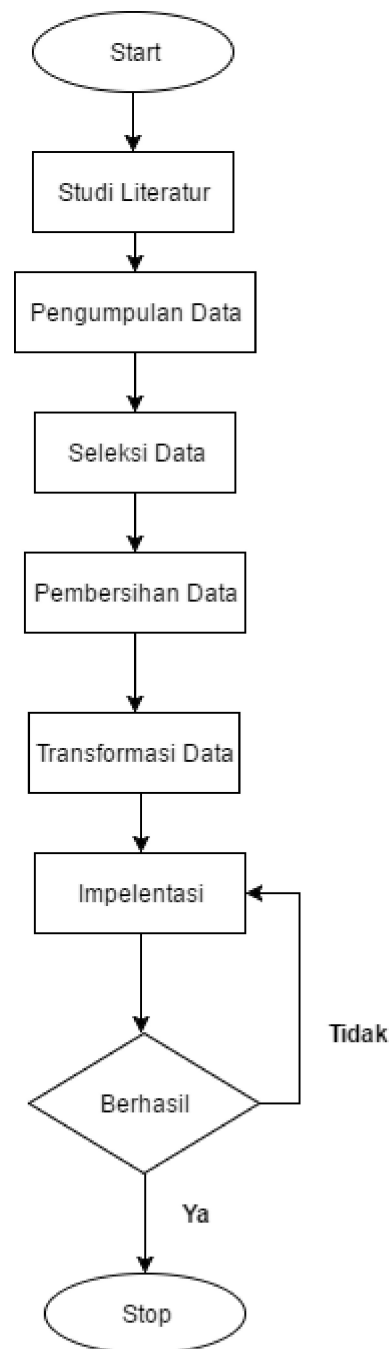


### 3.3. Alur Penelitian

Dalam melakukan penelitian ini, penulis menerapkan alur penelitian yang dapat dilihat pada gambar 3.1



**Gambar 3. 1** Flowchart alur penelitian

### 3.4. Studi Literatur

Langkah pertama yang dilakukan dalam penelitian ini adalah studi literatur. Studi Literatur dilakukan untuk mengumpulkan informasi yang diperlukan dalam penelitian ini. Informasi tersebut diperoleh dengan mempelajari dan membaca literatur - literatur yang ada hubungannya dengan permasalahan yang menjadi

### 3.5 Pengumpulan Data

Tahapan selanjutnya adalah pengumpulan data. Tahap ini merupakan tahapan yang penting bagi penulis karena dapat mempengaruhi proses implementasi dan kesimpulan apabila pengumpulan data tidak dilakukan dengan benar. Data yang digunakan dalam penelitian diperoleh dari Biro Sistem Informasi (BSI) Universitas Muhammadiyah Yogyakarta.

### 3.6 Seleksi Data

Data yang ada di *database* tidak semuanya dipakai untuk penelitian ini, oleh karena hanya data yang sesuai untuk dianalisis dan kemudian di *filter* di *database*. Sebagai contoh faktor kecenderungan orang membeli dua barang sekaligus dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan cukup dengan id pelanggan untuk menemukan pola-pola pembelian produk.

### 3.7 Pembersihan Data

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten. Pada umumnya data yang diperoleh, baik dari *database* Universitas maupun hasil eksperimen, mempunyai isi yang tidak sempurna seperti data yang hilang, data yang tidak valid dan data yang duplikat. Selain itu ada juga atribut-

atribut data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga juga akan mempengaruhi kinerja dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

### **3.8 Transformasi Data**

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining* karena beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Dalam penelitian ini data akan diubah ke dalam format *csv* melalui *excel*. Transformasi data dapat memudahkan proses *mining* di *software Rapidminer*.

### **3.9 Implementasi**

Merupakan suatu proses utama saat metode ini dilakukan untuk menemukan pengetahuan berharga atau sebuah informasi dari data. Teknik yang digunakan pada proses *mining* ini adalah prediksi dengan menggunakan algoritma *Naïve Bayes* dan *software* yang digunakan dalam proses *mining* ini yaitu *RapidMiner*.

### **4.0. Analisis Data**

Data yang didapatkan dari proses tahapan-tahapan *data mining* kemudian diolah kembali menggunakan *Rapidminer* untuk mendapatkan hasil analisis yang lebih akurat. Atribut yang diuji adalah nilai IPK semester I-IV mahasiswa angkatan 2010 yang telah dinyatakan lulus, jenis kelamin, jurusan sekolah dan asal pulau sebagai data *training*. Kemudian atribut yang digunakan untuk data *testing* adalah nilai IPK semester I-IV mahasiswa angkatan 2014 yang belum lulus, jenis kelamin, asal pulau, jurusan sekolah dan angkatan 2011 yang telah lulus yang

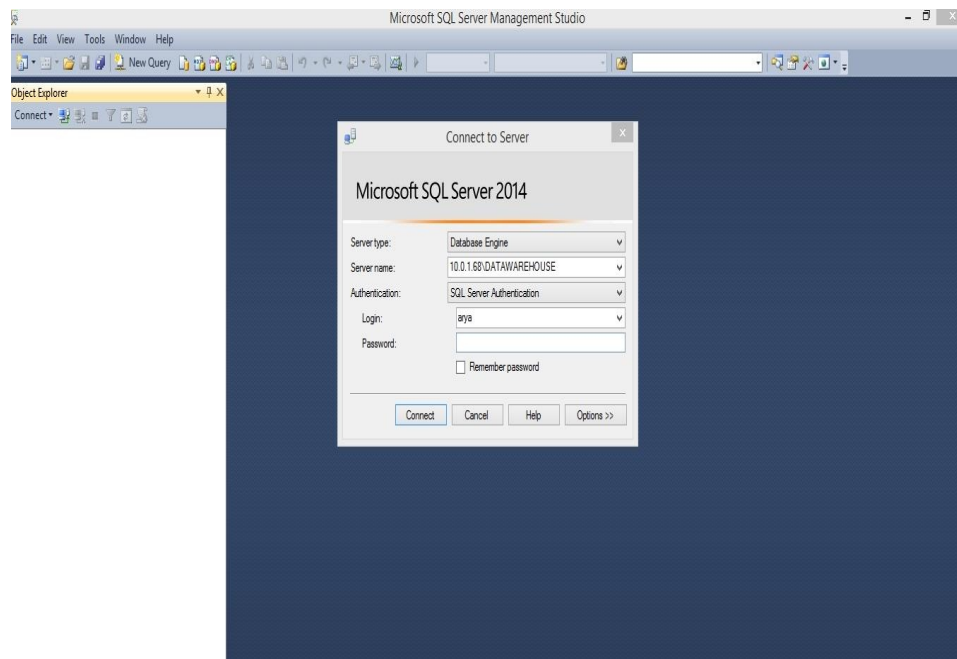
diambil secara acak. Hasil dari analisis prediksi menggunakan *Rapid Miner* dengan algoritma *Naive bayes* adalah dalam bentuk persentase.

## BAB IV

### HASIL DAN PEMBAHASAN

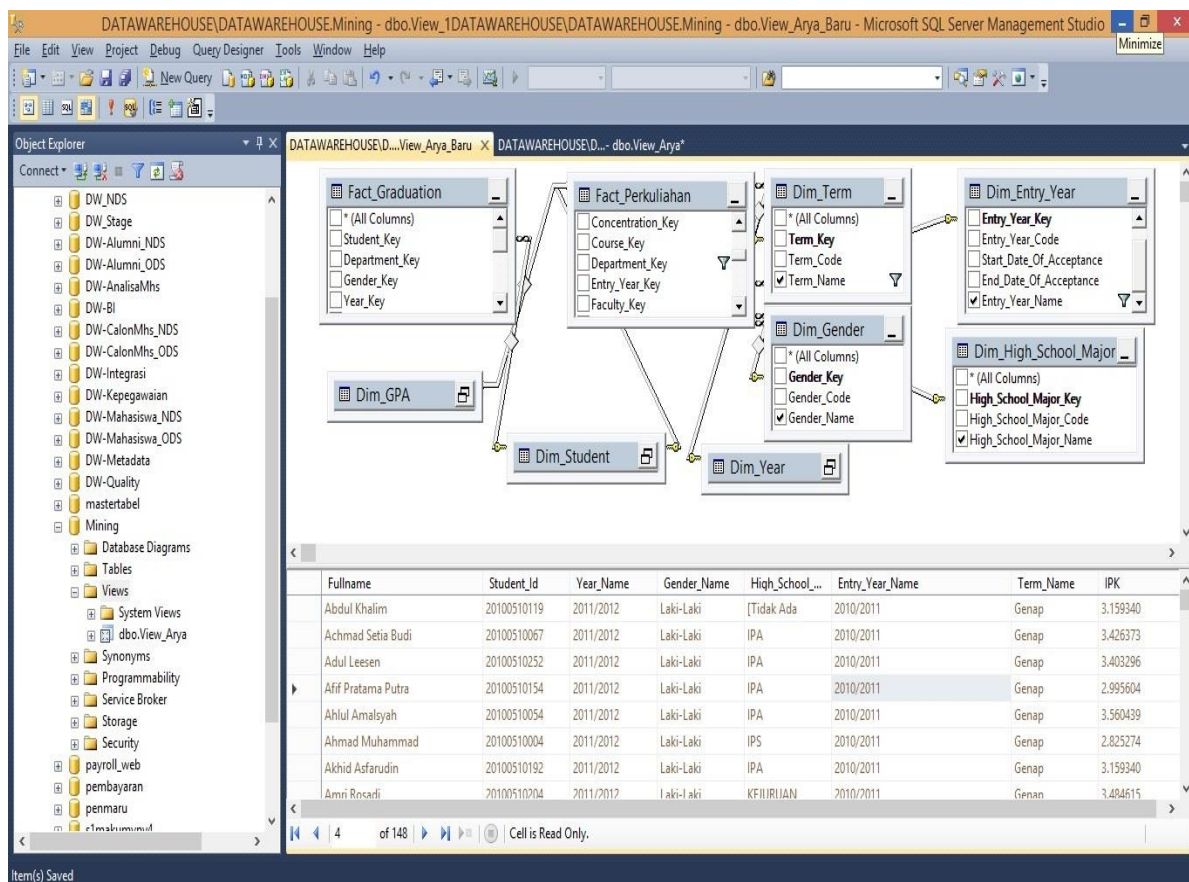
#### 4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah *data warehouse* Mahasiswa Universitas Muhammadiyah Yogyakarta prodi Hubungan Internasional angkatan 2010-2011 sebagai *data training* dan angkatan 2014 sebagai *data testing* pada pengujian pertama. Pengujian kedua *data* angkatan 2010-2011 yang telah lulus dan belum lulus menjadi *data training* dan *testing* menggunakan *cross validation* sebagai pembandingan tingkat akurasi. *Software* yang digunakan untuk mengakses *data warehouse* adalah *SQL Server Management 2014* dan untuk dapat mengakses *data warehouse* peneliti diberikan hak akses ke *database server* Universitas Muhammadiyah Yogyakarta. *Server name* dari *database server* BSI (Biro Sarana Informasi) seperti gambar 4.1 dibawah ini.



**Gambar 4. 1** *Connect Server*

Setelah terhubung ke *database server* penulis membuat *view* untuk memudahkan proses pengambilan data seperti gambar 4.2 dibawah ini.



**Gambar 4. 2** View Data

Data yang digunakan dalam penelitian ini adalah data dari jurusan Hubungan Internasional angkatan 2010-2011 yang sudah lulus dan belum lulus sebagai data training berjumlah 435 data . Data dari jurusan Hubungan Internasional angkatan 2014 yang belum lulus digabungkan dengan data 2011 yang sudah lulus sebagai data *testing* sebanyak 100 data dengan 4 *attribut*.

## 4.2. Seleksi Data

*Data selection* adalah proses menganalisis data-data yang relevan dari *database* karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses *data mining*. Data tersebut dipilih dan diseleksi dari *database* untuk di analisis. Dari semua data yang digunakan hanya *Province Of Birth, High School Major Name, Gender Name, GPA, Keterangan*. Karena informasi yang terkandung didalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan *indicator* penelitian.

## 4.3. Pembersihan Data

Setelah tahap pengumpulan data dan *filter* data maka tahap selanjutnya yaitu *cleaning data* agar tidak ada duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan cetak, sehingga data tersebut dapat diolah dan dilakukan proses *data mining*. Setelah semua data yang di butuhkan telah melalui tahap *cleaning data* maka data akan disimpan dalam *dataset* baru yang menggunakan *Microsoft Office Excel* dengan format *csv*. Data yang diambil dari jurusan Hubungan Internasional ini ada data yang *unknown* yaitu data pada atribut *Province\_Of\_Birth* sebanyak 1 data dan atribut *High\_School\_Major\_Name* sebanyak 80 data.

## 4.4. Tranfosrmasi Data

*Data Transformation* adalah tahap mengubah data menjadi bentuk yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa di aplikasikan. Dalam penelitian ini data yang akan diproses dari *database SQL Server 2014 Management*

*Studio* akan diubah menjadi *file CSV (comma delimited)* yang dapat digunakan untuk pengolahan data pada *Software RapidMiner*. Dan nama atribut data juga di ubah dari *Province\_Of\_Birth, Senior\_High\_School, Gender\_Name GPA dan keterangan* di ubah menjadi *Pulau, Gender, Jurusan SMA, IPK, Lulus* dan nilai yang ada pada atribut *Pulau dan IPK* juga di ubah sesuai dengan kebutuhan. Gambar 4.3 adalah data yang belum di ubah oleh penulis atau yang belum di lakukan transformasi data.

1	Gender_NHigh_school_Major	IPK	City_of_Birth	Keterangan	Pulau
2	Laki-Laki [Tidak Ada	3.15934	Tuban	Yes	Jawa
3	Laki-Laki IPA	3.426373	Sorong	Yes	Papua
4	Laki-Laki IPA	3.84945	Ujung Pandang	Yes	Jawa
5	Laki-Laki IPS	3.580219	Sleman	Yes	Jawa
6	Laki-Laki IPA	3.403296	Satun, Thailand	Yes	Thailand
7	Laki-Laki IPA	2.995604	Kudus	Yes	Jawa
8	Laki-Laki IPS	2.825274	Jakarta	Yes	Jawa
9	Laki-Laki IPA	3.15934	Sleman	Yes	Jawa
10	Laki-Laki IPA	3.240659	Tarakan	Yes	Kalimantan
11	Laki-Laki BAHASA	3.314285	Kalteng	Yes	Kalimantan
12	Laki-Laki IPA	3.003296	Tenggarong	Yes	Kalimantan
13	Perempuan [Tidak Ada	3.61791	Yogyakarta	Yes	Jawa
14	Perempuan IPA	3.28172	Sleman	Yes	Jawa
15	Laki-Laki IPS	3.127777	Sungai Tarab	Yes	Sumatera
16	Laki-Laki IPA	3.587912	Takengon	Yes	Sumatera
17	Perempuan IPA	2.285714	Ternate	Yes	Maluku
18	Perempuan IPS	3.063541	Jakarta	Yes	Jawa
19	Perempuan BAHASA	3.684615	Pasuruan	Yes	Jawa
20	Perempuan [Tidak Ada	3.637362	gunung bathin	Yes	Sumatera
21	Perempuan IPA	3.785714	Wamena	Yes	Papua
22	Perempuan IPA	2.656043	jakarta	Yes	Jawa
23	Perempuan IPA	3.68901	Bandung	Yes	Jawa
24	Perempuan IPA	3.785714	Wamena	Yes	Papua

**Gambar 4. 3** Transformasi Data

Selanjutnya penulis melakukan transformasi dan inisialisasi data agar penelitian ini bisa berjalan dengan baik. Data yang akan diinisialisasi merupakan data dari *atribut IPK dan Pulau* seperti tabel berikut ini:



<b>Nilai IPK</b>	<b>Inisialisasi</b>
IPK Asli > 3.00	A
$2 \leq \text{IPK Asli} \leq 3.00$	B
IPK Asli < 2	C

**Tabel 1** Inisialisasi nilai IPK

<b>Asal Pulau</b>	<b>Inisialisasi</b>
Sumatera, Riau, Bangka Belitung	A
Jawa, Bali, Madura	B
Kalimantan	C
Sulawesi	D
NTB, NTT, Papua	E
Luar negeri	F

**Tabel 2** Inisialisasi asal pulau

Selanjutnya data yang sudah di inisialisasi di transformasikan kedalam *excel* kembali seperti gambar 4.6 dibawah ini.

Gender_Name	Jurusan	IPK	Pulau	Lulus
Perempuan	[Tidak Ada]	3 A		Yes
Perempuan	IPA	3 C		Yes
Laki-Laki	IPS	2 B		Yes
Perempuan	[Tidak Ada]	3 B		Yes
Perempuan	IPA	3 B		Yes
Laki-Laki	IPA	2 B		Yes
Laki-Laki	IPS	3 D		Yes
Laki-Laki	IPS	3 B		Yes
Perempuan	AGAMA	3 B		Yes
Perempuan	IPS	3 C		Yes
Laki-Laki	IPS	2 B		Yes
Perempuan	IPA	2 E		Yes
Perempuan	KEJURUAN	2 E		Yes
Perempuan	[Tidak Ada]	3 B		Yes
Laki-Laki	IPS	3 B		Yes
Laki-Laki	IPS	2 E		Yes
Laki-Laki	IPS	3 B		Yes
Laki-Laki	[Tidak Ada]	3 B		Yes
Perempuan	IPS	2 B		Yes
Laki-Laki	IPA	3 B		Yes
Perempuan	IPA	3 B		Yes
Perempuan	IPS	2 B		Yes

**Gambar 4. 4** Data yang telah diinisialisasi

#### 4.5. Implementasi

Pada tahap ini dilakukan pemodelan data, metode yang dipakai pada penelitian ini adalah probabilitas (prediksi) dengan menggunakan algoritma *Naive Bayes*. Data yang telah dikumpul, diseleksi dan ditransformasi akan dikelola menggunakan metode probabilitas. Metode ini dapat digunakan dalam memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sebagai perbandingan.

Data yang akan diujikan dibagi menjadi dua bagian yaitu *training* dan *testing* kemudian dianalisis menggunakan *software Rapidminer*. Data *training* jurusan Hubungan Internasional untuk pengujian pertama memiliki 435 *record* data dan data *testing* memiliki 100 *record* data. Pengujian kedua menggunakan data *training* sebanyak 435, dan data tersebut juga menjadi data *testing* dengan

menggunakan metode *cross validation*. Gambar 4.5 adalah data *training* dan gambar 4.6 data *testing* dengan format .CSV yang akan diakses melalui *software Rapidminer*.

Gender_Name	Jurusan	IPK	Pulau	Lulus
Perempuan	[Tidak Ada]	3 A	Yes	
Perempuan	IPA	3 C	Yes	
Laki-Laki	IPS	2 B	Yes	
Perempuan	[Tidak Ada]	3 B	Yes	
Perempuan	IPA	3 B	Yes	
Laki-Laki	IPA	2 B	Yes	
Laki-Laki	IPS	3 D	Yes	
Laki-Laki	IPS	3 B	Yes	
Perempuan	AGAMA	3 B	Yes	
Perempuan	IPS	3 C	Yes	
Laki-Laki	IPS	2 B	Yes	
Perempuan	IPA	2 E	Yes	
Perempuan	KEJURUAN	2 E	Yes	
Perempuan	[Tidak Ada]	3 B	Yes	
Laki-Laki	IPS	3 B	Yes	
Laki-Laki	IPS	2 E	Yes	
Laki-Laki	IPS	3 B	Yes	
Laki-Laki	[Tidak Ada]	3 B	Yes	
Perempuan	IPS	2 B	Yes	
Laki-Laki	IPA	3 B	Yes	
Perempuan	IPA	3 B	Yes	
Perempuan	IPS	2 B	Yes	

**Gambar 4. 5 Data Training format csv**

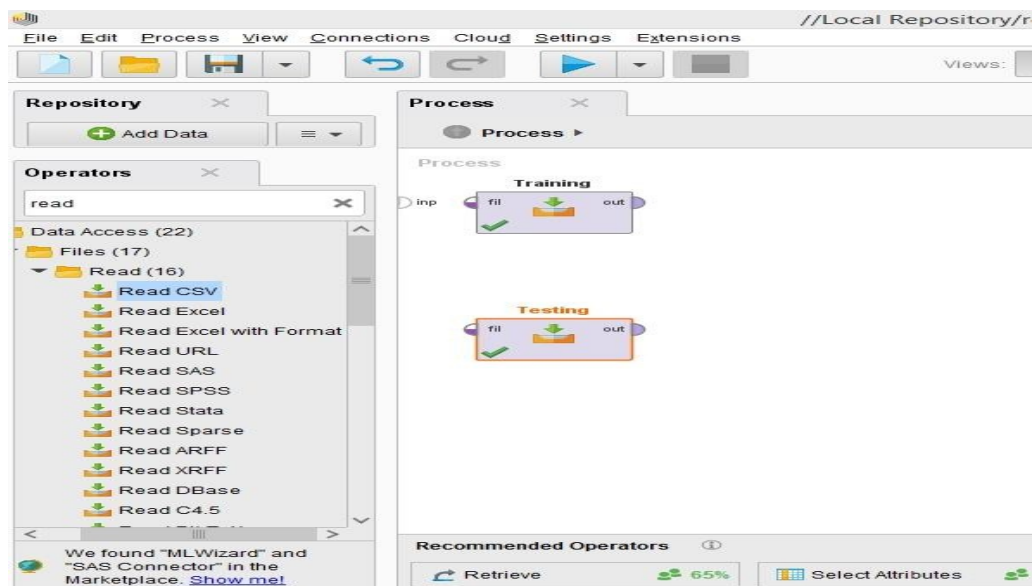
Gender	Jurusan	IPK	Pulau	Lulus
Laki-Laki	IPA	3 E	No	
Laki-Laki	IPA	1 C	No	
Laki-Laki	KEJURUAN	3 B	No	
Laki-Laki	IPA	2 A	No	
Laki-Laki	IPS	2 A	No	
Perempuan	IPA	3 C	No	
Perempuan	IPA	3 B	No	
Perempuan	IPA	3 B	No	
Perempuan	IPA	3 B	No	
Perempuan	IPA	3 A	No	
Perempuan	IPS	3 B	No	
Laki-Laki	IPS	2 E	No	
Perempuan	IPS	3 B	No	
Laki-Laki	IPS	3 B	No	
Laki-Laki	IPA	2 C	No	
Perempuan	IPA	3 B	No	
Perempuan	IPA	3 C	No	
Perempuan	IPA	1 B	No	
Perempuan	IPA	2 B	No	
Perempuan	IPA	1 B	No	
Perempuan	IPS	3 B	No	
Laki-Laki	IPA	2 E	No	

**Gambar 4. 6 Data Training format csv**

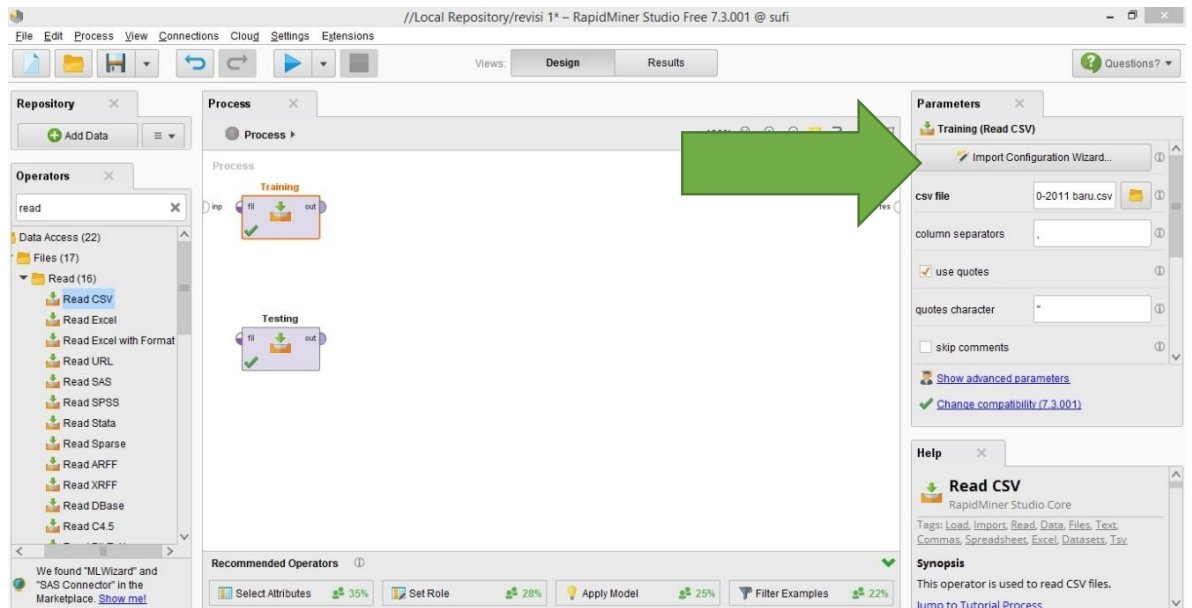
#### 4.6. Pengujian *software Rapidminer*

Atribut yang digunakan sebagai *label* adalah lulus, penulis akan menganalisis faktor yang mempengaruhi mahasiswa lulus tepat waktu dan memprediksi mahasiswa yang dapat lulus tepat waktu menggunakan data dari jurusan Hubungan Internasional. Data terlebih dahulu kita transformasi ke dalam format *.csv* agar bisa diakses menggunakan *software RapidMiner*.

Setelah data telah dibuat dalam bentuk format *.csv* selanjutnya lakukan *importing* data, untuk melakukan *importing* data dibutuhkan operator *read csv* lakukan *drag and drop* kedalam *view process* dan *rename* operator *read csv* menjadi *training* dan *testing* sehingga operator *read csv* muncul dalam *view process* seperti gambar 4.7. Setelah itu langkah selanjutnya langsung klik *Import Configuration Wizard* seperti gambar 4.8.

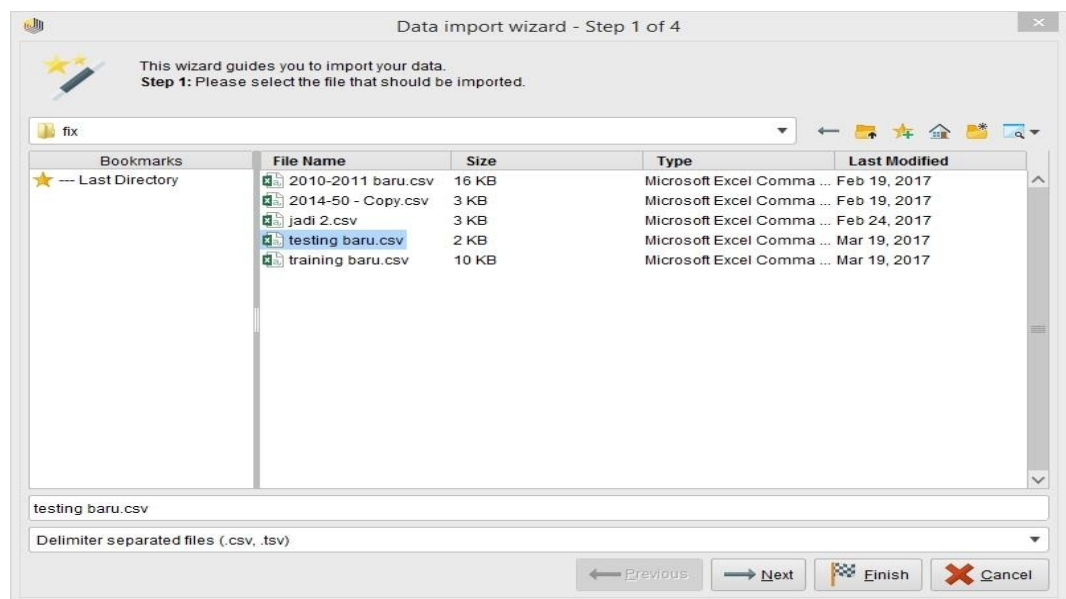


Gambar 4.7 Drag and drop csv



**Gambar 4. 8** *Importing file training*

Setelah klik *import configuration wizard* maka akan muncul form *data import wizard step 1* untuk memilih lokasi file yang digunakan untuk proses penelitian ini seperti gambar 4.9.



**Gambar 4. 9** *Form data import data training*

Setelah data yang telah yang akan digunakan dipilih langkah selanjutnya yaitu klik *Next* dan akan muncul form *data import wizard step 2* seperti gambar 4.10

Data import wizard - Step 2 of 4

This wizard guides you to import your data.  
Step 2: Please specify how the file should be parsed and how columns are separated.

File Reading

File Encoding: windows-1252

Trim Lines

Skip Comments: #

Column Separation

Comma ","

Semicolon ";"

Tab

Regular Expression: \s\*|\s\*

Escape Character: \


Use Quotes: "

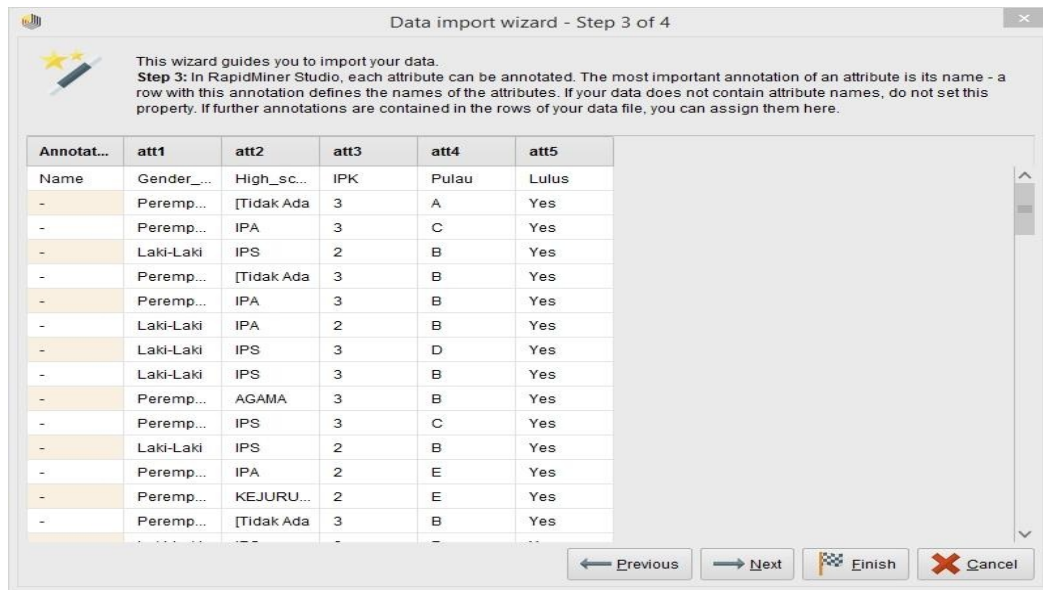
Gender_Name	High_school_Major	IPK	Pulau	Lulus
Perempuan	[Tidak Ada]	3	A	Yes
Perempuan	IPA	3	C	Yes
Laki-Laki	IPS	2	B	Yes
Perempuan	[Tidak Ada]	3	B	Yes
Perempuan	IPA	3	B	Yes

Row, Column	Error	Original value	Message
-------------	-------	----------------	---------


Previous Next Finish Cancel

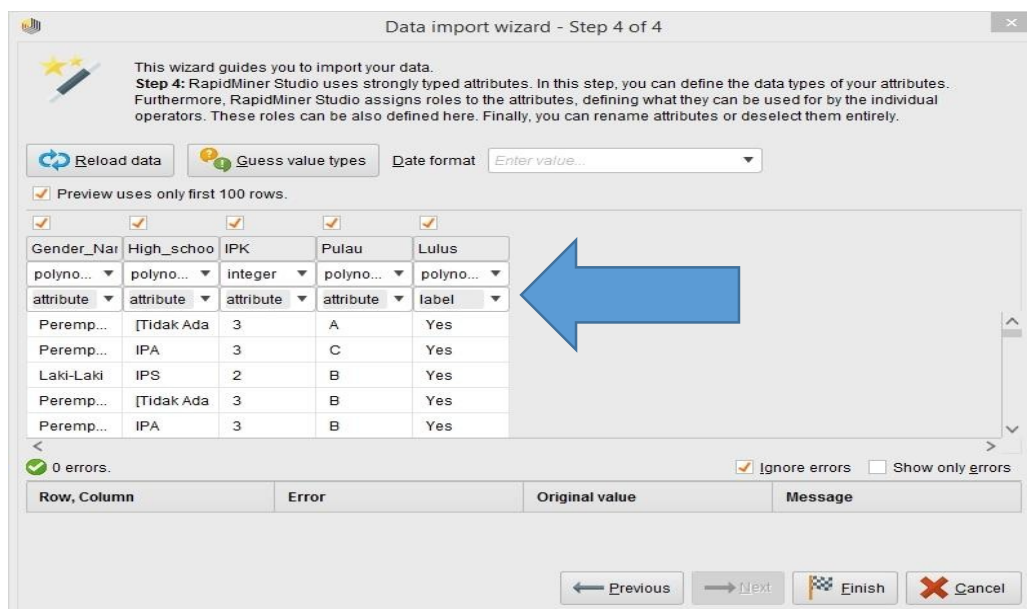
**Gambar 4. 10** *Data importing wizard step 2 data training*

Setelah muncul *form* step ke 2 seperti gambar diatas lalu pilih *Column Separation Comma* karena *Comma* di gunakan untuk memisahkan atribut satu dengan *atribut* yang lainnya yang sebelumnya *atribut* tersebut masih menyatu. Untuk melanjutkan ke *step* selanjutnya klik  *Next* maka akan muncul form *data import wizard step 3* seperti gambar 4.11.



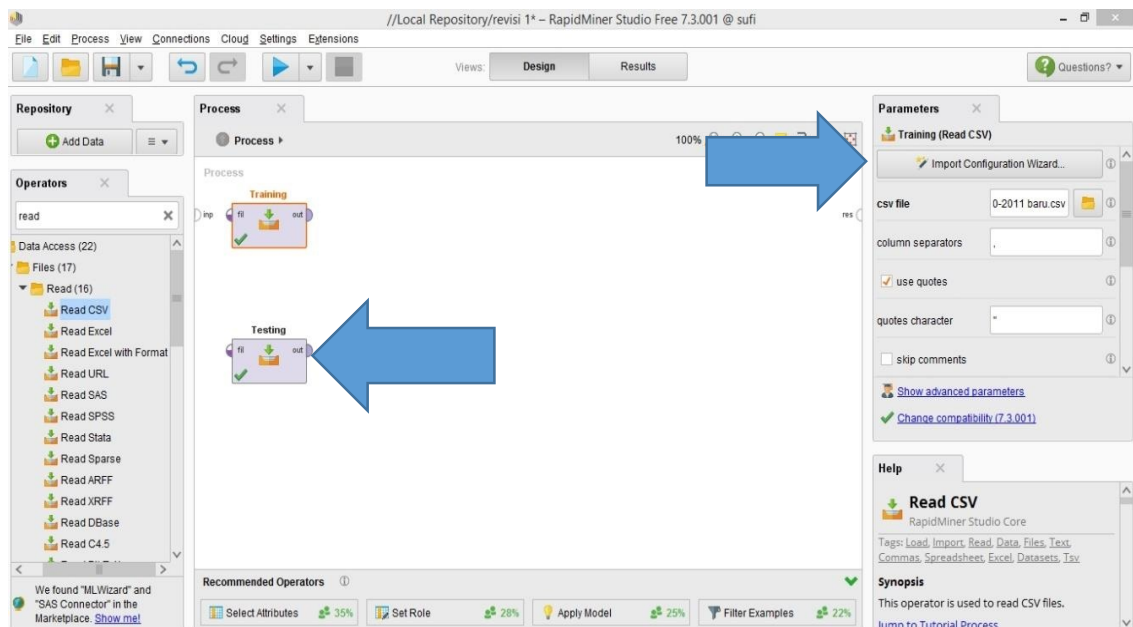
**Gambar 4. 11** Data importing wizard step 3 data training

Pada step ke 3 ini tidak ada dilakukan apapun maka dari itu langsung ke data import wizard step 4 dengan cara klik  Next maka akan muncul form data import wizard step 4 seperti gambar 4.12.



**Gambar 4. 12** Data importing wizard step 4 data training

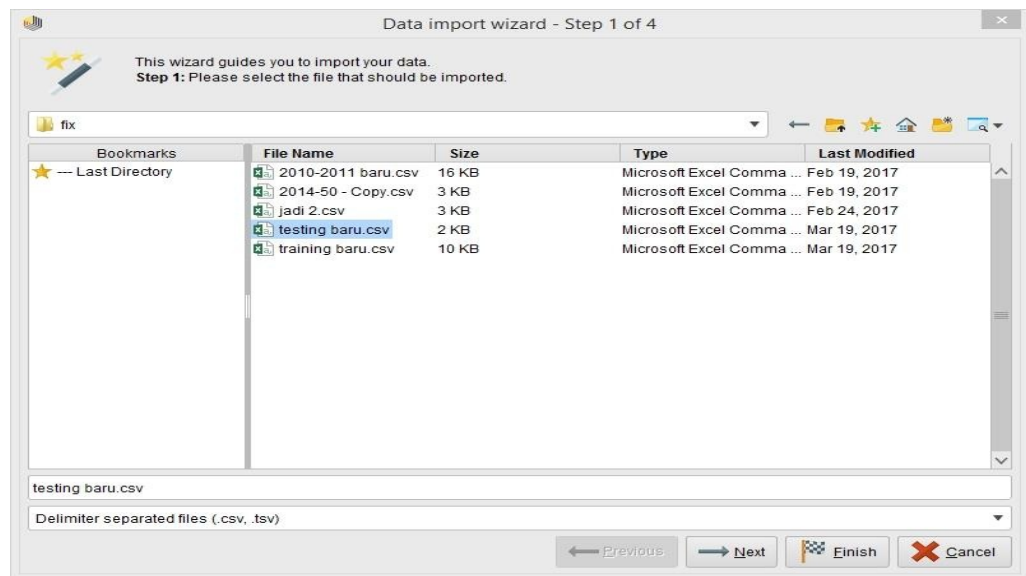
Setelah muncul *form data import wizard step 4* seperti gambar diatas lalu pilih salah satu atribut target karena pada klasifikasi tentu ada atribut target atau *label* dan atribut yang dipilih sebagai *label* yaitu atribut lulus. Setelah ditentukan label yang dipilih maka klik *Finish*. Setelah data selesai di *import* maka selanjutnya melakukan *data importing wizard* untuk data *testing* dengan mengikuti langkah-langkah yang sama dengan data *training* tetapi menggunakan *file.csv testing* seperti pada gambar 4.13 dibawah ini.




**Gambar 4. 13** *Importing file csv testing*

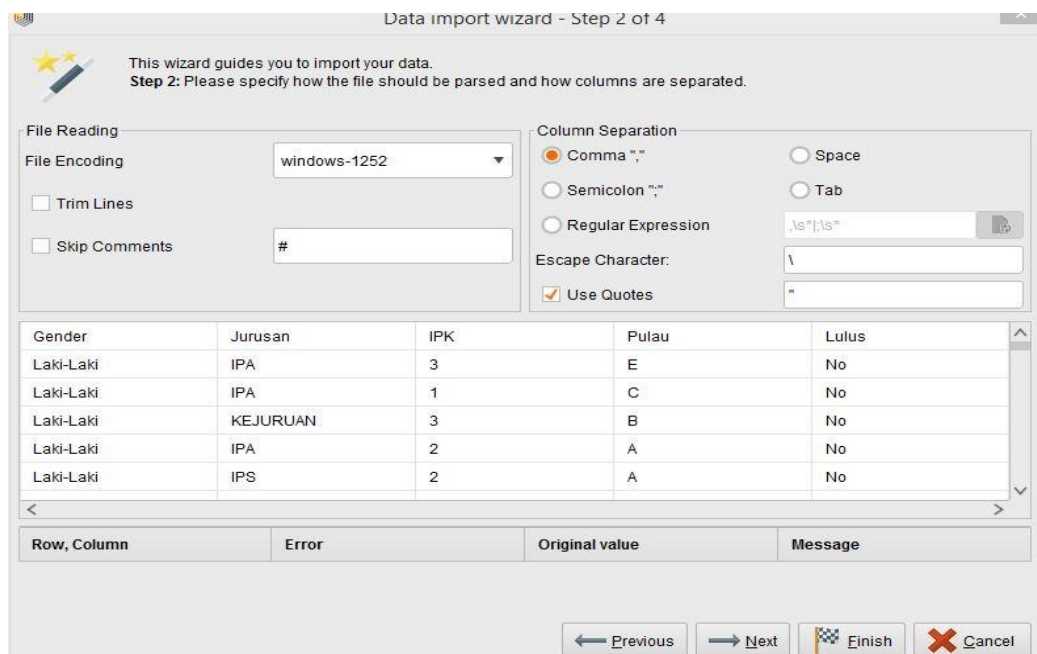
Setelah klik *import configuration wizard* maka akan muncul *form data import wizard step 1* untuk memilih lokasi *file* yang digunakan untuk proses penelitian ini seperti gambar 4.14.






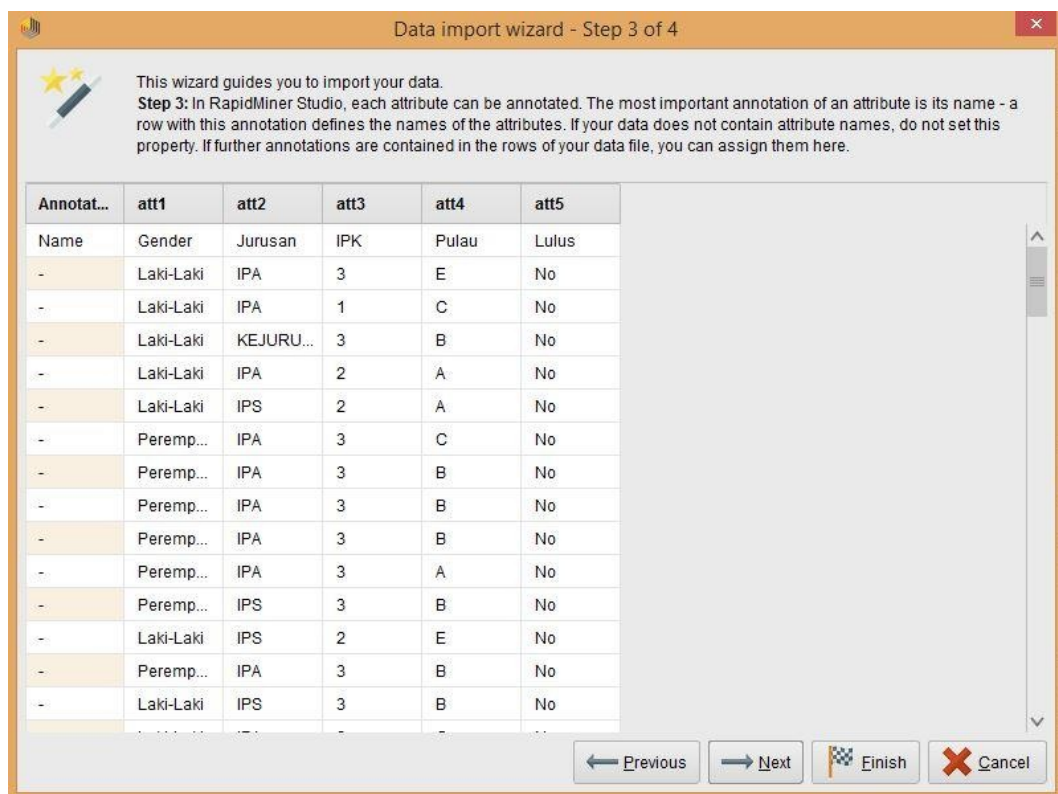
**Gambar 4. 14** Form data import data testing

Setelah data telah dipilih langkah selanjutnya yaitu klik  *Next* dan akan muncul form data import wizard step 2 seperti gambar 4.15




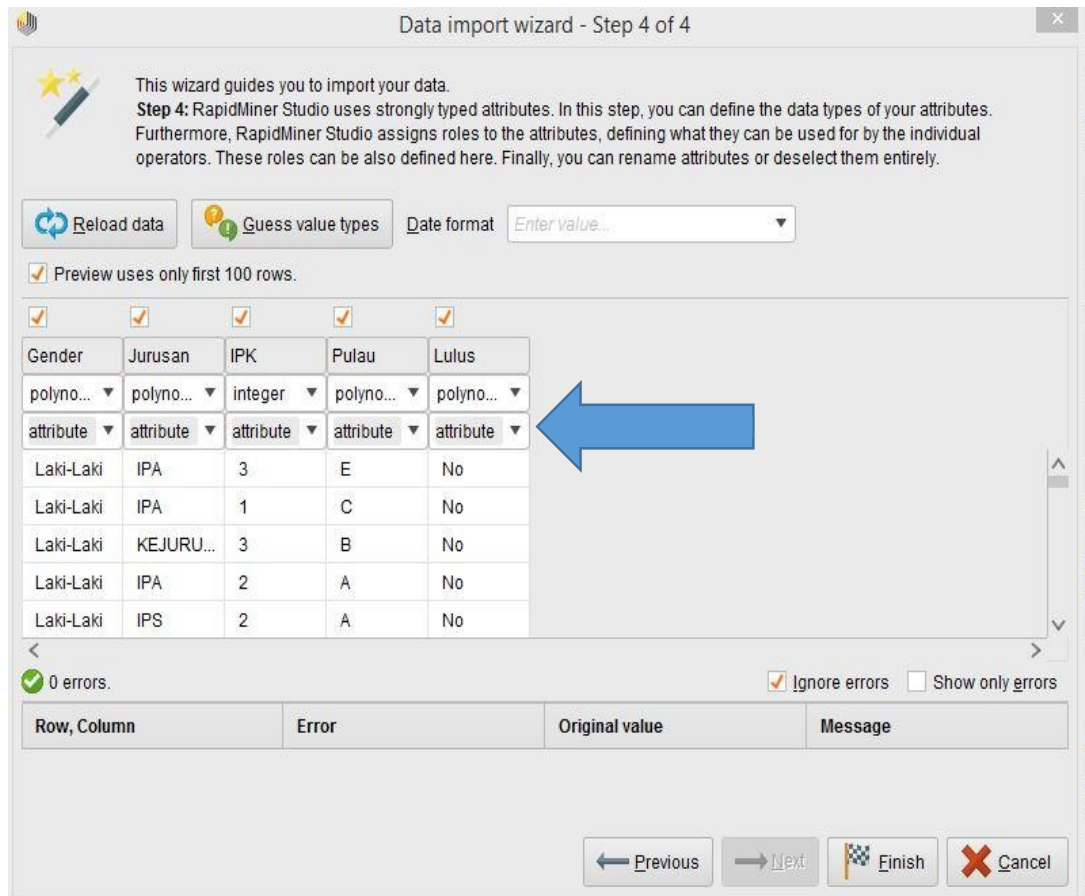
**Gambar 4. 15** Data importing wizard step 2 data training

Setelah muncul *form* step ke 2 seperti gambar diatas lalu pilih *Column Separation Comma* karena *Comma* di gunakan untuk memisahkan atribut satu dengan *atribut* yang lainnya yang sebelumnya *atribut* tersebut masih menyatu. Untuk melanjutkan ke step selanjutnya klik  *Next* maka akan muncul form *data import wizard step 3* seperti gambar 4.16.



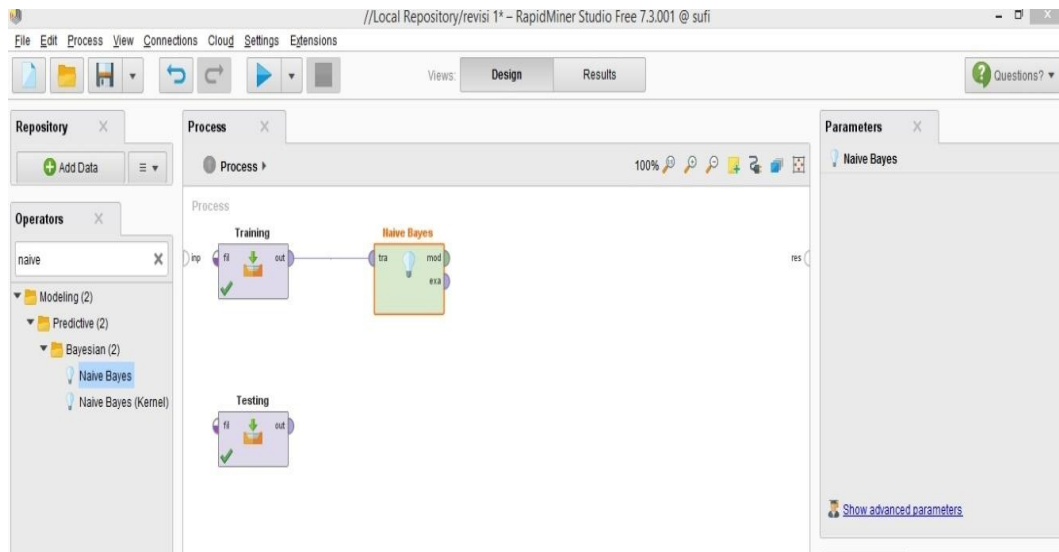
**Gambar 4. 16** *Data importing wizard step 3 data testing*

Pada step ke 3 ini tidak ada dilakukan apapun maka dari itu langsung ke *data import wizard step 4* dengan cara klik  *Next* maka akan muncul form *data import wizard step 4* seperti gambar 4.17



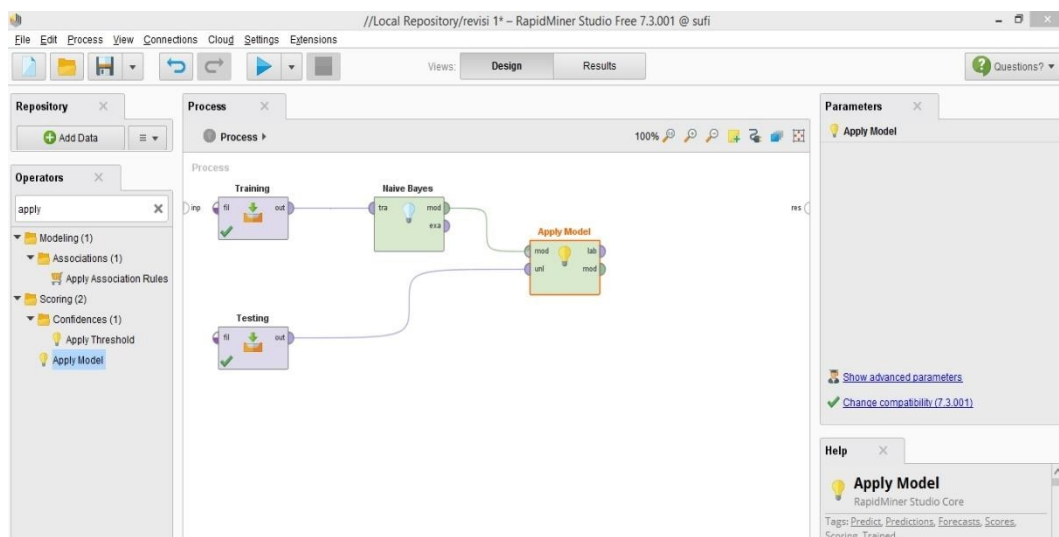
**Gambar 4. 17** Data importing wizard step 4 data testing

Setelah muncul form data import wizard step 4 seperti gambar diatas lalu pilih salah satu atribut target karena pada klasifikasi tentu ada atribut target atau *label* dan atribut yang dipilih sebagai *label* yaitu atribut lulus. Kemudian Setelah kedua data selesai di *import* maka selanjutnya *drag and drop Naïve bayes* dan kemudian hubungkan *read csv training* ke operator *naïve bayes* seperti gambar 4.18. Didalam operator *naïve bayes* ini data akan di proses dengan perhitungan algoritma *naïve bayes*.



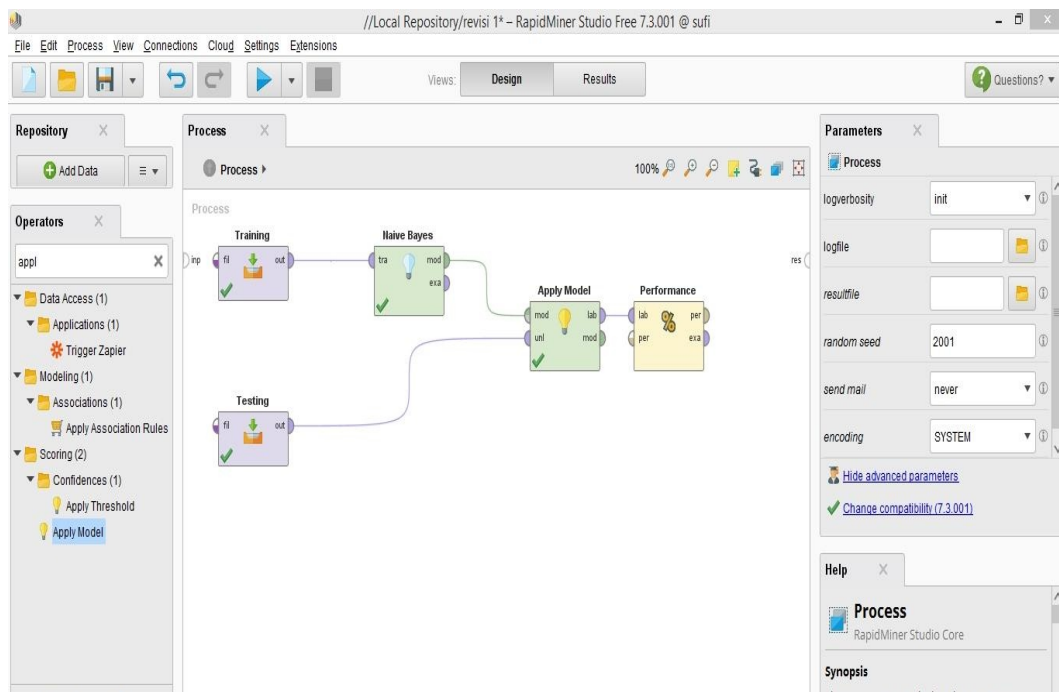
**Gambar 4. 18** File read csv training dengan operator naïve bayes

Langkah selanjutnya adalah memasukkan operator *apply model* kedalam *view process* dan hubungkan *port mod* operator *naïve bayes* dengan *port mod* *apply model* dan hubungkan juga *port ouput read csv testing* dengan *port unl* *apply model* seperti gambar 4.19.



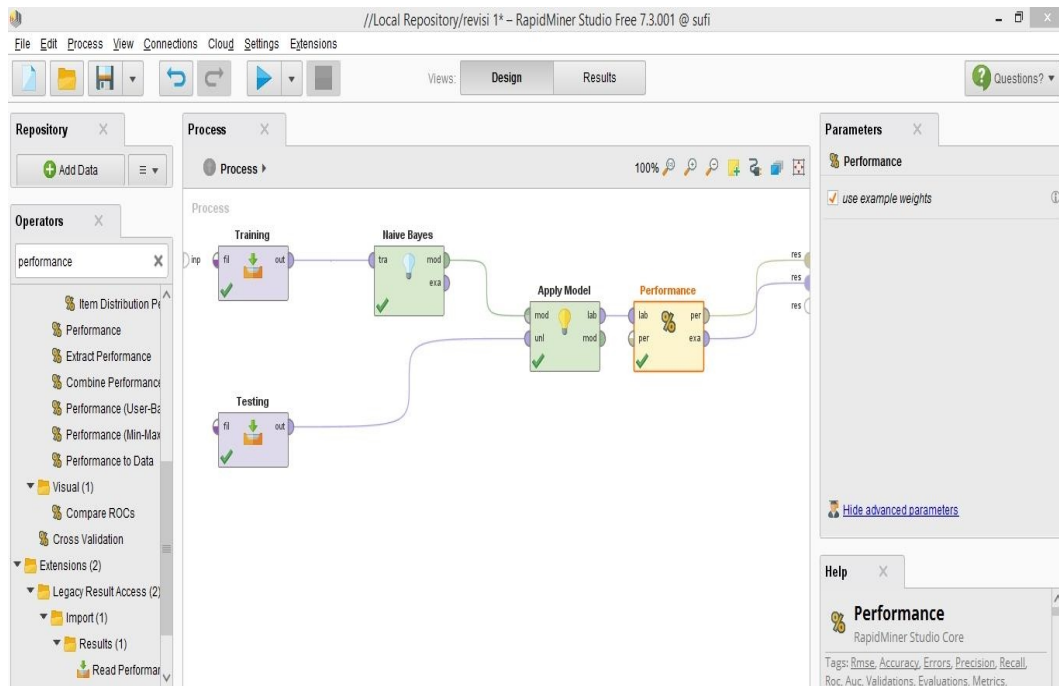
**Gambar 4. 19** Apply model

Dalam proses ini operator *apply model* berfungsi untuk menampung hasil pembelajaran data *training* yang telah diolah algoritma *naïve bayes* dan kemudian di validasi dengan data *testing* sehingga, mendapatkan hasil dari perhitungan data *training* yang menjadi tujuan prediksi bagi data *testing*. Selanjutnya adalah menambahkan operator *performance* untuk melihat tingkat *accuracy* dan *error* yang didapatkan dari hasil perhitungan di *apply model* seperti gambar 4.20.



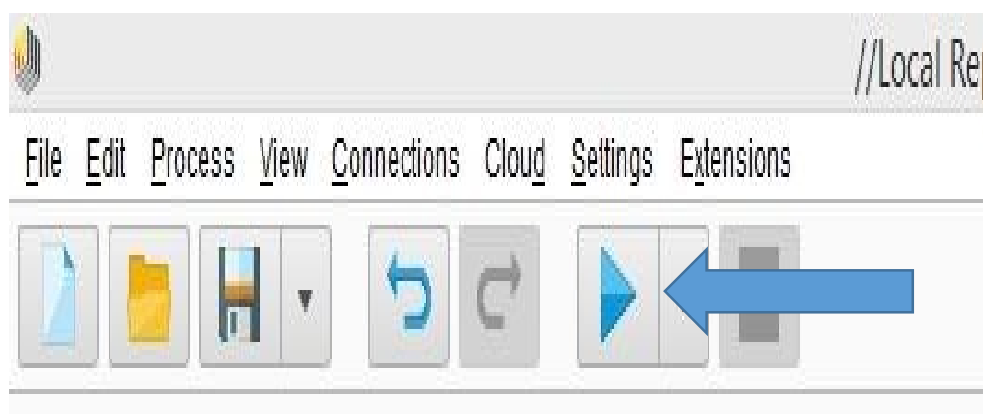
**Gambar 4. 20** Operator *performance*

Langkah selanjutnya adalah menghubungkan *output lab* dari *apply model* menuju *ouput lab* operator *performance*. Setelah semuanya terhubung, selanjutnya adalah hubungkan *port per* dan *exa* yang ada di operator *performance* menuju *port res* yang berada disebelah kanan *view process* seperti gambar 4.21.



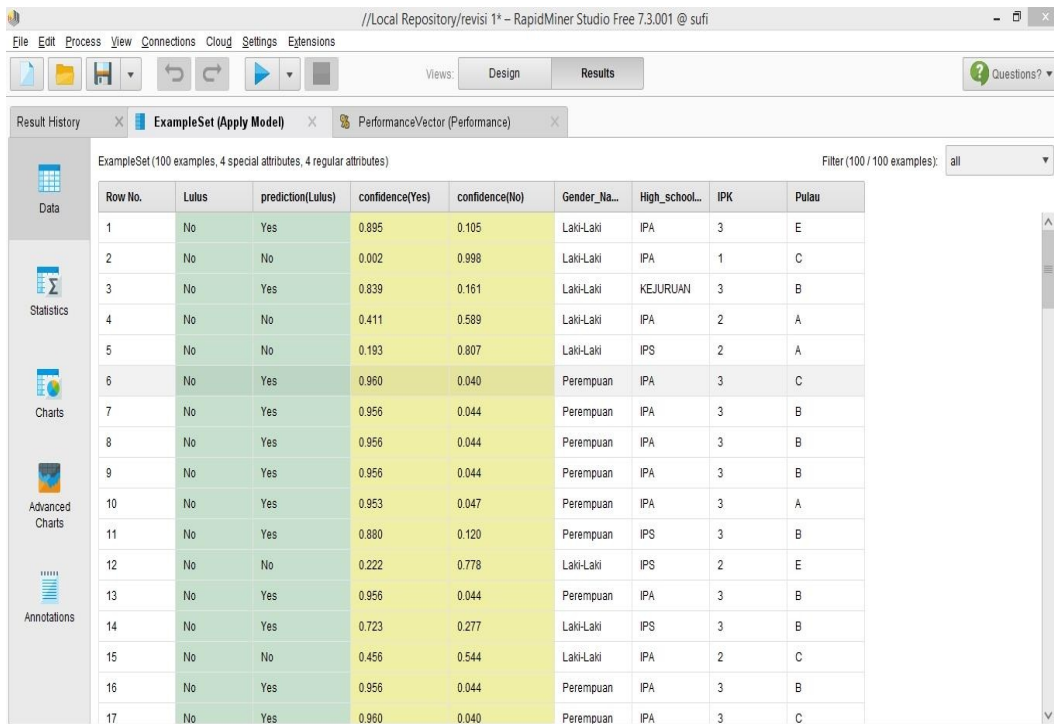
**Gambar 4. 21** Menghubungkan operator *performance*

Setelah semuanya terhubung, klik ikon *run* pada *toolbar*, seperti pada gambar 4.22 untuk menampilkan hasilnya. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



**Gambar 4. 22** Menjalankan *Rapidminer*

Setelah beberapa detik maka *RapidMiner* akan menampilkan hasil prediksi pada *view result*. Hasilnya berbentuk tabel seperti gambar 4.23.



Row No.	Lulus	prediction(Lulus)	confidence(Yes)	confidence(No)	Gender_Na...	High_school...	IPK	Pulau
1	No	Yes	0.895	0.105	Laki-Laki	IPA	3	E
2	No	No	0.002	0.998	Laki-Laki	IPA	1	C
3	No	Yes	0.839	0.161	Laki-Laki	KEJURUAN	3	B
4	No	No	0.411	0.589	Laki-Laki	IPA	2	A
5	No	No	0.193	0.807	Laki-Laki	IPS	2	A
6	No	Yes	0.960	0.040	Perempuan	IPA	3	C
7	No	Yes	0.956	0.044	Perempuan	IPA	3	B
8	No	Yes	0.956	0.044	Perempuan	IPA	3	B
9	No	Yes	0.956	0.044	Perempuan	IPA	3	B
10	No	Yes	0.953	0.047	Perempuan	IPA	3	A
11	No	Yes	0.880	0.120	Perempuan	IPS	3	B
12	No	No	0.222	0.778	Laki-Laki	IPS	2	E
13	No	Yes	0.956	0.044	Perempuan	IPA	3	B
14	No	Yes	0.723	0.277	Laki-Laki	IPS	3	B
15	No	No	0.456	0.544	Laki-Laki	IPA	2	C
16	No	Yes	0.956	0.044	Perempuan	IPA	3	B
17	No	Yes	0.960	0.040	Perempuan	IPA	3	C

**Gambar 4. 23** Hasil perhitungan

Dari hasil proses perhitungan menggunakan *rapidminer* dengan metode prediksi menampilkan hasil dari data *testing* yang telah diuji dengan data *training* berupa kolom *prediction* (lulus). Kolom ini memberikan informasi tentang data mahasiswa yang diambil dari data *testing* yang dapat diprediksi lulus tepat waktu (yes) dan mahasiswa yang tidak dapat lulus tepat waktu (no). Selanjutnya untuk mengetahui tingkat *accuracy* algoritma *naïve bayes*, klik tab *performanceVector* yang terletak disebelah kanan, sehingga menampilkan *tabel view* seperti gambar 4.24.

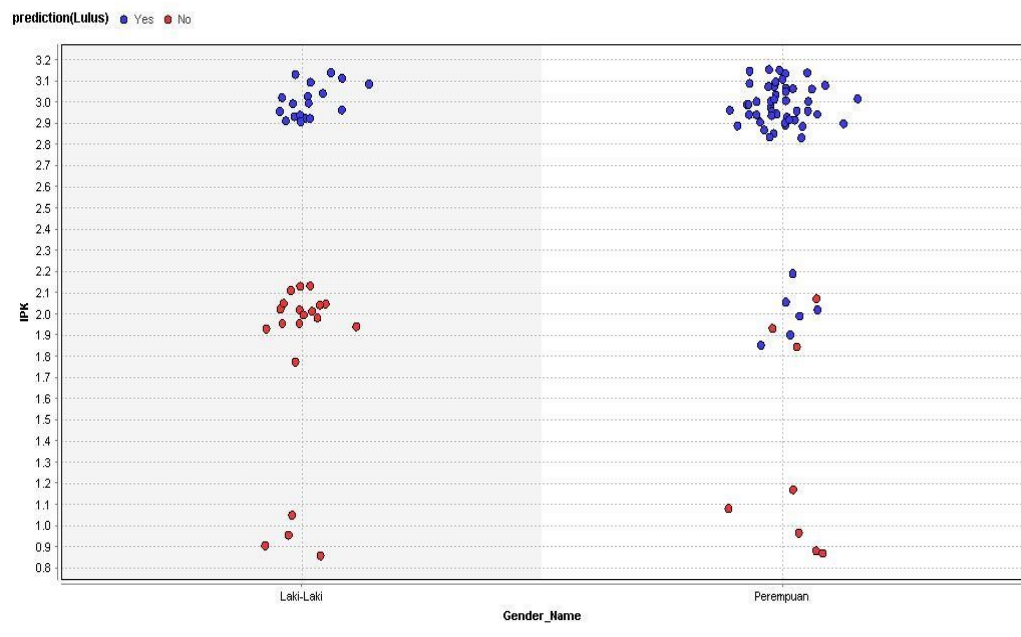
Table View Plot View

accuracy: 59.00%

	true No	true Yes	class precision
pred. No	18	10	64.29%
pred. Yes	31	41	56.94%
class recall	36.73%	80.39%	

**Gambar 4. 24 Accuracy**

Dapat dilihat pada gambar 4.24 tingkat *accuracy* dari *performance vector* yaitu 59.00%.

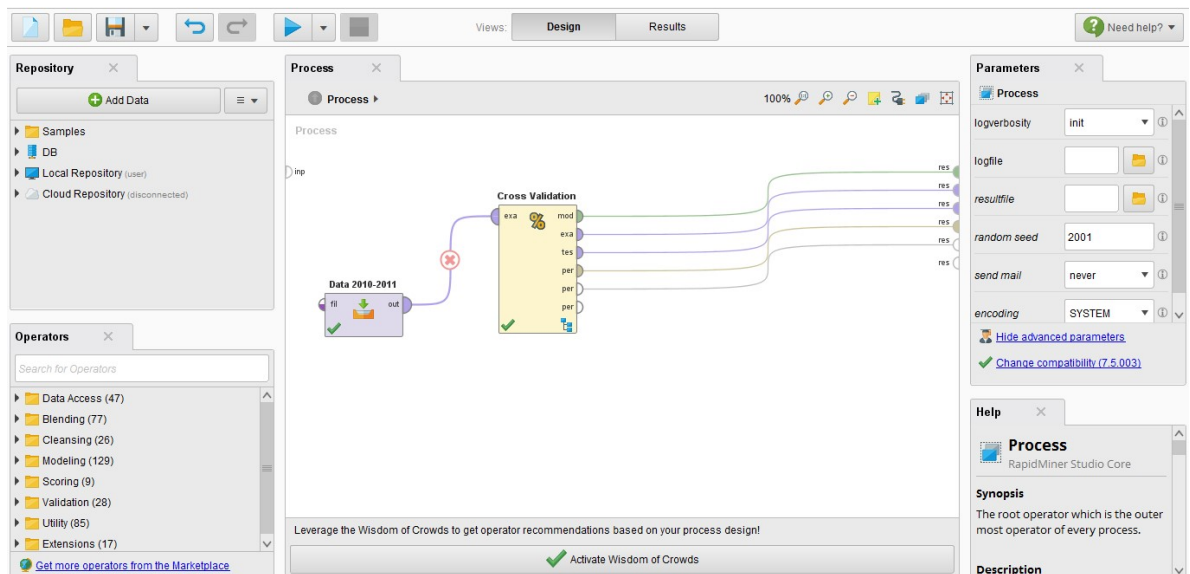


**Gambar 4. 25 Grafik Scatter**

Berdasarkan gambar 4.25 di atas bahwa jenis kelamin perempuan yang memiliki IPK 3 probabilitasnya lebih besar untuk diprediksi lulus tepat waktu dibandingkan laki-laki.

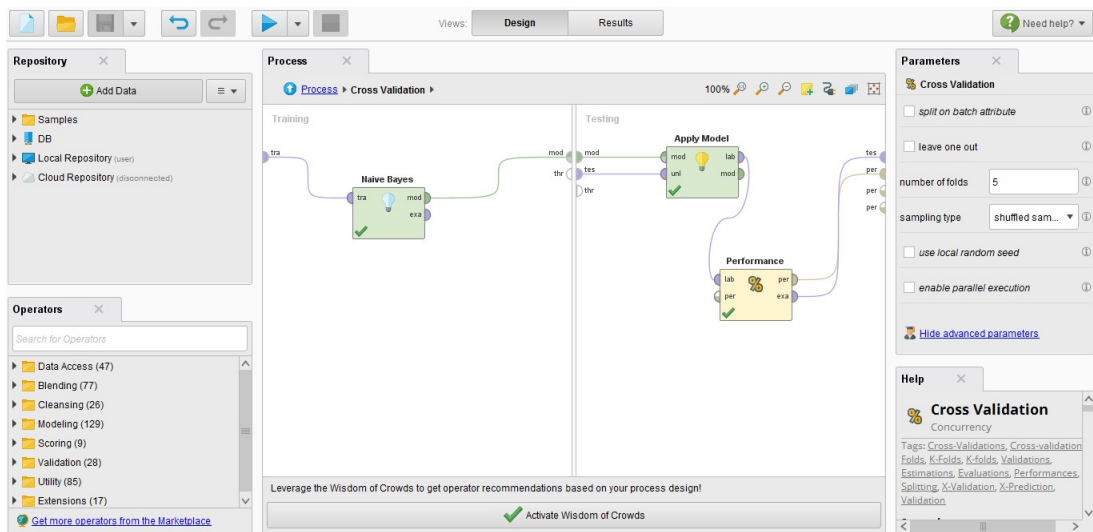


Pengujian kedua menggunakan data angkatan 2010-2011 yang telah lulus dan belum lulus sebagai *training* dan *testing*. Data tersebut kemudian diujikan dengan menggunakan operator *cross validation* di *rapidminer* seperti gambar 4.26



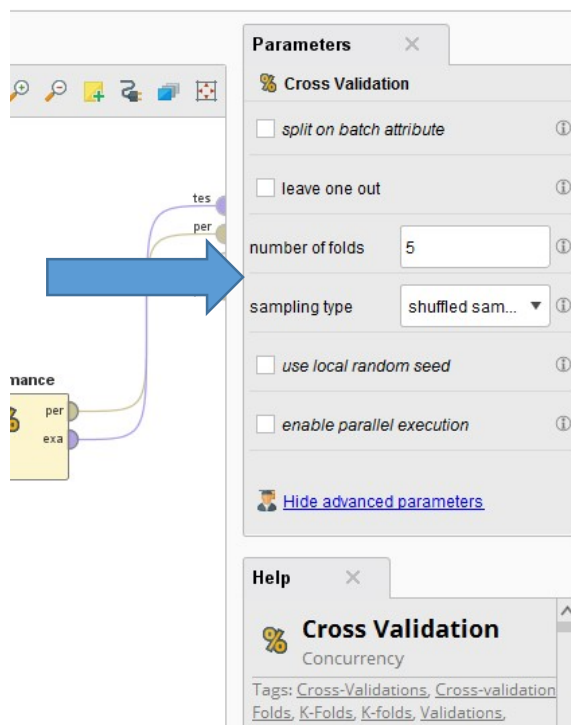
**Gambar 4. 26** *Cross validation*

Selanjutnya adalah mengatur operator *cross validation* dengan cara *double* klik operator *cross validation*. Kemudian *drag and drop* operator *naive bayes* dan hubungkan *port tra* dan *mod* pada kolom *training* . Selanjutnya *drag and drop* operator *apply model* dan *performance*, kemudian hubungkan *port mod*, *tes* pada operator *apply model* dan *lab* ke operator *performance* pada kolom *testing*. Pada operator *performance* hubungkan *port per* ke *port per* dan *port exa* dengan *port tes* pada kolom *testing* seperti gambar 4.27.



**Gambar 4. 27** *Cross validation view*

Langkah selanjutnya adalah memasukkan angka 5 pada kolom *number of fold* dan pilih *shuffled sampling* pada kolom *sampling type* seperti gambar 4.28.



**Gambar 4. 28** *Parameters cross validation*

Setelah semuanya telah terhubung dan *parameter operator cross validation* terisi, klik ikon *run* pada *toolbar* seperti gambar 4.29 untuk menampilkan hasilnya. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



**Gambar 4. 29** Menjalankan *rapidminer*

Setelah beberapa detik maka *rapidminer* akan menampilkan hasil perhitungan pada *view result*. Hasilnya berbentuk seperti gambar 4.30.

PerformanceVector (Performance)

accuracy: 77.35% +/- 5.18% (mikro: 77.37%)

	true Yes	true No	class precision
pred. Yes	283	55	83.73%
pred. No	55	93	62.84%
class recall	83.73%	62.84%	

**Gambar 4. 30** Hasil *shuffled sampling*

Dapat dilihat dari gambar 4.30 nilai akurasi meningkat ketika menggunakan operator *cross validation* dengan jumlah 5 *fold* dibandingkan dengan pengujian pertama menggunakan *data testing* angkatan 2014. Langkah selanjutnya adalah membandingkan tingkat akurasi tertinggi menggunakan metode *sampling type* yang berbeda dengan 5 *fold* dan hasilnya sebagai berikut:

Metode	Nilai Akurasi
<b>Automatic</b>	76.13 %
<b>Shuffled Sampling</b>	77.16 %
<b>Stratified Sampling</b>	76.13 %

**Tabel 3** Perbandingan akurasi

#### 4.7 Algoritma Naïve Bayes

Dalam penelitian algoritma yang digunakan adalah *Naïve bayes* dan untuk memudahkan menjelaskan tentang algoritma *Naïve bayes* dalam penelitian ini maka penulis membuat tabel data yang digunakan dalam penelitian seperti tabel 4.

**Tabel 4** Data *training* mahasiswa angkatan jurusan HI 2010 – 2011

No	Gender	Jurusan	IPK	Pulau	Lulus
1	Perempuan	[Tidak Ada	3	A	Yes
2	Perempuan	IPA	3	C	Yes
3	Laki-Laki	IPS	2	B	Yes

**Table 4.** Data *training* mahasiswa angkatan jurusan HI 2010 – 2011

No	Gender	Jurusan	IPK	Pulau	Lulus
4	Perempuan	[Tidak Ada	3	B	Yes
5	Perempuan	IPA	3	B	Yes
6	Laki-Laki	IPA	2	B	Yes
7	Laki-Laki	IPS	3	D	Yes
8	Laki-Laki	IPS	3	B	Yes
9	Perempuan	AGAMA	3	B	Yes
10	Perempuan	IPS	3	C	Yes
11	Laki-Laki	IPS	2	B	Yes
12	Perempuan	IPA	2	E	Yes
13	Perempuan	KEJURUAN	2	E	Yes
..	Perempuan	[Tidak Ada	3	B	Yes
..	Laki-Laki	IPS	3	B	Yes
..	Laki-Laki	IPS	2	E	Yes
..	Laki-Laki	IPS	3	B	Yes
..	Laki-Laki	[Tidak Ada	3	B	Yes
..	Perempuan	IPS	2	B	Yes
435	Laki-Laki	IPA	3	B	Yes

**Tabel 5** Data *testing* mahasiswa angkatan jurusan HI 2011 – 2014

No	Gender	Jurusan	IPK	Pulau	Lulus
1	Laki-Laki	IPA	3	E	No

**Table 5** Data *testing* mahasiswa angkatan jurusan HI 2011 – 2014

No	Gender	Jurusan	IPK	Pulau	Lulus
2	Laki-Laki	IPA	1	C	No
3	Laki-Laki	KEJURUAN	3	B	No
4	Laki-Laki	IPA	2	A	No
5	Laki-Laki	IPS	2	A	No
6	Perempuan	IPA	3	C	No
7	Perempuan	IPA	3	B	No
8	Perempuan	IPA	3	B	No
9	Perempuan	IPA	3	B	No
10	Perempuan	IPA	3	A	No
11	Perempuan	IPS	3	B	No
12	Laki-Laki	IPS	2	E	No
...	Perempuan	IPA	3	B	No
...	Laki-Laki	IPS	3	B	No
...	Laki-Laki	IPA	2	C	No
...	Perempuan	IPA	3	B	No
...	Perempuan	IPA	3	C	No
...	Perempuan	IPA	1	B	No
100	Perempuan	IPA	2	B	No

Tabel 4 diatas memiliki 435 *record* sebagai data *training* dan tabel 4 diatas memiliki 100 *record* sebagai data *testing* yang digunakan dalam penelitian ini.

Dalam kasus yang ada pada tabel 4 dan 5 akan dibuat prediksi untuk menentukan mahasiswa yang ada pada tabel 4 dapat lulus tepat waktu atau tidak menggunakan probabilitas. Untuk menghitung nilai probabilitas atau prediksi mahasiswa yang dapat lulus tepat waktu dan tidak tepat waktu dengan menggunakan cara seperti berikut:

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

Pertama menghitung jumlah dari *Yes* dan *No* dari tabel data *training*. Dari tabel data *training* didapatkan hasil sebagai berikut:

- Yes = 287
- No = 148

Setelah mengetahui jumlah mahasiswa yang lulus tepat waktu dan lulus tidak tepat waktu langkah selanjutnya adalah menghitung probabilitas masing-masing *attribute* dari data *testing* seperti berikut ini:

- *Gender* = Laki-laki, *Jurusan* = IPA, *IPK* = 3, *Pulau* = E,
- $P (\text{Lulus} = \text{Yes} / \text{Jumlah mahasiswa data Training}) = 287 / 435 = 0.660$
- $P (\text{Gender} = \text{Laki-laki} = \text{Yes} / \text{Tepat waktu}) = 138 / 287 = 0.481$
- $P (\text{Jurusan} = \text{IPA} = \text{Yes} / \text{Tepat waktu}) = 120 / 287 = 0.418$
- $P (\text{Pulau} = \text{E} = \text{Yes} / \text{Tepat waktu}) = 20 / 287 = 0.070$

Untuk menghitung nilai IPK yang memiliki tipe data yang kontinyu menggunakan rumus distribusi *Gaussian* seperti dibawah ini

$$\frac{1}{\sqrt{2 \pi \sigma}} e^{-\frac{(x-\mu)^2}{2 \sigma^2}}$$

Keterangan:

$\mu$  = Mean

$\sigma$  = Standar deviasi

Selanjutnya gantikan nilai x pada rumus diatas menjadi angka dari data *atribut* IPK yang dicari menjadi seperti dibawah ini:

$$\frac{1}{\sqrt{2 \pi 0.432}} e^{-\frac{(3-2.753)^2}{2 0.432^2}}$$

$$0.60697135032 \times 0.84920528533 = 0.51544327873$$

Selanjutnya menghitung probabilitas No dengan menggunakan data *testing* yang sama seperti perhitungan Yes diatas:

- *Gender* = Laki-laki, *Jurusan* = IPA, *IPK* = 3, *Pulau* = E,
- $P(\text{Lulus} = \text{No} / \text{Jumlah mahasiswa data Training}) = 148 / 435 = 0.340$



- $P(\text{Gender} = \text{Laki-laki} = \text{No} / \text{Tidak tepat waktu}) = 107 / 148 = 0.723$
- $P(\text{Jurusan} = \text{IPA} = \text{No} / \text{Tidak tepat waktu}) = 41 / 148 = 0.277$
- $P(\text{Pulau} = \text{E} = \text{No} / \text{Tidak tepat waktu}) = 9 / 148 = 0.061$

$$\frac{1}{\sqrt{2\pi} \cdot 0.432} e^{-\frac{(3 - 2.753)^2}{2 \cdot 0.432^2}}$$

$$0.46470390492 \times 0.37874081899 = 0.17600233754$$

Langkah selanjutnya menghitung semua nilai dari masing-masing *attribute* *Yes* dan *No*, kemudian melakukan normalisasi untuk masing-masing kelas seperti berikut ini:

$$\text{Yes} = 0.660 \times 0.481 \times 0.418 \times 0.070 \times 0.51544327873 = 0.00478789055$$

$$\text{No} = 0.340 \times 0.723 \times 0.277 \times 0.061 \times 0.17600233754 = 0.00073104692$$

### Proses perhitungan *confidence*

$$\text{Probabilitas Yes} = \frac{0.00478789055}{0.00478789055 + 0.00073104692} = 0.867$$

$$\text{Probabilitas No} = \frac{0.00073104692}{0.00478789055 + 0.00073104692} = 0.132$$

Dari perhitungan diatas diketahui bahwa mahasiswa dengan kategori Laki-laki, Jurusan IPA, IPK 3, Pulau E diprediksi lulus tepat waktu karena hasil perhitungan kelas *Yes* lebih besar dari kelas *No* dengan hasil probabilitas sebesar 0.867. Untuk perhitungan dan hasil lainnya akan dilampirkan pada lampiran.

