

# **DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI KELULUSAN MAHASISWA UNIVERSITAS DIAN NUSWANTORO**

Oleh: Yuda Septian Nugroho  
Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro  
Email: [yudaseptian92@gmail.com](mailto:yudaseptian92@gmail.com)

## **ABSTRAK**

Data mahasiswa dan data kelulusan mahasiswa Universitas Dian Nuswantoro menghasilkan data yang sangat berlimpah berupa data profil mahasiswa dan data akademik. Hal tersebut terjadi secara berulang dan menimbulkan penumpukan terhadap data mahasiswa sehingga mempengaruhi pencarian informasi terhadap data tersebut. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap data mahasiswa Universitas Dian Nuswantoro Fakultas Ilmu Komputer angkatan 2009 berjenjang DIII dan S1 dengan memanfaatkan proses *data mining* dengan menggunakan teknik klasifikasi. Metode yang digunakan adalah *CRISP-DM* dengan melalui proses *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Algoritma yang digunakan untuk klasifikasi kelulusan adalah algoritma *Naïve Bayes*. *Naïve Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Implementasi menggunakan *RapidMiner* 5.3 digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan adalah NIM, Nama, Jenjang, Progd, Provinsi Asal, Jenis Kelamin, SKS, IPK, dan Tahun Lulus. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan kebijakan oleh pihak Fasilkom.

Kata kunci : NIM, SKS, IPK, Tahun Lulus, *Naïve Bayes Classifier*, Kelulusan Mahasiswa

## **ABSTRACT**

Student data and data Dian Nuswantoro University student graduation produce data that is very abundant in the form of student profile data and academic data. This happens repeatedly and cause a build up of the student data that affect information retrieval to the data. This study aims to perform the classification of student data Dian Nuswantoro University of Computer Science faculty class of 2009 tiered Diploma and S1 by using data mining process using classification techniques. The method used is the *CRISP-DM* with a through understanding of business processes, understanding data, the data preparation, modeling, evaluation and deployment. The algorithm used for graduation classification is *Naïve Bayes* algorithm. *Naïve Bayes* is a simple probabilistic based prediction technique on the application of Bayes theorem or rule with a strong independence assumption on feature, meaning that a feature is not data relating to the presence or absence of other features in the same data. Implementation using *RapidMiner* 5.3 is used to help find an accurate value. Attributes used is NIM, Name, Qualification, courses, Province of Origin, Gender, credits, GPA, and Graduation Year. The results of this study are used as one basis for determining policy decisions by the computer science faculty

Keywords: NIM, credits, GPA, Graduation Year, *Naïve Bayes Classifier*, Graduation Students

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai siswa dan proses pembelajaran yang dihasilkan.

Pada institusi pendidikan perguruan tinggi, data mahasiswa dan data jumlah kelulusan mahasiswa dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan setiap tahunnya, profil, dan hasil akademik mahasiswa selama menempuh proses kegiatan belajar mengajar di perguruan tinggi. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi.

Berdasarkan berlimpahnya data mahasiswa dan data jumlah kelulusan mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data mahasiswa sehingga berguna bagi pihak universitas[1]. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (knowledge Discovery), misalnya informasi mengenai pengklasifikasian data mahasiswa berdasarkan profil dan data akademik. Pengetahuan baru tersebut dapat membantu pihak universitas untuk melakukan klasifikasi mengenai tingkat kelulusan mahasiswa guna menentukan strategi untuk meningkatkan kelulusan pada tahun - tahun berikutnya.

Berdasarkan data yang diperoleh dari UPT Data dan Informasi (PSI) UDINUS, pada tahun 2011 total jumlah kelulusan mahasiswa Fakultas Ilmu Komputer berjumlah 1115 mahasiswa, pada tahun 2012 berjumlah 612 sedangkan pada tahun 2013 berjumlah 829 mahasiswa. Pada tahun 2011 hingga 2012 total jumlah kelulusan mahasiswa Fakultas Ilmu Komputer mengalami penurunan berjumlah 503 mahasiswa. Pada tahun 2012 hingga 2013 total jumlah kelulusan mahasiswa Fakultas Ilmu Komputer mengalami peningkatan sejumlah 217 mahasiswa.

Total jumlah kelulusan mahasiswa pada tahun 2011, 2012, 2013 yang jumlah kelulusannya fluktuatif dijadikan sebagai dasar acuan dilakukannya proses klasifikasi, maka

penulis akan melakukan penelitian mengenai klasifikasi kelulusan mahasiswa yaitu dengan menggunakan data mahasiswa Fakultas Ilmu Komputer angkatan 2009. Penelitian ini akan melakukan analisis data secara ilmiah dengan menggunakan metode klasifikasi kelulusan mahasiswa Fakultas Ilmu Komputer angkatan 2009. Jika hasil klasifikasi kelulusan mahasiswa angkatan 2009 menunjukkan tingkat peningkatan maupun penurunan, maka hasil klasifikasi tersebut dapat dijadikan sebagai salah satu bahan evaluasi dalam menentukan kebijakan pihak FASILKOM dengan menggunakan teknik data mining.

Penelitian ini akan melakukan pengklasifikasian berdasarkan data mahasiswa UDINUS Fakultas Ilmu Komputer angkatan 2009. Atribut yang akan digunakan dalam melakukan klasifikasi kelulusan angkatan 2009 adalah Nomor Induk Mahasiswa (NIM), nama, jenjang, program studi, nama provinsi, jenis kelamin, SKS yang telah ditempuh, IPK, dan tahun kelulusan.

### 1.2 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah:

- Mengklasifikasi kelulusan mahasiswa Udinus Fakultas Ilmu Komputer Angkatan 2009.
- Mendapatkan akurasi yang tepat untuk melakukan klasifikasi kelulusan mahasiswa UDINUS Fakultas Ilmu Komputer angkatan tahun 2009 dengan menggunakan metode klasifikasi naïve bayes.

## 2 LANDASAN TEORI

### 2.1 Kelulusan Studi

Mahasiswa yang memenuhi persyaratan kelulusan ditetapkan dalam yudisium kelulusan Fakultas / Program Studi dan ditetapkan dengan keputusan Rektor. Tanggal kelulusan ditetapkan berdasarkan tanggal yudisium kelulusan dan merupakan tanggal penetapan IPK akhir (transkrip nilai). Berikut syarat kelulusan mahasiswa UDINUS pada Fakultas Ilmu Komputer[2].

### 2.2 Data Mining

*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi

informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [10].

### 2.3 Pengelompokan Data Mining

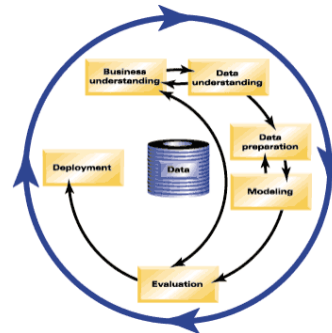
Menurut Larose, *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat di lakukan, yaitu [11]:

- a. Deskripsi  
Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
- b. Estimasi  
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori.
- c. Prediksi  
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
- d. Klasifikasi  
Dalam klasifikasi, terdapat target variabel kategori.
- e. Pengklusteran  
*Clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*).
- f. Asosiasi  
Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

### 2.4 Metode Data Mining

Menurut Larose, data mining memiliki enam fase CRISP-DM ( *Cross Industry Standard Process for Data Mining*) [11].

- a. Fase Pemahaman Bisnis ( *Business Understanding Phase* )
- b. Fase Pemahaman Data ( *Data Understanding Phase* )
- c. Fase Pengolahan Data ( *Data Preparation Phase* )
- d. Fase Pemodelan ( *Modeling Phase* )
- e. Fase Evaluasi ( *Evaluation Phase* )
- f. Fase Penyebaran ( *Deployment Phase* )



Gambar 2.1 Proses CRISP-DM [11]

### 2.5 Algoritma K-Means

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [5].  
Prediksi Bayes didasarkan pada formula teorema Bayes dengan formula umum sebagai berikut :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)}$$

### 2.6 RapidMiner

*RapidMiner* adalah sebuah lingkungan *machine learning data mining*, *text mining* dan *predictive analytics* [13].

## 3 HASIL DAN PEMBAHASAN

### 3.1 Pemahaman Bisnis (Business Understanding)

Pemahaman bisnis (*business understanding*), tahap pertama dalam proses CRISP-DM yang juga dapat disebut sebagai tahap pemahaman bisnis (penelitian).

#### 3.1.1 Menentukan Tujuan Bisnis

Tujuan bisnis berdasarkan pengolahan data mahasiswa angkatan 2009 antara lain untuk meningkatkan jumlah mahasiswa pada tahun-tahun berikutnya. Pengolahan data dilakukan karena telah faktor kelulusan yang fluktuatif dari kelulusan tahun 2011 hingga 2013. Tujuan pengolahan data dapat dijadikan sebagai salah satu dasar pengambilan keputusan dalam menentukan kebijakan oleh pihak fakultas.

### 3.1.2 Melakukan Penilaian Situasi

Pada tahapan ini dibutuhkan pemahaman terhadap tujuan bisnis dan menerjemahkan ke dalam tujuan *data mining*.

### 3.1.3 Menentukan strategi awal *data mining*

Strategi awal dalam menerapkan tujuan dilakukannya *data mining* adalah melakukan permintaan data mahasiswa terlebih dahulu ke bagian UPT Data dan Informasi (PSI) UDINUS.

## 3.2 Pemahaman Data (*Data Understanding*)

*Dataset* mahasiswa yang didapatkan dari UPT Data dan Informasi (PSI) UDINUS berupa dokumen *excel* sejumlah 1477 *record*.

### 3.2.1 Pengumpulan data awal

Adapun sumber data utama yang digunakan dalam penelitian ini adalah *dataset* mahasiswa UDINUS yang berjenjang DIII dan S1 pada tahun 2009 dan data jumlah kelulusan tahun 2011, 2012, 2013 dengan format *.xlsx* karena data yang diberikan berupa dokumen *excel*.

### 3.2.2 Mendeskripsikan data

*Dataset* mahasiswa terdiri dari beberapa atribut antara lain NIM, nama, jenjang, program studi, alamat, kota asal, nama provinsi, jenis kelamin, SKS yang telah ditempuh, IPK, dan tahun kelulusan.

### 3.2.3 Evaluasi kualitas data

Hasil evaluasi terhadap kualitas data yaitu menemukan banyak nilai kosong / *null* yang disebut dengan *missing value* pada atribut dalam *dataset* mahasiswa.

### 3.2.4 Pemilihan Atribut

Atribut yang digunakan adalah NIM, nama, jenjang, program studi, nama provinsi, jenis kelamin, SKS yang telah ditempuh, IPK dengan standar  $\geq 2,00$  dan tahun kelulusan.

## 3.3 Persiapan Data (*Data Preparation*)

Persiapan data mencakup semua kegiatan untuk membangun *dataset* mahasiswa yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa *dataset* mahasiswa dan selanjutnya akan melakukan proses *data mining*.

### 3.3.1 Seleksi data

Atribut yang digunakan adalah NIM, nama, jenjang, program studi, nama provinsi, jenis kelamin, SKS yang telah ditempuh, IPK dengan standar  $\geq 2,00$  dan tahun kelulusan.

### 3.3.2 Pengolahan data mentah (*Preprocessing Data*)

Pada tahap ini merupakan tahap untuk memastikan data mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan.

### 3.3.3 Transformasi data

Data yang berjenis numerikal seperti tahun lulus harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk nominal. Untuk melakukan inialisasi tahun lulus dapat dilakukan dengan:

1. Mahasiswa angkatan 2009 berjenjang Sarjana (S1) dengan tahun kelulusan 2012 dan 2013 dinyatakan lulus tepat waktu, dan diberi inisial "YES".
2. Mahasiswa angkatan 2009 berjenjang Sarjana (S1) belum terdapat tahun kelulusan dinyatakan tidak lulus tepat waktu, dan diberi inisial "NO".
3. Mahasiswa angkatan 2009 berjenjang Diploma III (D3) dengan tahun kelulusan 2011 dan 2012 dinyatakan lulus tepat waktu, dan diberi inisial "YES".
4. Mahasiswa angkatan 2009 berjenjang Diploma III (D3) belum terdapat tahun kelulusan dinyatakan tidak lulus tepat waktu, dan diberi inisial "NO".

Tabel 3.1 Inialisasi data tahun lulus

	A	B
1	Jenjang	Tahun Lulus
2	S1	YES
3	S1	YES
4	S1	YES
5	S1	YES
6	S1	YES
7	S1	YES
8	S1	YES
9	S1	NO
10	S1	NO
11	S1	YES
12	S1	YES
13	S1	NO
14	S1	NO
15	D3	YES
16	D3	NO
17	D3	YES
18	D3	YES
19	D3	YES
20	D3	NO
21	D3	YES
22	D3	NO
23	D3	NO
24	D3	YES
25	D3	YES

Berikut adalah contoh dataset mahasiswa yang telah dilakukan inialisasi berdasarkan jenjang dan tahun kelulusan yang terdapat pada Tabel 3.2.

Tabel 3.2 Contoh *dataset* mahasiswa yang telah dilakukan inialisasi

#	A	B	C	D	E	F	G	H	I
1	NIM	Nama	Jenjang	Program Studi	Provinsi	Sex	Sks	IPK	Tahun Lulus
2	A11.200 9.04635	HENDRA ANDHYK A	S1	Teknik Informatika	JAWA TENGAH	L	146	3.29	YES
3	A11.200 9.04636	YONATH ANGANI PANJAIT AN	S1	Teknik Informatika	JAWA TENGAH	L	150	2.87	YES
4	A11.200 9.04637	ABEDNE GO JAYA SETIAWA N	S1	Teknik Informatika	JAWA TENGAH	L	148	3.36	YES
5	A11.200 9.04640	BAGUS KURNIA WAN	S1	Teknik Informatika	JAWA TENGAH	L	146	2.96	YES
6	A11.200 9.04642	ARIE KURNIA WAN	S1	Teknik Informatika	JAWA TENGAH	L	146	3.01	YES
7	A11.200 9.04643	ALFON INDRA WUJAYA	S1	Teknik Informatika	JAWA TENGAH	L	146	3.18	YES
8	A11.200 9.04644	R. M. ALEXAND ER ELLDO SEPTIAN	S1	Teknik Informatika	JAWA TENGAH	L	148	3.11	YES
9	A11.200 9.04645	VINSENSI US RINDA RESI	S1	Teknik Informatika	JAWA TENGAH	L	140	2.87	NO
10	A11.200 9.04647	MUHAM MAD PARSTA NURISNA INI	S1	Teknik Informatika	JAWA TENGAH	L	134	2.25	NO

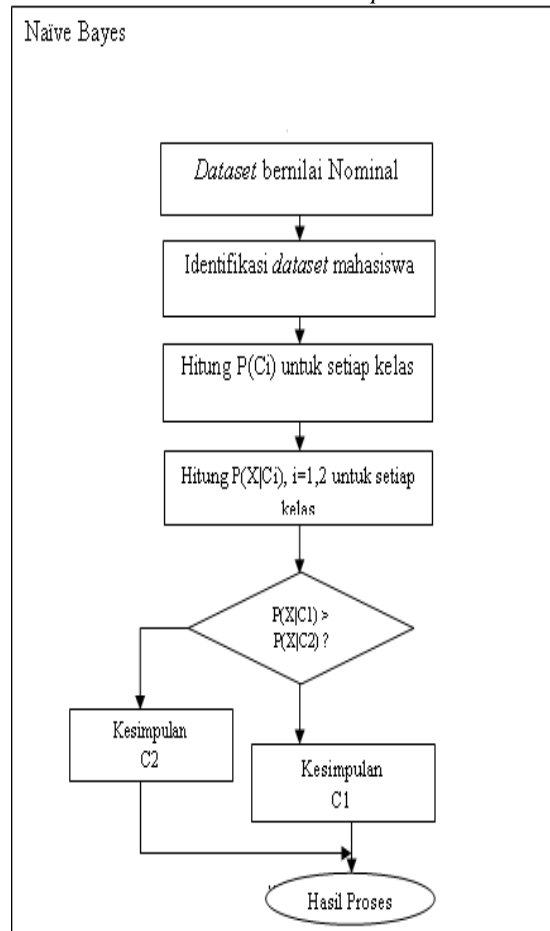
### 3.4 Pemodelan (*Modeling*)

Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan pemilihan teknik *data mining* dan menentukan algoritma yang akan digunakan.

### 3.4.1 Pemilihan teknik pemodelan

*Tool* yang digunakan adalah *RapidMiner* versi 5.3:

Gambar 3.1. Flowcart *RapidMiner*



Dari data atribut nominal, kemudian identifikasi dataset mahasiswa. Hitung  $P(C_i)$  untuk setiap atribut, dalam kasus dataset pada penelitian ini yaitu atribut tahun kelulusan yang terdiri dari 2 kelas yaitu kelas lulus tepat waktu dinyatakan “YES” dan tidak lulus tepat waktu dinyatakan “NO”.

Kemudian hitung  $P(X|C_i)$ ,  $i=1,2$  untuk setiap kelas atau atribut. Setelah itu bandingkan, jika  $P(X|C_1) > P(X|C_2)$  maka kesimpulannya C1 adalah kelas lulus tepat waktu = “YES”. Jika  $P(X|C_1) < P(X|C_2)$  maka kesimpulannya C2 tidak lulus tepat waktu = “NO”.

### 3.4.2 Perhitungan Data Mining

Berikut perhitungan manual naïve bayes dengan menggunakan dataset pada tabel

3.3 jika data training dan data testing di pilih secara acak :

Tabel 3.3 Penjelasan data training dan data testing

A	B	C	D	E	F	G	H	I	J
No	NIM	Nama	Jenjang	Program Studi	Provinsi	Sex	SKS	IPK	Tahun Lulus
2	A11.2009.04635	HENDRA ANDHYKA	S1	Teknik Informatika	JAWA TENGAH	L	146	3.29	YES
3	A11.2009.04636	YONATHAN GANI	S1	Teknik Informatika	JAWA TENGAH	L	150	2.87	YES
4	A11.2009.04637	SETIAWAN	S1	Teknik Informatika	JAWA TENGAH	L	148	3.36	YES
5	A11.2009.04640	BAGUS	S1	Teknik Informatika	JAWA TENGAH	L	146	2.96	YES
6	A11.2009.04642	ARIE KURNIAWAN	S1	Teknik Informatika	JAWA TENGAH	L	146	3.01	YES
7	A11.2009.04643	WIJAYA	S1	Teknik Informatika	JAWA TENGAH	L	146	3.18	YES
8	A11.2009.04644	ELDDO SEPTIAN	S1	Teknik Informatika	JAWA TENGAH	L	148	3.11	YES
9	A11.2009.04645	RESI	S1	Teknik Informatika	JAWA TENGAH	L	140	2.87	NO
10	A11.2009.04647	PARSTA	S1	Teknik Informatika	JAWA TENGAH	L	134	2.25	NO
11	A21.2009.06134	HERLAMBARANG S	D3	Manajemen Informatika	JAWA TENGAH	L	112	2.79	YES
12	A22.2009.01836	WRESTA PERMANA	D3	Teknik Informatika	JAWA TENGAH	L	113	3.09	NO
13	A22.2009.01841	EVAN FAJAR	D3	Teknik Informatika	JAWA TENGAH	L	111	3.14	YES
14	A22.2009.01842	NIKO BUDI	D3	Teknik Informatika	JAWA TENGAH	L	111	3	NO
15	A22.2009.01843	LUDYA AVILA	D3	Teknik Informatika	JAWA TENGAH	P	111	2.95	NO
16	A22.2009.01844	LAYA CAHYA	D3	Teknik Informatika	JAWA TENGAH	L	109	2.31	NO
17	A22.2009.01845	FITRI JAMIANI	D3	Teknik Informatika	JAWA TENGAH	P	111	2.86	YES
18	A22.2009.01846	YANU SETIYANTO	D3	Teknik Informatika	JAWA TENGAH	L	119	3.25	NO
19	A22.2009.01856	ADITYA DWI	D3	Teknik Informatika	JAWA TENGAH	L	105	2.08	NO
20	A22.2009.01866	HANDI PURNAWAN	D3	Teknik Informatika	JAWA TENGAH	L	111	2.95	YES
21	A14.2009.00560	LAZUARDI	S1	Visual	SELATAN	L	97	2.19	NO
22	A14.2009.00561	ARIF HAN	S1	Visual	JAWA TENGAH	L	143	3.13	NO

Menghitung jumlah kelas dari tahun lulus berdasarkan klasifikasi yang terbentuk (prior probability) :

1. C1 ( Class Tahun Lulus = “yes”) = jumlah “yes” pada kolom J Tahun Lulus = 11/20 = 0.55
2. C2 (Class Tahun Lulus = “no”) = jumlah ‘no” pada kolom J Tahun Lulus = 9/20 = 0.45

### 3.4.3 Menghitung jumlah kasus yang sama pada setiap atribut dari kelas Tahun Lulus ( yes / no ) berdasarkan data testing.

Gambar 3.2 Mencari P(Jenjang = “S1” | Class Tahun Lulus = “yes”)

D	J
Jenjang	Tahun Lulus
S1	YES
S1	YES
S1	YES
S1	YES
S1	YES
S1	YES
S1	YES
S1	NO
S1	NO
D3	YES
D3	NO
D3	YES
D3	NO
D3	NO
D3	NO
D3	YES
D3	NO
D3	NO
D3	NO
D3	YES
S1	NO
S1	NO

1.  $P(\text{Jenjang "S1"} | \text{Class Tahun Lulus = "yes"}) = 7/11 = 0.63636$
2.  $P(\text{Jenjang="S1"} | \text{Class Tahun Lulus = "no"}) = 4/9 = 0.44444$

( Hitung Probabilitas dari Seluruh Atribut jenjang, progdi, provinsi, jenis kelamin, SKS, IPK seperti contoh gambar 3.2 ). \*\*

### 3.4.4 Kalikan semua hasil variable

- a. Untuk semua atribut Class Tahun Lulus = “yes”
  - $P(X | \text{Class Tahun Lulus = "yes"}) = 0.63636 \times 0 \times 1 \times 0.90909 \times 0 \times 0 = 0$
- b. Untuk semua atribut Class Tahun Lulus = “no”
  - $P(X | \text{Class Tahun Lulus = "no"}) = 0.44444 \times 0.22222 \times 1 \times 1 \times 0.11111 \times 0.11111 = 0.00122$
- c. Perkalian prior probability dengan semua atribut Class Tahun Lulus = “yes”
  - $P(C_i | \text{Class Tahun Lulus = "yes"}) \times P(X | \text{Class Tahun Lulus = "yes"}) = 0.55 \times 0 = 0$
- d. Perkalian prior probability dengan semua atribut Class Tahun Lulus = “no”
  - $P(C_i | \text{Class Tahun Lulus = "no"}) \times P(X | \text{Class Tahun Lulus = "no"}) = 0.45 \times 0.00122 = 0.00055$

### 3.4.5 Bandingkan hasil kelas

1.  $P(C_i | \text{Class Tahun Lulus = "yes"}) \times P(X | \text{Class Tahun Lulus = "yes"}) < P(C_i | \text{Class Tahun Lulus = "no"}) \times P(X | \text{Class Tahun Lulus = "no"})$

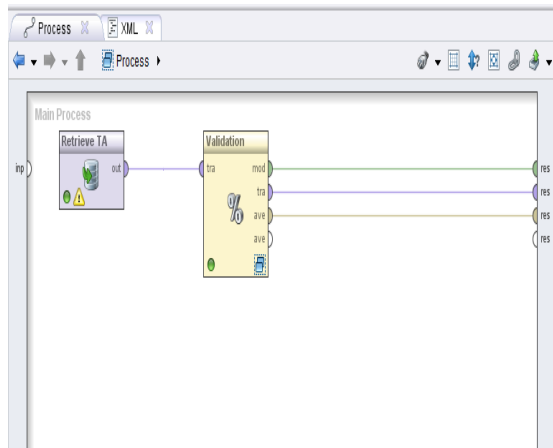
Kesimpulan :  
Class Tahun Lulus = “NO”

(Perhitungan antara perkalian Class Tahun Lulus “yes” dengan Class Tahun Lulus “no” menunjukkan bahwa nilai Class Tahun Lulus = “no” lebih besar dibandingkan kelas tahun lulus “yes”).



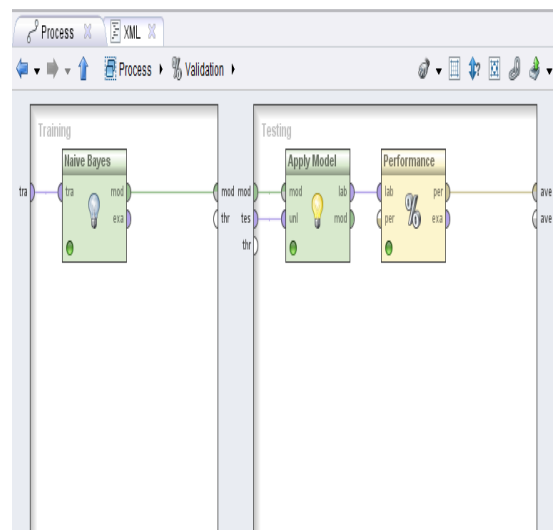
### 3.4.6 Implementasi dengan RapidMiner

Berikut adalah pengolahan data dengan menggunakan *naïve bayes* pada *RapidMiner* :



Gambar 3.3. Pemodelan *naïve bayes* pada *RapidMiner*

1. Pemodelan adalah tahapan (langkah) dalam membuat model dari suatu sistem nyata (realitas).
2. Rapidminer adalah sebuah lingkungan machine learning data mining, text mining dan predictive analytics.
3. Jumlah dataset mencapai 759 record maka penulis menggunakan tools Rapidminer untuk membantu proses perhitungan.
4. X Validation untuk membantu menghasilkan tingkat keakurasian berdasarkan dataset TA yang telah di lakukan proses klasifikasi.



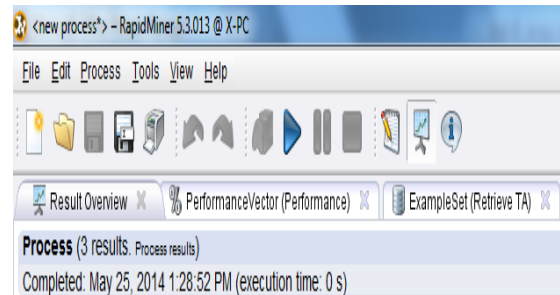
Gambar 3.4 Proses Training dan Testing

Tujuan utama penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma

*naïve bayes* yang digunakan untuk mengklasifikasi kelulusan.

Di dalam kolom training terdapat algoritma klasifikasi yang diterapkan yaitu *Naïve bayes*. Sedangkan di dalam kolom testing terdapat *Apply Model* untuk menjalankan model *naïve bayes* dan *Performance* untuk mengukur performa dari model *Naïve bayes* tersebut.

### 3.4.7 Hasil Pengujian dan Percobaan



Gambar 3.5 Result Overview

Pada percobaan dengan algoritma *naïve bayes* dengan menggunakan tools Rapidminer diperoleh waktu komputasi adalah 0 second. 0 second disini artinya komputasi menggunakan *naïve bayes* berjalan cukup cepat. Hal ini sesuai dengan kelebihan *naïve bayes* dibandingkan beberapa algoritma lain seperti neural network yang membutuhkan waktu cukup lama untuk melakukan komputasi data.

The screenshot shows the RapidMiner Performance Vector window. The title bar indicates the process is 'new process' - RapidMiner 5.3.013 @ X-PC. The window contains a menu bar (File, Edit, Process, Tools, View, Help) and a toolbar with various icons. Below the toolbar, there are several tabs: 'Result Overview', 'PerformanceVector (Performance)', 'ExampleSet (Retrieve TA)', and 'SimpleDistribution (Naive Bayes)'. The main area displays the following information:

Criterion Selector: Multiclass Classification Performance Annotations

precision recall

AUC (optimistic) accuracy: 82.00% +/- 3.44% (mikro: 82.00%)

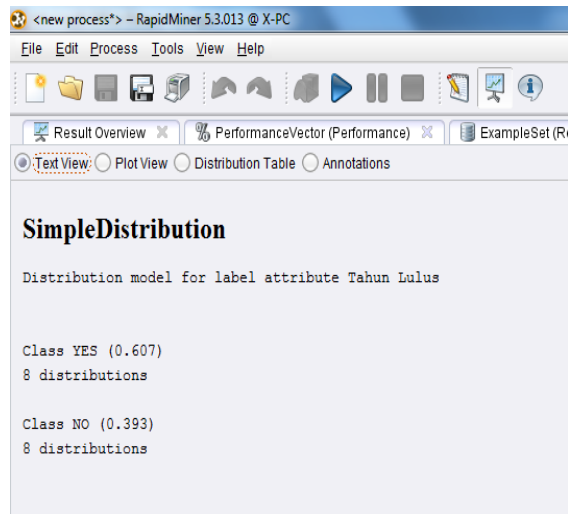
AUC (pessimistic)

	true YES	true NO	class precision
pred YES	391	66	85.56%
pred NO	70	232	76.82%
class recall	84.82%	77.85%	

Gambar 3.6 Performance Vector

Hasil akurasi model *naïve bayes* menunjukkan tingkat akurasinya 82.08% artinya model klasifikasi kelulusan menggunakan *naïve bayes* terbukti baik hal ini

dilihat dari tingkat akurasi yang mencapai 82.08% akan tetapi hal ini perlu di tinjau ulang dari sudut pandang kompleksitas dan jumlah datasetnya.



Gambar 3.7 Simple Distribution

Model distribusi untuk label atribut kelas Tahun Lulus adalah sebagai berikut:

Class No : 8 distributions  
Class Yes : 8 distributions

Percobaan pada penelitian ini menggunakan Rapidminer 5.3.008. Algoritma yang digunakan adalah *naive bayes*. Validasinya menggunakan *x-validation* dan untuk testing menggunakan *Apply Model* untuk menjalankan algoritma atau model *naive bayes* serta *Performance* untuk mengukur performa dari model *naive bayes* tersebut.

### 3.5 Evaluasi (Evaluation)

Evaluasi adalah fase lanjutan terhadap tujuan *data mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai dalam tahap *business understanding*.

#### 3.5.6 Evaluasi Hasil (Evaluation Results)

Tahap ini menilai sejauh mana hasil pemodelan *data mining* memenuhi tujuan *data mining* yang telah ditentukan pada tahap *business understanding*.

#### 3.5.7 Pengecekan Ulang Proses (Review Process)

Pada tahapan ini penulis memastikan bahwa semua tahapan / faktor penting yang telah dilakukan dalam pengolahan data tidak ada yang terlewatkan.

### 3.5.8 Menentukan Langkah Selanjutnya (Determine Next Steps)

Pada tahap ini adalah tahapan dalam menentukan langkah selanjutnya yang dilakukan. Terdapat 2 pilihan yaitu kembali pada tahap awal (*business understanding*) atau melanjutkan ke tahap akhir (*deployment*).

### 3.6 Persebaran (Deployment)

#### 3.6.6 Hasil Analisis

Deployment merupakan tahapan akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir yang berisi mengenai pengetahuan yang diperoleh atau pengenalan pola pada data dalam proses *data mining*.

Berdasarkan penelitian yang dilakukan, telah dihasilkan suatu pola, informasi, dan pengetahuan baru dalam proses data mining untuk klasifikasi kelulusan mahasiswa berdasarkan data mahasiswa Fakultas Ilmu Komputer Udinus angkatan 2009. Dari penelitian tersebut dihasilkan suatu pola, informasi, dan pengetahuan baru sesuai dengan tujuan *data mining* yaitu pola perhitungan *data mining* yang berisi data *training* dan data *testing* serta mencari probabilitas dari setiap atribut berdasarkan data *training* dan data *testing* untuk menghasilkan suatu informasi baru, apakah pada data mahasiswa Fakultas Ilmu Komputer angkatan 2009 lebih banyak kelas tahun lulus yang tepat waktu atau kelas tahun lulus tidak tepat waktu. Kemudian untuk menguji tingkat keakurasiannya maka digunakan Rapidminer sebagai alat bantu dalam proses pengujian tingkat akurasi dari klasifikasi tersebut.

Dari proses perhitungan *data mining* menggunakan algoritma *naive bayes* dan tingkat keakurasiannya, dihasilkan suatu informasi baru yaitu perhitungan *data mining* berdasarkan mahasiswa Fakultas Ilmu Komputer angkatan 2009, menunjukkan kelas tahun lulus “yes” / tepat waktu dengan total perkalian *prior probability* senilai 0, sedangkan kelas tahun lulus “no” / tidak tepat waktu dengan total perkalian *prior probability* senilai 0.00055. Untuk tingkat akurasi berdasarkan proses klasifikasi menggunakan algoritma *naive bayes*, dengan melalui semua tahapan dipastikan tidak ada bagian – bagian penting yang terlewatkan, dihasilkan tingkat akurasi sebesar 82.08 %.

Berdasarkan hasil perhitungan data mining dan proses pengujian tingkat akurasi dengan menggunakan Rapidminer, dapat



ditarik kesimpulan bahwa angkatan 2009 kelas tahun lulus “no” / tidak tepat waktu lebih besar dari kelas tahun lulus “yes” / tepat waktu. Sedangkan analisa yang dilakukan terhadap tingkat akurasi menggunakan algoritma naïve bayes menunjukkan bahwa nilai yang dihasilkan oleh algoritma *naïve bayes* memiliki tingkat kekuatan yang cukup tinggi. Hal ini di buktikan dengan hasil perhitungan yang mencapai nilai 82.08 %, Nilai 82. 08 % membuktikan bahwa model yang dibangun dapat digunakan untuk melakukan klasifikasi kelulusan mahasiswa. Nilai 82.08 % bisa juga di sebabkan oleh kekurangan kompleksitas data yang mengakibatkan model dapat memprediksi dengan akurat.

#### **4 KESIMPULAN DAN SARAN**

##### **4.1 Kesimpulan**

- a. Berdasarkan perhitungann data mining menggunakan algoritma naïve bayes, dapat ditarik kesimpulan bahwa kelas tahun lulus “no” / tidak lulus tepat waktu lebih besar daripada kelas tahun lulus “yes” / lulus tepat waktu.
- b. Dari hasil observasi terhadap dataset mahasiswa Udinus Fakultas Ilmu Komputer angkatan 2009 dan melalui proses perhitungan menggunakan metode klasifikasi naïve Bayes dengan atribut yang telah dijelaskan di pembahasan sebelumnya, didapatkan sebuah

hasil bahwa nilai akurasi terhadap klasifikasi kelulusan sebesar 82.08 %. Dimana 82.08 % bisa juga disebabkan oleh kurang kompleksitas data yang mengakibatkan model dapat memprediksi cukup akurat.

##### **4.2 Saran**

- a. Peneliti telah membahas penggunaan metode Naïve Bayes dalam penelitian klasifikasi kelulusan mahasiswa, diharapkan dalam penelitian selanjutnya dapat dibandingkan dengan memanfaatkan metode klasifikasi lainnya seperti metode C.4.5, metode nearest neighbor guna menentukan kelas berdasarkan atribut-atribut yang telah ditentukan sehingga dengan menggunakan banyak metode dapat mengetahui kelebihan masing-masing metode dan metode mana yang menghasilkan nilai akurasi yang lebih baik.
- b. Pengklasifikasian terhadap data mahasiswa UDINUS sebaiknya dilakukan secara rutin setiap tahun sebagai langkah preventif untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

## DAFTAR PUSTAKA

- [1] Johan Oscar Ong, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI MARKETING PRESIDENT UNIVERSITY," Jurnal Ilmiah Teknik Industri, vol. 12, no. 1, pp. 10-13, Juni 2013.
- [2] Universitas Dian Nuswantoro, Buku Panduan Akademik Mahasiswa Tahun Ajaran 2012-2013. Semarang, Jawa Tengah: Universitas Dian Nuswantoro, 2012.
- [3] Irwan Budiman, Toni Prahasto, and Yuli Christyono, "DATA CLUSTERING MENGGUNAKAN METODOLOGI CRISP-DM UNTUK PENGENALAN POLA PROPORSI PELAKSANAAN TRIDHARMA," in Seminar Nasional Aplikasi Teknologi Informasi 2012 (SNATI 2012), Yogyakarta, 2012.
- [4] Florin Gorunescu, Data Mining Concepts, Models and Techniques. Chennai, India: Springer, 2011.
- [5] Eko Prasetyo, Data Mining : Konsep dan Aplikasi menggunakan MATLAB, 1st ed. Yogyakarta, Indonesia: Andi, 2012.
- [6] Sergio Moro and Raul M.S. Laureano, "Using Data Mining for Bank Direct Marketing: An Application of The CRISP-DM Methodology," Instituto Universitário de Lisboa, Lisboa, 2011.
- [7] Arief Jananto, "Algoritma Naïve Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa," Jurnal Teknologi Informasi DINAMIK, vol 18, no.1, Januari 2013.
- [8] Marselina Silvia Suhartinah, Ernastuti, "Graduation Prediction Of Gunadarma University Students Using Algorithm Naïve Bayes C4.5 Algorithm," Faculty Of Industrial Engineering, 2010.
- [9] John Fredrik Ulysses, "Data Mining Classification Untuk Prediksi Lama Masa Studi Mahasiswa Berdasarkan Jalur Penerimaan Dengan Metode Naïve Bayes," Magister Teknik Informatika Universitas Atma Jaya Yogyakarta.
- [10] Turban, E. dkk, Decicion Support Systems and Intelligent Systems.: Andi Offset, 2005.
- [11] Larose, Daniel T, Data Mining Methods and Models. Hoboken New Jersey : Jhon Wiley & Sons, Inc, 2006.
- [12] Larose, Daniel T, Discovering Knowledge in Data: An Introduction to Data Mining : John Willey & Sons. Inc, 2005.

- [13] Carlo Vercellis, *Business Intelligence : Data Mining and Optimization for Decision Making*. Milano, Italy: A John Wiley and Sons, Ltd., Publication.
- [14] Han,J. and Kamber,M. “Data mining: Concepts and Techniques”, 2nd edition. The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor. 2006.