

BAB IV HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pada penelitian ini data yang digunakan adalah data sekunder. Data sekunder sendiri merupakan sumber data yang didapatkan peneliti dengan media perantara atau tidak secara langsung. Data sekunder yang diperoleh dari BSI Universitas Muhammadiyah Yogyakarta yaitu *tracert study* alumni UMY. Data yang diambil sebanyak 698 data dan yang digunakan sebanyak 435 data. Kenapa data yang diambil berjumlah sekian karena banyak data yang ada pada variabel tertentu bernilai *null* sehingga data tersebut tidak dapat digunakan untuk proses *data mining*.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P			
1	id	nim	nama	tempat_lahir	jenis_kelamin	id_prodi	bulan_lahir	tahun_lahir	ipk	judul_skrripsi	password	alamat	email	nomor_hp	nomor_tlp_rumah	id_jenis_pekerjaan	nama_u	
2	761189	0001032006	Solikin		L	52		2009	NULL	NULL	20010120064	NULL	NULL	NULL	NULL	5	NULL	
3	762049	1998110294	Hana Tri Permiana		NULL	71		2009	3.34	NULL	1998110294	NULL	NULL	NULL	NULL	5	NULL	
4	767351	98020217	Lilivandhi Putri Alamsy		NULL	52		2009	3.04	NULL	98020217	NULL	NULL	NULL	NULL	5	NULL	
5	776646	2011104000	Samsudin		NULL	504		NULL	NULL	NULL	20111040009	NULL	NULL	NULL	NULL	5	NULL	
6	776652	2011104001	Abdul Basith		NULL	504		NULL	NULL	NULL	20111040011	NULL	NULL	NULL	NULL	5	NULL	
7	776659	2011104003	Budi Mulyadi		NULL	504		NULL	NULL	NULL	20111040031	NULL	NULL	NULL	NULL	5	NULL	
8	776664	2011104008	Masyuri		NULL	504		NULL	NULL	NULL	20111040088	NULL	NULL	NULL	NULL	5	NULL	
9	776655	2011104005	Jimmy Nazroen		NULL	504		NULL	NULL	NULL	20111040057	NULL	NULL	NULL	NULL	5	NULL	
10	776656	2011104003	Iskandar		NULL	504		NULL	NULL	NULL	20111040039	NULL	NULL	NULL	NULL	5	NULL	
11	767352	2000012106	Fahmi Ariandi	Bengkulu	L	51	1	2000	3.06	NULL	2000012106	2001154964	2001154964	2001154964	2001154964	5	Unnes	
12	767362	2002011006	Daryon Mulyadi	Bandung Agung	L	51	50	2011	2.34	Audi Keselamsantia	2002011006	Jin. Merai daryon_m	082137601	082137601	082137601	5	CV. Alm	
13	768009	2007013004	Dwi Atmaja Luhur Sayek	Yogyakarta	L	53	10	2011	3.48	Pengaruh Frasa	0813015	Sembung dwi atma	085443625	085443625	085443625	5	PT. NOK	
14	768012	2007022011	Rizki Nurhidana	Siemam	P	52	10	2011	3.7	NULL	50889	Papungan rizki nur	080443318	080443318	080443318	5	Cy Jangle	
15	768035	2007031004	Alhayer Sulaiman	Kebonkahi	L	51	50	2011	3.04	NULL	21573	Ji. Seran d'ahayer sul	085244604	085244604	085244604	5	Rumah	
16	768058	2007031005	Deffi Ashariini	Bantul	P	51	50	2011	3.48	NULL	502568	Ji. Brigjen deffi asha	081328183	081328183	081328183	5	puskes	
17	768085	2007031007	Ali Muslim Bahriy	Gresik	L	51	50	2011	3.04	NULL	566824	Jebungah ali muslim	085888881	085888881	085888881	5	CV. ALI	
18	768102	2007031005	Viona Winingsih C	Magelang	P	51	50	2011	3.39	NULL	933243	Ningsih viona wini	087205054	087205054	087205054	5	RELO M	
19	768124	2007031012	Agustina Puspitasari	Yogyakarta	P	51	50	2011	2.99	NULL	909212	Ji. Jaganah agustina	085404081	085404081	085404081	5	kinik s	
20	768164	2007031017	Yogi Pratama Mustarati	Labuan Sumbawa	L	51	50	2011	3.04	NULL	218673	Ji. Garuda yogi prat	081311383	081311383	081311383	5	RS man	
21	768191	2007031005	Dina Agnita Sari	Baliapapan	P	52	10	2011	3.01	NULL	607991	Ji. Diker 2 dina agni	085150602	085150602	085150602	5	RILO B	
22	768205	2007032002	Ahliyah Wardhana	Wonosobo	L	52	50	2011	2.94	NULL	546004	Larangan wardhana	081327535	081327535	081327535	5	NULL	
23	768276	2007032011	Januar Rizki	Merabuan	L	52	50	2011	3.11	NULL	544483	Ji. Prof M'y Januar riz	081324267	081324267	081324267	5	NULL	
24	768318	2007042004	Aulia Daulia	Martanam	P	52	50	2011	2.77	NULL	833209	Ji. Lidah Bu aulia daul	081343286	081343286	081343286	5	kinik e	
25	768377	2007042005	Citra Selviaha Febrina	Indramayu	P	51	50	2011	3.48	NULL	90101	Jin. Kapur citra febr	081322051	081322051	081322051	5	PT. Ben	
26	768404	2007042007	Milla Nuraida Fahni	Yogyakarta	P	52	50	2011	3.24	NULL	1111	Biendaha milla nur	085404081	085404081	085404081	5	CV. TRI	
27	768535	2008011004	Razi Oktari	Way Unang	L	71	10	2011	3.51	Evaluasi Tinggi	0011081	Perum Bud razi okta	081278934	081278934	081278934	5	PT. Wij	
28	768554	2008012002	Muhammad Yusuf Musti	Blau Blau	L	52	50	2011	3.12	IMPLEMENTASI	000002002	Ji. Haraga muhamma	085251125	085251125	085251125	5	Unnes	
29	768559	2009072004	Muhammad Saleh	Suneneo	L	72	50	2011	3.36	NULL	stohic1398	Taljan Ang muhamm	081378988	081378988	081378988	5	Madar	
30	768615	20090530134	Rizki Yuthi Prianto	Jakarta	L	59	5	2012	3.34	NULL	20030530134	-	senyatarat	081378988	081378988	081378988	5	Usaha
31	768641	20040530146	Wahyu Setiawan Nugro	Siemam	L	59	5	2012	2.89	NULL	SUEMANIA	Wahyan S wahy_wah	087890052	087890052	087890052	5	DISPER	

Gambar 4. 1 View Data

Data asli yang didapatkan penulis memiliki 32 atribut diantaranya yaitu id, nim, nama, tempat lahir, jenis kelamin, id prodi, bulan lulus, tahun lulus, ipk, judul skripsi, password, alamat, email, nomor hp, nomor telepon rumah, id jenis pekerjaan, nama institusi, alamat institusi, bagia, masa tunggu, id salary range, tanggal lahir, tanggal mulai kerja, tanggal akhir kerja, nama institusi pertama, alamat institusi pertama, relevansi pendidikan, alasan relevansi, saran, angkatan, dan jabatan.

Selain itu pada data asli juga terdapat 4 keterangan yaitu master prodi, master fakultas, master jenis pekerjaan, dan master range gaji untuk menjelaskan id yang ada pada data kekeluruhan.

#	A	B	C
1	id	id_fakultas	nama_prodi
2	11	1	S1. TEKNIK SIPIL
3	12	1	S1. TEKNIK ELEKTRO
4	13	1	S1. TEKNIK MESIN
5	14	1	S1. TEKNOLOGI INFORMASI
6	21	2	S1. AGROTEKNOLOGI
7	22	2	S1. AGROBISNIS
8	31	3	S1. PENDIDIKAN DOKTER
9	32	3	S1. ILMU KEPERAWATAN
0	34	3	S1. KEDOKTERAN GIGI
1	35	3	S1. FARMASI
2	41	4	S1. EKONOMI MANAJEMEN
3	42	4	S1. AKUNTANSI
4	43	4	S1. ILMU EKONOMI
5	51	5	S1. ILMU HUBUNGAN INTERNASIONAL
6	52	5	S1. ILMU PEMERINTAHAN
7	53	5	S1. ILMU KOMUNIKASI
8	54	5	S1. PEND. BAHASA INGGRIS
9	61	6	S1. ILMU HUKUM
0	71	7	S1. DAKWAH (Komunikasi & Penyiaran Islam)
1	72	7	S1. TARBIAH (Pendidikan Agama Islam)
2	73	7	S1. SYARIAH (Ekonomi & Perbankan Islam)
3	74	7	D-III Pendidikan Agama Islam
4	81	8	S1. Pendidikan Bahasa Inggris
5	82	8	S1. Pendidikan Bahasa Arab
6	83	8	S1. Pendidikan Bahasa Jepang
7	101	10	S2. Magister Studi Islam
8	102	10	S2. Magister Manajemen
9	103	10	S2. Magister Manajemen Rumah Sakit
0	104	10	S2. Magister Ilmu Pemerintahan
1	105	10	S2. Magister Keperawatan
2	106	10	S2. Magister Ilmu Hubungan Internasional
3	107	10	S2. Magister Ilmu Hukum
4	201	20	S3. Psikologi Pendidikan Islam
5	202	20	S3. Politik Islam
6	301	30	D3. Teknik Elektromedik
7	302	30	D3. Teknik Mesin Otomotif dan Produksi
8	303	30	D3. Akuntansi Terapan
9	304	30	D3. Teknik Komputer Jaringan

Gambar 4. 2 ID Prodi

1	id	nama_fakultas
2	1	TEKNIK
3	2	PERTANIAN
4	3	KEDOKTERAN
5	4	EKONOMI
6	5	ISIPOL
7	6	HUKUM
8	7	AGAMA ISLAM
9	8	PENDIDIKAN BAHASA
10	10	PASCA SARJANA
11	20	PASCA S3
12	30	VOKASI D3

Gambar 4. 3 ID Fakultas

1	id	keterangan
2	1	Belum Bekerja
3	2	Pemerintah (Pusat / Departemen)
4	3	Pemerintah (Daerah)
5	4	Pemerintah (BUMN, BHMN)
6	5	Swasta (Jasa)
7	6	Swasta (Manufaktur)
8	7	Wiraswasta
9	8	Lain-Lain

Gambar 4. 4 ID Jenis Pekerjaan

1	id	range
2	1	< 1.000.000
3	2	1.000.000 - 3.000.000
4	3	3.000.000 - 5.000.000
5	4	5.000.000 - 7.500.000
6	5	7.500.000 - 10.000.000
7	6	10.000.000 - 12.500.000
8	7	12.500.000 - 15.000.000
9	8	> 15.000.000

Gambar 4. 5 Range Gaji

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 *gender*, *fakultas*, *ipk*, *tahun lulus*, & *tanggal mulai kerja*. Karena informasi yang terkandung didalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan *indicator* penelitian.

jenis_kelamin	id_prodi	tahun_lulus	ipk	tanggal_mulai_kerja
L	52	2003	NULL	NULL
L	11	2009	3.14	NULL
L	52	2003	3.04	NULL
L	104	NULL	NULL	NULL
L	104	NULL	NULL	NULL
L	104	NULL	NULL	NULL
L	104	NULL	NULL	NULL
L	104	NULL	NULL	NULL
L	104	NULL	NULL	NULL
L	51	2011	3.05	2013-11-01 00:00:00.000
L	11	2011	2.34	NULL
L	13	2011	3.48	2012-04-14 00:00:00.000
P	22	2011	3.7	NULL
L	51	2011	3.04	2015-12-01 00:00:00.000
P	51	2011	3.46	2015-10-01 18:01:00.000
L	51	2011	3.04	2015-09-01 00:00:00.000
P	51	2011	3.33	NULL
P	51	2011	2.99	2015-12-15 00:00:00.000
L	51	2011	3.04	2016-01-01 00:00:00.000
P	52	2011	3.01	2015-02-02 00:00:00.000
L	52	2011	2.84	2016-03-28 00:00:00.000
L	52	2011	3.11	NULL
P	54	2011	2.77	2016-04-01 00:00:00.000
P	41	2011	3.48	2013-01-01 00:00:00.000
P	42	2011	3.24	2012-06-21 00:00:00.000
L	11	2011	3.51	2012-01-04 00:00:00.000
L	12	2011	3.12	2015-02-02 00:00:00.000
L	72	2011	3.36	2012-07-31 00:00:00.000
L	53	2012	3.34	2014-10-06 00:00:00.000
L	53	2012	2.65	2016-01-04 00:00:00.000
L	51	2012	3.14	2015-08-10 21:21:00.000
L	52	2012	3.12	NULL
L	53	2012	3.04	NULL
L	61	2012	2.81	2013-01-15 00:00:00.000
L	12	2012	3.13	2016-12-01 00:00:00.000

Gambar 4. 6 Seleksi Data

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 dibutuhkan telah melalui tahap *cleaning data* maka penulis mendapatkan 435 data yang akan di olah menjadi 2 bagian yaitu *data training* dan *data testing* yang siap dimasukan pada software yang digunakan pada penelitian ini.

4.4. Transformasi Data

Data Transformation adalah tahap mengubah data menjadi bentuk yang sesuai untuk diproses dalam *data mining*. Dalam penelitian ini data yang akan diproses dari *Microsoft Excel 2016* akan digunakan untuk pengolahan pada *Software RapidMiner*. Dan atribut data juga di ubah dari *id_prodi* menjadi *fakultas*, *tanggal_mulai_kerja* diubah menjadi *status_mulai_kerja*. Gambar 4.7 adalah data yang sudah ditransformasikan kedalam *excel*.

jenis_kelam	Fakultas	ipk	tahun_lulus	status_mulai_kerja
L	HUKUM	3.06	2011	2013-11-01 00:00:00.000
L	TEKNIK	3.48	2011	2012-04-14 00:00:00.000
L	KEDOKTERAN	3.04	2011	2015-12-01 00:00:00.000
P	KEDOKTERAN	3.46	2011	2015-10-01 18:01:00.000
L	KEDOKTERAN	3.04	2011	2015-09-01 00:00:00.000
P	KEDOKTERAN	2.99	2011	2015-12-15 00:00:00.000
L	KEDOKTERAN	3.04	2011	2016-01-01 00:00:00.000
P	KEDOKTERAN	3.01	2011	2015-02-02 00:00:00.000
L	KEDOKTERAN	2.94	2011	2016-03-28 00:00:00.000
P	KEDOKTERAN	2.77	2011	2016-04-01 00:00:00.000
P	EKONOMI	3.48	2011	2013-01-01 00:00:00.000
P	EKONOMI	3.24	2011	2012-06-21 00:00:00.000
L	TEKNIK	3.51	2011	2012-01-04 00:00:00.000
L	TEKNIK	3.12	2011	2015-02-02 00:00:00.000
L	PENDIDIKAN AGAMA	3.36	2011	2012-07-31 00:00:00.000
L	ISIPOL	3.34	2012	2014-10-06 00:00:00.000
L	ISIPOL	2.65	2012	2016-01-04 00:00:00.000
L	ISIPOL	3.14	2012	2015-08-10 21:21:00.000

Gambar 4. 7 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 status_mulai_kerja seperti tabel berikut ini:

Tabel 1 Inisialisasi nilai IPK

Nilai IPK	Inisialisasi
IPK < 3	KURANG DARI 3
IPK 3 > 3.50	DARI 3 SAMPAI 3.50
IPK > 3.50	LEBIH DARI 3.50

Tabel 2 Inisialisasi Status Mulai Kerja

Status_Mulai_Kerja	Inisialisasi
Dari tahun lulus – 2 tahun	CEPAT
Lebih dari 2 tahun	LAMBAT

Selanjutnya data yang sudah di inisialisasi di transformasikan kedalam *excel* kembali seperti gambar 4.8 dibawah.

jenis_kelamin	fakultas	ipk	tahun_lulus	status_mulai_kerja
P	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2012	CEPAT
L	TEKNIK	LEBIH DARI 3.50	2011	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
P	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
L	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2012	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2011	CEPAT
P	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
P	EKONOMI	LEBIH DARI 3.50	2012	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
P	PENDIDIKAN AGAMA	LEBIH DARI 3.50	2012	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2012	CEPAT
P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
L	TEKNIK	LEBIH DARI 3.50	2012	CEPAT
P	EKONOMI	DARI 3 SAMPAI 3.50	2011	CEPAT
L	TEKNIK	KURANG DARI 3	2012	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
L	PENDIDIKAN AGAMA	LEBIH DARI 3.50	2012	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
L	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2011	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2012	CEPAT
P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
P	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2012	CEPAT
L	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
P	PENDIDIKAN AGAMA	LEBIH DARI 3.50	2012	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT

Gambar 4. 8 Data yang Sudah 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 dikumpulkan, diseleksi dan ditransformasi akan dikelola menggunakan 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 alumni universitas muhammadiyah yogyakarta memiliki 435 *record* , untuk data *training* memiliki 250 *record* data (tahun 2011,2012,2014) dan data *testing* memiliki 185 *record* data (tahun 2013).

jenis_kel	fakultas	ipk	tahun_lul	status_mulai_kerja
P	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2012	CEPAT
L	TEKNIK	LEBIH DARI 3.50	2011	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
P	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
L	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2012	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2011	CEPAT
P	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
P	EKONOMI	LEBIH DARI 3.50	2012	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
P	PENDIDIKAN AGAMA	LEBIH DARI 3.50	2012	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2012	CEPAT
P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
L	TEKNIK	LEBIH DARI 3.50	2012	CEPAT
P	EKONOMI	DARI 3 SAMPAI 3.50	2011	CEPAT
L	TEKNIK	KURANG DARI 3	2012	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
L	PENDIDIKAN AGAMA	LEBIH DARI 3.50	2012	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
L	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2011	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2012	CEPAT
P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT

Gambar 4. 9 Data Training

jenis_kelamin	fakultas	ipk	tahun_lul	status_mulai_kerja
L	ISIPOL	KURANG DARI 3	2013	CEPAT
P	HUKUM	DARI 3 SAMPAI 3.50	2013	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
P	EKONOMI	LEBIH DARI 3.50	2013	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2013	CEPAT
P	PENDIDIKAN AGAMA	LEBIH DARI 3.50	2013	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2013	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2013	CEPAT
L	EKONOMI	DARI 3 SAMPAI 3.50	2013	CEPAT
P	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2013	CEPAT
L	PERTANIAN	KURANG DARI 3	2013	CEPAT
L	TEKNIK	DARI 3 SAMPAI 3.50	2013	CEPAT
L	ISIPOL	DARI 3 SAMPAI 3.50	2013	CEPAT
L	EKONOMI	LEBIH DARI 3.50	2013	CEPAT
L	TEKNIK	LEBIH DARI 3.50	2013	CEPAT
P	EKONOMI	LEBIH DARI 3.50	2013	CEPAT
L	EKONOMI	LEBIH DARI 3.50	2013	CEPAT
P	EKONOMI	LEBIH DARI 3.50	2013	CEPAT
L	PERTANIAN	KURANG DARI 3	2013	CEPAT
L	ISIPOL	LEBIH DARI 3.50	2013	CEPAT

Gambar 4. 10 Data Testing

4.6. Hasil dan Pembahasan Perhitungan Berdasarkan Jumlah Data

Pada tahap ini penulis melakukan perhitungan data testing berdasarkan jumlah data yang digunakan, yaitu data sebanyak 50, 75, 100, 125, 150, dan 175 data. Sedangkan data testing yang di pakai sebanyak 435 data. Hasil yang didapatkan berbeda-beda, seperti pada tabel 3 di bawah.

Tabel 3 Hasil Perhitungan Berdasarkan Jumlah Data

Jumlah Data	Akurasi	CLASS			
		CEPAT (tepat)	CEPAT (tidak tepat)	LAMBAT (tepat)	LAMBAT (tidak tepat)
50	84%	25 data	8 data	17 data	0 data
75	77%	34 data	14 data	18 data	9 data
100	71%	47 data	26 data	24 data	3 data
125	69%	48 data	30 data	38 data	9 data
150	67%	52 data	39 data	49 data	10 data
175	63%	63 data	38 data	48 data	26 data

Berdasarkan tabel 3 di atas, data yang diambil secara acak. Untuk jumlah data 50 memiliki akurasi 84% dengan ketepatan prediksi sebanyak 42 data dan yang tidak tepat sebanyak 8 data. Untuk jumlah data 75 memiliki tingkat akurasi sebesar 77% dengan ketepatan prediksi sebanyak 52 data dan yang tidak tepat sebanyak 20 data. Untuk jumlah data 100 memiliki tingkat akurasi sebesar 71% dengan ketepatan prediksi sebanyak 71 data dan yang tidak tepat sebanyak 29 data. Untuk jumlah data 125 memiliki tingkat akurasi sebesar 69% dengan ketepatan prediksi sebanyak 86 data dan yang tidak tepat sebanyak 39 data. Untuk jumlah data 150 memiliki tingkat akurasi sebesar 67% dengan ketepatan prediksi sebanyak 101 data dan yang tidak tepat sebanyak 49 data. Terakhir untuk jumlah data 175 memiliki tingkat akurasi sebesar 63% dengan ketepatan prediksi sebanyak 111 data dan yang tidak tepat sebanyak 64 data.

Dapat dilihat dari hasil perhitungan tabel 3 diatas, bahwa semakin besar jumlah data yang digunakan maka semakin kecil tingkat akurasinya dan besar kecilnya tingkat akurasi juga berdasarkan isi data dari tiap atribut.

4.7. Hasil dan Pembahasan Perhitungan Menggunakan 50 Data Testing

Berdasarkan hasil pada tabel 4 diatas. Tingkat akurasi untuk tahun lulus 2011 dengan data training 420 data & testing 15 data yaitu sebesar 93,33% serta jumlah prediksi yang tepat sebanyak 14 data dan yang tidak tepat sebanyak 1 data. Untuk tahun lulus 2012 memiliki data training sebanyak 217 data & testing sebanyak 50 data, memiliki tingkat akurasi sebesar 64,00% dan jumlah prediksi yang tepat sebanyak 32 data, sedangkan yang tidak tepat sebanyak 18 data. Untuk tahun lulus 2013 memiliki jumlah data training sebanyak 250 data dan data testingnya sebanyak 50 data, dengan tingkat akurasi sebesar 62,00% dan jumlah prediksi yang tepat sebanyak 31 data, sedangkan yang tidak tepat sebanyak 19 data. Untuk tahun lulus 2014 memiliki jumlah data training sebanyak 418 data dan data testingnya sebanyak 17 data, dengan tingkat akurasi sebesar 88,24%, dan untuk jumlah prediksi yang tepat sebanyak 15 data, sedangkan yang tidak tepat sebanyak 2 data.

4.8. Hasil dan Pembahasan Perhitungan Menggunakan Data Keseluruhan

Pada tahap ini, penulis melakukan hal yang sama seperti tahap 4.7 di atas. Tapi bedanya pada tahap ini, penulis menggunakan jumlah data keseluruhan. Hasil perhitungan pada tabel 5, tingkat akurasi untuk tahun lulus 2011 dengan data training 420 data & testing 15 data yaitu 93,33% & jumlah prediksi yang tepat sebanyak 14 data, sedangkan yang tidak tepat sebanyak 1 data. Untuk tahun lulus 2012 data training 217 data & data testing 218 data dengan tingkat akurasi sebesar 65,14% dan jumlah prediksi yang tepat sebanyak 142 data, sedangkan yang tidak tepat sebanyak 76 data. Untuk tahun lulus 2013 data training 250 data & data testing sebanyak 185 data dengan tingkat akurasi sebesar 71,89% dan jumlah prediksi yang tepat sebanyak 133 data, sedangkan yang tidak tepat sebanyak 52 data. Untuk tahun lulus 2014 data training 418 data & testing 17 data, dengan tingkat akurasi sebesar 88,24% dan untuk jumlah prediksi yang tepat sebanyak 15 data, sedangkan yang tidak tepat sebanyak 2 data.

Tabel 4 Hasil Perhitungan Berdasarkan Tahun Lulus (50 data)

TAHUN LULUS	CLASS						CLASS PREDICTION		CLASS RECALL	
	JUMLAH DATA	AKURASI	CEPAT (tepat)	CEPAT (tidak tepat)	LAMBAT (tepat)	LAMBAT (tidak tepat)	CEPAT	LAMBAT	CEPAT	LAMBAT
2011	15	93,33%	6 data	1 data	8 data	0 data	85,71%	100,00%	100,00 %	88,89%
2012	50	64,00%	20 data	13 data	12 data	5 data	60,61%	70,59%	80,00%	48,00%
2013	50	62,00%	25 data	19 data	6 data	0 data	56,82%	100,00%	100,00 %	24,00%
2014	17	88,24%	15 data	0 data	0 data	2 data	88,24%	00,00%	100,00 %	00,00%

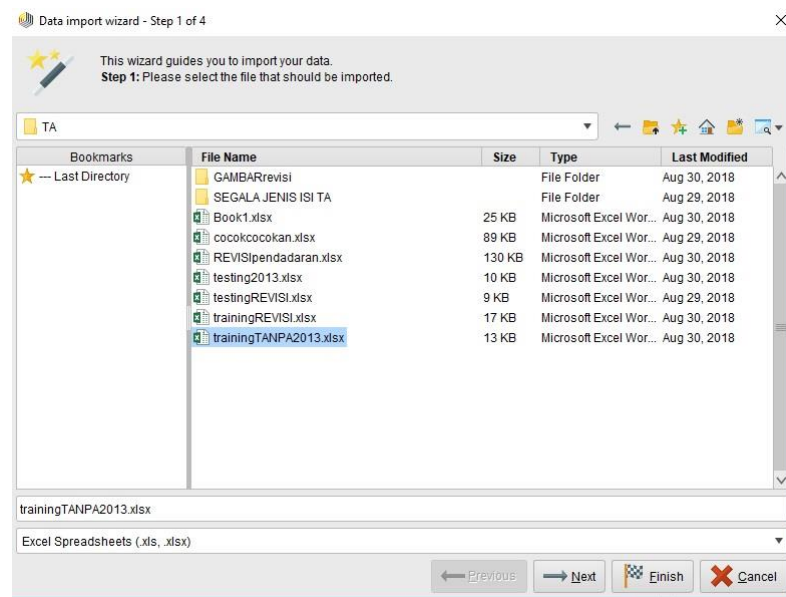
Tabel 5 Hasil Pengujian Berdasarkan Data Asli

TAHUN LULUS	CLASS							CLASS PREDICTION		CLASS RECALL	
	JUMLAH DATA	AKURASI	CEPAT (tepat)	CEPAT (tidak tepat)	LAMBAT (tepat)	LAMBAT (tidak tepat)	CEPAT	LAMBAT	CEPAT	LAMBAT	
88,89%	15	93,33%	6 data	1 data	8 data	0 data	85,71%	100,00%	100,00 %	88,89%	
2012	218	65,14%	71 data	59 data	71 data	17 data	54,62%	80,68%	80,68%	54,62%	
2013	185	71,89%	116 data	40 data	17 data	12 data	74,36%	58,62%	90,52%	29,82%	
2014	17	88,24%	15 data	0 data	0 data	2 data	88,24%	00,00%	100,00 %	00,00%	

4.9. Pengujian *software Rapidminer*

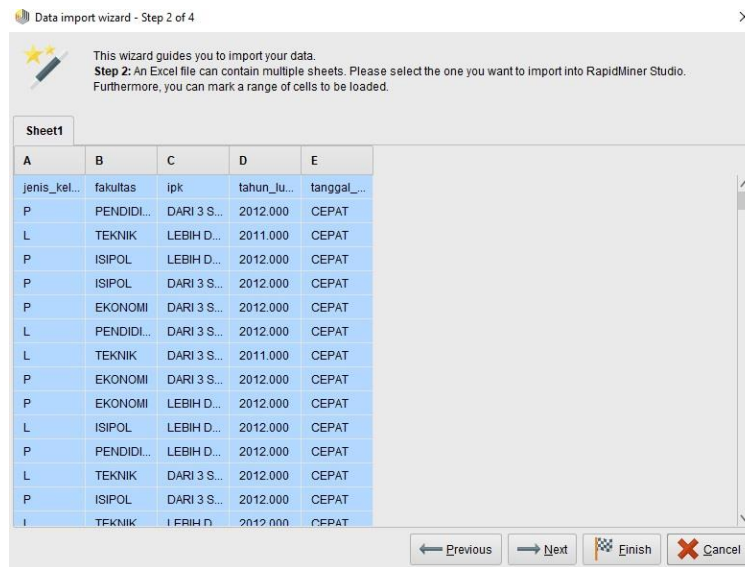
Pada tahap ini penulis hanya mengambil data training tahun 2011, 2012, 2014 dan data testing tahun 2013 sebagai contoh. Atribut yang digunakan sebagai *label* adalah *status_mulai_kerja*, penulis menganalisis 185 data (tahun 2013) yang dipakai sebagai data testing dengan format *excel* untuk memprediksi alumni yang mendapatkan pekerjaan cepat dan lambat agar bisa diakses menggunakan *software RapidMiner*.

Selanjutnya lakukan importing data terlebih dahulu yaitu *data training* dan *data testing* yang ada pada komputer atau laptop seperti pada gambar 4.11.



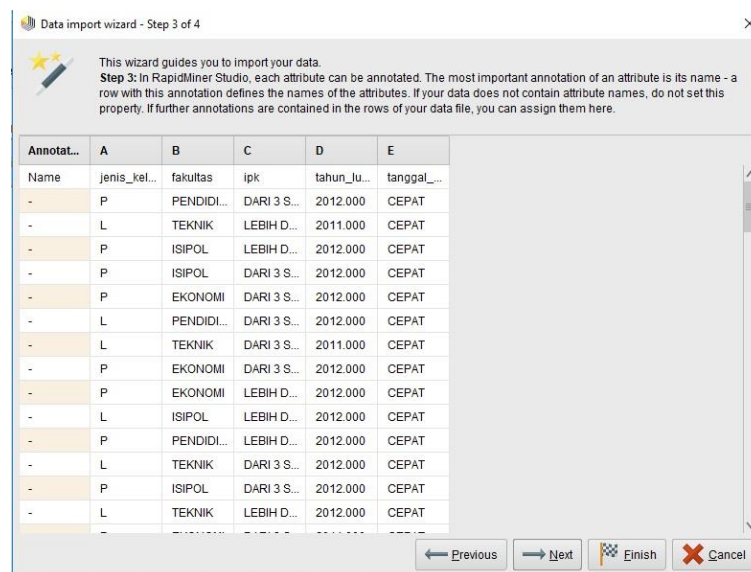
Gambar 4. 11 Import Data Training

Setelah data training telah dipilih langkah selanjutnya yaitu klik *Next* dan akan muncul form *data import wizard step 2* terlihat pada gambar 4.12.



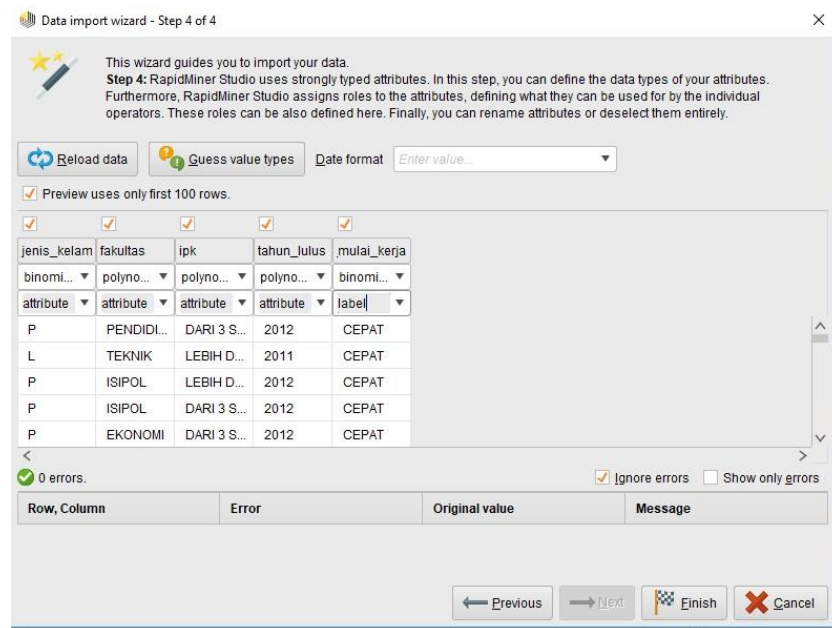
Gambar 4. 12 Data Import Wizard step 2 (data training)

Untuk melanjutkan ke step selanjutnya klik *Next* maka akan muncul *data import wizard step 3*, seperti pada gambar 4.13.



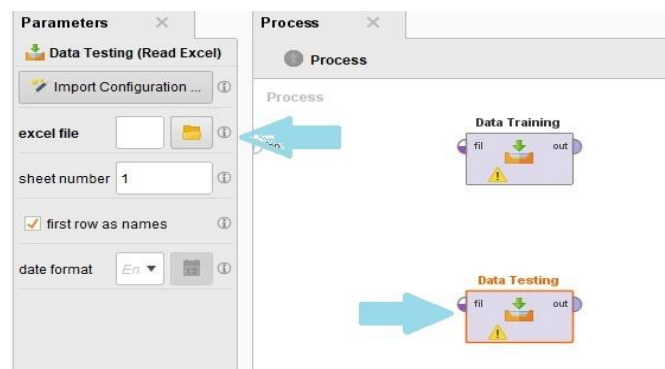
Gambar 4. 13 Data import wizard step 3 (data training)

Pada *step 3* ini tidak dilakukan perubahan apapun maka langsung ke data import wizard *step 4* dengan cara klik *Next* maka akan muncul form *data import wizard step 4* seperti pada gambar 4.14.



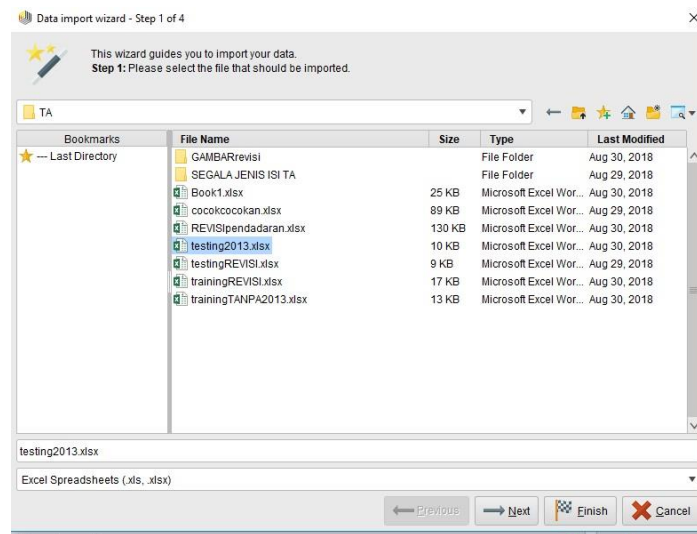
Gambar 4. 14 Data import wizard step 4 (data training)

Setelah muncul form seperti gambar 4.14 di atas lalu pilih atribut target atau *label* yang digunakan yaitu atribut *status mulai kerja*. Setelah ditentukan *label* yang dipilih maka klik *Finish*. Setelah import data trainig selesai makan langkah selanjutnya yaitu melakukan *import wizard data testing* dnegan menggunakan langkah-langkah yang sama.



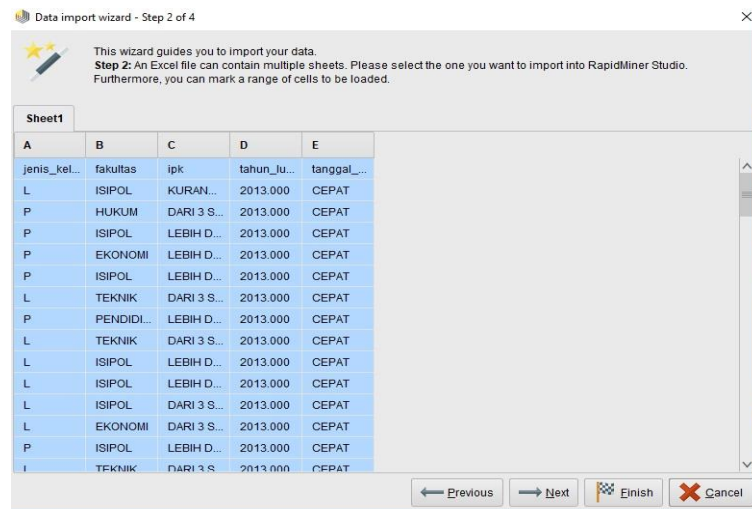
Gambar 4. 15 Impport File Tetsing

Setelah klik *import configuration wizard* maka akan muncul form *data import wizard step 1* untuk memilih lokasi file data *testing* yang digunakan untuk penelitian ini. Seperti pada gambar 4.16.



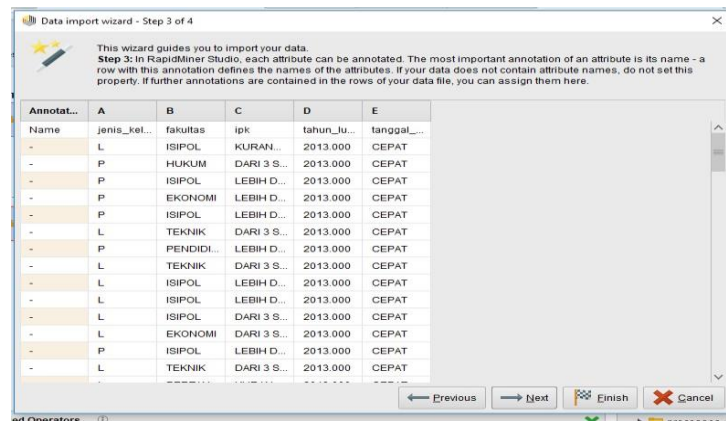
Gambar 4. 16 Form data Import File Testing

Setelah data dipilih step selanjutnya yaitu klik *Next* dan akan muncul *form data import wizard step 2* pada gambar 4.17.



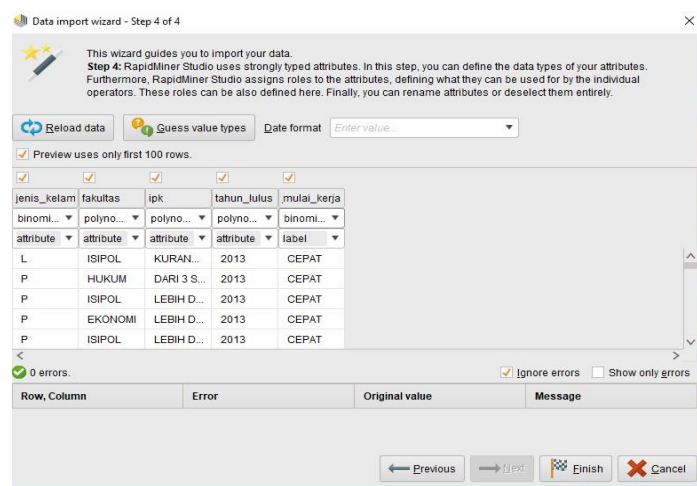
Gambar 4. 17 Data import wizard step 2 (data testing)

Setelah muncul seperti gambar 4.17 diatas lalu pilih *Next* untuk melakukan step berikutnya. Setelah klik *Next* maka akan muncul form data import wizard step 3 pada gambar 4.18.



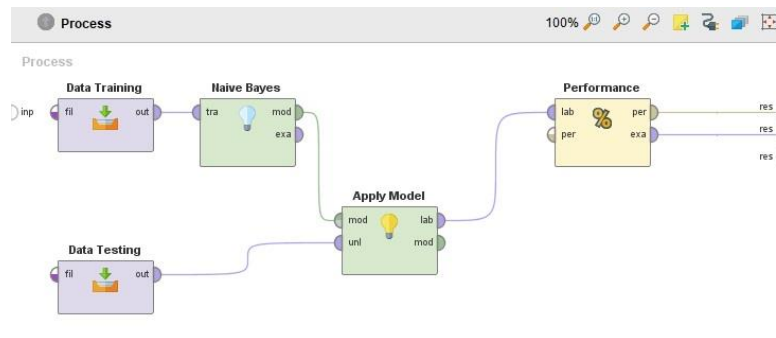
Gambar 4. 18 Data import wizard step 3 (data testing)

Pada *step 3* ini tidak dilakukan perubahan apapun maka langsung ke data import wizard *step 4* dengan cara klik *Next* maka akan muncul form *data import wizard step 4* seperti pada gambar 4.19.



Gambar 4. 19 Data import wizard step 4 (data testing)

Setelah muncul form seperti gambar 4.14 di atas lalu pilih atribut target atau *label* yang digunakan yaitu atribut *status mulai kerja*. Setelah ditentukan *label* yang dipilih maka klik *Finish*. Kemudian setelah kedua data sudah di import maka step selanjutnya yaitu drag and drop operators *Naive bayes*, *apply model*, dan *performace* lalu hubungjan dengan data *training* dan *testing*. Seperti pada gambar 4.20.



Gambar 4. 20 Menghubungkan Operator Performance

Setelah semuanya terhubung, klik icon *run* pada *toolbar*, seperti pada gambar 4.21 untuk menampilkan hasil. Tunggu beberapa saat, komputer memerlukan waktu untuk menyelesaikan perhitungan.



Gambar 4. 21 Menjalankan RapidMiner

Setelah beberapa detik maka *RapidMiner* akan menampilkan hasil prediksi pada *view result*. Hasilnya berbentuk tabel seperti gambar 4.22.

ExampleSet (185 examples, 4 special attributes, 4 regular attributes) Filter (185 / 185 examples): a

Row No.	status_mula...	prediction(s...	confidence(...	confidence(...	jenis_kelamin	fakultas	ipk	tahun_lulus
1	CEPAT	CEPAT	0.797	0.203	L	ISIPOL	KURANG DA...	2013
2	CEPAT	CEPAT	0.999	0.001	P	HUKUM	DARI 3 SAMP...	2013
3	CEPAT	CEPAT	0.581	0.419	P	ISIPOL	LEBIH DARI ...	2013
4	CEPAT	CEPAT	0.519	0.481	P	EKONOMI	LEBIH DARI ...	2013
5	CEPAT	CEPAT	0.581	0.419	P	ISIPOL	LEBIH DARI ...	2013
6	CEPAT	CEPAT	0.629	0.371	L	TEKNIK	DARI 3 SAMP...	2013
7	CEPAT	CEPAT	0.654	0.346	P	PENDIDIKAN...	LEBIH DARI ...	2013
8	CEPAT	CEPAT	0.629	0.371	L	TEKNIK	DARI 3 SAMP...	2013
9	CEPAT	CEPAT	0.708	0.292	L	ISIPOL	LEBIH DARI ...	2013
10	CEPAT	CEPAT	0.708	0.292	L	ISIPOL	LEBIH DARI ...	2013
11	CEPAT	CEPAT	0.744	0.256	L	ISIPOL	DARI 3 SAMP...	2013
12	CEPAT	CEPAT	0.693	0.307	L	EKONOMI	DARI 3 SAMP...	2013
13	CEPAT	CEPAT	0.625	0.375	P	ISIPOL	DARI 3 SAMP...	2013
14	CEPAT	CEPAT	0.693	0.307	L	EKONOMI	DARI 3 SAMP...	2013
15	CEPAT	CEPAT	0.581	0.419	P	ISIPOL	LEBIH DARI ...	2013
16	CEPAT	CEPAT	0.629	0.371	L	TEKNIK	DARI 3 SAMP...	2013
17	CEPAT	CEPAT	0.753	0.247	L	PERTANIAN	KURANG DA...	2013
18	CEPAT	CEPAT	0.693	0.307	L	EKONOMI	DARI 3 SAMP...	2013

Gambar 4. 22 Hasil Perhitungan

Dari hasil proses perhitungan menggunakan *rapid miner* dengan metode prediksi menampilkan hasil dari data *testing* dan *training* yang telah diuji. Kolom ini memberikan informasi tentang data alumni yang mendapatkan pekerjaan dengan cepat ataupun lambat pada tahun 2013. Selanjutnya untuk mengetahui tingkat *accuracy* algoritma *naive bayes*, klik tab *performanceVector* yang terletak disebelah kanan, sehingga *table view* seperti gambar 4.23.

accuracy: 71.89%

	true CEPAT	true LAMBAT	class precision
pred. CEPAT	116	40	74.36%
pred. LAMBAT	12	17	58.62%
class recall	90.62%	29.82%	

Gambar 4. 23 Accuracy

Dapat dilihat pada gambar 4.23 tingkat *accuracy* dari *performanceVector* yaitu 71,89%, *class precision* yaitu CEPAT 74.36%, LAMBAT 58.62%, dan untuk *class recall* yaitu CEPAT 90,62% sedangkan LAMBAT 29,82%.

Secara Umum precision, recall, dan accuracy dapat dirumuskan sebagai berikut:

➤ Untuk Kelas CEPAT

$$\text{precision} = \frac{116}{116 + 40} = \frac{116}{156} = 0,7435897436 = 74,36\%$$

$$\text{recall} = \frac{116}{116 + 12} = \frac{116}{128} = 0,90625 = 90,62\%$$

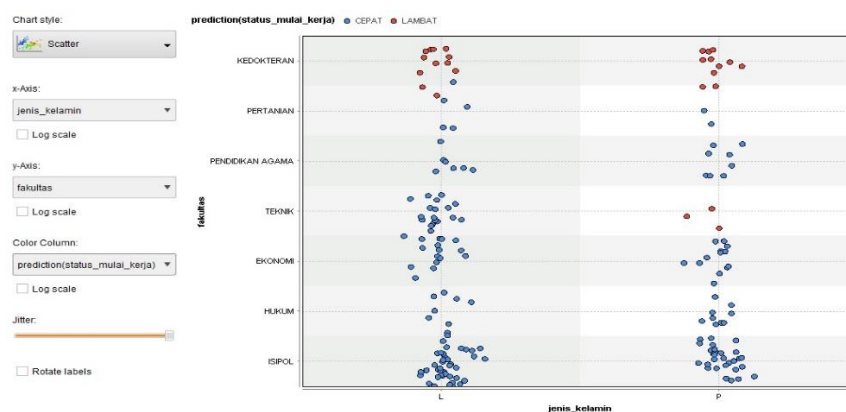
➤ Untuk Kelas LAMBAT

$$\text{precision} = \frac{17}{17 + 12} = \frac{17}{29} = 0,5862068966 = 58,62\%$$

$$\text{recall} = \frac{17}{17 + 40} = \frac{17}{57} = 0,298245614 = 29,82\%$$

➤ Untuk Tingkat Accuracy

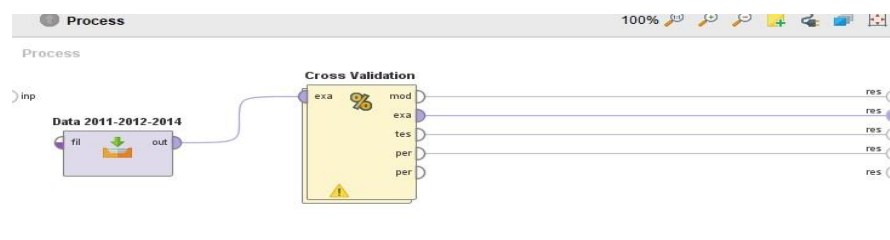
$$\text{accuracy} = \frac{116 + 17}{116 + 17 + 40 + 12} = \frac{133}{185} = 0,7189189189 = 71,89\%$$



Gambar 4. 24 Grafik Scatter

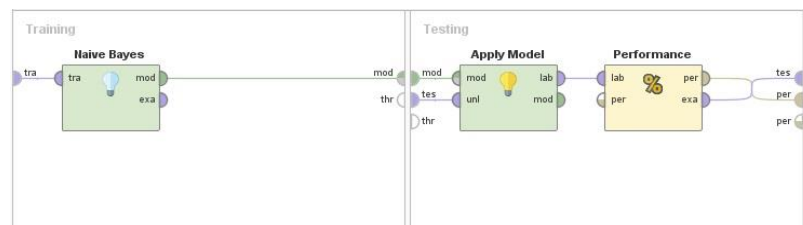
Berdasarkan gambar 4.24 di atas , grafik scatter diatas yaitu alumni Universitas Muhammadiyah Yogyakarta pada tahun 2013 dari 7 fakultas lebih banyak yang cepat mendapatkan pekerjaan setelah menyelesaikan studi S1 kecuali pada fakultas kedokteran.

Pengujian kedua menggunakan data angkatan 2011, 2012, dan 2014 yang *cepat* dan *lambat* mendapatkan pekerjaan sebagai *training* dan *testing*. Data tersebut kemudian diujiakan dengan menggunakan operator *cross validation* di *rapidminer* seperti pada gambar 4.25.



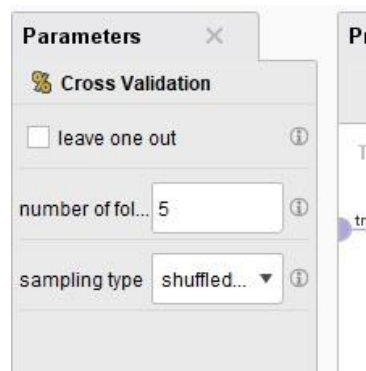
Gambar 4. 25 Cross Validation

Selanjutnya adalah mengatur operator *cross validation* dengan cara *double klik* operator *cross validation* Kemudian *drag* dan *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 pada gambar 4.26.



Gambar 4. 26 Cross Validation View

Langkah selanjutnya adalah masukan angka 5 pada kolom *number of fold* dan pilih *shuffled sampling* pada kolom *sampling type* seperti pada gambar 4.27.



Gambar 4. 27 Parameters Cross Validation

Setelah semuanya telah terhubung dan *parameter* operator *cross validation* terisi, klik *Run* pada toolbar. Maka akan muncul hasil perhitungan seperti pada gambar 4.28.

Table View
 Plot View

accuracy: 65.60% +/- 4.45% (micro average: 65.60%)

	true CEPAT	true LAMBAT	class precision
pred. CEPAT	78	55	58.65%
pred. LAMBAT	31	86	73.50%
class recall	71.56%	60.99%	

Gambar 4. 28 Hasil Shuffled Sampling

Dapat dilihat pada gambar 4.28 nilai akurasi menurun ketika menggunakan operator *cross validation* dengan jumlah *5 fold* dibanding perhitungan pertama menggunakan data *testing* angkatan 2013. Langkah selanjutnya yaitu membandingkan tingkat akurasi tertinggi menggunakan metode *sampling type* yang berbeda dengan *5 fold* dan hasilnya sebagai berikut:

Tabel 6 Perbandingan Akurasi

Metode	Nilai Akurasi
Menggunakan lulusan 2013 (data testing)	71, 89%
Shuffled Sampling	65.60%

Jadi Hasil Perbandingan pada tabel 6 di atas yaitu tingkat akurasi yang menggunakan Shuffled Sampling lebih rendah dibanding , metode yang tidak menggunakan metode shuffled Sampling.

4.10. Pengujian dengan Algoritma Naive Bayes

Pengujian pada penelitian ini menggunakan algoritma *Naive Bayes*, pada penelitian ini penulis membuat tabel data yang digunakan dalam penelitian seperti tabel 4 dan tabel 5.

Tabel 7 Data Training

No	Jenis_kelamin	fakultas	ipk	Tahun_lulus	Status_mulai_kerja
1	P	AGAMA ISLAM	DARI 3 SAMPAI 3.50	2012	CEPAT
2	L	TEKNIK	LEBIH DARI 3.50	2011	CEPAT
3	P	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
4	P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
5	P	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
6	L	AGAMA ISLAM	DARI 3 SAMPAI 3.50	2012	CEPAT
7	L	TEKNIK	DARI 3 SAMPAI 3.50	2011	CEPAT
8	P	EKONOMI	DARI 3 SAMPAI 3.50	2012	CEPAT
9	P	EKONOMI	LEBIH DARI 3.50	2012	CEPAT
10	L	ISIPOL	LEBIH DARI 3.50	2012	CEPAT
12	P	AGAMA ISLAM	LEBIH DARI 3.50	2012	CEPAT

No	Jenis_kelamin	fakultas	ipk	Tahun_lulus	Status_mulai_kerja
13	L	TEKNIK	DARI 3 SAMPAI 3.50	2012	CEPAT
14	P	ISIPOL	DARI 3 SAMPAI 3.50	2012	CEPAT
15	L	TEKNIK	LEBIH DARI 3.50	2012	CEPAT
.....	L	KEDOKTERAN	LEBIH DARI 3.0	2012	LAMBAT
.....	P	KEDOKTERAN	DARI 3 SAMPAI 3.50	2012	LAMBAT
.....	L	TEKNIK	KURANG DARI 3	2014	LAMBAT
.....	P	ISIPOL	LEBIH DARI 3.50	2014	LAMBAT
.....	L	TEKNIK	KURANG DARI 3	2014	LAMBAT
250	P	ISIPOL	LEBIH DARI 3.50	2014	LAMBAT

Tabel 8 Data Testing

No	Jenis_kelamin	fakultas	ipk	Tahun_lulus	Status_mulai_kerja
1	P	ISIPOL	DARI 3 SAMPAI 3.50	2013	CEPAT

No	Jenis kelamin	fakultas	ipk	Tahun_lulus	Status_mulai_kerja
2	P	EKONOMI	DARI 3 SAMPAI 3.50	2013	CEPAT
3	L	AGAMA ISLAM	DARI 3 SAMPAI 3.50	2013	CEPAT
4	L	ISIPOL	KURANG DARI 3	2013	CEPAT
5	P	HUKUM	DARI 3 SAMPAI 3.50	2013	CEPAT
6	P	EKONOMI	LEBIH DARI 3.50	2013	CEPAT
7	P	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
8	L	TEKNIK	DARI 3 SAMPAI 3.50	2013	CEPAT
9	L	ISIPOL	LEBIH DARI 3.50	2013	CEPAT
10	L	TEKNIK	KURANG DARI 3	2013	CEPAT
...	L	ISIPOL	DARI 3 SAMPAI 3.50	2013	CEPAT
...	L	TEKNIK	DARI 3 SAMPAI 3.50	2013	LAMBAT

No	Jenis kelamin	fakultas	ipk	Tahun_lulus	Status_mulai_kerja
...	L	HUKUM	DARI 3 SAMPAI 3.50	2013	LAMBAT
...	P	KEDOKTERAN	LEBIH DARI 3.50	2013	LAMBAT
...	L	KEDOKTERAN	DARI 3 SAMPAI 3.50	2013	LAMBAT
185	L	EKONOMI	LEBIH DARI 3.50	2013	LAMBAT

Tabel 7 diatas memiliki 250 *record* sebagai data *training* dan tabel 8 diatas memiliki 92 *record* sebagai data testing yang akan digunakan dalam penelitian ini. Dalam kasus yang ada pada tabel 7 dan 8 akan dibuat prediksi untuk menentukan alumni yang ada pada tabel 8 mendapatkan pekerjaan cepat atau lambat menggunakan probabilitas. Untuk menghitung nilai probabilitas atau prediksi alumni yang dapat pekerjaan cepat atau lambat dengan menggunakan cara sebagai berikut:

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

Pertama menghitung jumlah dari *Cepat* dan *Lambat* dari tabel data *training*. Dari tabel data *training* didapatkan hasil sebagai berikut:

- CEPAT = 237
- LAMBAT = 198

Setelah mengetahui jumlah alumni yang cepat dan lambat mendapatkan pekerjaan langkah selanjutnya adalah menghitung masing-masing *attribut* dari data testing seperti berikut :

- *Jenis Kelamin = P, Fakultas = ISIPOL, IPK = DARI 3 SAMPAI 3.50, Tahun Lulus = 2013*
- Tahap 1 : Menghitung jumlah class/label (probabilitas awal)
 - $P(\text{status_mulai_kerja} \mid \text{Jumlah data})$
 - $P(\text{CEPAT} / 435) = 237/435 = 0,54462759$
 - $P(\text{LAMBAT} / 435) = 198/435 = 0,45517241$
- Tahap 2 : Menghitung jumlah *record atribut* dengan class/label yang sama
 - $P(\text{Jenis Kelamin} \mid \text{jumlah data (cepat \& lambat)})$
 - $P(P / \text{CEPAT} = 95/237) = 0,4008438819$
 - (perhitungan di atas ialah jumlah data jenis kelamin “perempuan” dengan keterangan “cepat” dibagi jumlah data cepat)
 - $P(P / \text{LAMBAT} = 126/198) = 0,5151515152$
 - (perhitungan di atas ialah jumlah data jenis kelamin “perempuan” dengan keterangan “lambat” dibagi jumlah data lambat)
 - $P(\text{Fakultas} = \text{ISIPOL} \mid \text{CEPAT}) = 101/237 = 0,4261603376$
 - (perhitungan di atas ialah jumlah data fakultas “isipol” dengan keterangan “cepat” dibagi jumlah data cepat)
 - $P(\text{Fakultas} = \text{ISIPOL} \mid \text{LAMBAT}) = 72/198 = 0,2424242424$
 - (perhitungan di atas ialah jumlah data fakultas “isipol” dengan ketetapan “lambat” dibagi jumlah data lambat)
 - $P(\text{IPK} = \text{DARI 3 SAMPAI 3.50} \mid \text{CEPAT}) = 131/237 = 0,552742616$

(perhitungan di atas ialah jumlah data ipk “dari 3 sampai 3.50” dengan keterangan “cepat” dibagi jumlah data cepat)

$$P (\text{IPK} = \text{DARI 3 SAMPAI 3.50} \mid \text{LAMBAT}) = 108/198 \\ = 0,5454545455$$

(perhitungan di atas ialah jumlah data ipk “dari 3 sampai 3.50” dengan keterangan “lambat” dibagi jumlah data lambat)

$$\text{➤ } P (\text{Tahun Lulus} = 2013 \mid \text{CEPAT}) = 128/237 = \\ 0,5400843882$$

(perhitungan di atas ialah jumlah data tahun lulus “2013” dengan keterangan “cepat” dibagi jumlah data cepat)

$$P (\text{Tahun Lulus} = 2013 \mid \text{LAMBAT}) = 57/198 = \\ 0,2878787879$$

(perhitungan di atas ialah jumlah data tahun lulus “2013” dengan keterangan “lambat” dibagi jumlah data lambat)

- Tahap 3 : Dikalikan semua hasil variabel CEPAT & LAMBAT
 - $P(\text{Jenis Kelamin} = P \mid \text{CEPAT}) \times P(\text{Fakultas} = \text{ISIPOL} \mid \text{CEPAT}) \times P(\text{Ipk} = \text{DARI 3 SAMPAI 3.50} \mid \text{CEPAT}) \times P(\text{Tahun Lulus} = 2013 \mid \text{CEPAT})$

$$= 0,4008438819 \times 0,4261603376 \times 0,552742616 \times 0,5400843882$$

$$= 0,0509956181$$
 - $P(\text{Jenis Kelamin} = P \mid \text{LAMBAT}) \times P(\text{Fakultas} = \text{ISIPOL} \mid \text{LAMBAT}) \times P(\text{Ipk} = \text{DARI 3 SAMPAI 3.50} \mid \text{LAMBAT}) \times P(\text{Tahun Lulus} = 2013 \mid \text{LAMBAT})$

$$= 0,5151515152 \times 0,2424242424 \times 0,5454545455 \times 0,2878787879$$

$$= 0,0095079153$$
- Tahap 4 : Bandingkan hasil class CEPAT & LAMBAT
 CEPAT = 0,0509956181

$$\text{LAMBAT} = 0,0095079153$$

Tabel 9 Hasil Perhitungan Manual

Jenis_Kelamin	Fakultas	Ipk	Tahun_lulus	Status_mulai_kerja	Hasil prediksi
P	ISIPOL	Dari 3 sampai 3.50	2013	CEPAT	CEPAT

Karena hasil (P|CEPAT) lebih besar dari (P|LAMBAT) maka hasil prediksi untuk alumni fakultas isipol yang berjenis kelamin perempuan angkatan 2013 dengan IPK dari 3 sampai 3.50 adalah “CEPAT”.

TESTING DATA						CLASS PREDICTION	CEPAT	LAMBAT
jenis_kelamin	fakultas	ipk	tahun_lulus	status_mulai_kerja				
P	ISIPOL	DARI 3 SAMPAI 3.50	2013	CEPAT	CEPAT	2,8%	0,9%	
P	EKONOMI	DARI 3 SAMPAI 3.50	2013	CEPAT	CEPAT	1,3%	0,4%	
L	PENDIDIKAN AGAMA	DARI 3 SAMPAI 3.50	2013	CEPAT	CEPAT	1,0%	0,2%	

Gambar 4. 29 Hasil Prediksi Di Excel

Dari hasil perhitungan diatas diketahui bahwa alumni dengan kategori Perempuan, Fakultas ISIPOL, dengan IPK dari 3 sampai 3.50, tahun lulus 2013, dan status mulai bekerjanya *cepat* diprediksi mendapatkan pekerjaan *cepat* karena hasil perhitungan kelas *Cepat* lebih besar dari kelas *Lambat* dengan hasil probabilitas sebesar 0,0509956181. Untuk perhitungan dan hasil lainnya akan dilampirkan pada lampiran.

4.11. Pembahasan

Berikut adalah hasil dari pengujian yang telah dilakukan pada RapidMiner dan Algoritma Naive Bayes, sebagai berikut:

1. Pada penelitian ini penulis menggunakan 250 data sebagai data *training* dan 185 data untuk data *testing* sebagai pengujian manual di excel maupun pada pengujian menggunakan *software* RapidMiner.
2. Pengujian manual pada excel menggunakan model naive bayes yaitu pada tahun 2011 lebih banyak yang lambat mendapatkan pekerjaan, dengan perbandingan *cepat* 3% - *lambat* 5%, pada tahun 2012 lebih banyak yang lambat mendapatkan pekerjaan, dengan perbandingan *cepat* 37% - *lambat* 66%, pada tahun 2013 lebih banyak yang cepat mendapatkan pekerjaan, dengan perbandingan *cepat* 54% - *lambat* 29%, dan pada tahun 2014 lebih banyak yang cepat mendapatkan pekerjaan, dengan perbandingan *cepat* 6% - *lambat* 1%.

P/tahun_lulus	CEPAT	LAMBAT
2011	3%	5%
2012	37%	66%
2013	54%	29%
2014	6%	1%
	100%	100%

Gambar 4. 30 Perhitungan Model Naive Bayes (tahun lulus)

3. Untuk bagian *Class Prediction* menggunakan 185 data pada data *testing*. Terdapat data yang cocok antara class *cepat* dan class *predicted cepat* yaitu sebanyak 116 data dan untuk data lambat yang cocok antara class *lambat* dan class *predicted lambat* sebanyak 17 data. Sedangkan untuk prediction yang tidak tepat sebanyak 52 data.

Tabel 10 Hasil Perhitungan Class Prediction (confusion tabel)

ACCURACY:	71,89%		CLASS	
		Confutin Table	CEPAT	LAMBAT
		CEPAT	116	40
		LAMBAT	12	17

4. Tingkat *accuracy* pada perhitungan manual di excel dan pengujian menggunakan *software* RapidMiner sama yaitu sebesar 71,89%. Untuk kelas CEPAT pada *RapidMiner* predictionnya sebesar 74,36% , dan *recall* pada kelas CEPAT sebesar 90,62%. Sedangkan prediction untuk kelas LAMBAT sebesar 58,62% dan *recall* pada kelas LAMBAT sebesar 29,82%. (seperti pada gambar 4.23)
5. Pada Perhitungan pertama mendapatkan akurasi sebesar 71,89% sedangkan perhitungan kedua menggunakan cross validation yaitu sebesar 65,60%
6. Faktor yang memengaruhi hasil akhir *CEPAT* dan *LAMBAT* pada *software RapidMiner* alumni mendapatkan pekerjaan karena dari perhitungan manual model naive bayes pada excel lebih tinggi persennya pada class *cepat*, yang terdiri dari atribut jenis kelamin, fakultas, ipk, tahun lulan, dan status mulai kerja.