

Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes (Studi Kasus: Fasilkom Unilak)

Nurliana Nasution¹, Khairani Djahara², Ahmad Zamsuri³

^{1,2,3}Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning
Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015
e-mail: ¹nuliana_2006@yahoo.co.id, ²khairani.djahara@unilak.ac.id,
³ahmadzamsuri@unilak.ac.id

Abstrak

Evaluasi sukses tidaknya kinerja akademik Mahasiswa pada masa studinya salah satunya dapat dilihat dari ketepatannya dalam menyelesaikan masa studinya di universitas. Hal ini dapat dipantau sejak mereka berada pada tingkat 3 atau semester 6. Evaluasi dilakukan dengan membagi data menjadi dua yaitu sebagai data latih (training set) yang diambil dari data histori lulusan sebanyak 150 data; dan data uji (testing set) dari data mahasiswa di tingkat 3 atau semester 6 yang belum menyelesaikan masa studinya sebanyak 61 data. Hasil dari penelitian dengan pembangunan model Naïve Bayes terbaik berada pada akurasi 76%, kemudian model tersebut diterapkan untuk prediksi kelulusan mahasiswa pada tingkat 3, dengan perolehan hasil 48 mahasiswa lulus tepat waktu dan 13 mahasiswa lulus tidak tepat waktu. Pada penelitian ini atribut yang digunakan adalah jenis kelamin; program studi; asal sekolah; daerah asal; tempat tanggal lahir; indeks prestasi semester 1 sampai dengan 5; indeks prestasi kumulatif; dan target kelulusan.

Kata kunci: Data Mining, Kinerja Akademik, Klasifikasi, Naïve Bayes

Abstract

Evaluation of the success or failure of Student academic performance during her studies one of which can be seen from its accuracy in completing their study at the university. It can be monitored since they are at level 3 or semesters 6. Evaluation is done by dividing the data into two, as training data (training set) are taken from historical data as much as 150 graduates of the data; and test data (testing set) of the student data at level 3 or 6 semesters who have not completed their study as many as 61 data. Results of the study with the best development Naïve Bayes models are at 76% accuracy, then the model is applied to the prediction graduation at level 3, with the acquisition of the results of 48 students graduate on time and 13 students do not graduate on time. In this study attributes used is sex; study program; Which school are you from; place of Origin; date of birth; the semester grade of 1 to 5; grade-point average; and graduation targets.

Keywords: Data Mining, Academic Performance, Classification, Naïve Bayes

1. Pendahuluan

Salah satu misi Universitas Lancang Kuning ialah “Meningkatkan Kualitas Pendidikan untuk Menghasilkan Lulusan Kompetitif Berskala Nasional”. Kualitas pendidikan disini terkait langsung dengan kemampuan mahasiswa yang merupakan aspek penting dalam menentukan keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Salah satu indikator dari keberhasilan suatu perguruan tinggi dapat dilihat dari banyaknya tingkat kelulusan mahasiswa setiap tahunnya. Pada penelitian kali ini, peneliti mencoba untuk mengukur tingkat

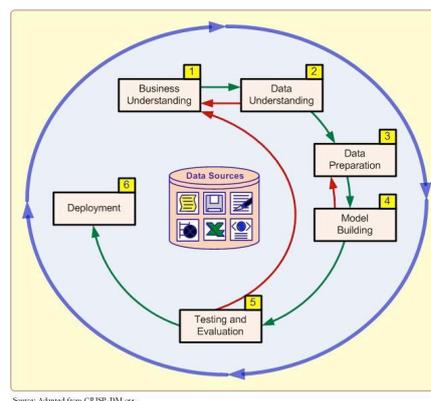
kelulusan mahasiswa Fakultas Ilmu Komputer Universitas Lancang Kuning melalui prediksi lulus tepat waktu atau lulus tidak tepat waktu berdasarkan evaluasi kinerja akademik di tingkat 3 atau semester 6 dengan menggunakan algoritma *Naïve Bayes Classifier*.

Beberapa penelitian terkait dengan topik penelitian pernah dilakukan [1] dengan mengevaluasi kinerja akademik mahasiswa menggunakan algoritma *Naïve Bayes* dengan memanfaatkan indikator Indeks Prestasi (IP) sebagai penentu mahasiswa lulus atau lulus tidak tepat waktu dengan hasil akurasi sebesar 70%. Penelitian yang dilakukan oleh [2] mengklasifikasikan seorang lulusan dari suatu perguruan tinggi dengan predikat memuaskan, sangat memuaskan atau *cumlaude* menggunakan algoritma C4.5.

Dari beberapa penelitian di atas, dapat disimpulkan bahwa atribut yang signifikan dalam penentuan kelulusan secara umum adalah Indeks Prestasi Kumulatif (IPK) dengan menggunakan algoritma *Naïve Bayes*. Sehingga pada penelitian ini akan dilakukan uji coba penerapan algoritma *Naïve Bayes* dengan beberapa atribut terpilih seperti NIM, Nama, Jenis Kelamin, Asal Sekolah, Kota Asal, Tempat Tanggal Lahir, Indeks Prestasi Semester 1 sampai dengan 6 dan Indeks Prestasi Kumulatif (IPK) yang dapat meningkatkan akurasi. Sementara untuk keakuratan akurasi klasifikasi data mining dapat diukur dengan menggunakan *cross validation*, *confusion matrix*, *ROC-Curve* dan *T-Test* seperti yang dilakukan oleh [3].

2. Metode Penelitian

Pada penelitian ini menggunakan model CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang memiliki 6 tahapan seperti yang diilustrasikan pada gambar 1 [4].



Source: Adapted from CRISP-DM.org

Gambar 1. Model CRISP-DM

2.1 Business Understanding (Pemahaman Proses Bisnis)

Fakultas Ilmu Komputer Universitas Lancang Kuning selama ini belum memanfaatkan data kelulusan yang ada untuk memprediksi data mahasiswa tingkat 3 atau semester 6 lulus tepat waktu atau lulus tidak tepat waktu. Prediksi kelulusan mahasiswa selama ini hanya ditentukan secara subjektif. Pada penelitian kali ini akan dilakukan prediksi terhadap kelulusan mahasiswa menggunakan algoritma *Naive Bayes Classifier*.

2.2 Data Understanding (Pemahaman Data)

Data latih yang digunakan pada penelitian ini diambil dari data lulusan tahun 2007 sampai dengan 2011 sebanyak 150 data terdiri dari 100 mahasiswa dari Prodi Teknik Informatika (60 lulus tepat waktu dan 40 tidak lulus tepat waktu) dan 50 mahasiswa Prodi Sistem Informasi (37 lulus tepat waktu dan 13 tidak lulus tepat waktu). Data Uji Mahasiswa diambil dari data mahasiswa tingkat 3 atau semester 6 yang belum lulus yaitu tahun 2012 sebanyak 61 data, 38 dari Prodi Teknik Informatika dan 23 dari Prodi Sistem Informasi. Data latih dan data uji memiliki jumlah atribut yang sama yaitu 14 buah dengan rincian sebagai berikut pada tabel 1.

Tabel 1. Atribut Data Latih dan Data Uji

No	Atribut	Keterangan
1.	NIM	Nomor Induk Mahasiswa
2.	Nama	Nama Asli Mahasiswa
3.	Jenis_Kelamin	Jenis Kelamin Mahasiswa
4.	Asal_Sekolah	Asal Sekolah Menengah Mahasiswa
5.	TTL	Tempat Tanggal Lahir Mahasiswa
6.	Kota_Asal	Kota Kelahiran Mahasiswa
7.	Program_Studi	Program Studi yang diambil Mahasiswa
8.	IP_Sem1	Indeks Prestasi Mahasiswa Semester 1
9.	IP_Sem2	Indeks Prestasi Mahasiswa Semester 2
10.	IP_Sem3	Indeks Prestasi Mahasiswa Semester 3
11.	IP_Sem4	Indeks Prestasi Mahasiswa Semester 4
12.	IP_Sem5	Indeks Prestasi Mahasiswa Semester 5
13.	IPK	Indeks Prestasi Mahasiswa Kumulatif
14.	Target_Kelulusan	Target Kelulusan Sarjana Mahasiswa

2.3 Data Preparation (Pengolahan Data)

Data yang diolah berasal dari data lulusan sebagai data latih dan data mahasiswa tingkat 3 atau semester 6 yang belum lulus sebagai data uji. Adapun tahapan *preprocessing* data yang akan dilakukan meliputi :

1. Validasi Data
Identifikasi dan menghapus data yang memiliki *outlier/noise*, data yang tidak konsisten dan data yang tidak lengkap (*missing value*). Pada data uji dan data latih, atribut TTL dan Kota_Asal memiliki *missing value* di beberapa data, sehingga pada penelitian ini atribut Asal_Sekolah dihapuskan agar tidak mempengaruhi tingkat akurasi yang dicapai.
2. Integrasi dan Transformasi Data
Integrasi data merupakan proses penyatuan data yang dibutuhkan. Pada data uji dan data latih, atribut Target_Kelulusan merupakan hasil integrasi dari data tahun mahasiswa masuk (tahun angkatan) dengan tahun mahasiswa lulus sarjana.
3. Pengurangan dan Diskretisasi Data
Proses pengurangan data ditujukan agar proses pelatihan tidak membutuhkan waktu yang lama, dengan catatan bahwa data yang terpilih adalah data yang benar-benar informatif. Atribut NIM dan Nama tidak memberikan nilai informatif yang dibutuhkan sehingga atribut tersebut dihapuskan. Kemudian proses selanjutnya adalah proses diskretisasi. Proses diskretisasi atribut data adalah proses perubahan nilai atribut menjadi nilai kategorikal, seperti JenisKelamin, Laki-laki dengan nilai 1 dan Perempuan dengan nilai 2; ProgramStudi, Teknik Informatika dengan nilai 1 dan Sistem Informasi dengan nilai 2; Asal_Sekolah, Negeri dengan nilai 1, Swasta dengan nilai 2, Kejuruan dengan nilai 3, Madrasah Aliyah dengan nilai 4 dan tanpa keterangan dengan nilai 5 serta Target_Kelulusan, lulus tepat waktu dengan nilai 1 dan lulus tidak tepat waktu dengan nilai 2.

Sehingga, jumlah atribut yang digunakan untuk penelitian setelah melalui tahapan pengolahan data (*preprocessing* data) adalah 10 buah dengan rincian pada tabel 2.

Tabel 2. Atribut Data Latih dan Uji Setelah di *Preprocessing*

No	Atribut	Keterangan
1.	Jenis_Kelamin	Jenis Kelamin Mahasiswa
2.	Asal_Sekolah	Asal Sekolah Menengah Mahasiswa
3.	Program_Studi	Program Studi yang diambil Mahasiswa
4.	IP_Sem1	Indeks Prestasi Mahasiswa Semester 1
5.	IP_Sem2	Indeks Prestasi Mahasiswa Semester 2
6.	IP_Sem3	Indeks Prestasi Mahasiswa Semester 3
7.	IP_Sem4	Indeks Prestasi Mahasiswa Semester 4
8.	IP_Sem5	Indeks Prestasi Mahasiswa Semester 5
9.	IPK	Indeks Prestasi Mahasiswa Kumulatif
10.	Target_Kelulusan	Target Kelulusan Sarjana Mahasiswa

2.4 Model Building (Pembanguna Model)

Model yang digunakan pada penelitian ini adalah klasifikasi, artinya data yang digunakan untuk pelatihan sudah memiliki target kelas. Target kelas yang dimaksud adalah pada atribut Target_Kelulusan yaitu lulus tepat waktu dan lulus tidak tepat waktu. Algoritma yang digunakan pada penelitian ini adalah *Naïve Bayes Classifier* yang memanfaatkan probabilitas sederhana dalam proses *mining* data. Secara sederhana ilustrasi perhitungan *Naïve Bayes Classifier* dapat dilihat pada gambar 2 untuk data uji dan data latih, sedangkan penyelesaiannya menggunakan algoritma *Naïve Bayes Classifier* dapat dilihat pada gambar 3.

Data Uji				
Diberikan Data Uji berikut : Nama Hewan = Musang; Penutup Kulit = Rambut; Melahirkan = Ya, maka diprediksi di kelas apakah hewan tersebut?				
Data Latih				
No	Nama Hewan	Penutup Kulit	Melahirkan	Kelas
1.	Ular	Sisik	Ya	Reptil
2.	Tikus	Bulu	Ya	Mamalia
3.	Kambing	Rambut	Ya	Mamalia
4.	Sapi	Rambut	Ya	Mamalia
5.	Kadal	Sisik	Tidak	Reptil
6.	Kucing	Rambut	Ya	Mamalia
7.	Bekicot	Cangkang	Tidak	Reptil
8.	Harimau	Rambut	Ya	Mamalia
9.	Rusa	Rambut	Ya	Mamalia
10	Kura-kura	Cangkang	Tidak	Reptil

Gambar 2. Data Uji dan Data Latih

Hitung Probabilitas Terhadap Kelas - Probabilitas Terbesar : Mamalia
Kelas Mamalia : $P(\text{Mamalia} \text{Hewan} = \text{Musang}) = \frac{P(\text{Rambut} \text{Mamalia}) \times P(\text{Lahir Ya} \text{Mamalia}) \times P(\text{Mamalia})}{P(\text{Hewan} = \text{Musang})}$ $P(\text{Mamalia} \text{Hewan} = \text{Musang}) = 5/6 \times 6/6 \times 6/10 = 0.5$
Kelas Reptil $P(\text{Reptil} \text{Hewan} = \text{Musang}) = \frac{P(\text{Rambut} \text{Reptil}) \times P(\text{Lahir Ya} \text{Reptil}) \times P(\text{Reptil})}{P(\text{Hewan} = \text{Musang})}$ $P(\text{Reptil} \text{Hewan} = \text{Musang}) = 0/4 \times 1/4 \times 4/10 = 0$

Gambar 3. Teknik Penyelesaian Menggunakan *Naïve Bayes Classifier*

2.5 Testing and Evaluation (Pengujian dan Evaluasi)

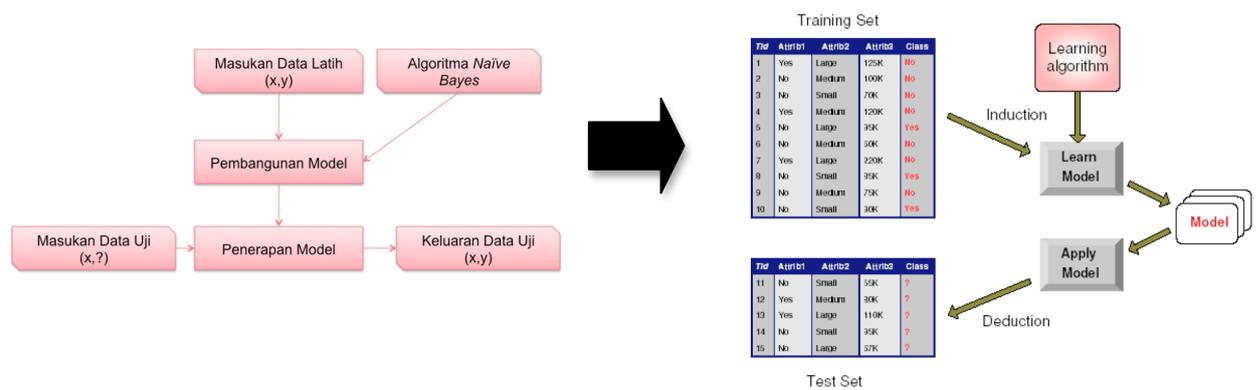
Pada tahapan pengujian ini akan dilakukan beberapa skenario percobaan untuk menentukan model manakah yang paling akurat untuk memprediksi tingkat kelulusan. Evaluasi yang digunakan adalah *accuracy* dengan skenario percobaan menggunakan teknik *holdout*, akan dibahas lebih lanjut pada hasil dan pembahasan.

2.6 Deployment (Penyebaran)

Setelah melalui proses pengujian dan evaluasi maka akan diperoleh model yang paling akurat untuk pengukuran tingkat kelulusan. Model tersebut akan digunakan untuk mengevaluasi data baru atau data uji, akan dibahas lebih lanjut pada bab 5.

3. Hasil dan Pembahasan

Proses pengerjaan klasifikasi untuk memprediksi mahasiswa lulus tepat atau lulus tidak tepat waktu dapat dilihat dari gambar 4.



Gambar 4. Proses Pengerjaan Klasifikasi Naïve Bayes

3.1 Pembangunan Model

Pembangunan model merupakan lanjutan pada bab 3 yaitu tahapan *testing* dan *evaluation* (pengujian dan evaluasi). Pada tahapan ini, data latih bersama dengan algoritma pelatihan yaitu *Naïve Bayes Classifier* akan membangun sebuah model dengan menggunakan teknik evaluasi *holdout*. *Holdout* adalah teknik evaluasi dengan membagi data latih (*training set*) dengan data uji (*testing set*) untuk mendapatkan tingkat akurasi tertinggi, berikut rincian hasil pengujiannya pada tabel 3.

Tabel 3. Hasil Pengujian Data Latih dan Data Uji dengan Teknik Evaluasi *HoldOut*

Percobaan	Data Latih : Data Uji								
	10 : 90	20 : 80	30 : 70	40 : 60	50 : 50	60 : 40	70 : 30	80 : 20	90 : 10
1	76.3	70	78.1	77.78	78.67	75	77.78	83.33	73.33
2	78.51	79.17	76.19	78.89	74.67	78.3	75.56	66.67	66.67
3	68.89	66.67	77.14	77.78	78.67	76.67	77.78	83.33	80
4	68.89	75	76.19	77.78	81.33	65	80	86.67	66.67
5	71.11	76.67	77.14	75.56	76	73.33	68.89	56.67	73.33
6	71.85	78.33	74.28	64.44	73.33	78.33	71.11	70	80
7	74.81	76.67	74.28	71.11	70.67	75	84.44	83.33	93.33
8	70.37	76.67	73.33	73.33	73.33	78.33	80	73.33	60
9	76.3	74.17	75.24	77.78	70.67	70	68.89	70	86.67
10	72.59	75.83	76.19	78.89	81.33	85	75.56	73.33	60
Total	729.62	749.18	758.08	753.34	758.67	754.96	760.01	746.66	740
Rata-Rata Akurasi	72.962	74.918	75.808	75.334	75.867	75.496	76.001	74.666	74

Dari tabel 3 pembangunan model terbaik adalah pada pembagian data latih : data uji , 70 : 30, dengan akurasi tertinggi yaitu 76.001%, dengan rincian data latih sebesar 70% dari 150 yaitu 105 data dan data uji sebesar 30% dari 150 yaitu 45 data.

3.2 Penerapan Model

Penerapan model merupakan lanjutan dari tahapan *deployment* (penyebaran). Tahapan ini akan menerapkan model terbaik yang diperoleh dari tahapan sebelumnya yaitu tahap pembangunan model. Tahap pengembangan model terbaik diperoleh 76.0001% dengan pembagian data latih : data uji terbaik adalah 70 : 30. Penerapan model terhadap data uji mahasiswa angkatan 2012 atau tingkat 3 sebanyak 61 data diperoleh hasil prediksinya yaitu sebanyak 15 mahasiswa lulus tidak tepat waktu pada tabel 4 dan 46 mahasiswa lulus tepat waktu pada tabel 5.

Tabel 4. Hasil Prediksi Mahasiswa Lulus Tidak Tepat Waktu (15 orang)

No	NIM	Nama	Program Studi	Target Kelulusan
1	1255201345	ROZA AGUSTINA	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
2	1255201347	ACHMAD RIZKI NURMANSYAH	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
3	1255201318	ERLISA	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
4	1255201308	FAHRUL AMRI	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
5	1255201292	ANTON ARDIANTORO	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
6	1255201205	ANGGA ARIFIAWAN HARTONO	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
7	1255201040	DESRIANTO	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
8	1255201283	ANANDA MUHAMMAD FITRAH	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
9	1255201195	RAFY	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
10	1255201199	IBAR MURANTO	TEKNIK INFORMATIKA	Lulus Tidak Tepat Waktu
11	1257201103	ANITA ROMALIA SINAGA	SISTEM INFORMASI	Lulus Tidak Tepat Waktu
12	1257201091	YOKI BAHARI	SISTEM INFORMASI	Lulus Tidak Tepat Waktu
13	1257201016	ERIX NANDES SAPUTRA	SISTEM INFORMASI	Lulus Tidak Tepat Waktu
14	1257201128	RAHMAT RIVALDI	SISTEM INFORMASI	Lulus Tidak Tepat Waktu
15	1257201004	AGUNG WAHYUDI	SISTEM INFORMASI	Lulus Tidak Tepat Waktu

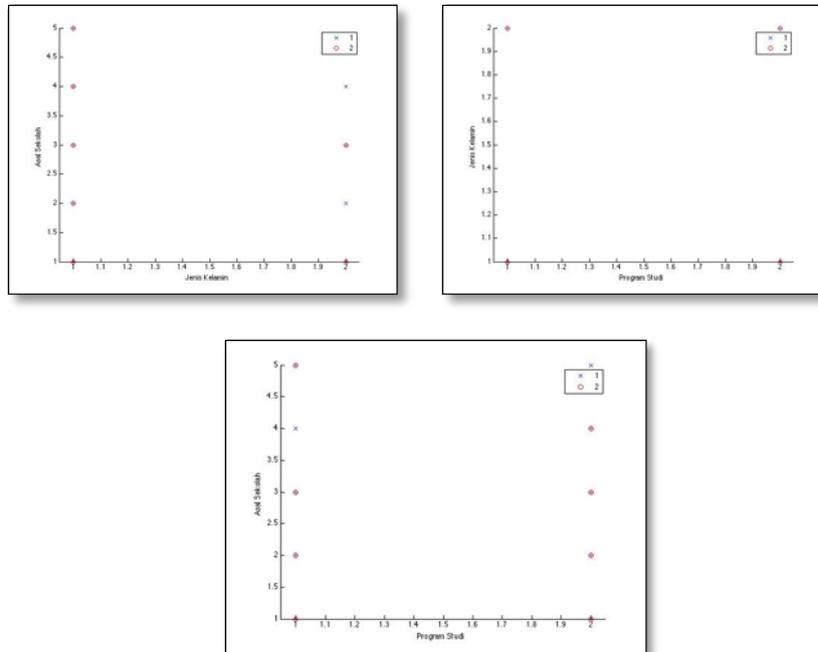
Tabel 5. Hasil Prediksi Mahasiswa Lulus Tepat Waktu (46 orang)

No	NIM	Nama	Program Studi	Target Kelulusan
1	1255201151	RISKA NESTARI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
2	1255201080	SYARIPAH VIVI ARISKA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
3	1255201180	SHOLECHAH	TEKNIK INFORMATIKA	Lulus Tepat Waktu
4	1255201203	NIA SAKINA ANGGREINI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
5	1255201233	RIZKI MISWAL FITRI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
6	1255201245	CITRA MARDENI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
7	1255201191	RAHMI KARTINI APRILIANI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
8	1255201192	IRMAWATI LESMANA DEWI DAMANIK	TEKNIK INFORMATIKA	Lulus Tepat Waktu
9	1255201229	ASTIKA WAHYUNI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
10	1255201030	ERIK LEO FERNANDO	TEKNIK INFORMATIKA	Lulus Tepat Waktu
11	1255201188	ROMMY ENDRAWAN	TEKNIK INFORMATIKA	Lulus Tepat Waktu
12	1255201181	YOGI ERSAN FADRIAL	TEKNIK INFORMATIKA	Lulus Tepat Waktu
13	1255201209	ABDUL RAHMAN	TEKNIK INFORMATIKA	Lulus Tepat Waktu
14	1255201247	DEBI RESLA AULIA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
15	1255201227	JOHAN FERNANDO	TEKNIK INFORMATIKA	Lulus Tepat Waktu
16	1255201374	EKO PUTRA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
17	1255201305	DIGDAYA ARIEF WICAKSANA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
18	1255201313	DARMAWINSYAH	TEKNIK INFORMATIKA	Lulus Tepat Waktu
19	1255201386	SUCI KURNIA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
20	1255201018	FEBY ANGGRAINI SAPUTRI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
21	1255201215	ERNI SARTIKA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
22	1255201234	ZIPORA PANJAITAN	TEKNIK INFORMATIKA	Lulus Tepat Waktu
23	1255201254	AGUNG ARIKINICO	TEKNIK INFORMATIKA	Lulus Tepat Waktu
24	1255201130	YULMI ASNI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
25	1255201084	WIDYA LESTARI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
26	1255201340	ADE ANGGREINI	TEKNIK INFORMATIKA	Lulus Tepat Waktu
27	1255201106	DANI DARMAWAN	TEKNIK INFORMATIKA	Lulus Tepat Waktu
28	1255201129	YULIAN DWI PUTRA	TEKNIK INFORMATIKA	Lulus Tepat Waktu
29	1257201009	DEBI TIAFANI	SISTEM INFORMASI	Lulus Tepat Waktu
30	1257201216	RISA AMELIA	SISTEM INFORMASI	Lulus Tepat Waktu
31	1257201119	RATH WULANDARI	SISTEM INFORMASI	Lulus Tepat Waktu
32	1257201354	RIANA AULIYA NISA'	SISTEM INFORMASI	Lulus Tepat Waktu
33	1257201081	VINY APRILA SARI	SISTEM INFORMASI	Lulus Tepat Waktu
34	1257201121	SONDANG DELVIN OISY MANULANG	SISTEM INFORMASI	Lulus Tepat Waktu
35	1257201207	APRIANI SAPUTRI	SISTEM INFORMASI	Lulus Tepat Waktu
36	1257201304	SELVIA MARTATI RIYADI	SISTEM INFORMASI	Lulus Tepat Waktu
37	1257201282	REVIYARNI	SISTEM INFORMASI	Lulus Tepat Waktu
38	1257201251	BOBY WARDANA GINTING	SISTEM INFORMASI	Lulus Tepat Waktu
39	1257201210	ARFIAN	SISTEM INFORMASI	Lulus Tepat Waktu
40	1257201358	IQBAL ROSADI	SISTEM INFORMASI	Lulus Tepat Waktu
41	1257201364	ROMI SANDY	SISTEM INFORMASI	Lulus Tepat Waktu
42	1257201088	HERU HARLIANSYAH	SISTEM INFORMASI	Lulus Tepat Waktu
43	1257201290	GITO SWIZAR	SISTEM INFORMASI	Lulus Tepat Waktu
44	1257201022	FERRY SANDRIA	SISTEM INFORMASI	Lulus Tepat Waktu
45	1257201302	AGUNG WICAKSONO	SISTEM INFORMASI	Lulus Tepat Waktu
46	1257201123	DESY SYAFITRI	SISTEM INFORMASI	Lulus Tepat Waktu

3.3 Keterkaitan antara Atribut

Pada tahapan ini akan dilakukan pembahasan mengenai keterkaitan antara atribut yang ada pada data mahasiswa dengan Target_Kelulusan yaitu :

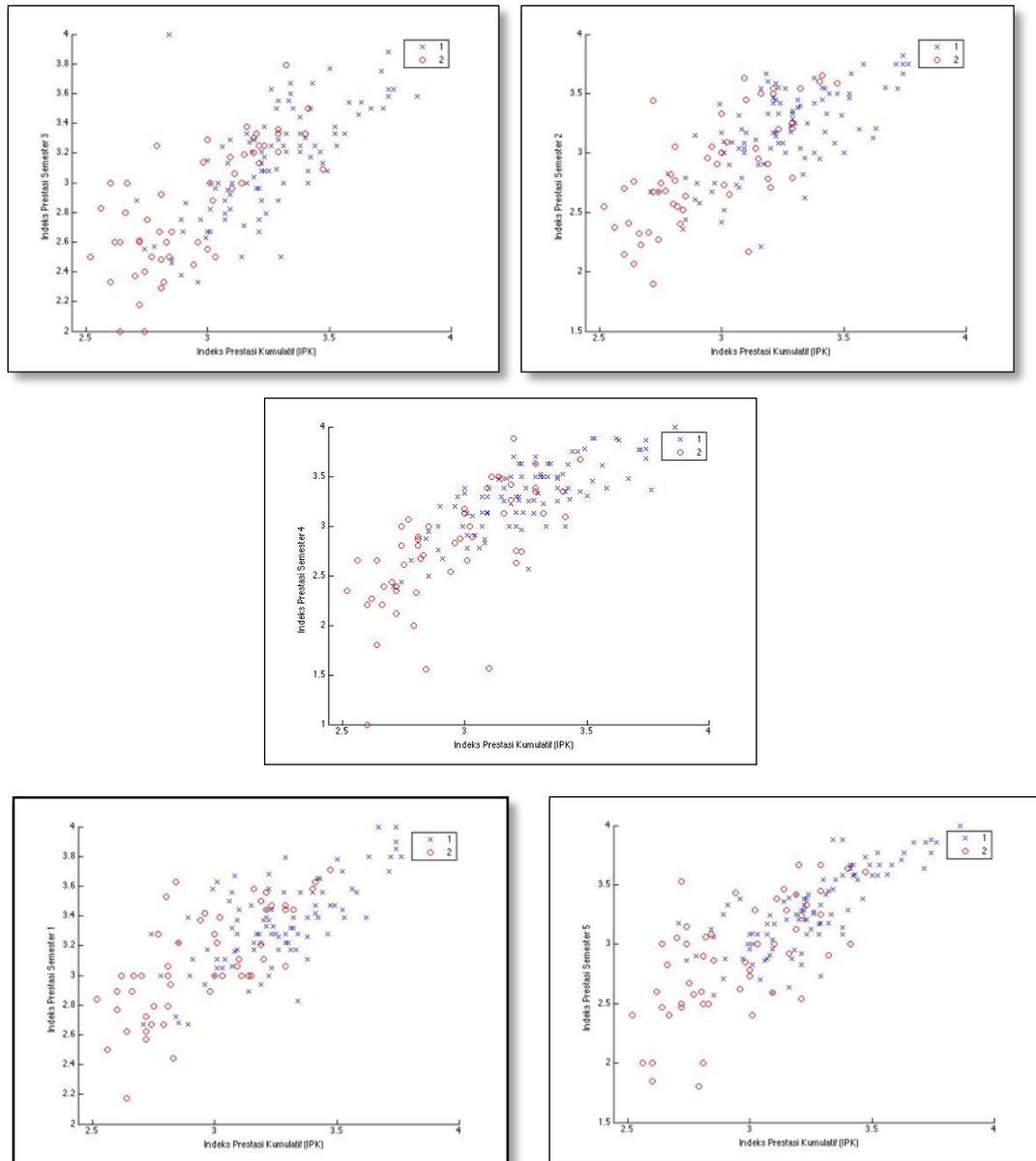
1. Atribut Jenis_Kelamin dan Asal_Sekolah; Program_Studi dan Jenis_Kelamin; Program_Studi dan Asal_Sekolah dengan Target_Kelulusan (Pola 1)



Gambar 5. Hasil Pemetaan Pola 1

Dari gambar 5 atas kiri, dapat dilihat bahwa lulusan tepat waktu adalah mahasiswa dengan jenis kelamin perempuan pada asal sekolah madrasah aliyah atau swasta. Dari gambar 5 atas kanan, dapat dilihat bahwa lulusan tepat waktu adalah seimbang antara jenis kelamin laki-laki dan perempuan. Dari gambar 5 bawah, dapat dilihat bahwa program studi teknik informatika dan sistem informasi seimbang dalam menghasilkan lulusan tepat waktu, sedangkan untuk asal sekolah kebanyakan mahasiswa lulus tepat waktu adalah berasal dari sekolah madrasah aliyah. Secara keseluruhan disimpulkan bahwa mahasiswa yang berasal dari madrasah aliyah memiliki kemungkinan terbesar untuk lulus tepat waktu, sehingga pola yang unik terbentuk disini adalah atribut Asal_Sekolah memiliki pengaruh yang cukup signifikan dalam menentukan kelulusan mahasiswa lulus tepat waktu atau lulus tidak tepat waktu.

2. Atribut Indeks Prestasi (Semester 1 – 5) dan IPK dengan Target_Kelulusan (Pola 2)



Gambar 6. Hasil Pemetaan Pola 2

Dari gambar 6, pemetaan data pola 2 menunjukkan bahwa mahasiswa yang memiliki Indeks Prestasi per semester dan Indeks Prestasi Kumulatif (IPK) di atas 3.5 memiliki kemungkinan paling besar untuk lulus tepat waktu. Sehingga pola unik yang terbentuk disini adalah Indeks Prestasi Mahasiswa per semester dan Indeks Prestasi Kumulatif Mahasiswa (IPK) memiliki peranan yang signifikan dalam menentukan mahasiswa lulus tepat waktu atau lulus tidak tepat waktu.

3.4 Pengujian Atribut Signifikan

Pada uji coba sebelumnya akurasi rata-rata tertinggi adalah 76.001% dengan atribut yang digunakan adalah Jenis_Kelamin, Asal_Sekolah, Program_Studi, Indeks_Prestasi (IP) semester 1 sampai dengan 5, Indeks_Prestasi_Kumulatif (IPK) dan Target_Kelulusan. Pada bagian keterkaitan atribut, Jenis_Kelamin dan Program_Studi tidak memiliki pengaruh yang signifikan, sehingga pada pengujian kali ini akan dilakukan pengujian hanya pada atribut Asal_sekolah,

Indeks_Prestasi (IP) semester 1 sampai dengan 5, Indeks_Prestasi_Kumulatif (IPK) dan Target_Kelulusan untuk melihat signifikansinya melalui perbandingan akurasi.

Tabel 6. Hasil Akurasi Data Latih dan Data Uji dengan Atribut Terpilih (Signifikan)

Percobaan Atribut	Data Latih : Data Uji								
	10 : 90	20 : 80	30 : 70	40 : 60	50 : 50	60 : 40	70 : 30	80 : 20	90 : 10
1	73.33	74.17	75.24	80	78.67	76.67	82.86	79.17	74.82
2	54.82	76.67	76.19	76.67	78.67	80	77.14	76.67	51.85
3	72.59	74.17	77.14	72.22	74.67	80	70.48	75.83	71.85
4	76.29	78.33	78.09	76.67	72	76.67	76.19	75.83	73.33
5	80	77.5	75.24	73.33	68	78.33	75.24	77.5	77.78
6	60	79.17	71.43	77.78	77.33	70	74.28	79.17	61.48
7	57.78	73.33	76.19	75.56	77.33	76.67	75.24	76.67	64.44
8	75.5	75.83	76.19	78.89	62.67	73.33	80	75	57.78
9	78.52	77.5	75.24	77.78	73.33	71.67	78.09	71.67	69.63
10	68.89	75.83	76.19	75.56	77.33	80	77.14	76.67	73.33
Total	697.72	762.5	757.14	764.46	740	763.34	766.66	764.18	676.29
Rata-Rata Akurasi	69.772	76.25	75.714	76.446	74	76.334	76.666	76.418	67.629

Dari tabel 3 dan tabel 6 terlihat bahwa akurasi tertinggi berada di tabel 6 sebesar 76.67%. Sehingga disimpulkan bahwa atribut terpilih (signifikan) mempengaruhi akurasi yang dihasilkan. Berikut perubahan hasil prediksi menggunakan atribut terpilih (signifikan) pada tabel 7 dan tabel 8.

Tabel 7. Hasil Prediksi Lulus Tepat Waktu dengan Atribut Terpilih (42 orang)

No	NIM	Nama	Program_Studi	Target_Kelulusan
1	1255201151	RISKA NESTARI	TEKNIK INFORMATIKA	LulusTepatWaktu
2	1255201080	SYARIPAH VIVI ARISKA	TEKNIK INFORMATIKA	LulusTepatWaktu
3	1255201180	SHOLECHAH	TEKNIK INFORMATIKA	LulusTepatWaktu
4	1255201203	NIA SAKINA ANGGREINI	TEKNIK INFORMATIKA	LulusTepatWaktu
5	1255201233	RIZKI MISWAL FITRI	TEKNIK INFORMATIKA	LulusTepatWaktu
6	1255201245	CITRA MARDENI	TEKNIK INFORMATIKA	LulusTepatWaktu
7	1255201229	ASTIKA WAHYUNI	TEKNIK INFORMATIKA	LulusTepatWaktu
8	1255201030	ERIK LEO FERNANDO	TEKNIK INFORMATIKA	LulusTepatWaktu
9	1255201188	ROMMY ENDRAWAN	TEKNIK INFORMATIKA	LulusTepatWaktu
10	1255201181	YOGI ERSAN FADRIAL	TEKNIK INFORMATIKA	LulusTepatWaktu
11	1255201209	ABDUL RAHMAN	TEKNIK INFORMATIKA	LulusTepatWaktu
12	1255201247	DEBI RESLA AULIA	TEKNIK INFORMATIKA	LulusTepatWaktu
13	1255201227	JOHAN FERNANDO	TEKNIK INFORMATIKA	LulusTepatWaktu
14	1255201374	EKO PUTRA	TEKNIK INFORMATIKA	LulusTepatWaktu
15	1255201305	DIGDAYA ARIEF WICAKSANA	TEKNIK INFORMATIKA	LulusTepatWaktu
16	1255201313	DARMAWINSYAH	TEKNIK INFORMATIKA	LulusTepatWaktu
17	1255201018	FEBY ANGGRAIN SAPUTRI	TEKNIK INFORMATIKA	LulusTepatWaktu
18	1255201234	ZIPORA PANJAITAN	TEKNIK INFORMATIKA	LulusTepatWaktu
19	1255201254	AGUNG ARIKINICO	TEKNIK INFORMATIKA	LulusTepatWaktu
20	1255201292	ANTON ARDIANTORO	TEKNIK INFORMATIKA	LulusTepatWaktu
21	1255201130	YULMI ASNI	TEKNIK INFORMATIKA	LulusTepatWaktu
22	1255201084	WIDYA LESTARI	TEKNIK INFORMATIKA	LulusTepatWaktu
23	1255201340	ADE ANGGREINI	TEKNIK INFORMATIKA	LulusTepatWaktu
24	1255201106	DANI DARMAWAN	TEKNIK INFORMATIKA	LulusTepatWaktu
25	1255201129	YULIAN DWI PUTRA	TEKNIK INFORMATIKA	LulusTepatWaktu
26	1257201216	RISA AMELIA	SISTEM INFORMASI	LulusTepatWaktu
27	1257201119	RATIH WULANDARI	SISTEM INFORMASI	LulusTepatWaktu
28	1257201081	VINY APRILA SARI	SISTEM INFORMASI	LulusTepatWaktu
29	1257201121	SONDANG DELVIN OISY MANULANG	SISTEM INFORMASI	LulusTepatWaktu
30	1257201207	APRILIANI SAPUTRI	SISTEM INFORMASI	LulusTepatWaktu
31	1257201304	SELVIA MARTATI RIYADI	SISTEM INFORMASI	LulusTepatWaktu
32	1257201282	REVIYARNI	SISTEM INFORMASI	LulusTepatWaktu
33	1257201251	BOBY WARDANA GINTING	SISTEM INFORMASI	LulusTepatWaktu
34	1257201210	ARFIAN	SISTEM INFORMASI	LulusTepatWaktu
35	1257201016	ERIX NANDES SAPUTRA	SISTEM INFORMASI	LulusTepatWaktu
36	1257201358	IQBAL ROSADI	SISTEM INFORMASI	LulusTepatWaktu
37	1257201364	ROMI SANDY	SISTEM INFORMASI	LulusTepatWaktu
38	1257201088	HERU HARLIANSYAH	SISTEM INFORMASI	LulusTepatWaktu
39	1257201290	GITO SWIZAR	SISTEM INFORMASI	LulusTepatWaktu
40	1257201022	FERRY SANDRIA	SISTEM INFORMASI	LulusTepatWaktu
41	1257201302	AGUNG WICAKSONO	SISTEM INFORMASI	LulusTepatWaktu
42	1257201123	DESY SYAFITRI	SISTEM INFORMASI	LulusTepatWaktu

Tabel 8. Hasil Prediksi Lulus Tidak Tepat Waktu dengan Atribut Terpilih (19 orang)

No	NIM	Nama	Program Studi	Target Kelulusan
1	1255201191	RAHMI KARTINI APRILIANI	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
2	1255201345	ROZA AGUSTINA	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
3	1255201192	IRMAWATI LESMANA DEWI DAMANIK	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
4	1255201347	ACHMAD RIZKI NURMANSYAH	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
5	1255201386	SUCI KURNIA	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
6	1255201215	ERNI SARTIKA	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
7	1255201318	ERLISA	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
8	1255201308	FAHRUL AMRI	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
9	1255201205	ANGGA ARIFIWAN HARTONO	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
10	1255201040	DESRIANTO	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
11	1255201283	ANANDA MUHAMMAD FITRAH	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
12	1255201195	RAFY	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
13	1255201199	IBAR MURANTO	TEKNIK INFORMATIKA	LulusTidakTepatWaktu
14	1257201009	DEBI TIAFANI	SISTEM INFORMASI	LulusTidakTepatWaktu
15	1257201103	ANITA ROMALIA SINAGA	SISTEM INFORMASI	LulusTidakTepatWaktu
16	1257201354	RIANA AULIYA NISA'	SISTEM INFORMASI	LulusTidakTepatWaktu
17	1257201091	YOKI BAHARI	SISTEM INFORMASI	LulusTidakTepatWaktu
18	1257201128	RAHMAT RIVALDI	SISTEM INFORMASI	LulusTidakTepatWaktu
19	1257201004	AGUNG WAHYUDI	SISTEM INFORMASI	LulusTidakTepatWaktu

Pada tabel 5 (46 orang) dan tabel 7 (42 orang) terlihat bahwa ada perubahan berupa pengurangan sekitar 4 orang yang diprediksi LulusTepatWaktu dan pada tabel 4 (15 orang) dan tabel 8 (19 orang) ada perubahan berupa penambahan sekitar 4 orang yang diprediksi LulusTidakTepatWaktu dengan ketepatan prediksi (akurasi) sebesar 76.67%. Hasil akurasi ini lebih tinggi dengan yang dilakukan oleh (Mujib dkk, 2013) yaitu 70%. Pada penelitian ini, data uji yang digunakan adalah tingkat 3 sedangkan (Mujib, dkk, 2013) adalah tingkat 2. Sementara atribut pada penelitian ini adalah Asal_Sekolah, Indeks_Prestasi_Semester 1 sampai dengan 5, Indeks_Prestasi_Kumulatif dan Target_Kelulusan sedangkan penelitian (Mujib dkk, 2013) menggunakan atribut Jenis_Kelamin, Indeks_Prestasi_Semester 1 sampai dengan 4, Indeks_Prestasi_Kumulatif dan Target_Kelulusan.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan diperoleh kesimpulan berikut :

1. Pembangunan model dengan menggunakan algoritma *Naïve Bayes Classifier* memiliki akurasi terbaik sebesar 76.001% kemudian setelah menggunakan atribut terpilih akurasi meningkat menjadi 76.67% dengan pembagian data latih dan data uji dengan porsi yang sama, 70 : 30.
2. Penerapan model dengan menggunakan algoritma *Naïve Bayes Classifier* menghasilkan prediksi mahasiswa yang lulus tepat waktu sebanyak 46 orang dan lulus tidak tepat waktu sebanyak 15 orang sedangkan dengan atribut terpilih diperoleh hasil prediksi mahasiswa lulus tepat waktu sebanyak 42 orang dan lulus tidak tepat waktu sebanyak 19 orang.
3. Atribut yang memiliki pengaruh yang paling signifikan untuk meningkatkan nilai akurasi adalah Asal Sekolah, Indeks Prestasi (IP) Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IP Semester 5 dan Indeks Prestasi Kumulatif (IPK).

Saran untuk pengembangan penelitian selanjutnya sebagai berikut :

1. Merancang *user interface* untuk implementasi algoritma *Naïve Bayes Classifier*
2. Mencari alternatif algoritma selain *Naïve Bayes* yang memiliki akurasi terbaik serta kombinasi atribut yang paling signifikan dalam penentuan kelas yang tepat.

Daftar Pustaka

- [1] Mujib, Hadi Suyono, M. Sarosa. Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal EECCIS*. 2013; vol 7(no 1): 59-64
 - [2] Nugroho, Sulistyoyo Yusuf. *Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Lulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta*. Prosiding Seminar Nasional Aplikasi Sains dan Teknologi (SNAST). Yogyakarta. 2014; vol 1: 1-6.
 - [3] Hastuti, Hafiizh. *Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa non Aktif*. Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (Semantik). Semarang. 2012; vol 1: 241-249.
 - [4] Larose T. Daniel. *Data Mining Methods dan Models*.. John Wiley & Sons, Inc Publication. 2006.
-