

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Hasil**

Penelitian ini menghasilkan sebuah klasifikasi dan sebuah informasi dalam penentuan lokasi promosi mahasiswa baru pada IIB Darmajaya Lampung. Dalam pengklasifikasian menggunakan tiga algoritma yang setara yaitu, *decision tree*, *naïve bayes*, dan *k-nearest neighbor* dengan menggunakan ketiga algoritma tersebut diharapkan mendapatkan algoritma yang memiliki nilai akurasi yang cukup baik sehingga informasi yang dihasilkan dapat digunakan sebagai bahan pertimbangan bagian pemasaran untuk melakukan promosi pada tahun-tahun berikutnya.

##### **4.1.1 Objek Penelitian**

Penelitian ini menggunakan data mahasiswa baru pada tahun 2016 hingga 2020 dengan semua jurusan. Data yang digunakan merupakan data yang terisi lengkap dalam basis data. Penelitian ini dilakukan di bagian pemasaran IIB Darmajaya Lampung, karena selama ini bagian pemasaran IIB Darmajaya menentukan lokasi promo secara manual dan kurang tepat sasaran.

##### **4.1.2 Pengumpulan Data**

Pengumpulan data dilakukan dengan meminta izin secara tertulis terlebih dahulu kebagian ICT Darmajaya, setelah mendapat balasan dari pihak terkait maka diizinkan untuk melakukan penelitian di bagian pengolahan data, dalam hal ini bagian ICT dan BAAK (Biro Administrasi Akademik Kemahasiswaan) Darmajaya. Kemudian data yang diminta diberikan melalui surat elektronik (electronic mail), data yang diterima berjumlah 1 file yang berisikan data mahasiswa baru tahun 2016 hingga 2020.

Jumlah data yang didapat 5685 record terdiri dari beberapa kolom atau variabel yaitu:

1. Nama
2. Jurusan
3. Jenis Kelamin

4. Kabupaten
5. Kecamatan
6. Sekolah
7. Pekerjaan Orang Tua

Jumlah data mahasiswa yang didapat 5685 mahasiswa namun, tidak semua data memiliki record atau tidak terekam dengan baik.

#### **4.1.3 Pre-Processing Data**

*Pre processing* merupakan tahapan awal dari *Data Mining*, pada tahapan ini data diolah menjadi data yang berkualitas agar data dapat diolah ketahapan selanjutnya. Dalam tahapan ini terdapat proses yang sangat penting dan teliti karena tahapan *pre processing* membutuhkan data yang benar dalam formatnya, konsisten dalam penulisan, tidak ada duplikasi data, tidak ada data yang kosong, dan sebagainya. Data yang tidak berkualitas maka akan menghasilkan informasi yang tidak akurat. Data mentah yang sudah diterima terdapat banyak data yang kurang bagus, seperti nilai angka yang error karena format yang salah, kehilangan salah satu huruf (typo), singkatan yang tidak konsisten, duplikasi data, data yang dibutuhkan berada di file yang berbeda. Dalam tahapan ini beberapa proses yaitu:

##### **4.1.3.1 Data Reduction**

Data reduction merupakan proses yang bertujuan untuk mengurangi atau mereduksi, atribut, dimensi, atau sejumlah data yang tidak dibutuhkan. Data reduction sangat dibutuhkan untuk mendapatkan atribut yang saling berkaitan.

Dalam penelitian ini atribut atau variabel yang dibutuhkan terdiri dari nama, kabupaten atau kota, pekerjaan orang tua, jumlah alumni yang mendaftar di IIB Darmajaya, dan asal sekolah mahasiswa baru. Dalam file yang diterima oleh peneliti maka dilakukan analisa kemudian menghasilkan banyak atribut yang tidak dibutuhkan dalam penelitian ini. Peneliti mengurangi atribut jenis kelamin, jurusan, tahun akademik dan kecamatan, penulis mereduksi atribut tersebut berpendapat bahwa atribut tersebut tidak terlalu berpengaruh terhadap proses pengklasifikasian.

Setelah melakukan reduksi atribut dilakukan reduksi data mahasiswa baru

angkatan 2016 dan 2017. Sehingga hasil akhir dari proses reduksi mendapatkan data mahasiswa baru tahun 2018, 2019, dan 2020 dengan jumlah 1281 data mahasiswa.

#### 4.1.3.2 Data Cleaning

Data Cleaning merupakan proses dalam mengisi missing value atau isian yang hilang, meminimalisir terjadinya penumpukan data atau duplikasi data dan mengecek kembali data yang tidak konsisten.

Proses cleaning peneliti menggunakan aplikasi *Microsoft Excel* dengan memfilter kolom satu perasatu dan mencari data yang kosong atau missing. Kolom sekolah terdapat banyak data yang tidak ter-record dengan baik, sehingga banyak data yang tidak dapat digunakan. Sebagai contoh pada kolom sekolah tahun 2016 dan 2017 tidak ada satupun asal sekolah yang tersimpan oleh database IIB Darmajaya sehingga data tersebut tidak dapat digunakan sebagai data set. Penulis menggunakan data 3 tahun terakhir yaitu, 2018,2019, dan 2020.

#### 4.1.3.3 Perubahan Data

Perubahan data atau data transformasi merupakan proses perubahan data agar mendapatkan data yang berkualitas dan sesuai dengan yang dibutuhkan. Dalam tahapan ini peneliti melakukan tranformasi data yang berdasar hasil wawancara (06/01/2021) dengan kepala bagian pemasaran IIB Darmajaya dengan hasil sebagai berikut:

**Tabel 4.1** Transformasi kabupaten dan kota.

Kabupaten/Kota	Dalam Kota
Kabupaten Lampung Barat	Tidak
Kabupaten Lampung Selatan	
Kabupaten Lampung Tengah	
Kabupaten Lampung Timur	
Kabupaten Lampung Utara	
Kabupaten Mesuji	
Kabupaten Pesawaran	
Kabupaten Pesisir Barat	
Kabupaten Pringsewu	
Kabupaten Tanggamus	
Kabupaten Tulang Bawang	
Kabupaten Tulang Bawang Barat	
Kabupaten Way Kanan	

Kota Bandar Lampung	Iya
Kota Metro	

**Tabel 4.2** Penghasilan Orang tua

Pekerjaan Orang Tua	Kelompok Penghasilan
Anggota Dewan	Tinggi
Pegawai BUMN	
Pensiunan	
PNS	Sedang
Wiraswasta	
TNI/Polri	
Guru	
Karyawan	Rendah
Petani	
Swasta	

**Tabel 4.3** Asal Sekolah.

No.	Asal Sekolah	Kelompok Asal Sekolah
1.	Sman Bhakti Mulya Bunga Mayang	Negeri
2.	Sman Ketibung	Negeri
3.	Smas Al Huda	Swasta
4.	Smas Assalam	Swasta
5.	Smas Bhakti Mulya	Swasta
6.	Smas Bpk Penabur	Swasta
7.	Smas Darul Fikri	Swasta
8.	Smas It Ar Raihan	Swasta
9.	Smas It Baitul Muslim	Swasta
10.	Sman 10 Bandar Lampung	Negeri
....	.....	.....
304.	Man 2 Bandar Lampung	Negeri

Dalam pengelompokan total alumni menggunakan aturan sebagai berikut:

- jika asal sekolah berada di dalam kota menggunakan 3 tingkatan yaitu,  $\geq 15$ ,  $\geq 9$  dan  $\leq 8$ .
- jika asal sekolah berada di luar kota menggunakan 3 tingkatan yaitu,  $\geq 12$ ,  $\geq 7$  dan  $< 7$ .

**Tabel 4.4** Total Alumni

No.	Asal Sekolah	Jumlah Alumni	Hasil Pengelompokan
1.	Smk Trisakti Jaya	2	$\leq 7$
2.	Smk Yadika Pagelaran	7	$\leq 7$

3.	Smkn 1 Abung Selatan	3	$\leq 7$
4.	Smkn 1 Banjit	3	$\leq 7$
5.	Smkn 1 Bukit Kemuning	3	$\leq 7$
6.	Smkn 1 Gadingrejo	3	$\leq 7$
7.	Smkn 1 Kalianda	8	$\leq 8$
8.	Smkn 1 Kotabumi	5	$\leq 7$
9.	Smkn 1 Metro	2	$\leq 7$
10.	Smkn 1 Ngambur	2	$\leq 7$
....	.....	.....	.....
304.	Sman 3 Metro	11	$\leq 8$

Tabel 4.5 Hasil Transformasi

No.	Nama	Dalam Kota	Total Alumni	Penghasilan Orang Tua	Asal Sekolah
1.	Aldo Erlansyah	tidak	$\geq 7$	sedang	swasta
2.	Aldo Fernando	iya	$\geq 9$	sedang	swasta
3.	Aldo Kusuma Natta	iya	$\geq 15$	sedang	swasta
4.	Aldo Valerio S	iya	$\geq 15$	sedang	swasta
5.	Aldy Bagus P.	iya	$\geq 15$	sedang	negeri
6.	Alfachrie Rachman	iya	$\geq 15$	sedang	negeri
7.	Alfayet Rifki A.	iya	$\leq 8$	sedang	swasta
8.	Alfin Rifaldo	tidak	$< 7$	sedang	negeri
9.	Alfira Dewi R.	tidak	$< 7$	rendah	negeri
10.	Alfito Dinova	tidak	$< 7$	sedang	negeri
....	.....	.....	.....	.....	.....
1281.	Zuzlifatul Adnan	tidak	$< 7$	rendah	negeri

#### 4.1.3.4 Penentuan Label

Dalam penentuan label pada penelitian ini, menggunakan dasar sebagai berikut:

Dalam Kota

Sangat Potensi, jika lokasi sekolah berada di dalam kota dan jumlah alumni  $\geq 15$ .

Potensi, jika lokasi sekolah berada di dalam kota dan jumlah alumni  $\geq 9$ , atau jumlah alumni  $\leq 8$  namun berasal dari sekolah negeri, atau jumlah alumni  $\leq 8$  tetapi penghasilan orang termasuk kedalam golongan sedang.

Kurang Potensi, jika lokasi sekolah berada di dalam kota, jumlah alumni  $\leq 8$  dan penghasilan orang tua tinggi, atau penghasilan orang tua rendah serta berasal dari sekolah swasta.

Luar Kota

Sangat Potensi, jika lokasi sekolah berada di luar kota dan jumlah alumni  $\geq 12$ .

Potensi, jika lokasi sekolah berada di luar kota dan jumlah alumni  $\geq 7$ , atau

jumlah alumni  $<7$  namun penghasilan orang tua dalam kategori sedang, atau jumlah alumni  $<7$ , penghasilan orang tua rendah, namun berasal dari sekolah negeri.

Kurang Potensi, jika lokasi sekolah berada di luar kota, jumlah alumni  $<7$  dan penghasilan orang tua tinggi, atau jumlah alumni  $<7$  penghasilan orang tua rendah serta berasal dari sekolah swasta.

Berdasarkan aturan-aturan di atas maka didapat hasil data dari *pre-processing* sebagai berikut:

**Tabel 4.6** Hasil Pelabelan

No.	Nama	Dalam Kota	Total Alumni	Penghasilan Orang Tua	Asal Sekolah	Ket.
1.	Aldo Erlansyah	tidak	$\geq 7$	sedang	swasta	P
2.	Aldo Fernando	iya	$\geq 9$	sedang	swasta	P
3.	Aldo Kusuma N.	iya	$\geq 15$	sedang	swasta	SP
4.	Aldo Valerio S	iya	$\geq 15$	sedang	swasta	SP
5.	Aldy Bagus P.	iya	$\geq 15$	sedang	negeri	SP
6.	Alfachrie R.	iya	$\geq 15$	sedang	negeri	SP
7.	Alfayet Rifki A.	iya	$\leq 8$	sedang	swasta	P
8.	Alfin Rifaldo	tidak	$< 7$	sedang	negeri	P
9.	Alfira Dewi R.	tidak	$< 7$	rendah	negeri	P
10.	Alfito Dinova	tidak	$< 7$	sedang	negeri	P
....	.....	.....	.....	.....	.....	....
1281.	Zuzlifatul Adnan	tidak	$< 7$	rendah	negeri	P

Keterangan:

SP : Sangat Potensi.

P : Potensi

KP : Kurang Potensi

#### 4.1.4 Penerapan Algoritma *Decision tree* Menggunakan *Microsoft Excel*

Penerapan algoritma *Decision tree* menggunakan *Microsoft Excel* membutuhkan beberapa tahapan, untuk mempermudah penghitungan nilai *entropy* maka dilakukan pengelompokan data dari keseluruhan, sehingga didapat data sebagai berikut:

**Tabel 4.7** Pengelompokan data untuk perhitungan nilai *Entropy*

	jumlah	Sangat potensi	Potensi	Kurang Potensi
--	--------	----------------	---------	----------------

Total		1281	798	374	110
Dalam kota	ya	686	514	143	29
	tidak	595	13	502	80
Penghasilan	tinggi	53	28	10	15
	sedang	878	418	460	0
	rendah	359	81	175	94
Sekolah	swasta	452	210	139	103
	negeri	829	317	506	6
Total	$\geq 15$	424	424	0	0
	$\geq 9$	129	0	129	0
	$\leq 8$	133	90	14	29
	$\geq 12$	13	13	0	0
	$\geq 7$	127	0	127	0
	$< 7$	455	271	375	80

Setelah dilakukan pengelompokan, maka tahapan selanjutnya yaitu, menentukan akar dari pohon berdasarkan nilai *index entropy* terendah atau menghitung nilai *gain* yang tertinggi dari masing-masing atribut.

Nilai *indek entropy* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (2) yaitu:

a. Formula untuk mendapatkan *entropy*

$$=((- \text{jumlah sangat potensi} / \text{total keseluruhan}) * \text{IMLOG2}(\text{jumlah sangat potensi} / \text{total keseluruhan}) + (- \text{jumlah potensi} / \text{total keseluruhan}) * \text{IMLOG2}(\text{jumlah potensi} / \text{total keseluruhan}) + (- \text{jumlah kurang potensi} / \text{total keseluruhan}) * \text{IMLOG2}(\text{jumlah kurang potensi} / \text{total keseluruhan}))$$

$$=((-D2/C2) * \text{IMLOG2}(D2/C2) + (-E2/C2) * \text{IMLOG2}(E2/C2) + (-F2/C2) * \text{IMLOG2}(F2/C2))$$

Setelah dilakukan perhitungan nilai *entropy* maka akan didapat data sebagai berikut:

**Tabel 4.8** Hasil Perhitungan nilai *entropy* pertama

	Jumlah	Sangat potensi	Potensi	Kurang potensi	Entropy
Total	1281	798	374	110	1,24805047
Dalam kota					
ya	686	514	143	29	0,97653405
tidak	595	13	502	80	0,7166201
Penghasilan					
tinggi	53	28	10	15	1,45568314
sedang	878	418	460	0	0
rendah	359	81	175	94	1,49616397
Sekolah					
swasta	452	210	139	103	1,52320126
negeri	829	317	506	6	1,01652142
Total Alumni					
≥15	424	424	0	0	0
≥9	129	0	129	0	0
≤8	133	90	14	29	1,20226592
≥12	13	13	0	0	0
≥7	127	0	127	0	0
<7	455	271	375	80	1,116114

Setelah mendapatkan nilai *entropy* maka dilanjutkan dengan menghitung nilai *gain*. Nilai *gain* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (3) yaitu:

b. Formula untuk mendapatkan nilai *gain*

$=(\text{entropy total} - ((\text{jumlah ya} / \text{total keseluruhan}) * \text{entropy ya}) - ((\text{jumlah tidak} / \text{total keseluruhan}) * \text{entropy tidak}))$

$$=(G3) - ((C4/C2)*G4) - ((C5/C2)*G5)$$

Sehingga didapat nilai *gain* sebagai berikut:

**Tabel 4.9** hasil perhitungan nilai *gain* node 1

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi	Entropy	Gain
Total	1281	798	374	110	1,24805047	
Dalam Kota						0,392241
Ya	686	514	143	29	0,97653405	
Tidak	595	13	502	80	0,7166201	

Penghasilan						0,768523
Tinggi	53	28	10	15	1,45568314	
Sedang	878	418	460	0	0	
Rendah	359	81	175	94	1,49616397	
Sekolah						0,052747
Swasta	452	210	139	103	1,52320126	
Negeri	829	317	506	6	1,01652142	
Total Alumni						1,123225
	$\geq 15$	424	424	0	0	0
	$\geq 9$	129	0	129	0	0
	$\leq 8$	133	90	14	29	1,20226592
	$\geq 12$	13	13	0	0	0
	$\geq 7$	127	0	127	0	0
	$< 7$	455	271	375	80	1,116114

Setelah mendapatkan nilai *gain* dari data tersebut, maka dapat dilihat nilai *gain* yang terbesar. Nilai *gain* yang terbesar akan menjadi node pertama. Berdasarkan data di atas, nilai *gain* terbesar terletak pada variabel total alumni.

Dalam menentukan node berikutnya, maka dapat dibuat tabel kembali untuk menentukan nilai *entropy* dan *gain* pada variabel total alumni.

**Tabel 4.10** Pengelompokan data perhitungan nilai *Entropy* perulangan kedua

		Jumlah	Sangat Potensi	Potensi	Kurang Potensi
Total		1281	798	374	110
Dalam Kota					
	Iya	133	90	14	29
	Tidak	0	0	0	0
Penghasilan Orang tua					
	Tinggi	8	0	0	8
	Sedang	90	90	0	0
	Rendah	35	0	14	21
Sekolah					
	Negeri	40	24	14	2
	Swasta	93	66	0	27

Setelah dilakukan pengelompokan, maka tahapan selanjutnya kembali menentukan nilai *entropy*. Nilai *entropy* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (2). Setelah dilakukan perhitungan nilai *entropy* maka akan didapat data sebagai

berikut:

**Tabel 4.11** Hasil Perhitungan Nilai *entropy* perulangan kedua

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi	<i>entropy</i>
Total	1281	798	374	110	1,24805047
Dalam Kota					
Iya	133	90	14	29	1,202266
Tidak	0	0	0	0	0
Penghasilan Orang tua					
Tinggi	8	0	0	8	0
Sedang	90	90	0	0	0
Rendah	35	0	14	21	0
Sekolah					
Negeri	40	24	14	2	1,188376
Swasta	93	66	0	27	0

Setelah mendapatkan nilai *entropy* maka dilanjutkan dengan menghitung nilai *gain*. Nilai *gain* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (3). Sehingga didapat nilai *gain* sebagai berikut:

**Tabel 4.12** Hasil perhitungan nilai *Gain* node 1.1

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi	<i>entropy</i>	<i>gain</i>
Total	1281	798	374	110	1,24805047	
Dalam Kota						1,123225
Iya	133	90	14	29	1,202266	
Tidak	0	0	0	0		
Penghasilan Orang tua						1,24805
Tinggi	8	0	0	8	0	
Sedang	90	90	0	0	0	
Rendah	35	0	14	21	0	
Sekolah						1,210943
Negeri	40	24	14	2	1,188376	
Swasta	93	66	0	27	0	

Setelah mendapatkan nilai *gain* dari data di atas, maka dapat dilihat nilai *gain* yang terbesar. Nilai *gain* yang terbesar akan menjadi node kedua. Berdasarkan data di atas, nilai *gain* terbesar terletak pada variabel penghasilan orang tua.

Dalam menentukan node berikutnya, maka dapat dibuat tabel kembali untuk menentukan nilai *entropy* dan *gain* pada variabel penghasilan orang tua.

**Tabel 4.13** Pengelompokan data perhitungan nilai *Entropy* perulangan ketiga

		Jumlah	Sangat Potensi	Potensi	Kurang Potensi
Total		1281	798	374	110
Dalam Kota					
	Iya	133	90	14	29
	Tidak	0	0	0	0
Sekolah					
	Negeri	40	24	14	2
	Swasta	93	66	0	27

Setelah dilakukan pengelompokan, maka tahapan selanjutnya kembali menentukan nilai *entropy*. Nilai *entropy* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (2).

Setelah dilakukan perhitungan nilai *entropy* maka akan didapat data sebagai berikut:

**Tabel 4.14** Hasil perhitungan nilai *Entropy* perulangan ketiga

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi	entropy
Total	1281	798	374	110	1,24805047
Dalam Kota					
Iya	133	90	14	29	1,202266
Tidak	0	0	0	0	0
Sekolah					
Negeri	40	24	14	2	1,188376
Swasta	93	66	0	27	0

Setelah mendapatkan nilai *entropy* maka dilanjutkan dengan menghitung nilai *gain* kembali, Nilai *gain* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (3). Sehingga didapat nilai *gain* sebagai berikut:

**Tabel 4.15** Hasil perhitungan nilai *Gain* node 1.1.1

	Jumlah	SP	P	KP	entropy	gain
Total	1281	798	374	110	1,24805047	
Dalam Kota						1,123225
Iya	133	90	14	29	1,202266	
Tidak	0	0	0	0	0	
sekolah						1,210943
Negeri	40	24	14	2	1,188376	
Swasta	93	66	0	27	0	

Setelah mendapatkan nilai *gain* dari data di atas, maka dapat dilihat nilai *gain* yang terbesar. Nilai *gain* yang terbesar akan menjadi node ketiga. Berdasarkan data di atas, nilai *gain* terbesar terletak pada variabel sekolah.

Tahapan selanjutnya yaitu penentuan node 1.2 Berdasarkan tabel 4.8 node 1.2 terletak pada variabel total alumni (<7).

Dalam penghitungan nilai entropy dan gain dapat dibuat tabel sebagai berikut:

**Tabel 4.16** Pengelompokan data perhitungan nilai *Entropy* perulangan ke-4

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi
Total	1281	798	374	110
Dalam Kota				
Iya	0	0	0	0
Tidak	455	271	104	80
Penghasilan Orang tua				
Tinggi	7	0	0	7
Sedang	271	271	0	0
Rendah	177	0	104	73
Sekolah				
Negeri	295	187	104	4
Swasta	160	84	0	76

Setelah dilakukan pengelompokan, maka tahapan selanjutnya kembali menentukan nilai *entropy*. Nilai *entropy* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (2). Setelah dilakukan perhitungan nilai entropy maka akan didapat data sebagai berikut:

**Tabel 4.17** Hasil perhitungan nilai *Entropy* perulangan ke-4

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi	entropy
Total	1281	798	374	110	1,24805
<b>Dalam Kota</b>					
Iya	0	0	0	0	0
Tidak	455	271	104	80	1,372882
<b>Penghasilan Orang tua</b>					
Tinggi	7	0	0	7	0
Sedang	271	271	0	0	0
Rendah	177	0	104	73	0
<b>Sekolah</b>					
Negeri	295	187	104	4	1,0313
Swasta	160	84	0	76	0

Setelah mendapatkan nilai *entropy* maka dilanjutkan dengan menghitung nilai *gain* kembali. Nilai *gain* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (3). Sehingga didapat nilai *gain* sebagai berikut:

**Tabel 4.18** Hasil perhitungan nilai *Gain* node 1.2

	Jumlah	SP	P	KP	entropy	gain
Total	1281	798	374	110	1,24805	
<b>Dalam Kota</b>						0,760415
Iya	0	0	0	0	0	
Tidak	455	271	104	80	1,372882	
<b>Penghasilan Orang tua</b>						1,24805
Tinggi	7	0	0	7	0	
Sedang	271	271	0	0	0	
Rendah	177	0	104	73	0	
<b>Sekolah</b>						1,010554
Negeri	295	187	104	4	1,0313	
Swasta	160	84	0	76	0	

Setelah mendapatkan nilai *gain* dari data di atas, maka dapat dilihat nilai *gain* yang terbesar. Nilai *gain* yang terbesar akan menjadi node kedua. Berdasarkan data di atas, nilai *gain* terbesar terletak pada variabel penghasilan orang tua.

Tahapan selanjutnya yaitu penentuan node 1.2.1 Berdasarkan tabel 4.18 penghitungan nilai *entropy* dan *gain* dapat dibuat tabel sebagai berikut:

**Tabel 4.19** Pengelompokan data perhitungan nilai *Entropy* perulangan ke-5

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi
Total	1281	798	374	110
Dalam Kota				
Iya	0	0	0	0
Tidak	455	271	104	80
Sekolah				
Negeri	295	187	104	4
Swasta	160	84	0	76

Setelah dilakukan pengelompokan, maka tahapan selanjutnya kembali menentukan nilai *entropy*. Nilai *entropy* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (2). Setelah dilakukan perhitungan nilai *entropy* maka akan didapat data sebagai berikut:

**Tabel 4.20** Hasil perhitungan nilai *Entropy* perulangan ke-5

	Jumlah	Sangat Potensi	Potensi	Kurang Potensi	entropy
Total	1281	798	374	110	1,24805
Dalam Kota					
Iya	0	0	0	0	0
Tidak	455	271	104	80	1,372882
Sekolah					
Negeri	295	187	104	4	1,0313
Swasta	160	84	0	76	0

Setelah mendapatkan nilai *entropy* maka dilanjutkan dengan menghitung nilai *gain* kembali, Nilai *gain* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (3). Setelah dilakukan perhitungan nilai *entropy* maka akan didapat data sebagai berikut:

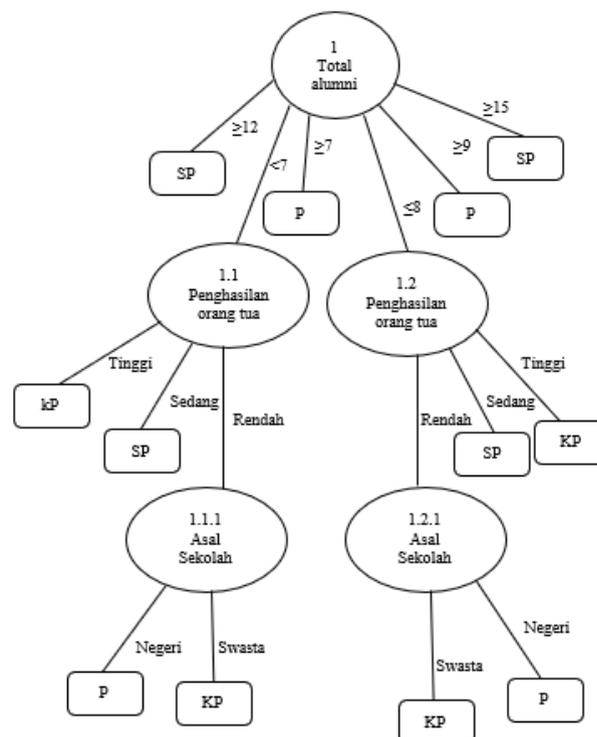
**Tabel 4.21** Hasil perhitungan nilai *Gain* node 1.2.1

	Jumlah	SP	P	KP	entropy	gain
Total	1281	798	374	110	1,24805	
Dalam Kota						0,760415
Iya	0	0	0	0	0	
Tidak	455	271	104	80	1,372882	
Sekolah						1,010554
Negeri	295	187	104	4	1,0313	
Swasta	160	84	0	76	0	

Berdasarkan tabel di atas maka node 1.2.1 terletak pada variabel sekolah.

#### 4.1.5 Hasil Penerapan Algoritma *Decision tree* Menggunakan *Microsoft Excel*

Setelah dilakukan perhitungan manual menggunakan *Microsoft Excel* maka hasil dari algoritma *Decision tree* dapat digambarkan sebagai berikut:



**Gambar 4.1** Hasil Algoritma *Decision tree* Menggunakan *Microsoft Excel*

#### 4.1.6 Penerapan Algoritma Naïve Bayes Menggunakan *Microsoft Excel*

Dalam penerapan algoritma naïve bayes data yang di gunakan sebagai data trening yaitu 70% dan 30% akan digunakan sebagai data *testing*.

$$\frac{70}{100} \times 1281 = 896,7$$

Berdasarkan perhitungan di atas maka data training yang akan digunakan berjumlah 897 *record* dan sisanya akan digunakan sebagai data *testing*.

Tahapan selanjutnya yaitu menghitung jumlah label, untuk mempermudah perhitungan maka dapat dibuat tabel sebagai berikut:

**Tabel 4.22** Menghitung jumlah label

Label	Jumlah Data	Jumlah Seluruh Data	Hasil
P(Y=SP)	625	897	0,69
P(Y=P)	217	897	0,24
P(Y=KP)	55	897	0,06

**Tabel 4.23** Mencari peluang dari variabel dalam kota

Dalam Kota	SP	P	KP
Iya	514	143	29
Tidak	111	74	26
Total	625	217	55

Berdasarkan tabel di atas, maka dapat dihitung masing-masing label dengan menggunakan rumus peluang:

Perhitungan untuk kolom peluang sangat potensi P(SP)

$$\frac{514}{625} = 0,82$$

$$\frac{111}{625} = 0,18$$

Perhitungan untuk kolom peluang potensi P(P)

$$\frac{143}{217} = 0,66$$

$$\frac{74}{217} = 0,34$$

Perhitungan untuk kolom peluang kurang potensi P(KP)

$$\frac{29}{55} = 0,53$$

$$\frac{26}{55} = 0,47$$

Hasil perhitungan peluang dari variabel dalam kota, dapat di buat tabel sebagai berikut:

**Tabel 4.24** Hasil peluang dari variabel dalam kota

Dalam Kota	SP	P	KP	P(SP)	P(P)	P(KP)
Iya	514	143	29	0,82	0,66	0,53
Tidak	111	74	26	0,18	0,34	0,47
Total	625	217	55			

**Tabel 4.25** Mencari peluang dari variabel total alumni

Total Alumni	SP	P	KP
$\geq 15$	424	0	0
$\geq 9$	0	129	0
$\leq 8$	90	14	29
$\geq 12$	4	0	0
$\geq 7$	0	48	0
$< 7$	107	26	26
Total	625	217	55

Berdasarkan tabel di atas, maka dapat dihitung masing-masing label dengan menggunakan rumus peluang:

Perhitungan untuk kolom peluang sangat potensi P(SP)

$$\frac{424}{625} = 0,68$$

$$\frac{0}{625} = 0,00$$

$$\frac{90}{625} = 0,14$$

$$\frac{4}{625} = 0,01$$

$$\frac{0}{625} = 0,00$$

$$\frac{107}{625} = 0,17$$

Perhitungan untuk kolom peluang potensi P(P)

$$\frac{0}{217} = 0,00$$

$$\frac{129}{217} = 0,59$$

$$\frac{14}{217} = 0,06$$

$$\frac{0}{217} = 0,00$$

$$\frac{48}{217} = 0,22$$

$$\frac{26}{217} = 0,12$$

Perhitungan untuk kolom peluang kurang potensi P(KP)

$$\frac{0}{55} = 0,00$$

$$\frac{0}{55} = 0,00$$

$$\frac{29}{55} = 0,53$$

$$\frac{0}{55} = 0,00$$

$$\frac{0}{55} = 0,00$$

$$\frac{26}{55} = 0,47$$

Hasil perhitungan peluang dari variabel total alumni, dapat di buat tabel sebagai berikut:

**Tabel 4.26** Hasil peluang dari variabel total alumni

Total Alumni	SP	P	KP	P(SP)	P(P)	P(KP)
$\geq 15$	424	0	0	0,68	0,00	0,00
$\geq 9$	0	129	0	0,00	0,59	0,00

$\leq 8$	90	14	29	0,14	0,06	0,53
$\geq 12$	4	0	0	0,01	0,00	0,00
$\geq 7$	0	48	0	0,00	0,22	0,00
$< 7$	107	26	26	0,17	0,12	0,47
Total	625	217	55			

**Tabel 4.27** Mencari peluang dari variabel penghasilan orang tua

Penghasilan orang tua	SP	P	KP
Tinggi	27	9	11
Sedang	520	130	0
Rendah	78	78	44
Total	625	217	55

Berdasarkan tabel di atas, maka dapat dihitung masing-masing label dengan menggunakan rumus peluang:

Perhitungan untuk kolom peluang sangat potensi P(SP)

$$\frac{27}{625} = 0,04$$

$$\frac{520}{625} = 0,83$$

$$\frac{78}{625} = 0,12$$

Perhitungan untuk kolom peluang potensi P(P)

$$\frac{9}{217} = 0,04$$

$$\frac{130}{217} = 0,60$$

$$\frac{78}{217} = 0,36$$

Perhitungan untuk kolom peluang kurang potensi P(KP)

$$\frac{11}{55} = 0,20$$

$$\frac{0}{55} = 0,00$$

$$\frac{44}{55} = 0,80$$

Hasil perhitungan peluang dari variabel penghasilan orang tua, dapat di buat tabel sebagai berikut:

**Tabel 4.28** Hasil peluang dari variabel penghasilan orang tua

Penghasilan orang tua	SP	P	KP	P(SP)	P(P)	P(KP)
Tinggi	27	9	11	0,04	0,04	0,20
Sedang	520	130	0	0,83	0,60	0,00
Rendah	78	78	44	0,12	0,36	0,80
Total	625	217	55			

**Tabel 4.29** Mencari peluang dari variabel asal sekolah

Asal Sekolah	SP	P	KP
Negeri	384	165	5
Swasta	241	52	50
Total	625	217	55

Berdasarkan tabel di atas, maka dapat dihitung masing-masing label dengan menggunakan rumus peluang:

Perhitungan untuk kolom peluang sangat potensi P(SP)

$$\frac{384}{625} = 0,61$$

$$\frac{241}{625} = 0,39$$

Perhitungan untuk kolom peluang potensi P(P)

$$\frac{165}{217} = 0,76$$

$$\frac{52}{217} = 0,24$$

Perhitungan untuk kolom peluang kurang potensi P(KP)

$$\frac{5}{55} = 0,09$$

$$\frac{50}{55} = 0,91$$

Hasil perhitungan peluang dari variabel asal sekolah, dapat di buat tabel sebagai berikut:

**Tabel 4.30** Hasil peluang dari variabel asal sekolah

Asal Sekolah	SP	P	KP	P(SP)	P(P)	P(KP)
Negeri	384	165	5	0,61	0,76	0,09
Swasta	241	52	50	0,39	0,24	0,91
Total	625	217	55			

Setelah mencari peluang dari setiap label dan variabel maka tahapan selanjutnya menghitung probabilitas dari setiap *record* dari data *testing*. Nilai peluang pada algoritma *naïve bayes* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (4). Tabel-tabel yang telah dibuat sebelumnya dimasukan kedalam *Microsoft Excel* sehingga akan tampil seperti gambar di bawah ini:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Dalam Kota	SP	P	KP	P(SP)	P(P)	P(KP)		Label	Jumlah Data	Jumlah Seluruh Data	
2	Iya	514	143	29	0,82	0,66	0,53		P(Y=SP)	625	897	0,69
3	Tidak	111	74	26	0,18	0,34	0,47		P(Y=P)	217	897	0,24
4	Total	625	217	55					P(Y=KP)	55	897	0,06
5												
6	Total Alumni	SP	P	KP	P(SP)	P(P)	P(KP)					
7	≥15	424	0	0	0,68	0,00	0,00					
8	≥9	0	129	0	0,00	0,59	0,00					
9	≤8	90	14	29	0,14	0,06	0,53					
10	≥12	4	0	0	0,01	0,00	0,00					
11	≥7	0	48	0	0,00	0,22	0,00					
12	<7	107	26	26	0,17	0,12	0,47					
13	Total	625	217	55								
14												
15	Penghasilan orang tua	SP	P	KP	P(SP)	P(P)	P(KP)					
16	Tinggi	27	9	11	0,04	0,04	0,20					
17	Sedang	520	130	0	0,83	0,60	0,00					
18	Rendah	78	78	44	0,12	0,36	0,80					
19	Total	625	217	55								
20												
21	Asal Sekolah	SP	P	KP	P(SP)	P(P)	P(KP)					
22	Negeri	384	165	5	0,61	0,76	0,09					
23	Swasta	241	52	50	0,39	0,24	0,91					
24	Total	625	217	55								
25												

**Gambar 4.2** Hasil perhitungan peluang dari setiap variabel menggunakan *Microsoft Excel*

Setelah tampil seperti pada gambar 4.2 kemudian memasukan data *testing* pada file dan *sheet* yang sama, tepat pada bagian bawa tabel asal sekolah, untuk mempermudah perhitungan manual pada *Microsoft Excel*.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Dalam Kota	SP	P	KP	P(SP)	P(P)	P(KP)		Label	Jumlah Data	Jumlah Seluruh Data			
2	Iya	514	143	29	0,82	0,66	0,53		P(Y=SP)	625	897	0,69		
3	Tidak	111	74	26	0,18	0,34	0,47		P(Y=P)	217	897	0,24		
4	Total	625	217	55					P(Y=KP)	55	897	0,06		
5														
6	Total Alumni	SP	P	KP	P(SP)	P(P)	P(KP)							
7	≥15	424	0	0	0,68	0,00	0,00							
8	≥9	0	129	0	0,00	0,59	0,00							
9	≤8	90	14	29	0,14	0,06	0,53							
10	≥12	4	0	0	0,01	0,00	0,00							
11	≥7	0	48	0	0,00	0,22	0,00							
12	<7	107	26	26	0,17	0,12	0,47							
13	Total	625	217	55										
14														
15	Penghasilan orang tua	SP	P	KP	P(SP)	P(P)	P(KP)							
16	Tinggi	27	9	11	0,04	0,04	0,20							
17	Sedang	520	130	0	0,83	0,60	0,00							
18	Rendah	78	78	44	0,12	0,36	0,80							
19	Total	625	217	55										
20														
21	Asal Sekolah	SP	P	KP	P(SP)	P(P)	P(KP)							
22	Negeri	384	165	5	0,61	0,76	0,09							
23	S swasta	241	52	50	0,39	0,24	0,91							
24	Total	625	217	55										
25														
26	nama	dalam kota	total alumni	penghasilan	asal sekolah	keterangan	Hasil predi	SANGAT POT	POTENSI	KURANG POTENSI				
27	FEBE MARGARETHA SIAHAAN	tidak	≥7	sedang	negeri	potensi	potensi	G523,5	0,82%	0,00%				
28	FEBRI LAFODI DANJALAN	tidak	≥7	sedang	swasta	potensi	potensi	G523,5	0,82%	0,00%				

**Gambar 4.3** Cara perhitungan naïve bayes menggunakan *Microsoft Excel*

Nilai probabilitas pada algoritma naïve bayes dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (4).

$$= \$L\$2 * VLOOKUP(B27; \$A\$2: \$G\$3; 5;) * VLOOKUP(C27; \$A\$7: \$G\$12; 5;) * VLOOKUP(D27; \$A\$16: \$G\$18; 5;) * VLOOKUP(E27; \$A\$22: \$G\$23; 5;)$$

Kolom L baris ke2 merupakan probabilitas dari label sangat potensi, kemudian menggunakan perulangan dan memasukan cell B27, dan diikuti cell A2 hingga G3, angka 5 di atas merupakan petunjuk bahwa probabilitas sangat potensi dari setiap variabel berada di kolom ke-5 dan di tutup dengan tanda kurung, formula di atas di lakukan perulangan sebanyak variabel yang ada, kemudian tekan enter. Nilai probabilitas dari data *testing* akan muncul kemudian ditarik kebawah, sehingga setiap record dari data *testing* akan terisi otomatis, untuk perhitungan cell I27 cukup mengganti angka 5 di atas menjadi angka 6, dikarenakan probabilitas potensi dari setiap variabel berada di kolom 6, dan mengganti cell L2 menjadi L3, dikarenakan probabilitas dari label potensi keseluruhan terletak pada

cell L3, peraturan ini dapat di lakukan juga untuk menghitung probabilitas dari label kurang potensi.

#### 4.1.7 Hasil Algoritma *Naïve Bayes* Menggunakan *Microsoft Excel*

Setelah dilakukan perhitungan peluang dari masing-masing variabel maka dari nilai peluang tersebut, data *testing* dapat diklasifikasikan berdasarkan nilai peluang tertinggi, sehingga dapat di simpulkan sebagaimana tabel di bawah ini:

**Tabel 4.31** Hasil perhitungan *naïve bayes* menggunakan *Microsoft Excel*

No .	dalam kota	total alumni	penghasilan orang tua	asal sekolah	Label	Sangat Potensi	Potensi	Kurang Potensi	Hasil prediksi
1.	tidak	$\geq 7$	sedang	negeri	P	0,00%	0,82 %	0,00%	P
2.	tidak	$< 7$	sedang	swasta	SP	0,67%	0,14 %	0,00%	SP
3.	tidak	$< 7$	rendah	swasta	KP	0,10%	0,08 %	0,98%	KP
4.	tidak	$< 7$	sedang	swasta	SP	0,67%	0,14 %	0,00%	SP
5.	tidak	$< 7$	sedang	swasta	SP	0,67%	0,14 %	0,00%	SP
6.	tidak	$\geq 7$	sedang	negeri	P	0,00%	0,82 %	0,00%	P
7.	tidak	$< 7$	sedang	negeri	SP	1,07%	0,45 %	0,00%	SP
8.	tidak	$< 7$	sedang	negeri	SP	1,07%	0,45 %	0,00%	SP
9.	tidak	$< 7$	rendah	negeri	P	0,16%	0,27 %	0,10%	P
10.	tidak	$< 7$	sedang	negeri	SP	1,07%	0,45 %	0,00%	SP
...	.....	.....	.....	.....	.....	...	...	...	...
38 4.	tidak	$< 7$	sedang	negeri	SP	1,07%	0,45 %	0,00%	SP

Berdasarkan tabel di atas maka dapat dilihat presentase dari setaiap baris, presentase tertinggi merupakan hasil keputusan dari data *testing* tersebut. Contoh, pada data *testing* ke-1 menunjukan presentase tertinggi terletak pada kolom potensi, dengan peluang 0,82% sehingga data *testing* ke-1 masuk kedalam klasifikasi potensi.

#### 4.1.8 Penerapan Algoritma *K- Nearest Neighbor* Menggunakan *Microsoft Excel*

Penerapan algoritma *K-nearest neighbor* dibutuhkan tranformasi data dari kategorikal menjadi numerikal untuk mempermudah proses perhitungan. Proses tranformasi dapat dilihat pada tabel di bawah ini:

Dalam Kota	Hasil Transformasi
Iya	1
Tidak	2

**Tabel 4.32** Tranformasi variabel dalam kota Algoritma *K- Nearest Neighbor*

**Tabel 4.33** Tranformasi Variabel Total Alumni Algoritma *K- Nearest Neighbor*

Total Alumni	Hasil Transformasi
$\geq 15$	115
$\geq 9$	19
$\leq 8$	18
$\geq 12$	212
$\geq 7$	28
$< 7$	27

Angka 1 pada hasil tranformasi melambangkan dari dalam kota, dan angka 2 melambangkan luar kota. Sebagai contoh 115, 1 melambangkan dalam kota dan 15 melambangkan lebih dari 15 alumni yang telah mendaftar di kampus Intitut Informatika dan Bisnis Darmajaya.

**Tabel 4.34** Tranformasi variabel penghasilan Algoritma *K- Nearest Neighbor*

Penghasilan Orang Tua	Hasil Transformasi
Rendah	1
Sedang	2
Rendah	3

**Tabel 4.35** tranformasi Asal Sekolah Algoritma *K- Nearest Neighbor*

Penghasilan Orang Tua	Hasil Transformasi
Negeri	1

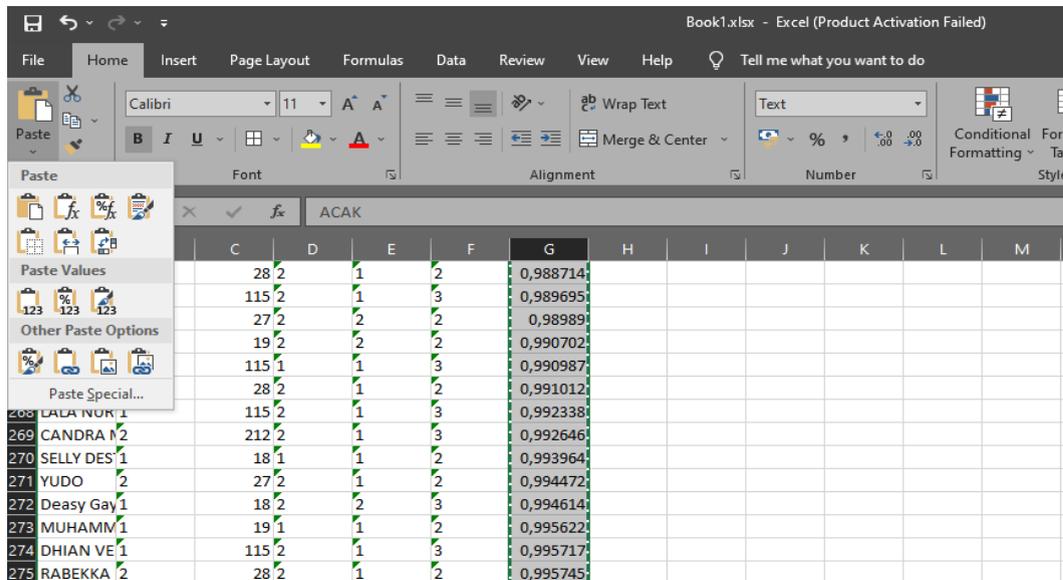
Swasta	2
--------	---

setelah dilakukan tranformasi dari kategorikal menjadi numerikal dari setiap variabel, maka tahapan selanjutnya memasukan hasil tranformasi kedalam software *Microsoft Excel*, sehingga akan tampil seperti gambar di bawah ini:

nama	dalam kot	total	penghasilan ortu	sekolah	keterangan
AARON ANG SAPUTRA	1	18	2	2	2
ABDATA DENIS TARADIKA	1	19	2	1	2
ABDUL HAMID	1	115	2	1	3
ABDULLATIF ANGGA SAPUTRA	2	27	1	2	1
ABDURAHMAN M SYARAFI	1	115	2	2	3
ABIZAL RAHMADAN	1	19	2	1	2
ACHMAD RIDO DINATA	2	27	2	2	2
ACHMAD RIYANSYAH	1	115	2	1	3
ACHMAD YUSTAN THORIQ	1	115	2	1	3
ADAM PUTRASYAWALDI	1	19	2	2	2
ADANG TARUNA	2	27	2	1	2
ADE DEL ALVI	1	115	1	1	3
ADE KURNIASIH	2	27	2	1	2
ADE SAPUTRO	2	27	2	1	2
ADE SETIYAWAN	1	115	2	1	3
ADELIA ALICA RIANDA	1	19	2	2	2
ADELILA PUTRI	1	115	3	1	3
ADELLA MARSHA SAFIRA	1	115	2	1	3
ADELLIA ERWAN	2	28	2	1	2
ADHELIA VIRANTI	1	115	2	1	3
ADHIT ARI SAPUTRA	1	115	3	2	3

**Gambar 4.4** Hasil Tansformasi *KKN* menggunakan *Microsoft Excel*

Setelah tampil seperti gambar 4.4 tahapan selanjutnya yaitu penentuan sampel yang akan dijadikan data *training* dalam perhitungan manual atau perhitungan menggunakan *Microsoft Excel*. Dalam penentuan data *testing* secara acak maka menggunakan formula `=RAND()` kemudian setelah semua data mendapat nilai dari formula ini, maka semua nilai di block dan menekan tombol `CTRL+C` kemudian menekan `PASTE` yang teletak di sebelah kiri atas, terlihat pada gambar:



**Gambar 4.5** Proses Sampling KKN menggunakan *Microsoft Excel*

Tahapan selanjutnya yaitu *block* semua angka dan data dengan menekan tombol CTRL+A dan pilih tools DATA yang berada di bagian atas, kemudian SHORT dan pilih yang ACAK dan tekan OK, maka akan tampil:

	A	B	C	D	E	F	G
1	nama	dalam kot	total	penghasil	sekolah	keteranga	ACAK
2	AGUNG ROMADHON	2	28	1	1	2	0,000476
3	SAPRI SETIAWAN	2	27	2	1	2	0,001559
4	SALSABILLA ANANDA PUTRI	1	115	2	1	3	0,001988
5	NICOLAS YUDA DINATA	2	27	2	1	2	0,002007
6	YUSRON DAROZAT	2	27	1	2	1	0,002023
7	AHMAD MURDIFIN	1	19	2	1	2	0,004098
8	FARREL IRGIA REVA	2	27	2	1	2	0,004192
9	ANITA SARI DWIYANI	2	28	2	1	2	0,00492
10	SYALSA ENGELICHA AGUSTIN	2	27	2	2	2	0,006791
11	DERJA ROBBANI	1	115	2	1	3	0,007037
12	RODHI FAISAL MUFID	2	27	1	1	2	0,007599
13	MAYA WIRLINDA KURNIAWATI	2	27	1	2	1	0,007632
14	ADELLIA ERWAN	2	28	2	1	2	0,007658
15	EDO PRATAMA PUTRA ARISDA	2	27	1	1	2	0,007733
16	PUTRI WULANDARI	1	115	2	2	3	0,008078
17	ANNISA YULIA ZARA TINI	2	27	2	1	2	0,008573

**Gambar 4.6** Hasil Sampling KKN menggunakan *Microsoft Excel*

Berdasarkan gambar *di atas* maka didapat 15 sample secara acak yang akan dijadikan data latih atau data *training* sebagai berikut:

**Tabel 4.36** Data *Training* Algoritma *K- Nearest Neighbor*

Nama	Dalam Kota	total	Penghasilan Orang Tua	Sekolah	Ket.
Agung Romadhon	2	28	1	1	2
Sapri Setiawan	2	27	2	1	2
Salsabilla Ananda Putri	1	115	2	1	3
Nicolas Yuda Dinata	2	27	2	1	2
Yusron Darozat	2	27	1	2	1
Ahmad Murdifin	1	19	2	1	2
Farrel Irgia Reva	2	27	2	1	2
Anita Sari Dwiyani	2	28	2	1	2
Syalsa Engelicha Agustin	2	27	2	2	2
Derja Robbani	1	115	2	1	3
Rodhi Faisal Mufid	2	27	1	1	2
Maya Wirlinda K.	2	27	1	2	1
Adellia Erwan	2	28	2	1	2
Edo Pratama Putra A.	2	27	1	1	2
Putri Wulandari	1	115	2	2	3

Setelah mendapatkan data *training di atas* maka, akan di lakukan data *testing*:

**Tabel 4.37** Data *Testing* Algoritma *K- Nearest Neighbor*

Nama	Dalam Kota	Total	Penghasilan Orang Tua	Asal Sekolah	Ket.
Annisa Yulia Zara T.	2	27	2	1	?

Dalam penentuan variabel keterangan pada data *testing di atas* maka tahapan pertama yaitu menentukan *Euclidean distance*.

Nilai *Euclidean distance* dapat juga ditentukan menggunakan formula pada *Microsoft Excel*, yang merupakan penerjemahan dari Persamaan (1) yaitu:

$$=\text{SQRT}((\text{B2}-\text{\$B\$18})^2+(\text{C2}-\text{\$C\$18})^2+(\text{D2}-\text{\$D\$18})^2+(\text{E2}-\text{\$E\$18})^2)$$

Dapat dijelaskan bahwa *cell* B2 dikurang dengan *cell* B18 yang merupakan data

*testing* kemudia di pangkatkan dengan 2, dan ditambah dengan cell C2 yang merupakan data trening dan di kurang cell C18 yang merupakan data *testing*, di lakukan perulangan sebanyak varibael yang ada sebanyak data *testing*, kemudian dapat di hasilkan nilai *Euclidean distance* dari setiap data *testing* yang ada, dalam penelitian ini menggunakan K=3 sehingga K3 yang mendekati yaitu *kluster* dari variabel data *testing*.

dapat dilihat pada gambar dibawah ini nilai *Euclidean distance* dari setiap data *record*:

	A	B	C	D	E	F	G	H	I
1	nama	dalam kot total		penghasil	sekolah	keterangan		DISTANCE	
2	AGUNG ROMADHON	2	28	1	1	2		(\$18)^2)	
3	SAPRI SETIAWAN	2	27	2	1	2		0	
4	SALSABILLA ANANDA PUTRI	1	115	2	1	3		88,00568	
5	NICOLAS YUDA DINATA	2	27	2	1	2		0	
6	YUSRON DAROZAT	2	27	1	2	1		1,414214	
7	AHMAD MURDIFIN	1	19	2	1	2		8,062258	
8	FARREL IRGIA REVA	2	27	2	1	2		0	
9	ANITA SARI DWIYANI	2	28	2	1	2		1	
10	SYALSA ENGELICHA AGUSTIN	2	27	2	2	2		1	
11	DERJA ROBBANI	1	115	2	1	3		88,00568	
12	RODHI FAISAL MUFID	2	27	1	1	2		1	
13	MAYA WIRLINDA KURNIAWATI	2	27	1	2	1		1,414214	
14	ADELLIA ERWAN	2	28	2	1	2		1	
15	EDO PRATAMA PUTRA ARISDA	2	27	1	1	2		1	
16	PUTRI WULANDARI	1	115	2	2	3		88,01136	
17									
18	ANNISA YULIA ZARA TINI	2	27	2	1	?			
19									

**Gambar 4.7** Hasil Nilai *Euclidean distance* menggunakan *Microsoft Excel*

Setelah mendapatkan nilai *Euclidean distance* dari setiap *record* data *testing* yaitu penentuan ranking dari *record* yang ada, dalam penelitian ini menggunakan K=3 yang artinya 3 nilai yang mendekati merupan hasil dari klasifikasi data *testing*, untuk mendapatkan ranking dari *Euclidean distance* dapat menggunakan formula:

$$=RANK(H2;$H$2:$H$16;1)$$

H2 atau *cell* H2 merupakan nilai *euclidean distance* dari *record* data *testing* pertama, kemudian *cell* H2 hingga H16 merupakan nilai *euclidean distance* dari setiap *record* data *testing*, dapat di lihat hasil ranking pada gambar di bawah ini:

	A	B	C	D	E	F	G	H	I	J
1	nama	dalam kot	total	penghasil	sekolah	keterangan		DISTANCE	RANK	
2	AGUNG ROMADHON	2	28	1	1	2		1,414214	1	
3	SAPRI SETIAWAN	2	27	2	1	2		0	1	
4	SALSABILLA ANANDA PUTRI	1	115	2	1	3		88,00568	13	
5	NICOLAS YUDA DINATA	2	27	2	1	2		0	1	
6	YUSRON DAROZAT	2	27	1	2	1		1,414214	9	
7	AHMAD MURDIFIN	1	19	2	1	2		8,062258	12	
8	FARREL IRGIA REVA	2	27	2	1	2		0	1	
9	ANITA SARI DWIYANI	2	28	2	1	2		1	4	
10	SYALSA ENGELICHA AGUSTIN	2	27	2	2	2		1	4	
11	DERJA ROBBANI	1	115	2	1	3		88,00568	13	
12	RODHI FAISAL MUFID	2	27	1	1	2		1	4	
13	MAYA WIRLINDA KURNIAWATI	2	27	1	2	1		1,414214	9	
14	ADELLIA ERWAN	2	28	2	1	2		1	4	
15	EDO PRATAMA PUTRA ARISDA	2	27	1	1	2		1	4	
16	PUTRI WULANDARI	1	115	2	2	3		88,01136	15	
17										
18	ANNISA YULIA ZARA TINI	2	27	2	1	?				
19										

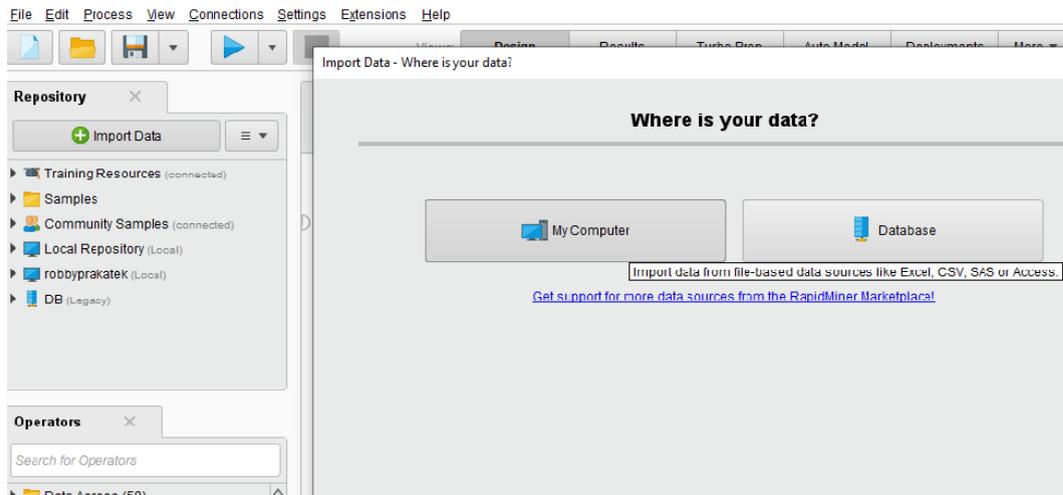
**Gambar 4.8** Hasil Rangking Algoritma *K- Nearest Neighbor*

#### 4.1.9 Hasil Algoritma *K- Nearest Neighbor* Menggunakan *Microsoft Excel*

Setelah dilakukan perhitungan nilai *euclidean distance* dan perangkingan maka data *testing* dengan nama Annisa Yulia Zara Tini yang berada di luar kota dengan kode numerik 2 dan total alumni kurang dari tujuh dengan kode numerik 27 penghasilan orang tua dalam kategori sedang di tujukkan dengan kode numerik 2 dan asal sekolah negeri ditunjukkan dengan kode numerik 1, termasuk kedalam klasifikasi lokasi promosi potensi ditunjukkan dengan kode klasifikasi numerik 2, secara terperinci dapat di lihat pada gambar 4.8 rangking 1 terlihat lebih dari 2 dengan menggunakan aturan awal yaitu  $K=3$ .

#### 4.1.10 Penerapan Algoritma *Decision tree* menggunakan *Rapid Miner*

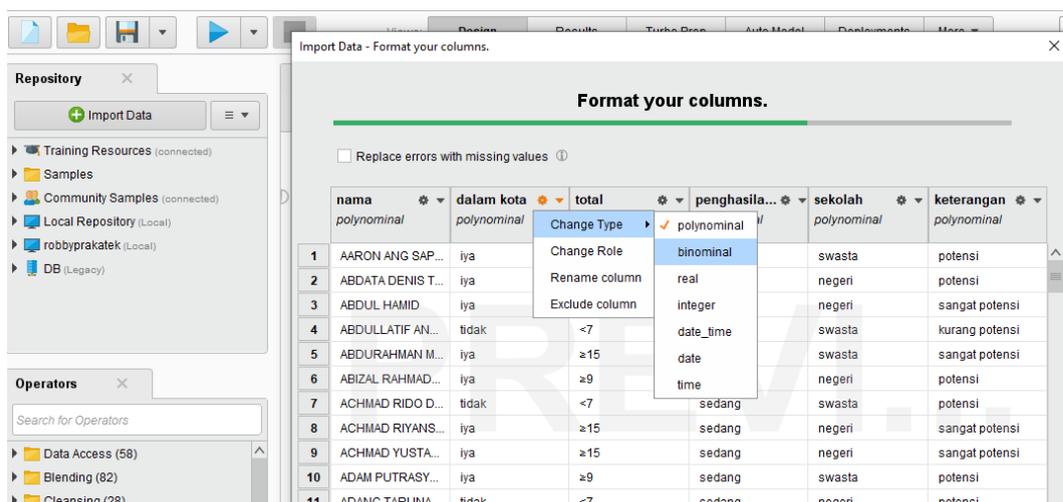
Tahapan pertama untuk menggunakan *Rapid miner 9.8* yaitu melakukan instalasi terlebih dahulu kedalam perangkat yang akan digunakan, jika *software Rapid miner 9.8* sudah terinstalasi, maka tahapan selanjutnya, melakukan *import data* seperti yang terlihat pada gambar di bawah ini:



**Gambar 4.9** Import data kedalam *Rapid miner 9.8*

Setelah tampilan terlihat seperti gambar 4.9 maka tekan opsi *My Computer*, karena data yang akan di olah terletak pada komputer, kemudian pilih *file* sesuai dengan *file* data yang tersimpan dan yang akan diolah.

Tahapan selanjutnya yaitu menyeleksi atau memperbaiki atribut dari setiap variabel. Dalam penyeleksian dapat dilihat pada gambar di bawah ini:

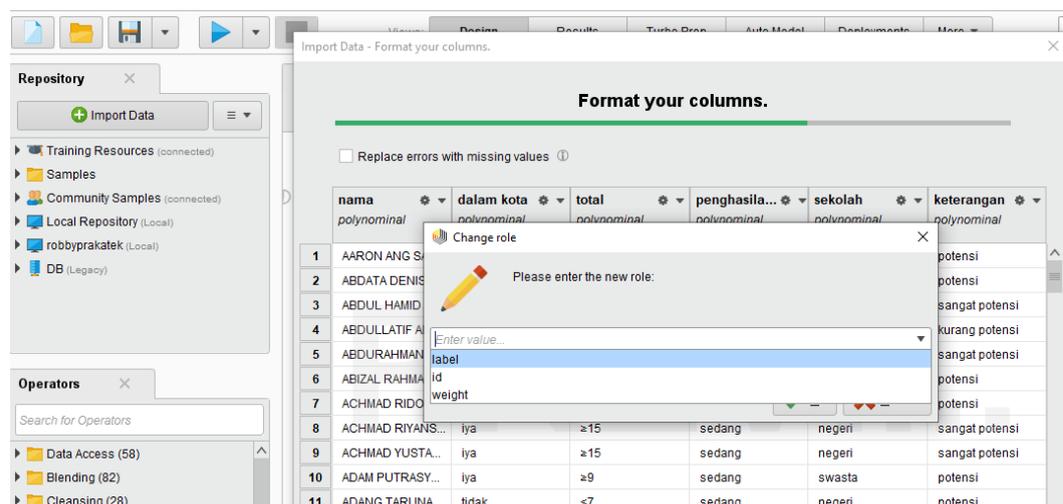


**Gambar 4.10** Proses Seleksi Atribut dari setiap Variabel

Dalam gambar 4.10 dapat dilihat bahwa variabel dalam kota terisi atribut secara otomatis, namun atribut tersebut tidak sesuai dengan yang diharapkan, cara untuk merubahnya yaitu menekan simbol segitiga yang berada tepat di sebelah nama variabel tersebut, kemudian pilih *Change Type*, dan pilih *Binominal*, dikarenakan

*record* data hanya memiliki 2 pilihan yaitu, ya dan tidak. Variabel total untuk atribut datanya menggunakan *Polynomial* dikarenakan record yang ada pada variabel tersebut memiliki lebih dari 2 pilihan, seleksi atribut ini dilakukan pada semua variabel atau kolom yang ada, sehingga atribut yang di pilih sesuai dengan *record* dari masing-masing variabel tersebut.

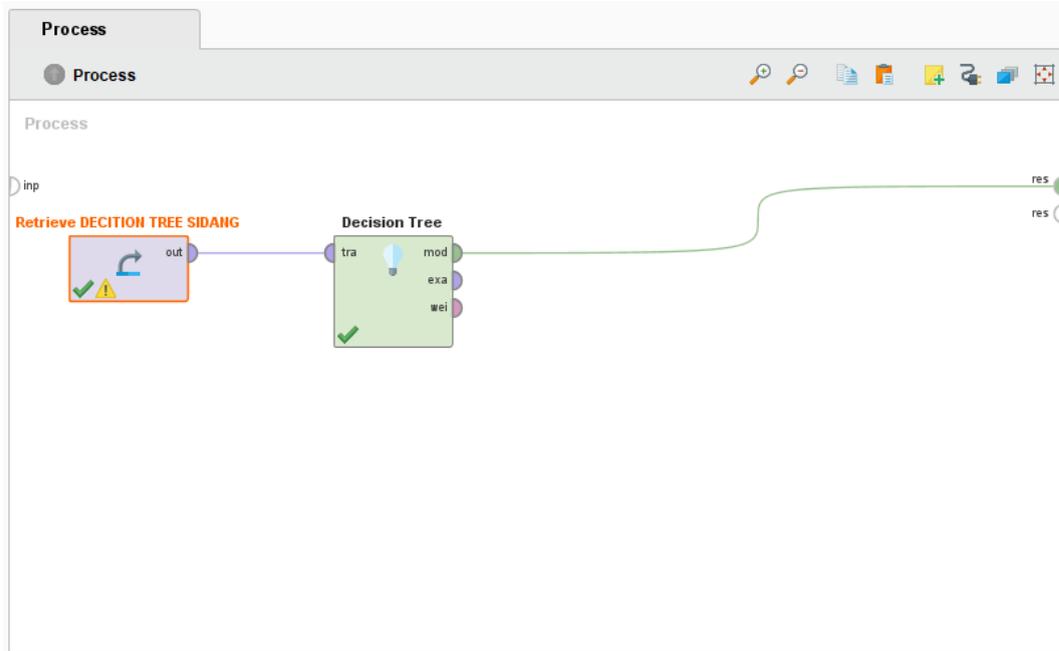
Langkah berikutnya yaitu menentukan label, dari data *training*. Tahapan dalam menentukan label dapat dilihat pada gambar di bawah ini:



**Gambar 4.11** Penentuan Label dari Data *Training*

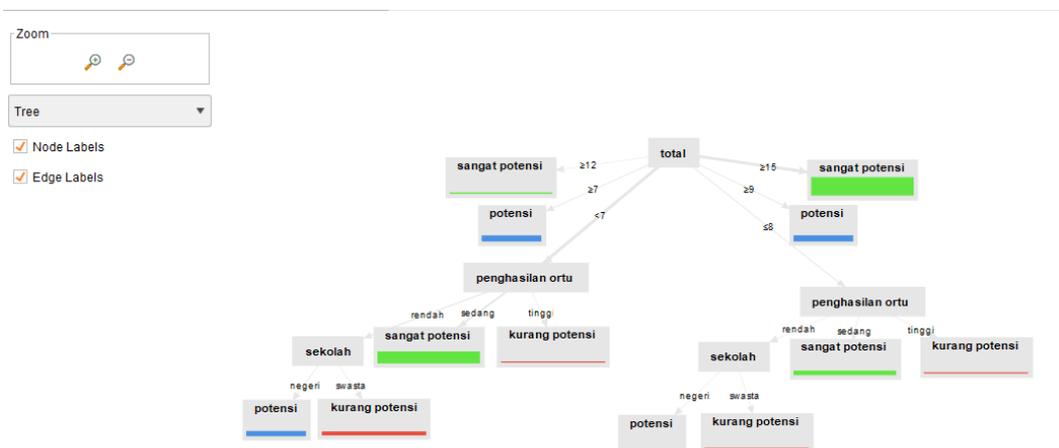
Dalam Gambar 4.11 dapat dijelaskan bahwa untuk memilih atribut label dapat menekan simbol segitiga yang berada tepat di samping nama variabel, kemudian pilih *change role* dan pilih label. Setelah *menentukan* label maka di lanjutkan menyimpan data pada *rapid miner*, pada bagian ini data disimpan pada *local repository*, kemudian pilih data.

Langkah berikutnya yaitu penerapan algoritma *decision tree*, pada tahapan ini pilih *design*, kemudian pilih data yang telah disimpan pada *local repository*, kemudian cari algoritma *decision tree* pada bagian *operators*, dan masukan kedalam *design*, serta hubungkan dari data, algoritma, dan outout, tampilan desain akan seperti gambar dibawah ini:



**Gambar 4.12** Design dari *Decision tree* menggunakan *Rapid miner*

Setelah tampilan sudah seperti pada gambar 4.12 maka langkah berikutnya yaitu menjalankan *design* yang sudah dibuat, dengan cara menekan simbol *run* yang ada pada *tools* bagian atas, sehingga hasil dari algoritma *decision tree* dapat dilihat pada gambar di bawah ini:



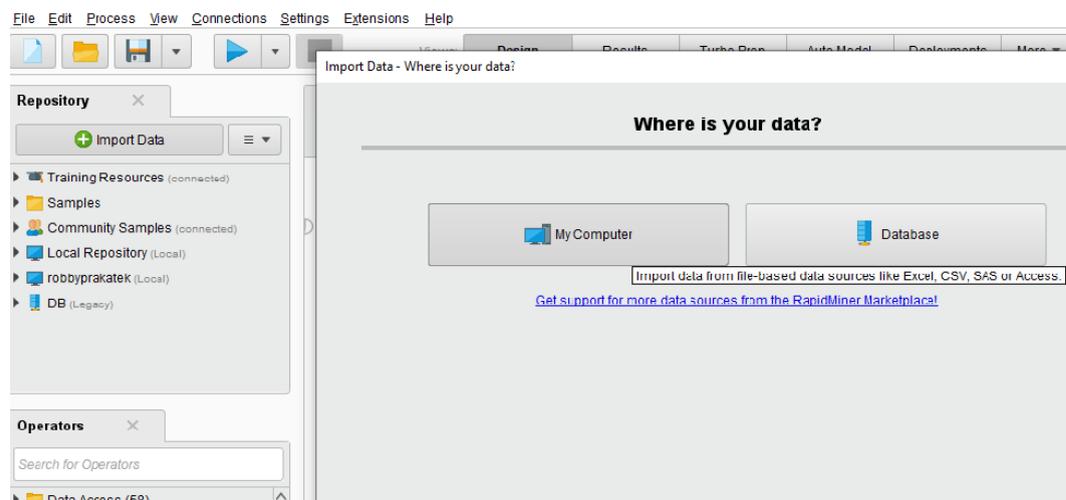
**Gambar 4.13** Output dari *decision tree* menggunakan *rapid miner*

Berdasarkan gambar di atas dapat dijelaskan bahwa node pertama merupakan

variabel total alumni, jika alumni lebih dari 12 dan 15 maka dapat disimpulkan bahwa asal sekolah tersebut sangat potensi, namun jika alumni lebih dari 9 dan lebih dari 7 maka asal sekolah mahasiswa tersebut masuk dalam kelompok potensi, dan jika alumni kurang dari 7 dan 9 maka dilihat kembali penghasilan orang tua, jika berpenghasilan tinggi maka masuk kedalam kelompok kurang potensi, dan jika penghasilan orang tua sedang masuk kedalam potensi dan jika penghasilan orang tua rendah maka dapat dilihat kembali asal sekolah mahasiswa tersebut, jika berasal dari sekolah negeri maka masuk kedalam kelompok potensi dan jika berasal dari sekolah swasta maka masuk kedalam kelompok kurang potensi, namun ada variabel yang tidak muncul yaitu variabel dalam kota, variabel tersebut tidak muncul dikarenakan nilai *gain* terlampau kecil, yang berarti variabel tersebut tidak terlalu berpengaruh dalam pohon keputusan.

#### 4.1.11 Penerapan Algoritma *Naïve Bayes* menggunakan *Rapid Miner*

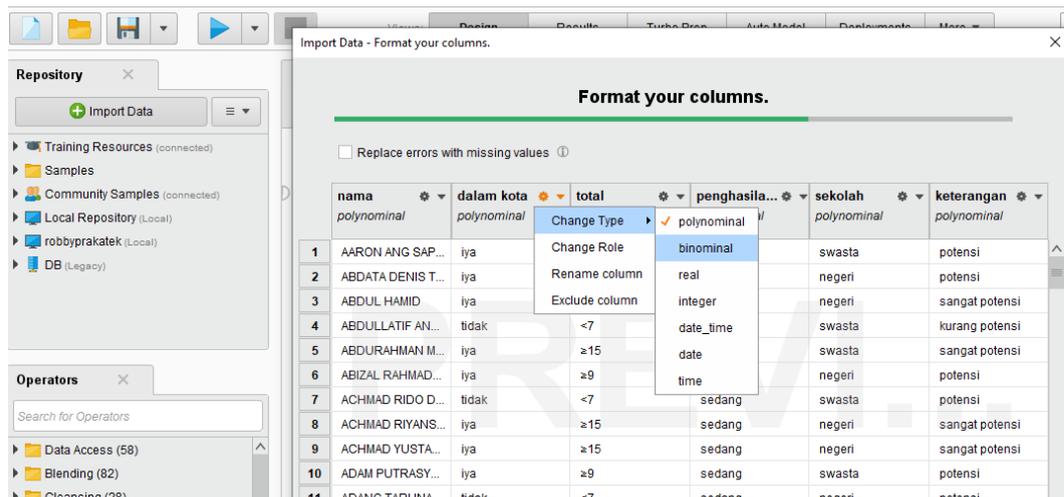
Tahapan pertama untuk menggunakan *Rapid miner 9.8* yaitu melakukan instalasi terlebih dahulu kedalam perangkat yang akan digunakan, jika *software Rapid miner 9.8* sudah terinstalasi, maka tahapan selanjutnya, melakukan *import* data seperti yang terlihat pada gambar di bawah ini:



**Gambar 4.14** Import data kedalam *Rapid miner 9.8*

Setelah tampilan terlihat seperti gambar 4.14 maka tekan opsi *My Computer*, karena data yang akan di olah terletak pada komputer, kemudian pilih *file* sesuai dengan *file* data yang tersimpan dan yang akan diolah.

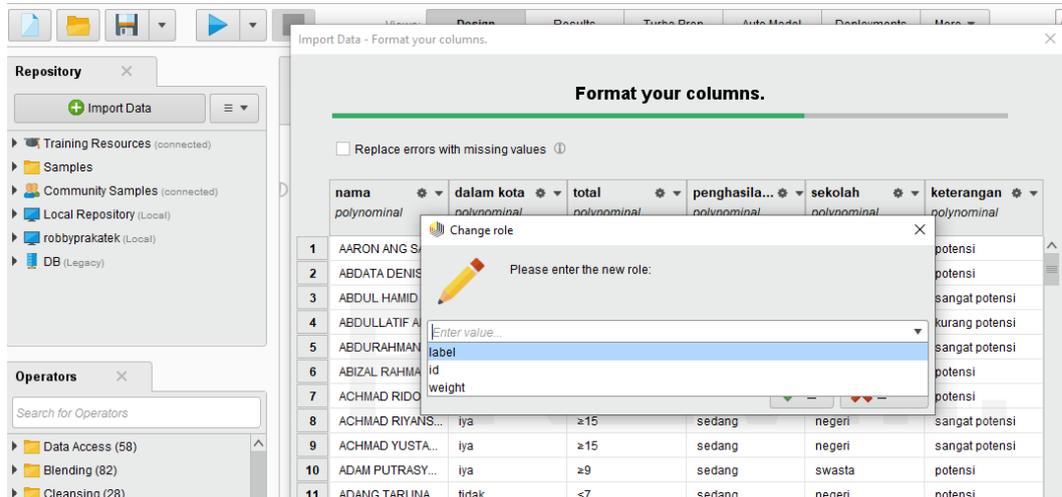
Tahapan selanjutnya yaitu menyeleksi atau memperbaiki atribut dari setiap variabel. Dalam penyeleksian dapat dilihat pada gambar di bawah ini:



**Gambar 4.15** Proses Seleksi Atribut dari setiap Variabel

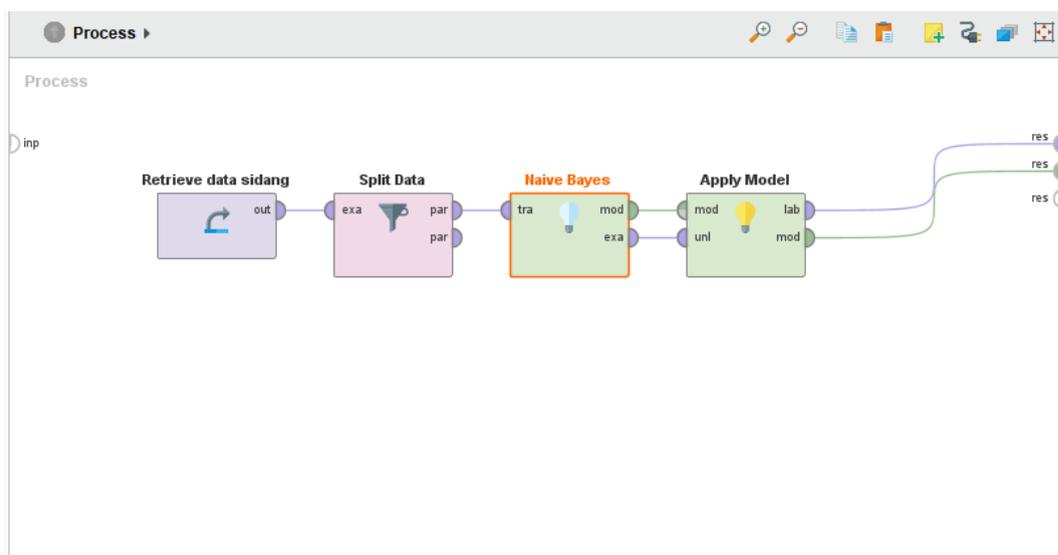
Dalam gambar 4.14 dapat dilihat bahwa variabel dalam kota terisi atribut secara otomatis, namun atribut tersebut tidak sesuai dengan yang diharapkan, cara untuk merubahnya yaitu menekan simbol segitiga yang berada tepat di sebelah nama variabel tersebut, kemudian pilih *Change Type*, dan pilih *Binominal*, di karena *record* data hanya memiliki 2 pilihan yaitu, ya dan tidak. Variabel total untuk atribut datanya menggunakan *Polynomial* dikarenakan record yang ada pada variabel tersebut memiliki lebih dari 2 pilihan, seleksi atribut ini dilakukan pada semua variabel atau kolom yang ada, sehingga atribut yang di pilih sesuai dengan *record* dari masing-masing variabel tersebut.

Langkah berikutnya yaitu menentukan label, dari data *training*. Tahapan dalam menentukan label dapat dilihat pada gambar di bawah ini:



**Gambar 4.16** Penentuan Label dari Data *Training*

Dalam Gambar 4.15 dapat dijelaskan bahwa untuk memilih atribut label dapat menekan simbol segitiga yang berada tepat di samping nama variabel, kemudian pilih *change role* dan pilih label. Setelah *menentukan* label maka di lanjutkan menyimpan data pada *rapid miner*, pada bagian ini data disimpan pada *local repository*, kemudian pilih data.



**Gambar 4.17** Design penerapan naïve bayes menggunakan *rapid miner*

Langkah berikutnya yaitu penerapan algoritma *naïve bayes*, pada tahapan ini pilih *design*, kemudian pilih data yang telah disimpan pada *local repository*, selanjutnya dalam *operators* gunakan *split data*, yang berguna untuk menentukan

data *testing* dan *trening*, dalam *split data* menggunakan perbandingan 70% dan 30%. Tahapan berikutnya cari algoritma *naïve bayes* pada bagian *operators*, dan masukan kedalam *design*, yang terakhir yaitu masukan *apply model* kedalam *design* untuk mendapatkan nilai *confidence* serta hubungkan dari data, *split data*, algoritma *naïve bayes*, *apply model* dan *output*, tampilan *design* dapat di lihat pada gambar 4.16 di atas.

Setelah tampilan sudah seperti pada gambar 4.16 maka langkah berikutnya yaitu menjalankan *design* yang sudah dibuat, dengan cara menekan simbol *run* yang ada pada *tools* bagian atas, sehingga hasil dari algoritma *naïve bayes* dapat dilihat pada gambar di bawah ini:

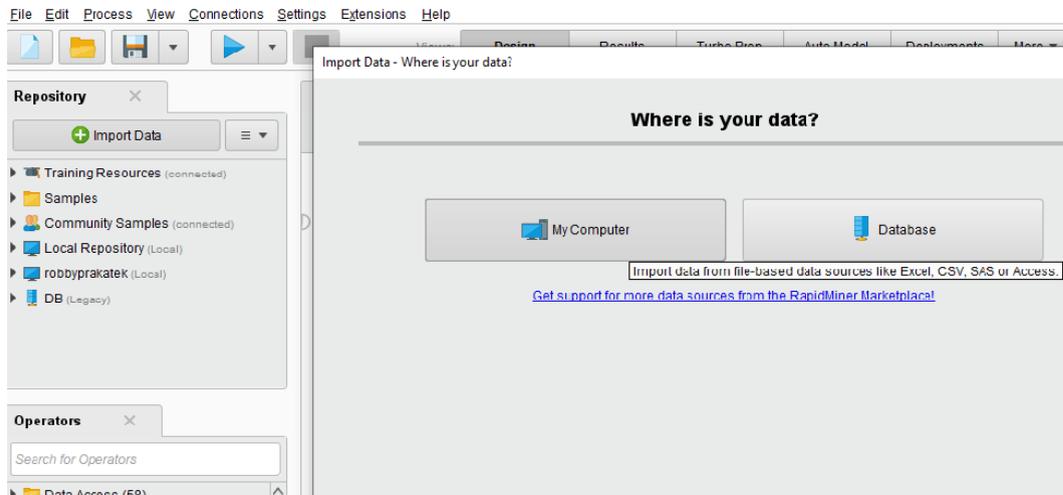
Row No.	keterangan	prediction(keterangan)	confidence(potensi)	confidence(sangat potensi)	confidence(kurang potensi)	nama	dalam kota
1	kurang potensi	kurang potensi	0.000	0.000	1.000	ABDULLATIF ...	tidak
2	sangat potensi	sangat potensi	0.000	1.000	0.000	ACHMAD YU...	iya
3	potensi	potensi	1.000	0.000	0.000	ADE KURNIA...	tidak
4	potensi	potensi	1.000	0.000	0.000	ADELIA ALIC...	iya
5	sangat potensi	sangat potensi	0.000	1.000	0.000	ADHIT ARI SA...	iya
6	kurang potensi	kurang potensi	0.007	0.000	0.993	ADI PRASET...	tidak
7	potensi	potensi	1.000	0.000	0.000	ADI SYUMAN ...	iya
8	sangat potensi	sangat potensi	0.000	1.000	0.000	ADILLA SETY...	iya
9	potensi	potensi	1.000	0.000	0.000	ADITYA PAN...	tidak
10	sangat potensi	sangat potensi	0.000	1.000	0.000	AELY GUSNITA	iya
11	sangat potensi	sangat potensi	0.000	1.000	0.000	AGUNG PRIM...	iya
12	potensi	potensi	1.000	0.000	0.000	AHMAD ARD...	iya
13	potensi	potensi	1.000	0.000	0.000	AHMAD IRPA...	iya
14	potensi	potensi	1.000	0.000	0.000	AJENG BELL...	iya

**Gambar 4.18** Output dari *Naïve bayes* menggunakan *rapid miner*

Berdasarkan gambar di atas dapat dijelaskan bahwa data *testing* atas nama Abdullatif Angga Saputra memiliki nilai *confidence* sangat potensi sebesar 0,000 dan nilai *confidence* potensi 0,000 serta nilai *confidence* kurang potensi sebesar 1,000 sehingga atas nama Abdullatif Angga Saputra dapat di kelompokkan dalam lokasi promosi kurang potensi.

#### 4.1.12 Penerapan Algoritma *K- Nearest Neighbor* menggunakan *Rapid Miner*

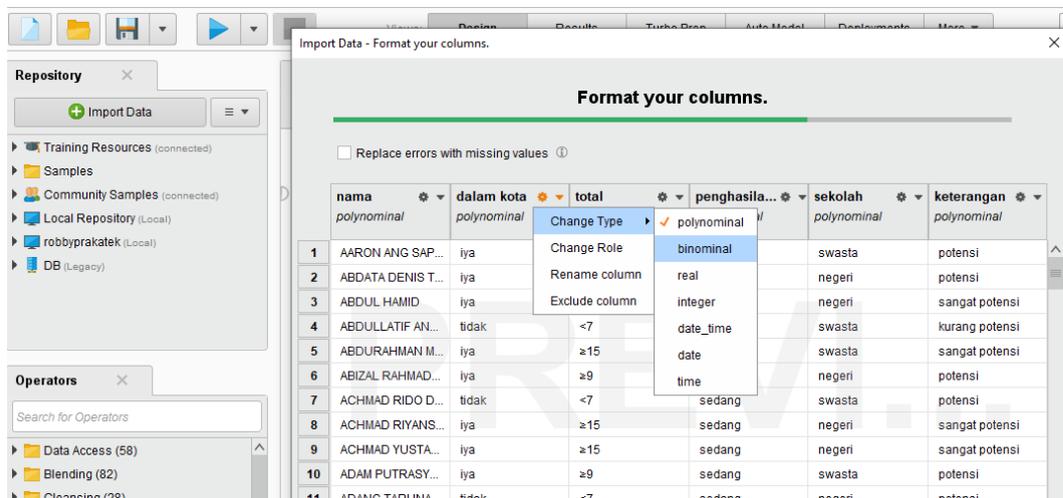
Tahapan pertama untuk menggunakan *Rapid Miner 9.8* yaitu melakukan instalasi terlebih dahulu kedalam perangkat yang akan digunakan, jika *software Rapid Miner 9.8* sudah terinstalasi, maka tahapan selanjutnya, melakukan *import* data seperti yang terlihat pada gambar di bawah ini:



**Gambar 4.19** Import data kedalam *Rapid miner 9.8*

Setelah tampilan terlihat seperti gambar 4.19 maka tekan opsi *My Computer*, karena data yang akan di olah terletak pada komputer, kemudian pilih *file* sesuai dengan *file* data yang tersimpan dan yang akan diolah.

Tahapan selanjutnya yaitu menyeleksi atau memperbaiki atribut dari setiap variabel. Dalam penyeleksian dapat dilihat pada gambar di bawah ini:

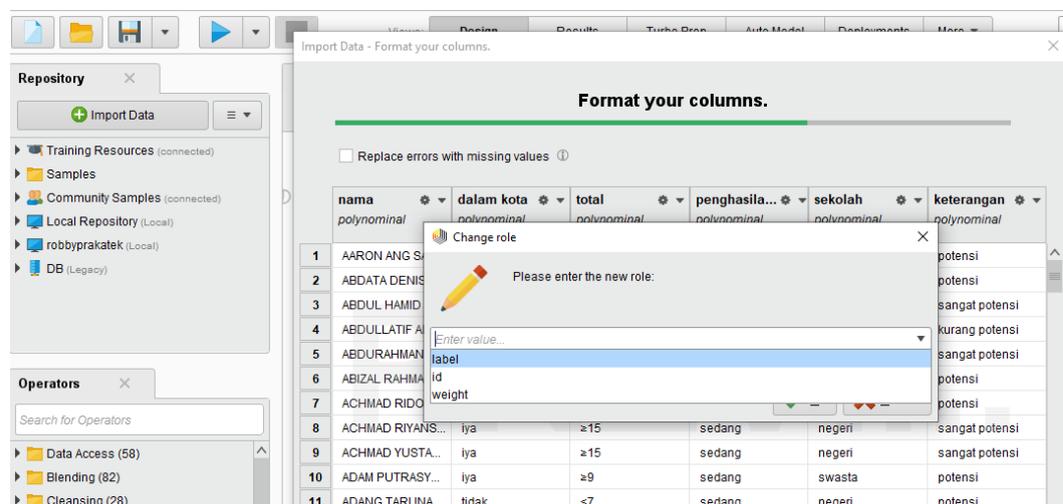


**Gambar 4.20** Proses Seleksi Atribut dari setiap Variabel

Dalam gambar 4.20 dapat dilihat bahwa variabel dalam kota terisi atribut secara otomatis, namun atribut tersebut tidak sesuai dengan yang diharapkan, cara untuk merubahnya yaitu menekan simbol segitiga yang berada tepat di sebelah nama

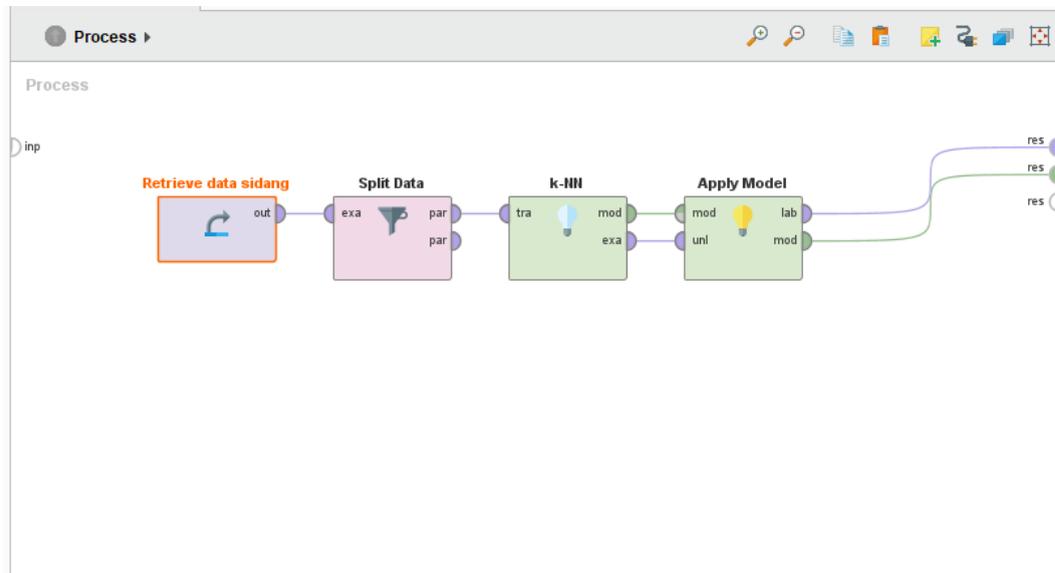
variabel tersebut, kemudian pilih *Change Type*, dan pilih *Binominal*, di karena *record* data hanya memiliki 2 pilihan yaitu, ya dan tidak. Variabel total untuk atribut datanya menggunakan *Polynomial* dikarenakan record yang ada pada variabel tersebut memiliki lebih dari 2 pilihan, seleksi atribut ini dilakukan pada semua variabel atau kolom yang ada, sehingga atribut yang di pilih sesuai dengan *record* dari masing-masing variabel tersebut.

Langkah berikutnya yaitu menentukan label, dari data *training*. Tahapan dalam menentukan label dapat dilihat pada gambar di bawah ini:



**Gambar 4.21** Penentuan Label dari Data *Training*

Dalam Gambar 4.21 dapat dijelaskan bahwa untuk memilih atribut label dapat menekan simbol segitiga yang berada tepat di samping nama variabel, kemudian pilih *change role* dan pilih label. Setelah *menentukan* label maka di lanjutkan menyimpan data pada *rapid miner*, pada bagian ini data disimpan pada *local repository*, kemudian pilih data.



**Gambar 4.22** Design penerapan *K- Nearest Neighbor* menggunakan *rapid miner*

Langkah berikutnya yaitu penerapan algoritma *k- nearest neighbor*, pada tahapan ini pilih *design*, kemudian pilih data yang telah disimpan pada *local repository*, selanjutnya dalam *operators* gunakan *split data*, yang berguna untuk menentukan data *testing* dan *trening*, dalam *split data* menggunakan perbandingan 70% dan 30%. Tahapan berikutnya cari algoritma *k- nearest neighbor* pada bagian *operators*, dan masukan kedalam *design*, yang terakhir yaitu masukan *apply model* kedalam *design* untuk mendapatkan nilai *confidence* serta hubungkan dari data, *spit data*, algoritma *naïve bayes*, *apply model* dan *output*, tampilan *design* dapat di lihat pada gambar 4.22 di atas.

Setelah tampilan sudah seperti pada gambar 4.21 maka langkah berikutnya yaitu menjalankan *design* yang sudah dibuat, dengan cara menekan simbol *run* yang ada pada *tools* bagian atas, sehingga hasil dari algoritma *naïve bayes* dapat dilihat pada gambar di bawah ini:

Row No.	keterangan	prediction(k...	confidence(...	confidence(...	confidence(...	nama	dalam kota	total	penghasilan...
1	potensi	potensi	1	0	0	AARON ANG ...	iya	≥9	sedang
2	potensi	potensi	1	0	0	ABDATA DEN...	iya	≥9	sedang
3	sangat potensi	sangat potensi	0	1	0	ABDUL HAMID	iya	≥15	sedang
4	kurang potensi	kurang potensi	0	0	1	ABDULLATIF ...	tidak	<7	rendah
5	sangat potensi	sangat potensi	0	1	0	ABDURAHMA...	iya	≥15	sedang
6	potensi	potensi	1	0	0	ABIZAL RAH...	iya	≥9	sedang
7	potensi	potensi	1	0	0	ACHMAD RID...	tidak	<7	sedang
8	sangat potensi	sangat potensi	0	1	0	ACHMAD RIY...	iya	≥15	sedang
9	sangat potensi	sangat potensi	0	1	0	ACHMAD YU...	iya	≥15	sedang
10	potensi	potensi	1	0	0	ADAM PUTRA...	iya	≥9	sedang
11	potensi	potensi	1	0	0	ADANG TAR...	tidak	<7	sedang

**Gambar 4.23** Output dari *K- Nearest Neighbor* menggunakan *rapid miner*

Berdasarkan gambar di atas dapat dijelaskan bahwa data *testing* atas nama Aaron Ang Saputra memiliki nilai *confidence* sangat potensi sebesar 0 dan nilai *confidence* potensi sebesar 1 serta nilai *confidence* kurang potensi 0 sehingga atas nama nama Aaron Ang Saputra dapat di kelompokkan dalam lokasi promosi potensi.

#### 4.1.13 Cross Validation

Dalam Cross Validation, jumlah tetap lipatan/partisi dari data ditentukan sendiri. Cara standar untuk memprediksi error rate dari teknik pembelajaran dari sebuah sampel data tetap adalah dengan menggunakan *tenfold cross validation*.

##### a. Akurasi

akurasi menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar, atau dalam pengertian lain akurasi merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya).

##### b. Presisi

Menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model, atau dapat disimpulkan bahwa presisi merupakan dari semua kelas positif yang telah diprediksi dengan benar, berapa banyak data yang benar-benar positif.

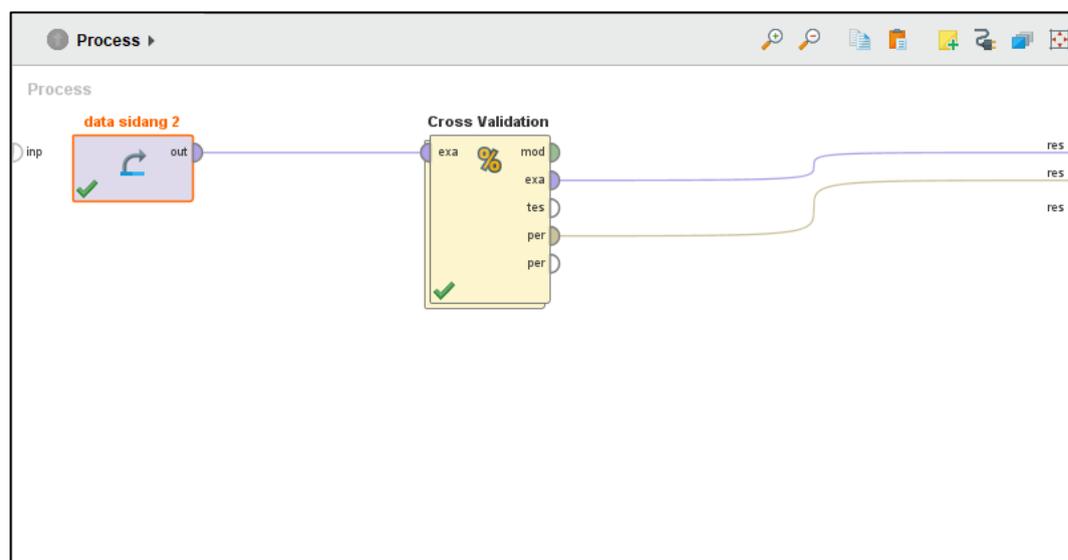
##### c. Recall

Recall merupakan keberhasilan model dalam menemukan kembali sebuah

informasi, maka recall dapat disimpulkan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

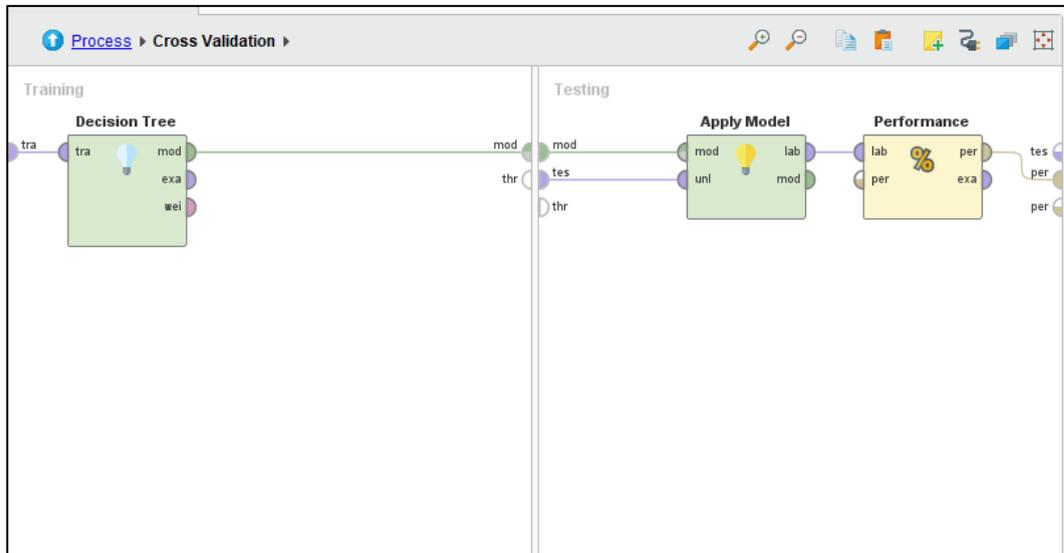
#### 4.1.13.1 *Cross Validation* pada algoritma *Decision tree*

Dataset yang telah tersimpan dalam *rapid miner* dapat digunakan kembali untuk tahapan *cross validation*. Tahapan pertama dataset dimasukkan ke dalam proses, kemudian gunakan operator *cross validation*, dengan ketentuan *10-fold cross validation* hal ini merujuk dengan penelitian terdahulu (Saifudin, 2017) susunan *cross validation* dapat dilihat di bawah ini:



**Gambar 4.24** Susunan *Cross Validation* algoritma *Decision tree*

Dalam melatih dan menguji pemodelan *Cross Validation* menggunakan tiga operator yaitu, pada bagian *training* digunakan untuk algoritma *decision tree*, dan untuk bagian *testing* digunakan fitur *Apply Model* yang berguna untuk untuk menampilkan *confusion table*, dan operator *performance* digunakan untuk menampilkan hasil dari *accuracy*, *precision*, dan *recall*.



**Gambar 4.25** Susunan operator pada algoritma *Decision tree*

Kemudian jika tampilan sudah seperti pada gambar 4.24 maka dapat dijalankan dengan menekan fitur *run* yang ada pada bagian kiri atas.

Table View Plot View

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true potensi	true sangat potensi	true kurang potensi	class precision
pred. potensi	374	0	0	100.00%
pred. sangat potensi	0	798	0	100.00%
pred. kurang potensi	0	0	109	100.00%
class recall	100.00%	100.00%	100.00%	

**Gambar 4.26** Hasil *Cross Validation* algoritma *Decision tree*

Dalam gambar 4.26 dapat di lihat bahwa algoritma *decision* memiliki nilai akurasi sebesar 100% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus persamaan (5)

$$Accuracy = \frac{374+798+109}{374+798+109} = \frac{1281}{1281} \times 100\% = 100\%$$

nilai presisi untuk prediksi label potensi sebesar 99,20% hal ini dapat dibuktikan

dengan perhitungan manual dengan rumus persamaan (6)

$$Presisi = \frac{374}{374+0} = \frac{374}{374} \times 100\% = 100\%$$

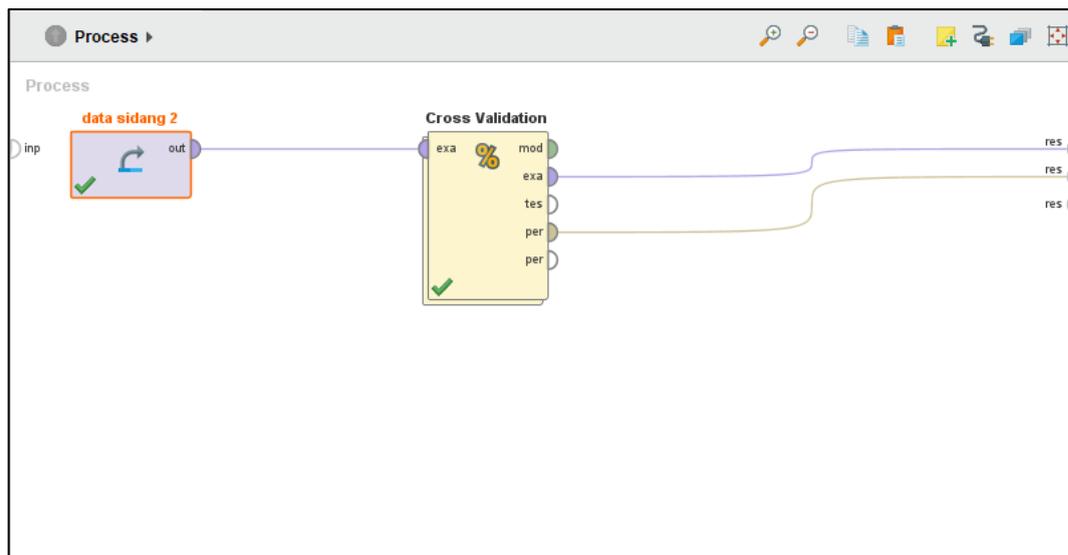
serta nilai *recall* untuk prediksi label potensi sebesar 97,98% hasil tersebut dibuktikan dengan perhitungan manual dengan rumus persamaan (7)

$$Recall = \frac{374}{374+0} = \frac{374}{374} \times 100\% = 100\%$$

Hasil tersebut menggunakan ketentuan *criterion gain ratio*, *apply pruning* dengan *confidence* 0,1 dan *apply prepruning* 0,01 dan *maximal depth* 10 serta *cross validation number of folds* sebesar 10 dan *sampling type automatic*.

#### 4.1.13.2 Cross Validation pada algoritma Naïve Bayes

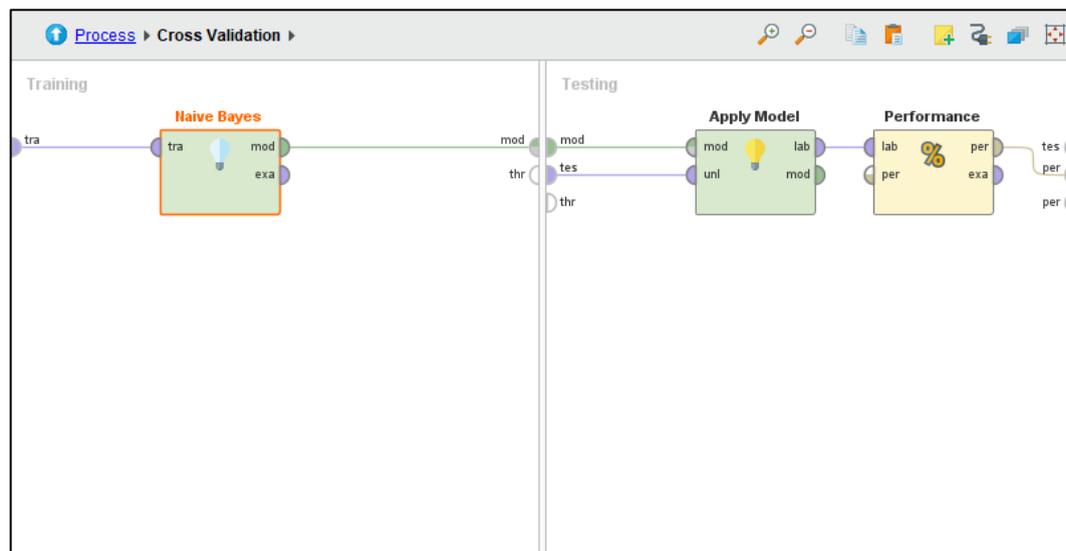
Dataset yang telah tersimpan dalam *rapid miner* dapat digunakan kembali untuk tahapan *cross validation*. Tahapan pertama dataset dimasukkan kedalam proses, kemudian gunakan operator *cross validation*, dapat dilihat pada gambar di bawah ini:



**Gambar 4.27** Susunan *Cross Validation* algoritma *Naïve bayes*

Dalam melatih dan menguji pemodelan *Cross Validation* menggunakan tiga operator yaitu, pada bagian *training* digunakan untuk algoritma *naïve bayes*, dan untuk bagian *testing* digunakan fitur *Apply Model* yang berguna untuk untuk

menampilkan *confusion table*, dan operator *performance* digunakan untuk menampilkan hasil dari *accuracy*, *precision*, dan *recall*.



**Gambar 4.28** Susunan operator pada algoritma *Naive Bayes*

Kemudian jika tampilan sudah seperti pada gambar 4.28 maka dapat dijalankan dengan menekan fitur *run* yang terletak pada bagian kiri atas.

Table View Plot View

accuracy: 84.78% +/- 4.09% (micro average: 84.78%)

	true potensi	true sangat potensi	true kurang potensi	class precision
pred. potensi	370	185	4	66.19%
pred. sangat potensi	4	613	2	99.03%
pred. kurang potensi	0	0	103	100.00%
class recall	98.93%	76.82%	94.50%	

**Gambar 4.29** Hasil *Cross Validation* algoritma *Naive Bayes*

Dalam gambar 4.29 dapat di lihat bahwa algoritma *naive bayes* memiliki nilai akurasi sebesar 84,78% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus persamaan (5)

$$Accuracy = \frac{370+613+103}{370+185+4+4+613+2+103} = \frac{1086}{1281} \times 100\% = 84,78\%$$

nilai presisi untuk prediksi label potensi sebesar 99,20% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus persamaan (6)

$$Presisi = \frac{370}{370+185+3} = \frac{370}{559} \times 100\% = 66,18\%$$

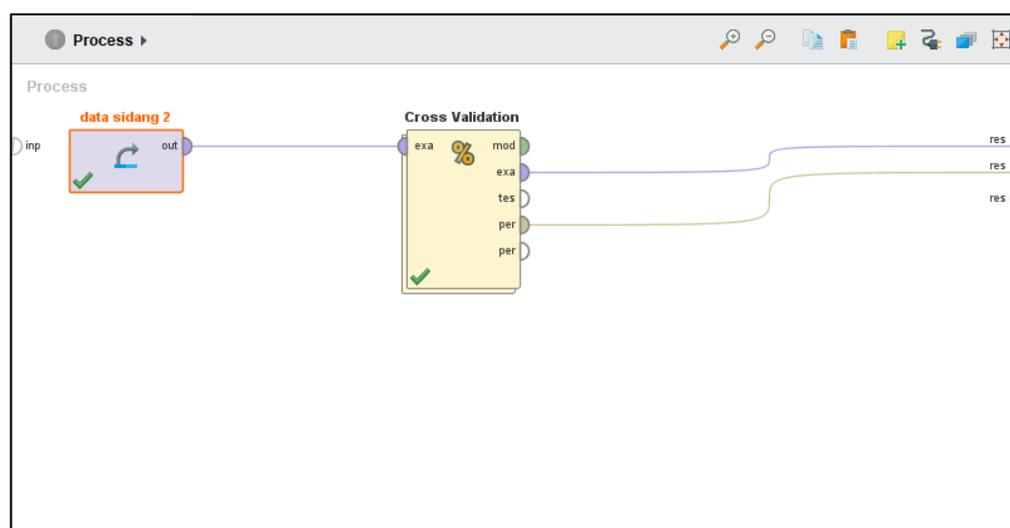
serta nilai *recall* untuk prediksi label potensi sebesar 97,98% hasil tersebut dibuktikan dengan perhitungan manual dengan rumus persamaan (7)

$$Recall = \frac{370}{370+4} = \frac{370}{374} \times 100\% = 98,93\%$$

hasil tersebut dengan ketentuan *cross validation number of folds* sebesar 10 dan *sampling type automatic*.

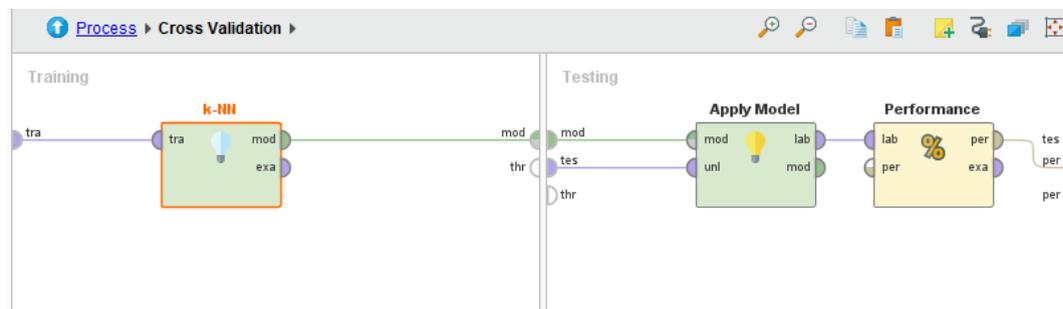
#### 4.1.13.3 Cross Validation pada algoritma K- Nearest Neighbor

Dataset yang telah tersimpan dalam *rapid miner* dapat digunakan kembali untuk tahapan *cross validation*. Tahapan pertama dataset dimasukan kedalam proses, kemudian gunakan operator *data split* untuk menentukan jumlah data trening dan data *testing* yang akan digunakan, dalam penelitian menggunakan rumus perbandingan 0,7 data trening dan 0,3 data *testing*, kemudian tahapan berikutnya menggunakan *cross validation*, dapat dilihat pada gambar di bawah ini:



**Gambar 4.30** Susunan *Cross Validation* algoritma K-Nearest Neighbor

Dalam melatih dan menguji pemodelan *Cross Validation* menggunakan tiga operator yaitu, pada bagian *training* digunakan untuk algoritma *k-nearest neighbor*, dan untuk bagian *testing* digunakan fitur *Apply Model* yang berguna untuk untuk menampilkan *confusion table*, dan operator *performance* digunakan untuk menampilkan hasil dari *accuracy*, *precision*, dan *recall*.



**Gambar 4.31** Susunan operator pada algoritma *K-Nearest Neighbor*

Kemudian jika tampilan sudah seperti pada gambar 4.31 maka dapat dijalankan dengan menekan fitur *run* yang terletak pada bagian kiri atas.

Table View Plot View

accuracy: 99.61% +/- 0.76% (micro average: 99.61%)

	true potensi	true sangat potensi	true kurang potensi	class precision
pred. potensi	374	3	0	99.20%
pred. sangat potensi	0	795	2	99.75%
pred. kurang potensi	0	0	107	100.00%
class recall	100.00%	99.62%	98.17%	

**Gambar 4.32** Hasil *Cross Validation* algoritma *K-Nearest Neighbor*

Dalam gambar 4.32 dapat di lihat bahwa algoritma *k-nearest neighbor* memiliki nilai akurasi sebesar 99,61% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus persamaan (5)

$$Accuracy = \frac{374+395+107}{374+3+795+2+107} = \frac{1276}{1281} \times 100\% = 99,61\%$$

nilai presisi untuk prediksi label potensi sebesar 99,20% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus persamaan (6)

$$Presisi = \frac{374}{374+3} = \frac{374}{377} \times 100\% = 99,20\%$$

serta nilai *recall* untuk prediksi label potensi sebesar 97,98% hasil tersebut dibuktikan dengan perhitungan manual dengan rumus persamaan (7)

$$Recall = \frac{374}{374+0} = \frac{374}{374} \times 100\% = 100\%$$

Hasil tersebut berdasarkan ketentuan K=5 dan weighted dengan type mixed measures, dan mixed measure mixed euclidean disatance, serta *cross validation number of folds* sebesar 10 dan *sampling type automatic*.

#### 4.2 Pembahasan

Penelitian ini menggunakan tiga algoritma yaitu, *decision tree*, *naïve bayes*, dan *k-nearest neighbor*. Ketiga algoritma tersebut memiliki nilai akurasi yang berbeda-beda, untuk algoritma *decision tree* memiliki nilai akurasi sebesar 100% dan dan yang kedua yaitu algoritma *k-nearest neighbor* memiliki nilai akurasi 99,61% serta yang terakhir yaitu algoritma *naïve bayes* mendapatkan nilai akurasi 84,78%. Berikut tabel dari hasil pengukuran akurasi, presisi, *recall* dari ketiga algoritma tersebut yaitu:

**Tabel 4.39** Hasil Pengukuran Algoritma

Algoritma	Akurasi	Presisi	<i>Recall</i>
<i>Decision tree</i>	100%	100%	100%
<i>K-Nearest Neighbor</i>	99,61%	99,20%	100%
<i>Naïve Bayes</i>	84,78%	66,18%	98,93%

berdasarkan tabel 4.39 maka dapat dilihat algoritma yang memiliki nilai akurasi, presisi, *recall*, yang cukup baik yaitu algoritma *decision tree*, namun pada penerapan algoritma *decision tree* variabel luar kota tidak dapat terbaca oleh *rapid miner*, setelah di analisa manual menggunakan *microsoft excel* variabel luar kota memiliki nilai *gain* yang cukup rendah, sehingga variabel luar kota dianggap tidak berpengaruh oleh pengolahan hasil *rapid miner*.