), Action ( $p=0,000$ ) and responsible ( $p=0,001$ ). Unrelated factor with excellent service ( $p=0,066$ ) The conclusion is the relation of capability, attitude, appearance, attention, action, and responsibility with Excellent service at the administration outpatient officer in RSUD Sultan Thaha Saifuddin Tebo in 2020. Hope the hospital can implement a servise schedule in accordance with the applicable schedule provisions.*

**Keywords : Excellent service; capability; attitude; appearance; attention; action; and responsibility**

## PENDAHULUAN

Kelulusan tepat waktu merupakan hal penting yang perlu disikapi dengan bijak oleh sebuah perguruan tinggi. Tingkat kelulusan merupakan salah satu parameter efektifitas institusi pendidikan. Penurunan tingkat kelulusan mahasiswa akan berpengaruh terhadap akreditasi perguruan tinggi. Oleh karena itu perlu adanya monitoring serta evaluasi terhadap kecenderungan kelulusan mahasiswa tepat waktu atau tidak. Database perguruan tinggi menyimpan data administrasi dan akademik mahasiswa, data tersebut apabila dieksplorasi dengan tepat menggunakan teknik Data Mining maka dapat diketahui pola atau pengetahuan untuk mengambil keputusan.

Penggunaan metode klasifikasi Data Mining untuk prediksi ketepatan waktu kelulusan mahasiswa dengan menggunakan algoritma Naïve Bayes dapat memberikan informasi mengenai keakuratan ketepatan waktu kelulusan mahasiswa. Data Mining adalah proses menganalisis data dan meringkas hasilnya menjadi informasi yang berguna. Secara teknis, Data Mining merupakan proses untuk menemukan korelasi diantara banyak bidang dalam dataset besar. Data Mining memiliki beberapa metode, salah satunya adalah metode klasifikasi yang merupakan teknik pembelajaran untuk mengklasifikasikan item data kedalam label kelas yang telah ditentukan. Metode klasifikasi memiliki beberapa algoritma salah satunya adalah Naive Bayes.

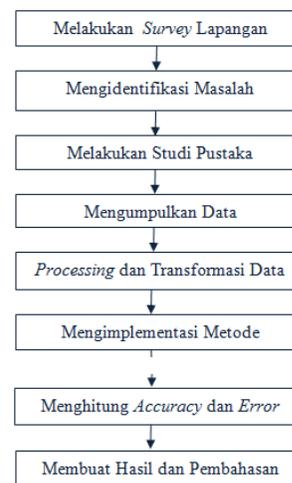
Metode Naïve Bayes merupakan penggolong probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam kumpulan data yang diberikan.

Dengan demikian penggunaan metode Data Mining akan memberikan hasil akurasi terbaik dalam pengklasifikasian data.

Penelitian sebelumnya mengkaji tentang perbandingan kinerja beberapa metode klasifikasi Data Mining dengan membandingkan algoritma Decision Tree dan Naive Bayes. Penelitian ini bertujuan untuk memprediksi mahasiswa yang dropout. Dari hasil pengujian ketepatan menggunakan kedua tersebut maka diperoleh keakuratan tertinggi yaitu pada algoritma Decision Tree.

Penelitian mengenai penggunaan algoritma Decision Tree seperti J48, Naïve Bayes, Random Tree, dan Decision Stump untuk mengidentifikasi siswa yang lemah dan kemungkinan gagal dalam ujian tinggi. Dari pengujian diperoleh bahwa algoritma J48 merupakan algoritma yang memiliki akurasi tertinggi dibandingkan keempat algoritma yang digunakan.

## METODE PENELITIAN



**Gambar 1. Kerangka Kerja**



Berdasarkan kerangka kerja pada Gambar 1, maka masing-masing langkahnya dapat diuraikan seperti berikut ini :

1. Melakukan Survey Lapangan  
Sebelum memulai penelitian terlebih dahulu dilakukan survey di lapangan untuk mendapatkan gambaran kualitatif mengenai ketepatan kelulusan mahasiswa di Universitas Negeri Padang.
2. Mengidentifikasi Masalah  
Tahap identifikasi masalah adalah tahap di mana objek penelitian merumuskan masalahnya..
3. Melakukan Studi Pustaka  
Untuk mencapai tujuan yang akan ditentukan, maka perlu dipelajari beberapa literatur-literatur yang digunakan.
4. Mengumpulkan Data  
Adapun dalam pengumpulan data dilakukan dengan beberapa cara yaitu:
  - a. Metode pengamatan langsung
  - b. Metode wawancara
  - c. Metode studi pustaka
  - d. Metode browsing
5. Processing dan Transformasi Data  
Pada tahap *Processing* dan Transformasi Data, data mentah akan diubah dan digabungkan kedalam format yang

## 2. Pengumpulan Data

Pada penelitian ini data yang digunakan adalah data mahasiswa Universitas Negeri Padang Program Studi Pendidikan Teknik Informatika dan Komputer tahun angkatan 2011 dan 2012. Data yang digunakan berjumlah 46 record.

### a. Pemilihan Variabel

Dari data mahasiswa, yang diambil sebagai variabel keputusannya adalah

sesuai untuk diproses ke dalam Data Mining.

## 6. Mengimplementasi Metode

Setelah proses analisa, maka selanjutnya dilakukan tahapan pengujian. Dalam pengujian diperlukan perangkat keras dan perangkat lunak komputer. Pada tahap ini akan dilakukan implmentasi metode yang telah diusulkan sebelumnya, akan diuji dengan menggunakan *Software RapidMiner*

## 7. Menghitung *Accuracy* dan *Error*

Pada tahap ini akan dihitung nilai *Accuracy* dan *Error* dari algoritma *Naive Bayes* untuk mengevaluasi nilai *Accuracy* dan *Error* dari pengukuran terhadap nilai sebenarnya atau nilai dianggap benar.

## 8. Membuat Hasil dan Pembahasan

Hasil dan pembahasan bertujuan untuk memberikan gambaran dan hasil yang diperoleh dari penelitian ini.

## ANALISA DAN PERANCANGAN

### 1. Analisa Data Mining

Merupakan suatu rangkaian proses yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

lulus tepat waktu dan terlambat. Sedangkan yang diambil sebagai variabel penentu dalam pembentukan keputusan adalah jenis kelamin, status masuk, IPK, daerah asal dan status pekerjaan.

### b. Melakukan Pra-Proses

Setelah melakukan pemilihan variabel, maka format data akan



ditransformasikan berdasarkan variabel-variabel yang sudah terpilih.

### 3. Metode Klasifikasi

Hasil klasifikasi yang diperoleh dapat memberikan informasi, mengenai tingkat

a. Proses Klasifikasi Menggunakan *Naive Bayes*

*accuracy* dan *error* ketepatan waktu kelulusan mahasiswa Universitas Negeri Padang. Penggunaan Algoritma *Naive Bayes* dilakukan dengan beberapa tahapan untuk mendapatkan informasi yang diinginkan.

**Tabel 1. Hasil Perhitungan Probabilitas Tepat dan Terlambat**

| No | NIM     | Probabilitas |           | Prediksi  |
|----|---------|--------------|-----------|-----------|
|    |         | Tepat        | Terlambat |           |
| 1  | 1102628 | 0,016        | 0         | Tepat     |
| 2  | 1102631 | 0,003        | 0         | Tepat     |
| 3  | 1102632 | 0,004        | 0         | Tepat     |
| 4  | 1102638 | 0,001        | 0,003     | Terlambat |
| 5  | 1102644 | 0            | 0,014     | Terlambat |
| 6  | 1102650 | 0,001        | 0,003     | Terlambat |
| 7  | 1102651 | 0,004        | 0         | Tepat     |
| 8  | 1102656 | 0,004        | 0         | Tepat     |
| 9  | 1102663 | 0            | 0,014     | Terlambat |
| 10 | 1102664 | 0            | 0,009     | Terlambat |
| 11 | 1102668 | 0            | 0,014     | Terlambat |
| 12 | 1102672 | 0,008        | 0         | Tepat     |
| 13 | 1102675 | 0,008        | 0         | Tepat     |
| 14 | 1102676 | 0            | 0,004     | Terlambat |
| 15 | 1102678 | 0            | 0,004     | Terlambat |
| 16 | 1102687 | 0            | 0,01      | Terlambat |
| 17 | 1102688 | 0,01         | 0         | Tepat     |
| 18 | 1102691 | 0            | 0,01      | Terlambat |
| 19 | 1102692 | 0,01         | 0         | Tepat     |
| 20 | 1102696 | 0,014        | 0,018     | Terlambat |
| 21 | 1102697 | 0,017        | 0,007     | Tepat     |
| 22 | 1102698 | 0,012        | 0         | Tepat     |
| 23 | 1102703 | 0,002        | 0,004     | Terlambat |
| 24 | 1102705 | 0            | 0,003     | Terlambat |
| 25 | 1102707 | 0,002        | 0,004     | Terlambat |
| 26 | 1106999 | 0            | 0,017     | Terlambat |
| 27 | 1107001 | 0            | 0,005     | Terlambat |
| 28 | 1107016 | 0,007        | 0         | Tepat     |
| 29 | 1107017 | 0,002        | 0,004     | Terlambat |



|    |         |       |       |           |
|----|---------|-------|-------|-----------|
| 30 | 1107025 | 0,008 | 0,012 | Terlambat |
| 31 | 1107033 | 0,003 | 0,006 | Terlambat |
| 32 | 1202175 | 0,012 | 0,001 | Tepat     |
| 33 | 1202183 | 0,012 | 0,001 | Tepat     |
| 34 | 1202191 | 0,012 | 0     | Tepat     |
| 35 | 1202196 | 0,002 | 0     | Tepat     |
| 36 | 1202197 | 0,002 | 0     | Tepat     |
| 37 | 1203244 | 0,015 | 0,002 | Tepat     |
| 38 | 1203237 | 0,007 | 0,003 | Tepat     |
| 39 | 1203238 | 0,003 | 0,001 | Tepat     |
| 40 | 1203239 | 0,003 | 0,001 | Tepat     |
| 41 | 1206507 | 0,016 | 0     | Tepat     |
| 42 | 1206519 | 0,006 | 0     | Tepat     |
| 43 | 1206520 | 0     | 0,005 | Terlambat |
| 44 | 1206522 | 0     | 0,042 | Terlambat |
| 45 | 1206538 | 0     | 0,008 | Terlambat |
| 46 | 1206545 | 0,017 | 0,007 | Tepat     |

#### 4. Tingkat Akurasi dan *Error* Algoritma *Naive Bayes*.

Pada perhitungan akurasi *Naive Bayes*, Diperoleh tingkat akurasi Algoritma *Naive Bayes* memiliki akurasi sebesar 93,48%.

**Tabel 2. Tabel Perbandingan Akurasi dan *Error* Algoritma C4.5 dan *Naive Bayes***

| Algoritma   | Akurasi | Error  |
|-------------|---------|--------|
| Naive Bayes | 93,48%  | 6,52 % |

Pada perhitungan *error Naive Bayes*, Diperoleh nilai *error* algoritma *Naive Bayes* memiliki sebesar 6,52 %.

## IMPLEMENTASI DAN HASIL

Pada Implementasi dan Hasil akan dijelaskan Implementasi atau pengujian untuk mengetahui hasil dari perhitungan manual dengan hasil menggunakan software pendukung algoritma *Naive Bayes*. Hal ini bertujuan untuk melihat data yang dianalisa dan diolah sudah benar atau belum. *Software*

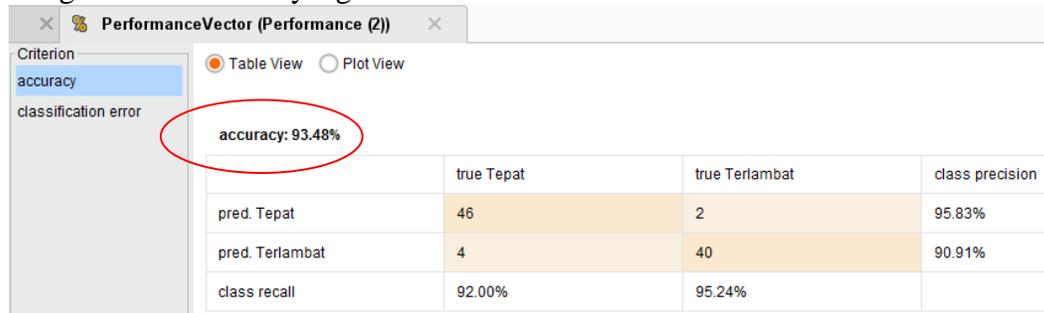
yang digunakan adalah *Rapidminer Studio 7.5.3*. *Rapidminer Studio* merupakan aplikasi *Data Mining open source*. Pada kasus memprediksi ketepatan waktu kelulusan mahasiswa ini, data yang akan digunakan pada *Rapidminer* berjumlah sebanyak 92 record

## 1. Tingkat Akurasi dan Error Algoritma Naive Bayes

### a. Naive Bayes

#### Tingkat Akurasi Naive Bayes

Pada perhitungan akurasi *Naive Bayes* diperoleh akurasi sebesar 93,48% karena menghasilkan 86 data yang diklasifikasikan secara benar.

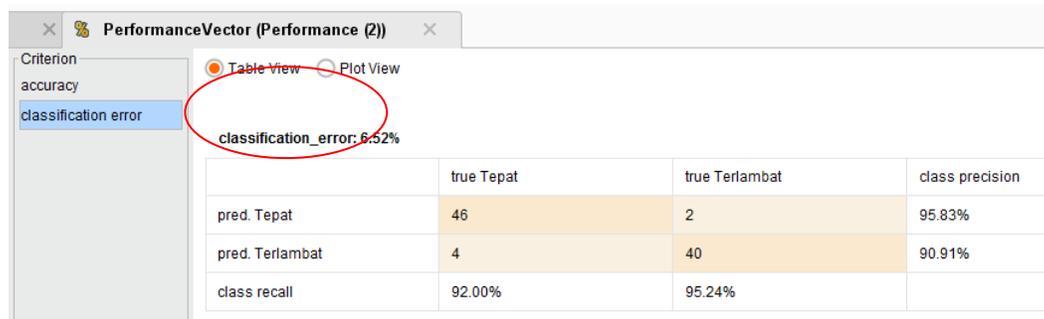


|                 | true Tepat | true Terlambat | class precision |
|-----------------|------------|----------------|-----------------|
| pred. Tepat     | 46         | 2              | 95.83%          |
| pred. Terlambat | 4          | 40             | 90.91%          |
| class recall    | 92.00%     | 95.24%         |                 |

**Gambar 2. Akurasi Naive Bayes**

#### Tingkat Error Naive Bayes

Pada perhitungan *Error Naive Bayes* diperoleh akurasi sebesar 6,52% karena menghasilkan 6 data yang diklasifikasikan secara tidak benar.



|                 | true Tepat | true Terlambat | class precision |
|-----------------|------------|----------------|-----------------|
| pred. Tepat     | 46         | 2              | 95.83%          |
| pred. Terlambat | 4          | 40             | 90.91%          |
| class recall    | 92.00%     | 95.24%         |                 |

**Gambar 3. Error Naive Bayes**

## KESIMPULAN

1. Pengukuran tingkat akurasi kinerja metode klasifikasi *Naive Bayes* menghasilkan nilai akurasi sebesar 93,48%.
2. Pengukuran tingkat error pada algoritma *Naive Bayes* menghasilkan tingkat error sebesar 6,52 %.
3. Dari hasil pengujian yang telah dilakukan, Algoritma *Naive Bayes*

memiliki kinerja yang bagus karena C4.5 memiliki nilai akurasi yang tinggi, semakin tinggi nilai akurasi maka pengklasifikasian data semakin mendekati benar. Algoritma *Naive Bayes* juga memiliki nilai error yang lebih rendah, semakin rendah nilai error maka pengklasifikasian semakin mendekati benar.

**DAFTAR PUSTAKA**

- Sumathi K., Kannan S. dan Nagaraan K. (2016), "Data Mining: Analysis of student database using Classification Techniques", International Journal of Computer Applications, Vol. 141, No. 8, Hal. 22-27.
- Agrawal Gaurav dan Gupta Hitesh (2013). "Optimization of C4.5 Decision Tree Algorithm for Data Mining Application". International Journal of Emerging Technology and Advanced Engineering, Vol.3, Hal. 341-345.
- Patil Tina dan Sherekar S. (2013). "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification", International Journal Of Computer Science And Applications, Vol. 6, No.2, Hal. 256-261.
- S. Ghadeer, Oda Abu dan M. El-Halees Alaa (2015). "Data Mining In Higher education: University Student Dropout Case Study", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, Hal. 15-27.
- Sharma Ritu, Kumar Shiv dan Maheshwari Rohit (2015), "Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets". IJCSMC, Vol. 4, Issue. 12, December 2015, Hal. 125-134.
- Lestari Handayani dan Eka Lona Maulida. (2015). "Perkiraan Waktu Studi Mahasiswa Menggunakan Metode Klasifikasi Dengan Algoritma Naive Bayes", Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 7, Hal. 138-147.
- Miftahul Chair, Yuki Novia Nasution dan Nanda Arista Rizki (2017). "Aplikasi Klasifikasi Algoritma C4.5 (Studi Kasus Masa Studi Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2008)", Jurnal Informatika Mulawarman, Vol. 12, No. 1, Hal. 51-55.
- Cipta Riang Sari (2016). "Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA Negeri 1 Polewali". Indonesian Journal on Networking and Security. 2016; Volume 5: Hal 48-54.
- Siska Haryati, Aji Sudarsono dan Eko Suryana (2015), "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)", Jurnal Media Infotama Vol. 11 No. 2, Hal. 130-138.
- Dicky Nofriansyah, Kamil Erwansyah, Mukhlis Ramadhan, (2016). "Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)". Jurnal SAINTIKOM, Vol. 15, No. 2, Mei 2016.