



KLASIFIKASI DATA MINING KELULUSAN MAHASISWA (STIMIKOM) STELLA MARIS SUMBA MENGGUNAKAN ALGORITMANAIVE BAYES

Rahel Danga Tamo ^{1*}, Trisno ², Titus Kurra ³ Uliyatunisa ⁴

^{1*2.3.} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Stella Maris Sumba

⁴Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

*email koresponden: trisnomtf@gmail.com

Abstrak

Dalam data mining, penelitian mengenai klasifikasi kelulusan mahasiswa sudah pernah dilakukan oleh peneliti lain. Sebagian besar penelitian tersebut difokuskan pada identifikasi variabel prediktor. Ada banyak penelitian dalam literatur terdahulu yang menjelaskan faktor-faktor apa saja yang dapat mensukseskan proses pengklasifikasian kelulusan mahasiswa. Pemilihan penggunaan algoritma Naive Bayes, dan pada penelitian ini didasarkan pada beberapa alasan, yaitu: Selain ketiga algoritma tersebut sama-sama mudah diimplementasikan dan sama-sama dapat memberikan hasil yang baik dalam kasus klasifikasi, ketiga algoritma tersebut juga mempunyai beberapa keunggulan masing-masing. Implementasi data training sebanyak 51 data dengan algoritma Naive Bayes berhasil memprediksi besarnya kelulusan mahasiswa dengan persentase keakuratan sebesar 74,67%, Sebanyak 184 mahasiswa sebagai data uji yang dihasilkan penelitian ini bahwa mahasiswa yang akan lulus tepat waktu sebanyak 42 mahasiswa atau sekitar 22,8% dari jumlah data testing dengan keakuratan sebesar 74,67%, Bagi Prang Studi Teknik Informatika berdasarkan hasil penelitian ini diharapkan dapat dimanfaatkan oleh pihak jurusan sebagai bahan pertimbangan bahwa dari jumlah data testing sebanyak 184 mahasiswa yang diprediksi lulus tepat waktu sekitar 22,8% dan jumlah tidak tepat waktu 77,2%.

Kata Kunci : Kelulusan Mahasiswa, Naïve Bayes, STIMIKOM Stella Maris Sumba

Abstract

In data mining, research on student graduation classifications has been carried out by other researchers. Most of these studies focused on identifying predictor variables. There is a lot of research in previous literature that explains what factors can make the student graduation classification process successful. The choice of using the Naive Bayes algorithm in this research was based on several reasons, namely: Apart from the three algorithms being equally easy to implement and both being able to provide good results in classification cases, the three algorithms also have several advantages. The implementation of 51 training data using the Naive Bayes algorithm succeeded in predicting the number of students graduating with an accuracy percentage of 74.67%. A total of 184 students as test data produced in this research showed that 42 students would graduate on time or around 22.8%. from the total testing data with an accuracy of 74.67%, for students studying Informatics Engineering, based on the results of this research, it is hoped that it can be used by the department as material for consideration. From the total testing data of 184 students, 22.8% are predicted to graduate on time and the number not on time 77.2%.

Keywords: Student Graduation, Naïve Bayes

1. PENDAHULUAN

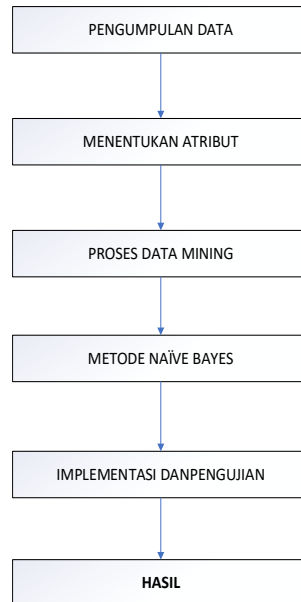
Dalam data mining, penelitian mengenai klasifikasi kelulusan mahasiswa sudah pernah dilakukan oleh peneliti lain. Sebagian besar penelitian tersebut difokuskan pada identifikasi variabel prediktor. Ada banyak penelitian dalam literatur terdahulu yang menjelaskan faktor-faktor apa saja yang dapat mensukseskan proses pengklasifikasian kelulusan mahasiswa. Data mining yang ada di dunia pada saat ini sangat rentan terhadap noisy data, data yang hilang atau tidak lengkap, dan data yang tidak konsisten karena biasanya ukuran dari data mining tersebut sangat besar serta sumber dari data-data tersebut biasanya lebih dari satu (heterogen). Untuk itu, menyiapkan data yang baik, memadai dan representatif merupakan langkah awal yang tidak dapat diabaikan begitu saja. Menggunakan algoritma Naive Bayes, sudah pernah digunakan untuk mengolah dan memining kelulusan mahasiswa kelebihan dan kekurangan. Namun dipastikan cepat dalam melakukan klasifikasi. Hal ini dikarenakan data mining yang digunakan oleh peneliti sebelumnya tidak sama, perlakuan yang diberikan kepada data tersebut juga berbeda. Semakin kompleks data, noise pada data, serta data yang tidak konsisten tentunya akan berdampak pada kinerja algoritma pengklasifikasi data mining kelulusan mahasiswa STIMIKOM Stella Maris Sumba. Berdasarkan identifikasi masalah di atas, maka pada penelitian ini berusaha menjawab klasifikasi data mining kelulusan mahasiswa STIMIKOM Stella Maris Sumba menggunakan algoritma naive bayes. Tujuan dari penelitian ini adalah untuk mendapatkan data yang akurat dan jelas dalam penentuan klasifikasi data mining kelulusan mahasiswa stimikom stella maris sumba. Manfaat teoritis dari penelitian ini diharapkan dapat memberikan masukan atas teori pemodelan algoritma Naive Bayes, khususnya untuk penelitian yang berhubungan dengan penelitian di bidang klasifikasi.

Penelitian yang dilakukan oleh David Hartanto Kamagi dan Seng Hansun (2014) yang berjudul "Implementasi Data Mining dengan Algoritma naïve bayes untuk Memprediksi Tingkat Kelulusan Mahasiswa (Studi Kasus: program studi teknik informatika universitas multimedia Nusantara)". Pada penelitian ini, menjelaskan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa adalah IPK, Indeks Prestasi Semester (IPS) 1 sampai 4, dan jenis kelamin (Ridwan, Suyono, & Sarosa, 2013). Dalam penelitiannya, penulis menggunakan algoritma naïve bayes untuk menentukan prediksi kelulusan berdasarkan atribut jenis kelamin, asal sekolah SMA, dan IPS 1 sampai dengan 6.

Penelitian yang dilakukan oleh Mutiara Ayu Banjarsari, H. Irawan Budiman, dan Andi Farmadi (2015) yang berjudul "Penerapan Koptimal Pada Algoritma KKn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi teknik informatika Berdasarkan IP Sampai Dengan Semester 4". Pada penelitian ini, peneliti mencoba untuk mengetahui nilai k-optimal dan tingkat akurasi pada algoritma kNN untuk prediksi kelulusan mahasiswa. Penentuan nilai k pada algoritma kNN menjadi hal yang sangat penting karena akan mempengaruhi kinerja algoritma kNN (Wu, 2009). Pencarian nilai k-optimal pada algoritma kNN yang dilakukan oleh peneliti yaitu dengan cara menggunakan metode k-fold cross validation dengan variabel input indeks prestasi mahasiswa sampai dengan semester 4.

2. METODE PENELITIAN

Dalam penelitian ini, metode pengumpulan data terdiri dari wawancara kepada Sekolah yang ada di tempat lokasi dan Observasi dilakukan turun lapangan secara langsung pada STIMIKOM Stella Maris Sumba.



Gambar Tahapan Penelitian

Variabel ekonomi adalah variabel yang berisi tentang keadaan ekonomi orang tua atau keluarga mahasiswa. Pilihan yang terdapat pada program ini antara lain dibedakan menjadi tiga bagian yaitu rendah, sedang dan tinggi.

Tabel Atribut Ekonomi

No	Penghasilan	Keterangan
1	Penghasilan Rp. 500.000 – 1.000.000 / bulan	Rendah
2	Penghasilan Rp. 1.100.000 – 2.500.000 / bulan	Sedang
3	Penghasilan Rp. 2.600.000 – 4.000.000 / bulan	Tinggi

Sumber : Data Mahasiswa STIMIKOM Stella Maris Sumba Barat Daya

Indeks Prestasi Kumulatif (IPK)

Variabel Indeks Prestasi Kumulatif (IPK) adalah nilai rata-rata Indeks Prestasi (IP) semester yang telah ditempuh oleh mahasiswa. Variabel Indeks Prestasi Kumulatif (IPK) dikelompokkan menjadi 3 bagian sebagai berikut:

Tabel Atribut Indeks Prestasi Kumulatif (IPK)

No	Penghasilan	Keterangan
1	$IPK \geq 3$	3
2	$2 \leq IPK < 3$	2
3	$IPK < 2$	1

Keputusan

Variabel keputusan merupakan data yang berfungsi untuk menentukan hasil keputusan. Dalam data keputusan hanya memiliki 2 buah nilai yaitu tepat waktu dan tidak tepat waktu.

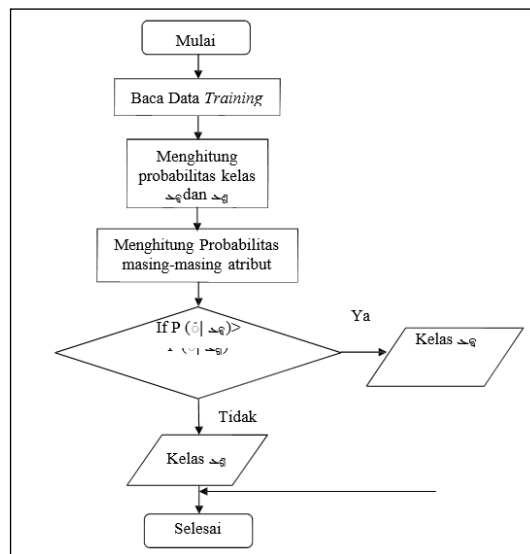
Transformasi Data

Tahap selanjutnya setelah data cleaning adalah transformasi data yaitu pada tahap ini data yang sudah bersih tersebut perlu ditransformasi terlebih dahulu. Data yang sudah ditransformasi dapat diolah menggunakan data mining. Berikut ini adalah hasil transformasi data dari data training.

181	TEKNIK INFORMATIK	S	181201296	Yulita Priska Mako	Takirin, 8 July 2000	P
182	TEKNIK INFORMATIK	S	181201104	Jumiati Bani	Parai Deta, 27 July 1999	P
183	TEKNIK INFORMATIK	S	181201347	Martha Bulu	Wanno Wunga, 10 July 1997	P
184	TEKNIK INFORMATIK	S	181201130	Magdalena Denga	Reda Mata, 24 February 1996	P
185	TEKNIK INFORMATIK	S	181201413	Berta Bela Rade Kaka	Seda, 10 March 1998	P
186	TEKNIK INFORMATIK	S	171200498	JEFRIANUS TANGGU EGE	POGO TENA, 5 December 1998	L
					MEHANG GALLU, 30 January 1998	L
187	TEKNIK INFORMATIK	S	171200495	JACK JANGA HABA KODI	PUU WOO, 8 December 1997	L
188	TEKNIK INFORMATIK	S	171200550	MARSELINUS TAMO AMA	Wee Tombo, 5 April 1999	P
189	TEKNIK INFORMATIK	S	181201185	Nilawati Kusuma Bili	KALEMBU DANGA, 30 November 1998	P
190	TEKNIK INFORMATIK	S	171200581	OKTAVIANA RAMBI	Puu Maowo, 10 July 1998	L
191	TEKNIK INFORMATIK	S	181201055	Dominikus Renda	Watu Mete, 26 September 1995	P
192	TEKNIK INFORMATIK	S	181201120	Lidia Tangu	WANNO DIKKI, 7 June 1997	P
193	TEKNIK INFORMATIK	S	171200392	ANNA GORETTI MAWO	WULUWAWI, 13 October 1998	P
194	TEKNIK INFORMATIK	S	171200719	ALIYANA L PORA	WATU KABUNGGUTA, 27 November 1995	L
195	TEKNIK INFORMATIK	S	171200571	NIKODEMUS NGONGO	GENGGOL, 12 January 1996	L
196	TEKNIK INFORMATIK	S	171200704	GABRIEL GHEDA RANGGA	Gollu Nuu, 23 July 1997	P
197	TEKNIK INFORMATIK	S	181201265	Yasinta Milla	Malata, 28 January 1999	P
198	TEKNIK INFORMATIK	S	181201036	Debora Umbu Lele	HOMBA KARIPIT, 12 May 1999	L
199	TEKNIK INFORMATIK	S	171200565	MEIRANTO FLORENZO FRIDS	Kori, 6 May 1997	P
200	TEKNIK INFORMATIK	S	181201175	Meryani Patnai	WANNO MARADA, 25 June	L
201	TEKNIK INFORMATIK	S	171200478	GULIELMUS FEBRUNIA	Kwiratu, 25 April 1998	P
202	TEKNIK INFORMATIK	S	181201442	Mariana Watti Riwu	BONDO RONGO, 11 February 1998	P
203	TEKNIK INFORMATIK	S	171200551	MARTA AWANG	Puu Delo, 12 May 1997	L
204	TEKNIK INFORMATIK	S	181201006	Antonius Buka Lede	Mali Mada, 7 March 1996	L
205	TEKNIK INFORMATIK	S	181201345	Markus Saba Kodi	Wanno Baru, 17 February 1996	L
206	TEKNIK INFORMATIK	S	181201398	Selfianus Ferdinand Paila	Puu Bila, 5 October 1999	P
207	TEKNIK INFORMATIK	S	181201208	Oncepani Bulu	Bondo Bulu, 7 November 1991	P
208	TEKNIK INFORMATIK	S	181201320	Yensi Awang Umbu Deta	Omba Nalo, 12 May 1998	L
209	TEKNIK INFORMATIK	S	181201233	Sepriadi Lende Poety		P

Gambar Hasil Data Transformasi

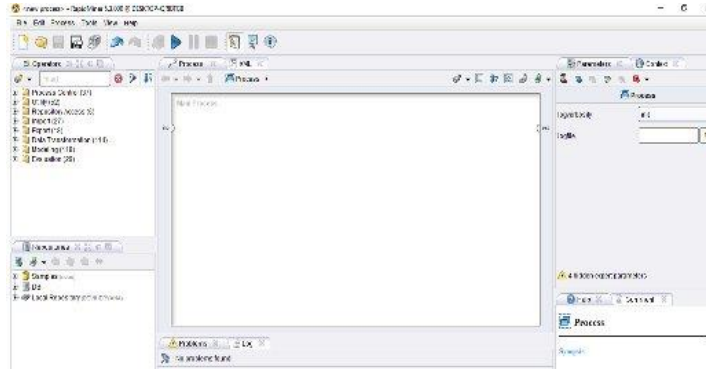
Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika Alur perhitungan Naive Bayes menurut Odi Nurdiawan dan Noval Salim.



Gambar Alur Naive Bayes

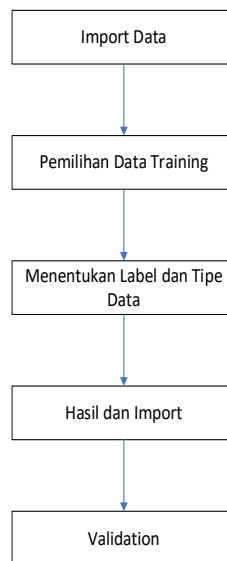
Implementasi dan Pengujian Metode

Implementasi dan pengujian keakurasian pada penelitian ini menggunakan alat RapidMiner 5.3. RapidMiner merupakan pemrograman lunak yang bekerja dalam pengolahan data.



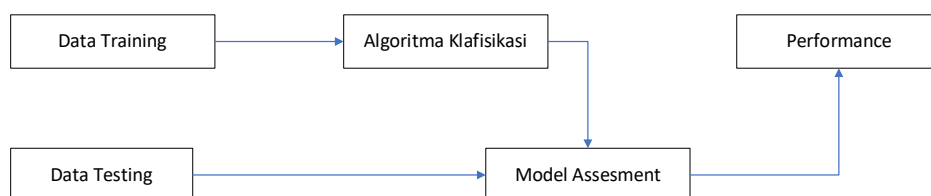
Gambar RapidMiner 5.3

Tahapan-tahapan pengujian akurasi menggunakan RapidMiner 5.3 metode Naive Bayes:



Gambar Pengujian menggunakan RapidMiner 5.

Tahapan-tahapan implementasi data testing menggunakan RapidMiner 5.3 metode Naive Bayes.



Gambar Implementasi RapidMiner Data Testing

3. HASIL DAN PEMBAHASAN

Data yang digunakan sudah dilakukan proses pembersihan dan transformasi dalam bentuk kategori. Pengujian menggunakan data mahasiswa angkatan 2017 Teknik Informatika. Data

mahasiswa sebanyak 134 record diambil 51 record untuk digunakan sebagai data training. Berdasarkan hasil pengolahan data dan jumlah data yang digunakan tersebut dapat dibagi menjadi dua kelas kategori lulus tepat waktu sebanyak 15 mahasiswa dan lulus tidak tepat waktu sebanyak 36 mahasiswa. Proses pengujian, data dibagi menjadi 2 bagian yaitu data training dan data testing dengan menggunakan Naive Bayes.

Analisis Metode Naive Bayes Perhitungan Naive Bayes:

Tabel Probabilitas Kelas

Kelas			
Tepat Waktu		Tidak Tepat Waktu	
Tepat Waktu	15/51	Tidak Tepat Waktu	36/51

Tabel Jenis Kelamin

Jenis Kelamin	Tepat Waktu	Tidak Tepat Waktu
Pria	2/15	6/36
Wanita	13/15	30/36

Tabel Kota Kelahiran

Kota Kelahira	Tepat Waktu	Tidak Tepat Waktu
Dalam Kota	3/15	3/36
Luar Kota	12/15	33/36

Tabel Tipe Sekolah

Kota Kelahira	Tepat Waktu	Tidak Tepat Waktu
Umum	15/15	31/36
Kejuruan	0/15	5/36

Tabel Lokasi Sekolah

Lokasi Sekolah	Tepat Waktu	Tidak Tepat Waktu
Dalam Kota	4/15	3/36
Luar Kota	11/15	36/36

Tabel Ekonomi

Ekonomi	Tepat Waktu	Tidak Tepat Waktu
Rendah	9/15	24/36
Sedang	5/15	8/36
Tinggi	1/15	3/36

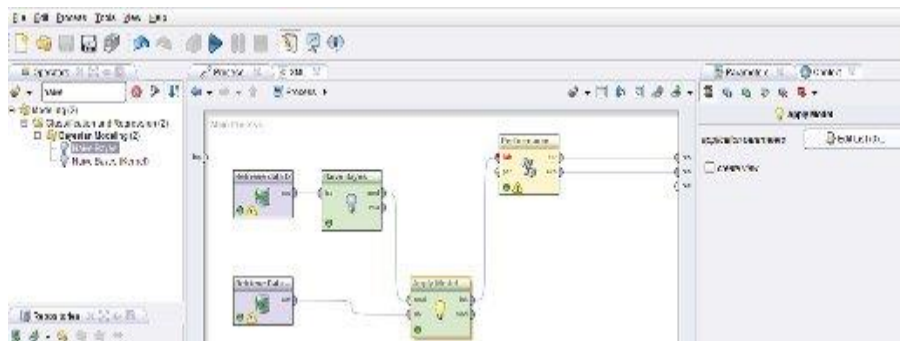
Tabel Indeks Prestasi Komulatif (IPK)

IPK	Tepat Waktu	Tidak Tepat Waktu
1	0/15	0/36
2	0/15	0/36
3	15/15	36/36

Tabel Data Testing

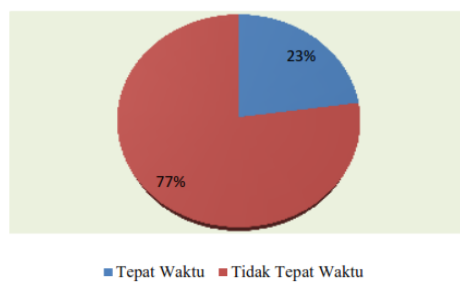
Jenis Kelamin	Kota Kelahiran	Tipe Sekolah	Lokasi Sekolah	Ekonomi	IPK	Keputusan
Wanita	Dalam Kota	Kejuruan	Dalam Kota	Sedang	3	?

Implementasi Data Testing:

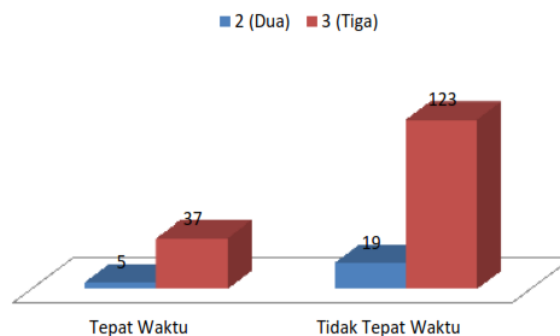


Gambar Susunan Operator Data training, Data Testing, Apply Model, dan Performance

Jumlah data testing yang digunakan sebanyak 184 mahasiswa prodi Teknik Informatika angkatan 2017 sebagai data testing yang menggunakan metode Naive Bayes didapatkan hasil penelitian ini bahwa mahasiswa yang akan lulus tepat waktu sebanyak 42 mahasiswa dari seluruh data testing dan mahasiswa tidak tepat waktu sebanyak 142 mahasiswa. Perbandingan hasil kelas lulus tepat waktu dan tidak tepat waktu dapat divisualisasikan dalam bentuk Gambar dibawah ini



Grafik Prediksi Kelulusan Berdasarkan IPK





5. DAFTAR PUSTAKA

- David Hartanto kamagi, Seng Hansun, “Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa”, „Vol VI. No1. Juni 2014”.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eeccis*, 7(1), 59–64.
- Mutiara Ayu Banjarsari, H. Irawan Budiman, dan Andi Farmadi (2015). Penerapan Koptimal Pada Algoritma KKn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi teknik informatika Berdasarkan IP Sampai Dengan Semester 4”. Program pascasarjana magister ilmu computer Sekolah tinggi manajemen informatika dan computer Nusa mandiri Jakarta 2017.
- Wu X, Kumar V. 2009. *The Top Ten Algorithms in Data Mining*. New York: CRC Press
- Nurdiawan, Odi. Noval Salim, 2018. Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Metode Naive Bayes Classifier Untuk Optimasi Strategi Pemasaran. (Online) , tersedia : [http:// www.jurnalstmiksubang.ac.id](http://www.jurnalstmiksubang.ac.id)