

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/316463938>

Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta

Conference Paper · April 2017

CITATIONS

0

READS

2,916

2 authors:



[Supardi Salmu](#)

Syarif Hidayatullah State Islamic University Jakarta

4 PUBLICATIONS 0 CITATIONS

[SEE PROFILE](#)



[Achmad Solichin](#)

Universitas Budi Luhur

27 PUBLICATIONS 37 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Doing a research [View project](#)



Achmatim.Net [View project](#)

Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta

Prediction of Timeliness Graduation of Students Using Naïve Bayes: A Case Study at Islamic State University Syarif Hidayatullah Jakarta

Supardi Salmu¹, Achmad Solichin²

¹Magister Ilmu Komputer, ²Prodi Teknik Informatika

Universitas Budi Luhur, Jakarta, 12260

Telp. (021) 6853753, Fax. (021) 5869225

E-mail: ¹supardi@uinjkt.ac.id, ²achmad.solichin@budiluhur.ac.id

Abstract

The academic process from learning activity until graduation at the State Islamic University (UIN) Syarif Hidayatullah Jakarta has been supported by an information system called Academic Information System (AIS). AIS is able to provide the necessary data and information related to academic activities of the students. Of the system, the data showed that the number of students entering more than the number of students who graduate. However, AIS has not been able to produce the deeper knowledge that cause this condition. One way to achieve the highest quality of the college system is to gain knowledge from the education data. So it is necessary to use data mining techniques to gain more knowledge from the data and information that is been provided by AIS. This study aims to find the factors that affect the student graduation predictions through the students academic performance data in the first semester to the fourth semester. The model used in this paper is Cross Industry Standard Process for Data Mining (CRISP-DM) and the algorithm to be implement is Naïve bayes for data classification. From the data understanding stage, has been obtained 12 predictor attributes that will be analyzed for the 1 class attribute of student graduation status. The result of accuracy is 80.72%, it can be concluded that this model can be used as a reference in predicting the student graduation. UIN can take advantage of this knowledge as a preventive action to avoid the decreasing of graduated student annually.

Keywords: Prediction, Graduate Students, Data Mining, UIN Jakarta, Naïve Bayes, CRISP-DM

Abstrak

Proses belajar mengajar hingga proses wisuda di Universitas Islam Negeri (UIN) Syarif Hidayatullah Jakarta telah didukung oleh sistem informasi yang bernama Academic Information System (AIS). AIS mampu memberikan data dan informasi yang dibutuhkan terkait kegiatan akademik mahasiswa. Melalui sistem ini ditemukan bahwa jumlah mahasiswa baru lebih banyak dibandingkan jumlah mahasiswa yang lulus. Namun AIS belum dapat menghasilkan pengetahuan (knowledge) yang lebih mendalam terkait kondisi ini. Salah satu cara untuk mencapai kualitas mutu tertinggi dari sistem perguruan tinggi adalah dengan menggali pengetahuan dari data bidang pendidikan. Sehingga diperlukan adanya pemanfaatan teknik data mining untuk menggali pengetahuan dari data mahasiswa yang tersedia di AIS. Penelitian ini bertujuan untuk menemukan faktor-faktor yang mempengaruhi prediksi kelulusan mahasiswa melalui data kinerja akademik mahasiswa pada semester satu sampai semester empat. Model yang digunakan yaitu Cross Industry Standard Process for Data Mining (CRISP-DM) dengan mengimplementasikan algoritma Naïve bayes untuk klasifikasi data. Melalui tahap data understanding didapatkan 12 atribut predictor yang akan dianalisis terhadap 1 atribut class kelulusan mahasiswa. Dengan perolehan hasil akurasi sebesar 80,72%, dapat disimpulkan bahwa model ini dapat dijadikan acuan dalam memprediksi kelulusan mahasiswa. Pengetahuan ini dapat dimanfaatkan oleh pihak UIN sebagai langkah preventif untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

Kata kunci: Prediksi, Kelulusan Mahasiswa, UIN Jakarta, Data Mining, Naïve bayes, CRISP-DM

1. PENDAHULUAN

UIN Syarif Hidayatullah Jakarta berdiri berdasarkan Surat Keputusan Presiden RI Nomor 031 tahun 2002. Sejarah pendirian UIN Syarif Hidayatullah Jakarta merupakan satu mata rantai sejarah perkembangan perguruan tinggi Islam Indonesia dalam menjawab kebutuhan pendidikan tinggi Islam modern yang telah dimulai jauh sebelum Indonesia merdeka. Pada buku pedoman akademik UIN Syarif Hidayatullah Jakarta tahun akademik 2015/2016, Program Strata 1 Reguler adalah program pendidikan akademik setelah pendidikan menengah yang memiliki beban studi sekurang-kurangnya 144 SKS dan sebanyak-banyaknya 160 SKS yang dijadwalkan untuk 8 semester dan paling lama 14 semester [1].

Berdasarkan hasil pengamatan di UIN Syarif Hidayatullah, ditemukan sebaran yang tak seimbang antara jumlah mahasiswa yang masuk dan keluar karena telah menyelesaikan masa studi. Mahasiswa masuk dalam jumlah besar, namun jumlah yang lulus tepat waktu sesuai dengan ketentuan jauh lebih kecil dibandingkan jumlah mahasiswa yang masuk ke UIN Syarif Hidayatullah Jakarta. Pada tahun 2012 jumlah mahasiswa masuk sebesar 4332 sedangkan jumlah mahasiswa lulus tepat waktu hanya sebesar 2631. Dalam jangka panjang, hal tersebut dapat menyebabkan terjadinya penumpukan jumlah mahasiswa di UIN Syarif Hidayatullah.

Jumlah mahasiswa lulus tepat waktu dapat ditingkatkan dengan meningkatkan kualitas pembelajaran dan layanan akademik untuk mahasiswa. Selain itu, jika waktu penyelesaian studi mahasiswa dapat diprediksikan maka penanganan mahasiswa akan lebih efektif. Prediksi merupakan proses keilmuan untuk memperoleh pengetahuan secara sistematis berdasarkan bukti fisik. Ilmuwan melakukan pengamatan serta membentuk hipotesis dalam usahanya untuk menjelaskan fenomena alam. Prediksi yang dibuat berdasarkan hipotesis tersebut diuji dengan melakukan eksperimen. Jika suatu hipotesis diuji berkali-kali, hipotesis tersebut dapat menjadi suatu teori ilmiah [4].

Salah satu teknik melakukan prediksi yang dapat digunakan adalah dengan teknik penggalian data atau *data mining*. Penggalian data berdasarkan data pendidikan di perguruan tinggi dapat meningkatkan kualitas pembelajaran mahasiswa di perguruan tinggi[2]. *Data mining* berkembang digunakan untuk menyelesaikan masalah menyangkut pendidikan. Beberapa penelitian terkait *data mining* digunakan sebagai upaya meningkatkan akreditasi perguruan tinggi [3].

Menurut Turban [5] dalam buku berjudul “*Decision Support System and Intellegent System*”, *data mining* diartikan sebagai suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data. *Data mining* adalah proses yang menggunakan teknik statistik, matematik, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar [6]. *Data mining* berkaitan dengan bidang-bidang ilmu lain, seperti *database system*, *data warehouse*, *statistics*, *machine learning*, *information retrieval* dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial analisis*, *image database*, *signal processing*[7]. Setiap organisasi atau perusahaan pasti bekerja dengan data, perusahaan yang sudah cukup lama tentunya juga memiliki data yang banyak. Data dalam perusahaan tersebut semakin lama akan semakin menumpuk dan disimpan tiap harinya, tiap bulannya dan tiap tahunnya.

Sementara itu, klasifikasi merupakan proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [7]. Proses klasifikasi didasarkan pada empat komponen [8], yaitu kelas, *predictor*, *training dataset* dan *testing dataset*.

Teknologi *data mining* dapat dimanfaatkan untuk menggali pengetahuan di basis data AIS UIN Syarif Hidayatullah Jakarta dalam menemukan model yang menggambarkan karakteristik kelulusan mahasiswa. Model ini selanjutnya dapat digunakan untuk memprediksi kelulusan mahasiswa guna membantu para pengambil kebijakan di UIN Syarif Hidayatullah Jakarta untuk meningkatkan kualitas universitas. Oleh karena itu, penelitian ini ditujukan untuk membuat sebuah sistem yang dapat memprediksi kelulusan mahasiswa tepat waktu dengan cara mengevaluasi kinerjanya pada tahun kedua, apakah berpeluang

menyelesaikan kuliah tepat waktu atau tidak. Algoritma *data mining* yang digunakan dalam penelitian ini adalah algoritma *Naïve Bayes*.

Algoritma *Naïve bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve bayes* merupakan pengklasifikasian metode probabilistik dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema bayes*. Teorema tersebut dikombinasikan dengan *Naïve* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naïve bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri kelas lainnya [9].

2. PENELITIAN TERKAIT

Sofi Defiyanti dengan judul penelitian “*Perbandingan: Prediksi Prestasi Belajar Mahasiswa Menggunakan Teknik Data Mining (Study Kasus Fasilkom UNSIKA)*”. Tujuan dari penelitian ini melakukan komparasi terhadap teknik *data mining* untuk memprediksi prestasi belajar mahasiswa. Hasil dari penelitian ini didapat *Naïve bayes* merupakan teknik *data mining* yang memiliki akurasi yang paling tinggi sebesar 63,56% jika dibandingkan dengan *decision tree* dan *artificial neural network* [10].

Heru Sulistiono dalam penelitiannya yang berjudul “*Kajian Penerapan Algoritma C4.5, Neural Network dan Naïve Bayes Untuk Klasifikasi Mahasiswa yang Bermasalah dalam Registrasi*”. Tujuan dari penelitian ini adalah analisis komparasi tiga algoritma klasifikasi *data mining* yaitu *C4.5*, *Neural network* dan *Naïve bayes* sehingga dapat diketahui algoritma yang paling akurat untuk memprediksi mahasiswa yang bermasalah dalam registrasi. Hasil dari penelitian ini adalah algoritma *Naïve bayes* memiliki tingkat akurasi yang paling tinggi, sehingga baik digunakan untuk klasifikasi mahasiswa yang bermasalah dalam registrasi dengan presentase 93,58% [11].

Sri Kusumadewi dalam penelitiannya yang berjudul “*Klasifikasi Status Gizi Menggunakan Naïve Bayesian Classification*”. Tujuan dari penelitian ini adalah untuk menilai status gizi seseorang, apabila ada dua yang memiliki berat badan dan tinggi badan yang sama bisa jadi memiliki status gizi yang berbeda. Hasil dari penelitian ini adalah algoritma NBC dapat menentukan status gizi seseorang menggunakan alat ukur antropometri dengan cukup baik dengan akurasi sebesar 93,2% dari 47 data yang diujikan [12].

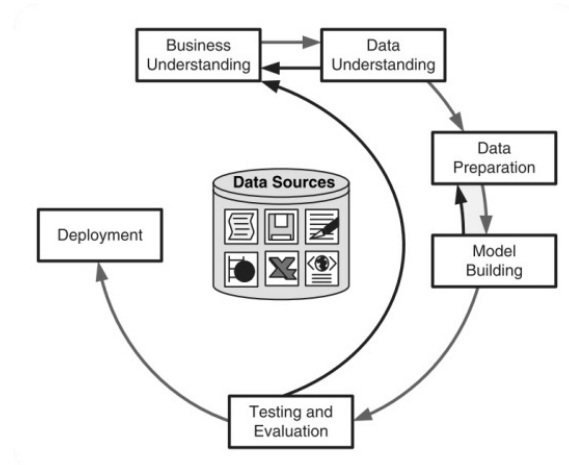
Penelitian yang telah dilakukan oleh Sulistiono dan Defiyanti menunjukkan bahwa algoritma *Naïve bayes* memiliki tingkat akurasi yang paling tinggi. Tingkat akurasi algoritma *Naïve bayes* sebesar 93,58% dibandingkan dengan algoritma *C4.5* sebesar 93,05 dan *Neural Network* sebesar 89,56% [11]. Sedangkan presentase tingkat akurasi pada penelitian yang dilakukan oleh Defiyanti untuk algoritma *Naïve bayes* sebesar 63,56%, *decision tree* sebesar 61,47% dan ANN sebesar 60,51% [10].

Oleh karena itulah penulis menggunakan metode klasifikasi *Naïve bayes* untuk melakukan penelitian ini. Metode klasifikasi *Naïve bayes* dipilih karena metode *Naïve bayes* merupakan metode probabilitas statistik yang sederhana tetapi menghasilkan hasil yang akurat.

3. RANCANGAN SISTEM

3.1. Metodologi Penelitian

Metodologi yang digunakan dalam mengembangkan *data mining* adalah CRISP-DM, CRISP-DM adalah metode yang dibentuk oleh komisi Eropa pada tahun 1996 yang menerapkan standar dalam proses *data mining*. Dalam CRISP-DM terdapat enam fase yang akan dilakukan dalam penelitian pengembangan *data mining* sesuai dengan ilustrasi pada Gambar 1.



Gambar 1: Proses CRISP-DM

1. Business Understanding

Adanya kesulitan untuk memprediksi kelulusan mahasiswa tepat waktu UIN Syarif Hidayatullah Jakarta. Untuk itu dalam penelitian ini akan dilakukan pembuatan aplikasi prediksi kelulusan mahasiswa tepat waktu UIN Syarif Hidayatullah Jakarta.

2. Data Understanding

Untuk pemilihan prediksi kelulusan mahasiswa tepat waktu, didapat data dari PUSTIPANDA UIN Syarif Hidayatullah Jakarta, yang terdiri dari 13 atribut, dimana 12 atribut *predictor* dan 1 atribut hasil. Atribut-atribut yang menjadi parameter terlihat pada tabel 1 yaitu :

Tabel 1: Atribut dan Nilai Kategori

Atribut	Nilai	Tipe
Jenis kelamin	Laki-laki, Perempuan	Diskrit
Jenis seleksi	1=SPMB Mandiri, 2=Lokal, 3=Lokal Non-Reguler, 4=PMDK, 5=Khusus, 6=Seleksi PGMI, 7=SPMB Bersama, 8=SBMPTN, 9=UM-PTKIN, 10=Kerjasama, 11=Double degree, 12=SNMPTN, 13=SPAN-PTKIN	Diskrit
Pendapatan ayah	1=0-1000000, 2=1000001-3600000, 3=3600001-7600000, 4=7600001-15000000, 5=15000001-25000000, 6=25000001-100000000	Diskrit
Pendidikan ibu	1=Tidak tamat SD/MI, 2=SD/MI, 3=SLTP=MTs, 4=SLTA/MA, 5=Diploma/SM, 6=S1, 7=S2, 8=S3	Diskrit
IP semester 1		Numerik
IP semester 2		Numerik
IP semester 3		Numerik
IP semester 4		Numerik
SKS semester 1		Numerik
SKS semester 2		Numerik
SKS semester 3		Numerik
SKS semester 4		Numerik
Status	True, False	

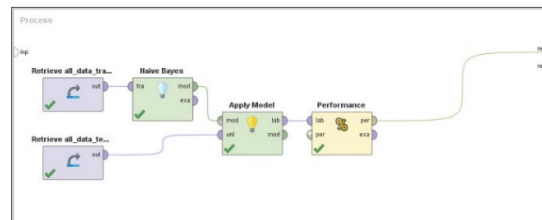
Referensi atribut Jenis kelamin, Pendapat ayah, Pendidikan Ibu, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, SKS 1, SKS 2, SKS 3 dan SKS 4 mengacu pada penelitian Khafiizh Hastuti [3]. Untuk referensi atribut Jenis Seleksi mengacu pada penelitian Dwi Untari [13]. Sedangkan atribut untuk Pendapatan ayah dan Pendidikan ibu merupakan pemikiran penulis sendiri.

3. Data Preparation

Data yang diperoleh untuk penelitian ini sebanyak 1743 *record* mahasiswa. Untuk mendapatkan data yang berkualitas, ada beberapa teknik *preprocessing* yang digunakan, yaitu *data validation* dan *data size reduction and discretization*.

4. Modeling Fase

Tahap ini juga disebut tahap *learning* karena pada tahap ini *data training* klasifikasi oleh model kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan algoritma *Naïve bayes*.



Gambar 2: Penerapan *Naïve Bayes*

5. Evaluation Fase

Pada tahap *evaluation* dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi menggunakan *precision*, *recall*, *f-measure* dan akurasi.

6. Deployment

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini juga diterapkan model yang paling akurat untuk penentuan klasifikasi prediksi kelulusan mahasiswa tepat waktu.

3.2. Teknik Pengujian

Metode *confusion matrix* merepresentasikan hasil evaluasi model dengan menggunakan tabel matriks, jika *dataset* terdiri dari dua kelas, kelas pertama dianggap positif dan kelas kedua dianggap negative. Evaluasi menggunakan *confusion matrix* menghasilkan nilai akurasi, *precision* dan *recall*.

Tabel 2: *Confusion Matrix*

Correct Classification	Classified as	
	+	-
+	True positives (tp)	False negatives (fn)
-	False positives (fp)	True negatives (tn)

Berikut adalah persamaan model *confusion matrix* :

1. *Precision* digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif, yang dihitung dengan menggunakan persamaan:

$$precision = \frac{tp}{tp+fp} \dots\dots\dots (1)$$

2. *Recall* digunakan untuk menunjukkan presentase kelas data positif yang berhasil diprediksi benar dari keseluruhan data kelas positif, yang dihitung dengan persamaan:

$$recall = \frac{tp}{tp+fn} \dots\dots\dots (2)$$

3. *F-Measure* merupakan gabungan dari *precision* dan *recall* yang digunakan untuk mengukur kemampuan algoritma dalam mengklasifikasikan kelas minoritas. Perhitungan *f-measure* menggunakan fungsi sebagai berikut:

$$F - Measure = \frac{2 \times recall \times precision}{recall + precision} \dots (3)$$

4. Akurasi adalah jumlah perbandingan data yang benar dengan jumlah keseluruhan data. Dapat dihitung dengan menggunakan persamaan:

$$akurasi = \frac{tp+tn}{tp+tn+fp+fn} \times 100\% \dots\dots\dots (4)$$

4. HASIL PEMBAHASAN

4.1. Penglompokan data

Data yang digunakan bersumber dari PUSTIPANDA UIN Syarif Hidayatullah Jakarta, Jenjang strata 1 dengan program Reguler dan tahun angkatan 2011 yang telah lulus sebanyak 1743. Untuk *datatraining* digunakan 1162 data dan *data testing* digunakan 581 data. Atribut yang menjadi parameter sebanyak 13 atribut, dimana 12 atribut *predictor* dan 1 atribut hasil.

Tabel 3: Data Training

Jenis Kelamin	Jenis Seleksi	Pendapatan Ayah	Pendidikan Ibu	IP Semester 1	IP Semester 2	IP Semester 3	IP Semester 4	SKS Semester 1	SKS Semester 2	SKS Semester 3	SKS Semester 4	Status Kelulusan
P	8	2	4	3.09	3.52	3.43	3.29	22	23	21	21	F
L	1	1	1	3.86	3.43	3.62	3.57	22	23	21	21	F
P	8	4	4	2.96	3.06	2.59	3.06	24	17	22	16	F
L	8	2	2	3.21	3.0	3.22	3.57	19	19	18	21	F
L	1	2	6	3.14	2.91	3.19	2.57	22	25	21	21	F
P	1	6	6	3.4	3.38	3.4	3.65	20	21	20	20	F
P	4	2	6	2.7	2.95	2.85	2.7	20	21	20	20	F
L	1	1	2	3.14	3.3	3.57	3.57	22	23	21	21	F
P	8	1	4	3.29	3.14	3.23	3.64	21	21	22	22	F
L	1	1	1	3.19	3.5	3.4	3.5	21	20	20	20	F

4.2. Perhitungan Algoritma Naïve Bayes

4.2.1. Prior Probabilitas

Proses untuk melakukan perhitungan kelulusan mahasiswa tepat waktu menggunakan data yang telah dipilih. Langkah pertama mencari nilai prior probabilitas untuk semua kasus yang Tepat dan Terlambat. Jumlah Tepat sebanyak 140 dan jumlah Terlambat sebanyak 1022 sehingga jumlah data yang dipakai adalah 1162 data.

$$P(Tepat) = \frac{140}{1162} = 0,12$$

$$P(Terlambat) = \frac{1022}{1162} = 0,89$$

4.2.2. Rata – rata

Tabel 4: Perhitungan rata-rata atribut

	Rata-rata	
	Tepat	Terlambat
IP Semester 1	3,45	3,28
IP Semester 2	3,45	3,30
IP Semester 3	3,46	3,31
IP Semester 4	3,49	3,33
SKS Semester 1	21,06	21,64
SKS Semester 2	21,59	21,35
SKS Semester 3	22,81	21,45
SKS Semester 4	21,76	21,19

4.2.3. Standar Deviasi

Tabel 5: Perhitungan standar deviasi atribut

	Standar Deviasi	
	Tepat	Terlambat
IP Semester 1	0,2913	0,2890
IP Semester 2	0,2793	0,2923
IP Semester 3	0,2885	0,3309
IP Semester 4	0,2984	0,3533
SKS Semester 1	1,8738	1,5227
SKS Semester 2	2,2328	1,9362
SKS Semester 3	1,4667	1,7972
SKS Semester 4	2,0619	1,9279

4.2.4. Posterior Probabilitas

Jika terdapat kasus dengan Jenis kelamin Perempuan, Jenis seleksi 1, Pendapatan ayah 1, Pendidikan ibu 4, IP semester 1 3.45, IP semester 2 3.5, IP semester 3 3.29, IP semester 4 3.65, SKS Semester 1 22, SKS Semester 2 24, SKS Semester 3 24, dan SKS Semester 4 23. Maka dapat dihitung nilai posterior probabilitas sebagai berikut.

Tabel 6: Probabilitas Posterior Data

Kasus Baru		Probabilitas	
Atribut	Nilai	Tepat	Terlambat
Jenis kelamin	Perempuan	0,7	0,66
Jenis seleksi	1	0,51	0,43
Pendapatan ayah	1	0,21	0,28
Pendidikan ibu	4	0,36	0,38
IP semester 1	3,45	1,37	1,16
IP semester 2	3,5	1,40	1,07
IP semester 3	3,29	1,15	1,20
IP semester 4	3,65	1,15	0,75
SKS semester 1	22	0,19	0,18
SKS semester 2	24	0,10	0,08
SKS semester 3	24	0,20	0,08
SKS semester 4	23	0,16	0,13

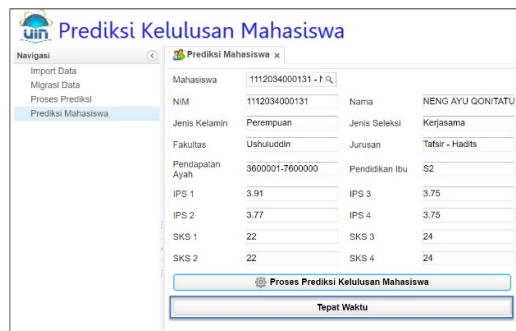
Sehingga didapat nilai sebagai berikut :

Prediksi (Tepat) = 0,53

Prediksi (Terlambat) = 0,47

Untuk menentukan data tersebut apakah termasuk Tepat atau Terlambat adalah dari nilai *posterior* yang paling besar. Pada kasus ini prediksinya adalah Tepat.

4.3. Prototipe Aplikasi



Gambar 3: Hasil Prediksi Kelulusan Mahasiswa

4.4. Evaluasi

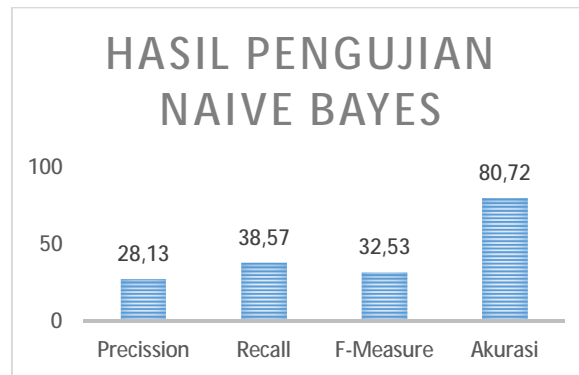
Pengujian dilakukan dengan *confusion matrix* yang terdiri dari *precision*, *recall*, *f-measure* dan akurasi dilakukan pada *dataset* yang diolah dengan menggunakan algoritma *Naïve bayes*. Pegujian *confusion*

matrix untuk *dataset* yang diolah menggunakan algoritma *Naïve bayes* untuk nilai akurasi dapat dilihat pada Tabel 7.

Tabel 7: Hasil Pengujian Naive Bayes

	True Terlambat	True Tepat	Class precision
Pred. Terlambat	442	43	91,13%
Pred. Tepat	69	27	28,12%
Class recall	86,50%	38,57%	

Dari hasil pengujian sistem yang telah dilakukan diketahui bahwa nilai akurasi yang ada menghasilkan presentase 80,72%.



Gambar 4: Presentase Hasil Pengujian Naive Bayes

5. KESIMPULAN

Dari hasil penelitian yang telah dilaksanakan, maka dapat disimpulkan bahwa:

1. Atribut yang digunakan untuk memprediksi kelulusan mahasiswa tepat waktu adalah Jenis kelamin, Jenis seleksi, Pendapatan ayah, Pendidikan ibu, IP semester 1 sampai 4 dan SKS semester 1 sampai 4.
2. Penerapan algoritma *data mining* menggunakan *Naïve bayes* dapat dilakukan untuk memprediksi kelulusan mahasiswa tepat waktu pada UIN Syarif Hidayatullah Jakarta.
3. Akurasi pengujian data yang diperoleh dalam penelitian ini sebesar 80,72% dari 1162 data yang digunakan untuk *data training* dan 587 data untuk *data testing*.

DAFTAR PUSTAKA

- [1] Tim Penyusun, “Pedoman Akademik Program Strata 1 2015/2016”, UIN Press: Jakarta, 2015.
- [2] Tair, M. Mohammed Abu dan El-halees, Alaa M., “Mining Educational Data to Improve Students’ Performance: A Case Study”, International Journal of Information and Communication Technology Research, Vol. 2, No. 2, 140-146, 2012.
- [3] Hastuti, Khafizh, “Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif”, Seminar Nasional Teknologi Informasi & Komunikasi Terapan (Semantik), 241-249, 2012.
- [4] Fajar, Ardi S, “Kajian Penerapan Model Naïve Bayes dan C4.5 dalam Prediksi Penawaran Produk pada Bank XYZ Tbk”, Universitas Budi Luhur: Jakarta, 2014.
- [5] Turban, Efraim, Aronson, Jay E. dan Liang, Ting-Peng, “Decision Support System and Intelligent System”, Prentice-Hall of India: New Delhi, 7th ed, 2007.
- [6] Kusriani dan Luthfi, E. T., “Algoritma Data Mining”, Andi Offset: Yogyakarta, 2009.

-
- [7] Han, Jiawei, Kamber, Micheline dan Pei, Jian. “*Data Mining Concepts and Techniques*”, Morgan Kaufmann, 3th ed, 2011.
 - [8] Gorunescu, Florin, “*Data Mining Concepts, Models and Techniques*”, Springer: Heidelberg, Vol. 12, 2011.
 - [9] Bustami, “*Penerapan Algoritma Naïve Bayes untuk Mengklasifikasi Data Nasabah Asuransi*”, Jurnal Informatika, hal 884-898, 2014.
 - [10] Defianti, Sofi, “*Perbandingan: Prediksi Prestasi Belajar Mahasiswa Menggunakan Teknik Data Mining (Study Kasus Fasilkom UNSIKA)*”, Konferensi Nasional Sistem Informasi, 2014.
 - [11] Sulistiono, Heru, “*Kajian Penerapan Algoritma C4.5, Neural Network dan Naïve Bayes untuk Klasifikasi Mahasiswa yang Bermasalah dalam Registrasi*”, Faktor Exacta, hal 305-315, 2015.
 - [12] Kusumadewi, Sri, “*Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification*”, Universitas Islam Indonesia: Yogyakarta, 2009.
 - [13] Untari, Dwi, “*Data Mining untuk Analisa Prediksi Mahasiswa Non-Aktif Menggunakan Metode Decision Tree C4.5*”, Universitas Dian Nuswantoro: Semarang, 2014.