FORECASTING KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES (STUDI KASUS : TEKNIK INFORMATIKA UNIROW)

Siti Maslakah*, Asfan Muqtadir*, Amaludin Arifia*

*Program Studi Teknik Informatika, Universitas PGRI Rongolawe

Correspondence Author: asfanme@gmail.com

Info Artikel :

Sejarah Artikel :

Menerima :
20 Mei 2021

Revisi :
27 Mei 2021

Diterima :
1 Juni 2021

Online :
31 Juli 2021

Keyword :

Data Mining, peramalan, ***Naïve Bayes, PHP, MySql.***

ABSTRACT (in English)

Informatics Engineering is one of the Department at the University PGRI Ronggolawe Tuban which have larger student data compared to other courses. Each year the graduation data growing much. On the other hand the owned data if not managed properly, it would just be a pile of data is not useful, so that the resulting information is also not much. Therefore, the purpose of this research is to make an application to do forecasting/prediction against a long study of graduation of students by applying techniques of data mining. Applications that are created using the programming language PHP, and MySql database. Methods used in forecasting applications this is the naïve bayes algorithm. From the results of tests performed using the data as much as 60 to 30 as the training data and 30 as test data. The accuracy of the pattern of 70 % and 30 % with the error in the right amount of data as much as 42 and that was not right 18.

.

Kata Kunci :

Data mining, Forecasting , ***Naïve Bayes, PHP, MySql.***

INTISARI (in Indonesia)

Program Studi teknik Informatika merupakan salah satu program studi di Universitas PGRI Ronggolawe Tuban yang memiliki data mahasiswa lebih besar dibandingkan dengan program studi lainnya. Setiap tahunnya data kelulusan tersebut semakin bertambah banyak. Disisi lain data yang dimiliki tersebut jika tidak dikelola dengan baik, maka hanya akan menjadi tumpukan data yang tidak bermanfaat, sehingga informasi yang dihasilkan juga tidak banyak. Oleh karena itu, tujuan penelitian ini adalah membuat suatu aplikasi untuk melakukan peramalan / prediksi terhadap lama studi kelulusan mahasiswa dengan menerapkan teknik data mining. Aplikasi yang dibuat, menggunakan bahasa pemrograman PHP, dan basis data MySql. Metode yang digunakan dalam aplikasi peramalan ini adalah algoritma naïve bayes. Dari hasil pengujian yang dilakukan menggunakan data sebanyak 30 dengan data training dan 30 sebagai data testing, akurasi polanya sebesar 70% dan errornya 60% jadi jumlah data yang tepat sebanyak 42 dan yang tidak tepat 18

1. PENDAHULUAN

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. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Berdasarkan buku panduan akademik tahun 2012, Program Studi Teknik Informatika UNIROW Tuban memiliki beban studi keseluruhan 148 (seratus empat puluh delapan) sks (satuan kredit semester) dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester dan paling lama 14 (empat belas) semester. Dari tahun 2011 sampai tahun 2016, peserta wisuda Program Sarjana (S1) di Fakultas Teknik Informatika menempuh masa studi lebih dari 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) di Fakultas Teknik Informatika yang menempuh lama studi lebih dari 8 semester dari yang dijadwalkan 8 semester. Maka dari itu manajemen dari Pihak Program Studi Teknik Informatika Unirow Tuban membutuhkan aplikasi yang dapat meramalkan atau memprediksi kelulusan mahasiswanya.

Penemuan pengetahuan dalam database (*Knowledge Discovery* in Databases/KDD), sering disebut Data Mining (Penambangan Data), mengacu pada penemuan informasi yang berguna dari kumpulan data yang besar (Goele & Chanan, 2012). Dengan memanfaatkan data mining pada data bidang pendidikan, sebuah institusi perguruan tinggi bisa memperoleh suatu informasi yang berguna, dimana selanjutnya informasi tersebut dapat menjadi suatu landasan untuk melakukan perbaikan untuk meningkatkan kualitas perguruan tinggi.

Algoritma *Naive Bayes* telah banyak digunakan untuk menyelesaikan masalah prediksi. Diantaranya digunakan untuk mencari prediksi waktu studi mahasiswa, yang menjelaskan bahwa faktor-faktor yang berpengaruh dalam klasifikasi kinerja akademik mahasiswa adalah faktor yang berkaitan dengan latar belakang sekolah sebelumnya dan data akademik pribadi saat berada di perguruan tinggi (Jananto, 2013). Sedangkan pada penelitian lain, algoritma *naïve bayes* digunakan untuk menampilkan informasi tingkat kelulusan mahasiswa, yang menjelaskan bahwa untuk mengetahui informasi tingkat kelulusan mahasiswa dapat diukur dari data master mahasiswa dan data kelulusan mahasiswa (Albab & Eviyanti, 2015).

Berdasarkan hasil dari penelitian yang telah dilakukan tersebut, maka algoritma *naïve bayes* layak untuk diimplementasikan dalam meramalkan kelulusan mahasiswa Program Studi Teknik Informatika Universitas PGRI Ronggolawe (Unirow) Tuban.

 ISSN Online : 2775-1147

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005): (1)Deskripsi; (2) Estimasi; (3)Prediksi; (4) Klasifikasi; (5)Clusterisasi; dan (5) Asosiasi.

2.2.Peramalan (*Forecasting*)

Peramalan adalah upaya mempekirakan apa yang terjadi di masa depan, berbasis pada metode ilmiah (ilmu dan teknologi) serta dilakukan secara sistematis (Santoso, 2009). Walaupun demikian, kegiatan peramalan tidaklah semata-mata berdasarkan prosedur ilmiah atau teroranisir, karena ada kegiatan peramalan yang menggunakan intuisi (perasaan) atau lewat diskusi informal dalam sebuah grup. Berikut beberapa ciri sebuah kegiatan peramalan dapat dilihat pada tabel berikut.

Tabel 2.1 Ciri Sebuah Kegiatan Peramalan (Santoso, 2009)

No	Aspek	Peramalan
1	Fokus	Data di masa lalu
2	Tujuan	Menguji perkmbangan saat ini dan relevansinya di masa mendatang
3	Metode	Proyeksi berdasar ilmu statistik, diskusi dan review program
4	Orang yang terlibat	Pembuat keputusan, petugas administrasi, praktisi, analis
5	Frekuensi	Reguler (teratur)
6	Keberhasilan	Tidak sekedar akurasi, namun bersifat pembelajaran

Dari kriteria diatas, terlihat bahwa peramalan adalah kegiatan yang bersifat teratur, berupaya memprediksi masa depan dengan menggunakan tidak hanya metode ilmiah, namun juga mempertimbangkan hal-hal yang bersifat kualitatif, seperti perasaan, pengalaman seseorang dan lainnya. Peramalan yang dibuat selalu diupayakan agar dapat meminimumkan pengaruh ketidakpastian ini terhadap perusahaan. Dengan kata lain peramalan bertujuan mendapatkan peramalan yang bisa meminimumkan kesalahan (Subagyo, 1986).

2.3.Naive Bayes

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (aturan *Bayes*) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naive Bayes* model yang digunakan adalah "model fitur independen" (Nurrohman & Nugroho, 2015).

Dalam sebuah aturan yang mudah, sebuah klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidaknya ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Untuk contohnya, buah akan dianggap sebagai sebuah apel jika berwarna merah, berbentuk bulat dan berdiameter sekitar 6 cm. Walaupun jika ciri-ciri tersebut bergantung satu sama lainnya, dalam Bayes hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apapun. Berdasarkan ciri alami dari sebuah model probabilitas, klasifikasi *Naive Bayes* bisa dibuat lebih efisien dalam bentuk pembelajaran. Dalam beberapa bentuk praktiknya, parameter untuk perhitungan model *Naïve Bayes* menggunakan metode maximum likehood, atau kemiripan tertinggi.

Prediksi *Naïve Bayes* didasarkan pada teorema Bayes dengan formula umum (Bustami, 2014) :

$$P(H|X)=\frac{P(X \mid H).P(H)}{P(X)} \quad (2.1)$$

Klasifikasi *naïve bayes* yang mengacu pada teorema *bayes* mempunyai persamaan sebagai berikut:

$$P(C_i|X)=\frac{P(X \mid Ci).P(Ci)}{P(X)} \quad (2.2)$$

Keterangan :

$P(C_i|X)$: Probabilitas akhir bersyarat suatu class berdasarkan kondisi atribut X

$P(X|C_i)$: Probabilitas munculnya X berdasarkan kondisi hipotesis dari class tertentu.

$P(C_i)$: Probabilitas munculnya suatu class tertentu

P(X) : Probabilitas munculnya nilai X

Proses dari pengklasifikasian *naïve bayes* adalah sebagai berikut (Han dkk., 2012):

1. Variabel *D* adalah kumpulan dari data dan label yang terkait dengan *class*.
   Setiap data diwakili oleh vector atribut *n*-dimensi, $X=(x_1,x_2,\ldots,x_n)$ dengan *n* dibuat dari data n atribut, berturut-turut, $A_1,A_1,\ldots,A_n$.
2. Misalkan terdapat *i class*, $C_1,C_2,\ldots,C_i$. Diberikan sebuah data *X*, kemudian pengklasifikasian akan memprediksi *X* ke dalam kelompok yang memiliki probabilitas posterior tertinggi berdasarkan kondisi *X*. Artinya pengklasifikasian *naïve bayes* memprediksi bahwa data *X* termasuk *class* C_i, jika dan hanya jika

   $$P(C_i|X) > P(C_j|X) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (2.3)$$

   Maka nilai $P(C_i|X)$ harus lebih dari nilai $P(C_j|X)$ supaya diperoleh hasil akhir $P(C_i|X)$,

3. Ketika *P(X)* konstan untuk semua class maka hanya $P(X|C_i)\ P(C_i)$ yang dihitung. Jika probabilitas class prior sebelumnya tidak diketahui, maka diasumsikan bahwa class-nya sama, yaitu $P(C_1)= P(C_2)=\cdots= P(C_m)$, untuk menghitung $P(X|C_i)$ *dan* $P(X|C_i)\ P(C_i)$. Perhatikan bahwa probabilitas *class prior* dapat diperkirakan oleh

$$P(C_i) == \frac{(|C(i,D)|)}{|D|} \quad (2.4)$$

Dimana $|C_{i,D}|$ adalah jumlah data *training* dari *class* C_i an *D* adalah jumlah total data training yang digunakan.

4. Apabila diberikan kumpulan data yang mempunyai banyak atribut, maka mengurangi perhitungan $P(X|C_i)$, *naïve bayes* mengasumsikan pembuatan *class independen* yang bersyarat. Anggap bahwa nilai-nilai stribut tersebut bersifat independen satu sama lain dan diantara atribut tidak terdapat relasi depensi, maka

$$P(X|C_i) = \prod_{k=1}^{n} P(xk|Ci) = P(x_1|C_i) \times P(x_2|C_i) \times \cdots \times P(x_n|C_i) \quad (2.5)$$

Perhitungan *P(X|Ci)* pada setiap atribut mengikuti hal-hal berikut :

a. Jika A_k adalah kategorikal, maka $P(x_k|C_i)$ adalah jumlah data dari *class* C_i di *D* yang memiliki nilai x_k untuk atribut A_k dibagi dengan $|C_{i,D}|$ yaitu jumlah data dari *class* C_i di *D*.
b. Jika A_k adalah numeric, biasanya diasumsikan memiliki distribusi Gauss dengan rata-rata μ dan standar deviasi σ, didefinisikan oleh

$$g(x,\mu,\sigma) = = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(Xi-\mu)^2}{2\sigma^2}} \quad (2.6)$$

sehingga diperoleh

$$P(x_k|C_i) = g\ (x_k, \mu_{ci,}\ \sigma_{ci}) \quad (2.7)$$

Setelah itu akan dihitung μ_{ci} *dan* σ_{ci} yang merupakan deviasi mean (rata-rata) dan standar deviasi masing-masing nilai atribut A_k untuk *training tuple class* C_i.

5. $P(X|Ci)\ P(Ci)$ dievaluasi pada setiap *class Ci* untuk memprediksi pengklasifikasian label *class* data *X* dengan menggunakan

$$P(X|C_i)\ P(Ci) > P(X|C_j)\ P(C_j) \text{ untuk } 1 \leq j \leq m,\ j \neq I \quad (2.8)$$

Label *class* untuk data *X* yang diprediksi adalah *class* Ci jika nilai $P(X|C_i)\ P(Ci)$ lebih dari nilai $P(X|C_j)\ P(C_j)$.

 ISSN Online : 2775-1147

3. METODE PENELITIAN

3.1. Penentuan Nilai Class Data

Untuk proses perhitungan, maka selanjutnya data yang ada dilakukan proses konversi kedalam bentuk yang dapat diolah dengan algoritma *naïve bayes*. Adapun proses konversi data telah diperoleh dari tahap persiapan data adalah sebagai berikut :

1. Jenis Kelamin

   Untuk jenis kelamin, dikarenakan hanya berisi dua nilai yaitu laki-laki dan perempuan maka tidak dilakukan konversi.

2. Kota Asal

   Untuk alamat mahasiswa dikelompokan hanya menjadi dua nilai yaitu alamat yang berasal dari Tuban dikonversikan menjadi ’Dalam Kota’ dan yang berasal dari luar Tuban dikonversikan menjadi ’Luar Kota’.

3. Tipe Sekolah

   Untuk tipe sekolah dilakukan pengelompokan yaitu dari sisi tipe sekolahnya. Untuk sekolah berkategori SMU atau SMA di konversikan menjadi ’Umum’ sedangkan selain SMU atau SMA dikonversikan menjadi ’Kejuruan’.

4. IPK

   Rentang IPK adalah sebagai berikut:

a. Memuaskan : IPK 2,00 – 2,75
b. Sangat memuaskan : IPK 2,76 – 3,50
c. Dengan Pujian/Cumlaude : IPK 3,51 –4,00

   Berikut konversi nilai IPK dilakukandengan membuat range pada setiap predikat nilai kelulusan

5. Lama Studi

   Program Studi Teknik Informatika memiliki 8 semester dengan beban kredit 144 – 148 sks dapat diselesaikan selama 4 tahun. Untuk mengetahui lama studi berdasarkan data yang di dapat yaitu hasil dari pengurangan tahun lulus dengan tahun masuk, dalam penulisan skripsi ini penulis membagi beberapa class untuk lama studi antara lain :

a. Tepat waktu: Lama Studi 4 Tahun
b. Tidak tepat waktu : Lama Studi lebih dari 4 Tahun

   Nilai class pada atribut lama studi dikategorikan berdasarkan semester yang ditempuh pada saat lulus, yaitu :

a. Tepat waktu, jika lama studi 8 semester atau kurang dari 8 semester.
b. Terlambat, jika lama studi lebih dari 8 semester.

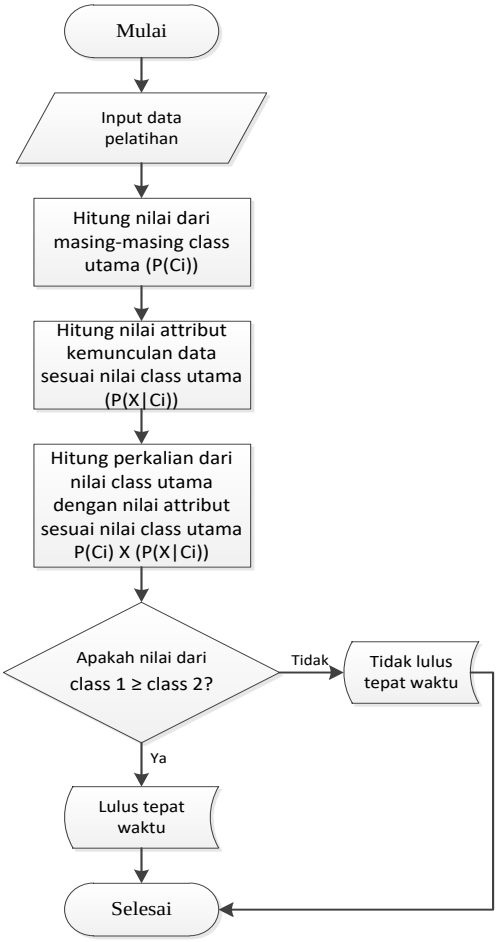
 ISSN Online : 2775-1147

Tabel 3.1 Penentuan Atribut dan Nilai *Class* Dalam Data

No	Atribut	*Class* dalam Data
1	Jenis Kelamin (X_1)	a. Laki-laki b. Perempuan
2	Alamat (X_2)	a. Dalam Kota b. Luar Kota
3	Tipe Sekolah (X_3)	a. Umum b. Kejuruan
5	IPK (X_4)	a. IPK 2,00 – 2,75 b. IPK 2,76 – 3,50 c. IPK 3,51 – 4,00
6	Lama Studi ($C_{1,2}$)	a. Lulus tepat waktu b. lulus terlambat

3.2. Alur Proses Pemrosesan Data

Untuk lebih memahami alur dari proses perhitungan data mining algoritma *naïve bayes* dapat digambarkan dengan flowchart yang dapat dilihat pada Gambar 3.1. gambar tersebut adalah alur algoritma *naïve bayes* diawali dengan menginputkan data pelatihan berupa data-data mahasiswa pada suatu sistem. Setelah data diinputkan hitung jumlah class sesuai yang telah ditentukan pada data pelatihan. Selanjutnya hitung jumlah kasus yang sama dengan kelas yang sama. Kemudian hitunglah perkalian hasil dari kemungkinan kelas terhadap kasus tadi. Setelah hasilnya diperoleh maka bandingkan nilai keduanta jika nilai kelas 1 lebih besar dari pada kelas 2 maka hasil prediksi lulus tepat waktu dan jika tidak maka tidak lulus tepat waktu dan proses selesai.



Gambar 3.1 Flowchart prediksi *Naïve Bayes*

3.3. Data Training

Algoritma *naïve bayes* digunakan untuk melakukan perhitungan terhadap probabilitas nilai *class* data dalam data pelatihan / *testing* untuk setiap variable tetap (C) berdasarkan data pengujian / *training*. Jumlah keseluruan data pelatihan / training sebanyak 110 data dari angkatan 2007-2010. Terdapat dua *Class* target yaitu mahasiswa yang lulus tepat waktu atau "lama studi= lulus tepat waktu" dan mahasiswa yang tidak lulus tepat waktu "lama studi= lulus tidak tepat waktu".

4. HASIL DAN ANALISA

4.1.Hasil

Algoritma *naïve bayes* digunakan untuk melakukan perhitungan terhadap probabilitas nilai *class* data dalam data pelatihan / *testing* untuk setiap variable tetap (C) berdasarkan data pengujian / *training*. Berdasarkan jumlah keseluruan data training sebanyak 110 data dari

angkatan 2007-2010. Terdapat dua *class* target yaitu mahasiswa yang lulus tepat waktu atau "lama studi= lulus tepat waktu" dan mahasiswa yang tidak lulus tepat waktu "lama studi= lulus tidak tepat waktu" Lihat Tabel 4.1.

Tabel 4.1. data training tahun masuk 2007

No	Mahasiswa	Jenis Kelamin	Tahun Lulus	Kota Asal	Tipe Sekolah	IPK	Lama Studi
1	A1	Perempuan	2013	Tuban	Umum	2.77	Tidak Tepat Waktu
2	A2	Perempuan	2013	Lamongan	Umum	3.06	Tidak Tepat Waktu
3	A3	Laki - Laki	2013	Lamongan	Umum	3.06	Tidak Tepat Waktu
4	A4	Laki - Laki	2013	Lamongan	Umum	2.94	Tidak Tepat Waktu
5	A5	Laki - Laki	2013	Tuban	Kejuruan	3.29	Tidak Tepat Waktu
6	A6	Laki - Laki	2013	Tuban	Umum	2.84	Tidak Tepat Waktu
7	A7	Laki - Laki	2013	Tuban	Umum	2.98	Tidak Tepat Waktu
8	A8	Laki - Laki	2013	Tuban	Umum	3.1	Tidak Tepat Waktu
9-27	……	…..	……	…..	……	…..	……..
28	A28	Laki - Laki	2011	Tuban	Kejuruan	2.75	Tepat Waktu

Tabel 4.1 adalah data yang digunakan untuk data training mahasiswa tahun masuk 2007, jumlah mahasiswa sebanyak 28 dengan presentase jenis kelamin laki-laki 75% dan perempuan sebanyak 25%, jika dilihat dari tahun kelulusan maka lulasan 2011 memperoleh 32,14%, tahun 2012 memperoleh 7,14%, dan tahun 2013 sebanyak 60,7%. Dari presentase kota asal maka dikelompokkan dengan besaran dalam kota sebesar 21,43% dan luar kota sebesar 78,6%, sedangkan variabel type (asal sekola) dari sekolah kejuruan sebesar 42,9%) dan dari umum sebesar 57,1%. Index Prestasi Mahsiswa (IPK) dikelompokkan ke dalam 3 kategori yaitu IPK 2,00 - 2,75 sebesar 14,3%, IPK 2, 76 – 3,50 sebesar 85.7% , sedangkan IPK 3,51 – 4,00 tidak ada mahasiswa ankatan 2007 yang mempunya IPK tersebut sedangkan kalkulasi untuk lulus tepat waktu sebesar 35,7% dan tidak tepat waktu sebesar 64,3%. Data training tahun masuk 2008 dapat dilihat pada Tabel 4.2.

Tabel 4.2 data training tahun masuk 2008

No	Mahasiswa	Jenis Kelamin	Tahun Lulus	Kota Asal	Tipe Sekolah	IPK	Lama Studi
1	A29	Laki - laki	2013	Tuban	Umum	2.86	Tidak tepat waktu
2	A30	Laki - laki	2013	Tuban	Umum	3.09	Tidak tepat waktu
3	A31	Perempuan	2013	Tuban	Umum	3.68	Tidak tepat waktu
4	A32	Laki - laki	2013	Tuban	Kejuruan	2.98	Tidak tepat waktu
5	A33	Laki - laki	2013	Tuban	Umum	3.45	Tidak tepat waktu
6	A34	Perempuan	2013	Tuban	Umum	3.07	Tidak tepat waktu
7	A35	Laki - laki	2013	Lamongan	Kejuruan	3.28	Tidak tepat waktu

 ISSN Online : 2775-1147

8	A36	Laki - laki	2012	Tuban	Kejuruan	2.6	Tepat waktu
9-40	……..	…..	…	……	….	..	……
41	A69	Perempuan	2013	Tuban	Umum	2.99	Tidak tepat waktu

Tabel 4.2 adalah data yang digunakan untuk data training mahasiswa tahun masuk 2008, jumlah mahasiswa sebanyak 41 dengan presentase jenis kelamin laki-laki 68,3% dan perempuan sebanyak 31,7%, tahun kelulusan maka lulasan 2012 memperoleh 12,2%, tahun 2013 memperoleh 87,8%. Dari presentase kota asal dalam kota sebesar 75,6% dan luar kota sebesar 24,4%, variabel type (asal sekolah) dari sekolah kejuruan sebesar 26,8% dan dari umum sebesar 74,2%. IPK 2,00 - 2,75 sebesar 12,2%, IPK 2, 76 – 3,50 sebesar 83% , sedangkan IPK 3,51 – 4,00 sebesar 4,88%. Dan kalkulasi untuk lulus tepat waktu sebesar 9,75% dan tidak tepat waktu sebesar 95,25%. Data training tahun masuk 2009 dapat dilihat pada Tabel 4.3.

Tabel 3.4. Data Training Tahun Masuk 2009

No	Mahasiswa	Jenis Kelamin	Tahun Lulus	Kota Asal	Tipe Sekolah	IPK	Lama Studi
1	A70	Perempuan	2014	Lamongan	Umum	3.4	Tidak tepat waktu
2	A71	Perempuan	2014	Tuban	Umum	2.5	Tidak tepat waktu
3	A72	Laki - Laki	2015	Tuban	Umum	2.5	Tidak tepat waktu
4	A73	Laki - Laki	2013	Rembang	Umum	3.53	Tepat waktu
5	A74	Laki - Laki	2015	Lamongan	Umum	2.25	Tidak tepat waktu
6	A75	Laki - Laki	2013	Rembang	Kejuruan	2.77	Tepat waktu
7	A76	Perempuan	2015	Tuban	Umum	3.2	Tidak tepat waktu
8	A77	Laki - Laki	2014	Lamongan	Umum	3.15	Tidak tepat waktu
9	A78 – A89	…..	……	…….	…	…	…
10	A90	Laki - Laki	2013	Lamongan	Umum	3.09	Tepat waktu

Pada Tabel 4.4 adalah data yang digunakan untuk data training mahasiswa tahun masuk 2009, jumlah mahasiswa sebanyak 21 dengan presentase jenis kelamin laki-laki 61,90% dan perempuan sebanyak 38,09%, tahun kelulusan maka lulasan 2012 memperoleh 9,52%, tahun 2013 memperoleh 28,57%, 2014 sebesar 23,80% , dan 2015 sebesar 38,09%. Presentase kota asal yaitu dalam kota sebesar 61,9% dan luar kota sebesar 38,09%, variabel type (asal sekolah) dari sekolah kejuruan sebesar 26,8% dan dari umum sebesar 66,66%. IPK 2,00 - 2,75 sebesar 38,09%, IPK 2, 76 – 3,50 sebesar 47,61% , sedangkan IPK 3,51 – 4,00 sebesar 14,28% sedangkan kalkulasi untuk lulus tepat waktu sebesar 33,33% dan tidak tepat waktu sebesar 66,66%. Data training tahun masuk 2010 dapat dilihat pada Tabel 4.5.

Tabel 4.4. Data Training Tahun Masuk 2010

No	Mahasiswa	Jenis Kelamin	Tahun Lulus	Kota Asal	Tipe Sekolah	IPK	Lama Studi
1	A91	Laki - Laki	2015	Tuban	Kejuruan	3.2	Tidak tepat waktu
2	A92	Perempuan	2015	Tuban	Umum	2.75	Tidak tepat waktu
3	A93	Laki - Laki	2015	Rembang	Umum	3.45	Tidak tepat waktu
4	A94	Laki - Laki	2015	Bojonegoro	Umum	3.15	Tidak tepat waktu
5	A95	Laki - Laki	2014	Lamongan	Umum	3.56	Tepat waktu
6	A96	Laki - Laki	2016	Blora	Umum	2.75	Tidak tepat waktu
7	A97	Laki - Laki	2014	Jatirogo	Umum	3.65	Tepat waktu
8	A98	Laki - Laki	2015	Tuban	Umum	3.00	Tidak tepat waktu
9	A99 – A109	….	….	….	…	….	….
10	A110	Laki - Laki	2016	Lamongan	Umum	3.45	Tidak tepat waktu

Pada Tabel 4.5 adalah data yang digunakan untuk data training mahasiswa tahun masuk 2009, jumlah mahasiswa sebanyak 20 dengan presentase jenis kelamin laki-laki 65% dan perempuan sebanyak 35%, tahun kelulusan 2014 memperoleh 25%, tahun 2015 memperoleh 55%, dan 2016 sebesar 20% , Presentase kota asal yaitu dalam kota sebesar 60% dan luar kota sebesar 40%, variabel type (asal sekolah) dari sekolah kejuruan sebesar 15% dan dari umum sebesar 65%. IPK 2,00 - 2,75 sebesar 10%, IPK 2, 76 – 3,50 sebesar 65% , sedangkan IPK 3,51 – 4,00 sebesar 25% sedangkan kalkulasi untuk lulus tepat waktu sebesar 25% dan tidak tepat waktu sebesar 75%.

Setelah melakukan pemrosesan data menggunakan Algoritma Naïve Bayes maka didapatkan hasil peramalan seperti pada Tabel 4.6.

Tabel 4.6 Hasil prediksi menggunakan 20 data testing

No	Mahasiswa	X_1	X_2	X_3	X_4	$C_{(i)}$	Keterangan Prediksi
1	A111	Perempuan	Tuban	Umum	3.00	Tidak Tepat Waktu	Benar
2	A112	Laki - Laki	Plumpang	Kejuruan	3.06	Tepat Waktu	Salah
3	A113	Laki - Laki	Semanding	Kejuruan	3.06	Tidak Tepat Waktu	Benar
4	A114	Laki - Laki	Tuban	Kejuruan	2.94	Tidak Tepat Waktu	Benar
5	A115	Perempuan	Rembang	Umum	3.29	Tepat Waktu	Salah
6	A116	Laki - Laki	Tuban	Kejuruan	2.84	Tepat Waktu	Salah
7	A117	Perempuan	Tuban	Umum	2.98	Tidak Tepat Waktu	Salah
8	A118	Perempuan	Rengel	Umum	3.01	Tidak Tepat Waktu	Salah
9	A120	Perempuan	Tuban	Umum	3.09	Tidak Tepat Waktu	Benar
10	A121	Perempuan	Tuban	Umum	2.99	Tidak Tepat Waktu	Benar
11	A122	Perempuan	Bojonegoro	Umum	3.20	Tidak Tepat Waktu	Benar

 ISSN Online : 2775-1147

12	A123	Laki - Laki	Rembang	Kejuruan	3.45	Tidak Tepat Waktu	Benar
13	A124	Laki - Laki	Lamongan	Umum	3.85	Tepat Waktu	Benar
14	A125	Laki - Laki	Tuban	Umum	3.08	Tidak Tepat Waktu	Benar
15	A126	Perempuan	Mojokerto	Umum	3.54	Tepat Waktu	Salah
16	A127	Perempuan	Lamongan	Umum	2.55	Tepat Waktu	Salah
17	A128	Laki - Laki	Tuban	Kejuruan	3.15	Tidak Tepat Waktu	Benar
18	A129	Perempuan	Tuban	Umum	2.45	Tepat Waktu	Benar
19	A130	Laki - Laki	Tuban	Kejuruan	3.5	Tidak Tepat Waktu	Salah
20	A131	Perempuan	Rengel	Umum	3.2	Tidak Tepat Waktu	Benar

4.2. Analisa

Berdasarkan Tabel 4.6. didapatkan Pada data diatas P(Ci) merupakan data target,kemudian akan ditentukan *class* dan atribut yang digunakan dengan ketentuan :

C_1 = (Lama studi =”tepat waktu”)

C_2 = (Lama studi =”tidak tepat waktu”)

X_1= (Jenis kelamin =”perempuan”)

X_2 = (Kota asal =”tuban”)

X_3 = (Tipe Sekolah =”umum”)

X_4 = (Ipk =”3.30”)

Sesuai dengan persamaan persamaan (2.1) dan (2.2) telah diketahui nilai dari class utama dan sekarang hitung nilai atribut kemunculan data berdasarkan *class* utama

$P(X_1| C_1)$ = P(Jenis kelamin =”perempuan”) | C_1 = (Lama studi =”tepat waktu”)

$= \frac{4}{26} = 0.154$

$P(X_1 |C_2)$ = P(Jenis kelamin =”perempuan”) | C_2 = (Lama studi = “tidak tepat waktu”)

$= \frac{31}{84} = 0.369$

$P(X_2| C_1)$ = P(asal kota =”dalam kota”) | C_1 = (Lama studi =”tepat waktu”)

$= \frac{15}{26} = 0.577$

$P(X_2 |C_2)$ = P(asal kota =”dalam kota”) | C_2 = (Lama studi = “tidak tepat waktu”)

$= \frac{57}{84} = 0.679$

$P(X_3| C_1)$ = P(tipe sekolah =”umum”) | C_1 = (Lama studi =”tepat waktu”)

$= \frac{15}{26} = 0.577$

$P(X_3 |C_2)$ = P(tipe sekolah =”umum”) | C_2 = (Lama studi = “tidak tepat waktu”)

$= \frac{62}{84} = 0.739$

$P(X_4| C_1)$ = P(ipk =”3.30”) | C_1 = (Lama studi =”tepat waktu”)

$= \frac{11}{26} = 0.423$

$P(X_4 |C_2)$ = P(ipk =”3.30”) | C_2 = (Lama studi = “tidak tepat waktu”)

$= \frac{68}{84} = 0.810$

1. Hitung nilai P(X|Ci) untuk *i*=1=tepat waktu, *i*=2= tidak tepat waktu, maka digunakan rumus (2.5) yaitu:

   $P(X|C_i) = \prod_{k=1}^{n} P(xk|Ci)$

   $= P(x_1|C_i) \times P(x_2|C_i) \times \cdots \times P(x_n|C_i)$

   Perhitungan untuk *i* =1 bernilai “tepat waktu”

   P(X|lama studi =”tepat waktu”) = P(Jenis kelamin =”perempuan”) x
   P(asal kota =”dalam kota”) x
   P(tipe sekolah =”umum”) x
   P(ipk =”3.30”)
   = 0.154 x 0.577 x 0.577 x 0.423
   = 0.0217

   P(X|lama studi =”tidak tepat waktu”) = P(Jenis kelamin =”perempuan”) x
   P(asal kota =”dalam kota”) x
   P(tipe sekolah =”umum”) x
   P(ipk =”3.30”)
   = 0.369 x 0.679 x 0.738 x 0.81
   = 0.1497

2. Selanjutnya menghitung $P(X|C_i)$ $P(C_i)$ yaitu:

   P(X|lama studi =”tepat waktu”) P(lama studi=” tepat waktu)
   = 0.0217 x 0.236
   = 0.005
   P(X|lama studi =”tidak tepat waktu”) P(lama studi=” tidak tepat waktu)
   = 0.1497 x 0.764
   = 0.114

3. Jika dilihat dari nilai yang diperoleh Pada perhitungan diatas, diketahui bahwa nilai lama studi =” tidak tepat waktu” lebih besar dari nilai lama studi =” tepat waktu

Dari data pada Tabel 4.6, maka presentasi kesalahan dalam prediksi adalah

Jumlah data : 20

Jumlah data dengan keterangan “Benar” : 12

$$Persentase\ Akurasi = \frac{Jumlah\ data\ dengan\ keterangan\ "Benar"}{Jumlah\ Data} \times 100\% = \frac{12}{20} \times 100\% = 60\%$$

Jumlah data dengan keterangan “Salah” : 8

$$Persentase\ error = \frac{Jumlah\ data\ dengan\ keterangan\ "Salah"}{Jumlah\ Data} \times 100\% = \frac{8}{20} \times 100\% = 40\%$$

Pada pengujian acak 20 alumni Program studi Teknik Infomatika maka didapatkan nilai presentase akurasi hasil prediksi kelulusan sebesar 60% dan presentase *error* sebesar 40%.

5. KESIMPULAN

5.1. Kesimpulan

Berdasarkan uraian, implementasi, dan pengujian sistem pada bab-bab sebelumnya, beberapa kesimpulan yang dapat diambil dari penelitian skripsi ini adalah sebagai berikut:

1. Variabel penentu yang digunakan dalam penelitian ini adalah jenis kelamin, status pernikhan, pekerjaan, kota asal, tipe sekolah asal, dan ipk.
2. Tingkat kesalahan dari algoritma *naïve bayes* yang digunakan untuk peramalan kelulusan mahasiswa berkisar pada 40% sedangkan tingkat kecocokan sebesar 60%.

5.2. Saran

Beberapa saran untuk pengembangan lebih lanjut terhadap aplikasi peramalan kelulusan mahasiswa menggunakan algoritma *naïve bayes* dapat dijabarkan sebagai berikut:

1. Peneliti telah membahas penggunaan algoritma *naïve bayes* dalam penelitian skripsi peramalan kelulusan mahasiswa, diharapkan dalam penelitian selanjutnya dapat menggunakan metode yang berbeda atau mengkombinasikan algoritma *naïve bayes* dengan metode lain.
2. Sebaiknya jumlah data yang digunakan dalan data training maupun data testing ditambah hingga dapat hasil akurasi fungsi algoritma yang lebih baik.
3. Variabel dan jangkauan data yang digunakan dalam proses peramalan menggunakan algoritma *naïve bayes* sebaiknya ditambah lagi, sehingga dapat dicapai tingkat keakuratan hasil peramalan yang lebih tinggi.

DAFTAR PUSTAKA

Albab, U dan Eviyanti, A. 2015. *Aplikasi Datamining Untuk Menampilkan Informasi Tingkat Kelulusam Mahasiswa*. Jurnal Ilmiah Mahasiswa Universitas Muhammadiyah Sidoarjo.

Bustami. 2014. *Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*. Jurnal Ilmiah Mahasiswa Universitas Malikussaleh Aceh No.1 Vo. 8.

Goele, S. dan Chanana N. 2012. *Data Mining Trend in Past, Current and Future*. International Journal of Computing & Business Research.

Han, J., Kamber,M., dan Pei, J., 2012. *Data Mining*: *Concepts and Techniques Third Edition*. Elsevier Inc.

Jananto, A., 2013. *Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa*. Jurnal Ilmiah Mahasiswa Universitas Stikubank No.1 Vol. 18.

Larose, D. T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey and Sons, Inc New York.

Nurrohman, N dan Nugroho, Y., 2015. *Aplikasi Pemprediksi Masa Studi dan Predikat Kelulusan Mahasiswa Menggunakan Metode Naive Bayes*. Universitas Muhammadiyah Surakarta.