

BAB IV

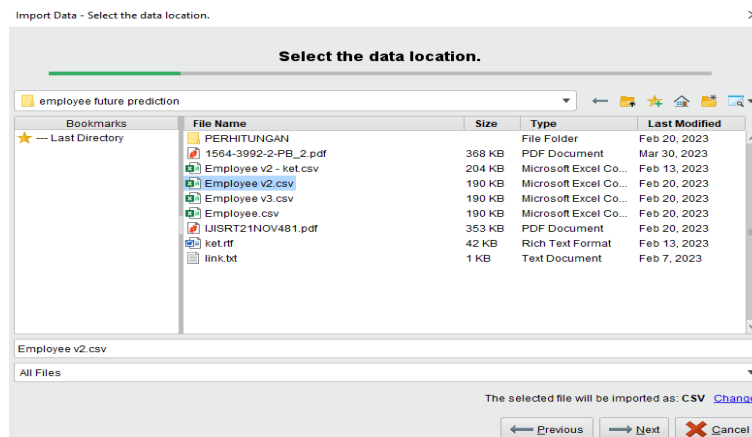
HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada penelitian ini digunakan seleksi fitur PSO dan Algoritma *Decision Tree* C4.5. kasus yang diambil adalah data *employee future prediction* yang berjumlah 4.653 data dan dilakukan kross cek apakah ada *missing value* atau *redudance* data atau tidak, ternyata data dari Kaggle baik dan tidak harus dilakukan preprocesing lagi sehingga data yang di oleh tetap berjumlah 4.653 data yang siap diolah. Data sample ini yang akan dihitung probabilitasnya, kemudian akan dicari tingkat akurasi tertinggi dari data tersebut dan akan diimplementasikan pada tools *RapidMiner*.

4.1.1 Dataset

Dataset berupa excel yang telah siap akan di *import* ke dalam *tools* rapidminer, ambil dataset dari kaggle yang berada di penyimpanan komputer lalu pilih dan *import*.



Kemudian setelah di pilih next maka muncul tampilan seperti gambar berikut ini. Pastikan nama *sheet* dan *cell range dataset* sudah sesuai dengan data yang kita miliki.

Import Data - Specify your data format

Specify your data format

☒ Header Row File Encoding ☒ Use Quotes

Start Row Escape Character ☐ Trim Lines

Column Separator Decimal Character

☒ Skip Comments

	Education	JoiningYe...	City	Payment...	Age	Gender	EverBenc...	Experien...	LeaveOr...
1									
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0
3	Bachelors	2013	Pune	1	28	Female	No	3	1
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0
5	Masters	2016	Bangalore	3	27	Male	No	5	1
6	Masters	2017	Pune	3	24	Male	Yes	2	1
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1
10	Bachelors	2016	Pune	3	23	Male	No	1	0
11	Masters	2017	New Delhi	2	37	Male	No	2	0

no problems.

Previous Next Cancel

Gambar 4. 2 Kolom *Dataset*

Kemudian atur type atribut sesuai dengan *value* atau nilai yang ada didalamnya serta atur *role* atribut seperti label pada *dataset*, jangan lupa centang *replace errors missing value* seperti gambar berikut ini.

Import Data - Format your columns.

Format your columns.

Date format ☒ Replace errors with missing values ⓘ

	PaymentTier	Age	Gender	EverBenchd	Experience...	LeaveOrNot
	integer	integer	binominal	binominal	integer	binominal label
1	3	34	Male	No	0	0
2	1	28	Female	No	3	1
3	3	38	Female	No	2	0
4	3	27	Male	No	5	1
5	3	24	Male	Yes	2	1
6	3	22	Male	No	0	0
7	3	38	Male	No	0	0
8	3	34	Female	No	2	1
9	3	23	Male	No	1	0
10	2	37	Male	No	2	0
11	3	27	Male	No	5	1

8 warnings. View Details ⓘ

Previous Next Cancel

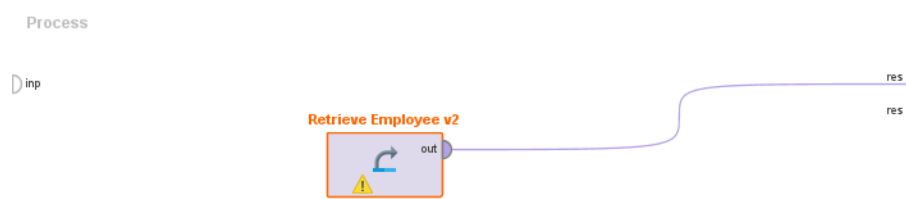
Gambar 4. 3 Format *Type Dataset*

Tabel 4. 1 *Type Data Atribut*

Atribut	Deskripsi	Kriteria
Education	Sekolah	Bachelors
		Masters
		PHD

JoiningYear	Tahun Bergabung	2012-2018
City	Kota	Bangalore
		Pune
		New Delhi
PaymentTier	Range Gaji	1= Highest (Tinggi)
		2= Mid Level (Sedang)
		3= Lowest (Rendah)
Age	Umur	22-41
Gender	Jenis Kelamin	Male = Laki-laki
		Female = Perempuan
EverBenched	Tidak pernah keluar dai proyek selama 1 bulan	Yes = Iya
		No = Tidak
ExperienceInCurrentDomain	Pengalaman di lapangan saat ini	0-7 (palingg tinggi 7)
LeaveOrNot	Apakah karyawan tinggalkan perusahaan dalam 2 tahun mendatang	0 = No
		1 = Yes

Untuk dapat melihat *dataset* pada tools rapidminer bisa dilakukasn seperti gambar dibawah ini



Gambar 4. 4 Tampilan Dataset Di Rapidminer

Type atribut juga dapat dilihat dari *statistics* pada tools rapidminer, selain type data kita juga dapat melihat *missing* data dan nilai min dan max pada setiap atribut.

Name	Type	Missing	St...	Filter (9 / 9 attributes):
Label LeaveOrNot	Binominal	0	Least 1 (1600)	Most 0 (3053)
Education	Polynominal	0	Least PHD (179)	Most Bachelors (3601)
JoiningYear	Polynominal	0	Least 2018 (367)	Most 2017 (1108)
City	Polynominal	0	Least New Delhi (1157)	Most Bangalore (2228)
PaymentTier	Polynominal	0	Least 1 (243)	Most 3 (3492)
Age	Integer	0	Min 22	Max 41
Gender	Binominal	0	Least Female (1875)	Most Male (2778)

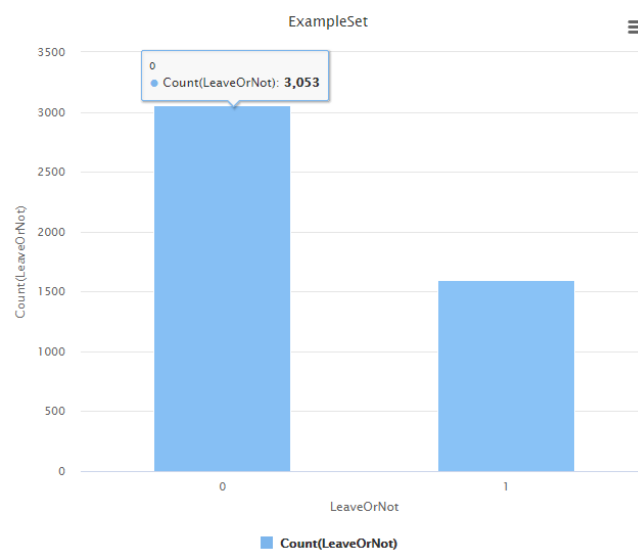
Showing attributes 1 - 9 Examples: 4,653 Special Attributes: 1 Regular Attributes: 8

Gambar 4. 5 Statistik *Dataset*

4.1.2 Model Data Masukkan

a. *LeaveOrNot*

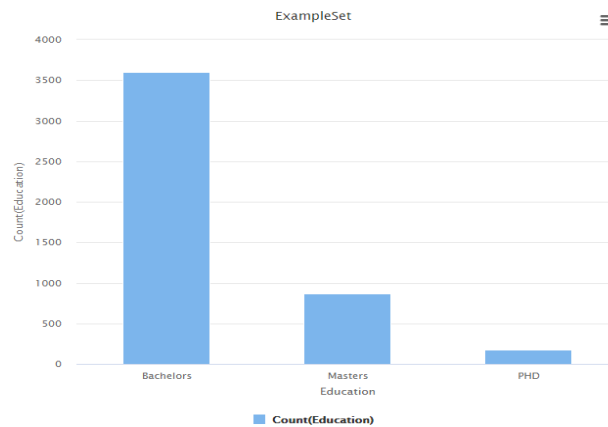
Gambar grafik dibawah merupakan gambaran karakteristik data *leaveornot* yaitu sebagai label *dataset*, dari keseluruhan dataset sebanyak 4.653 *record*, digambarkan terdapat 3053 *no* dan 1600 *yes*.



Gambar 4. 6 Data *LeaveOrNot*

b. *Education*

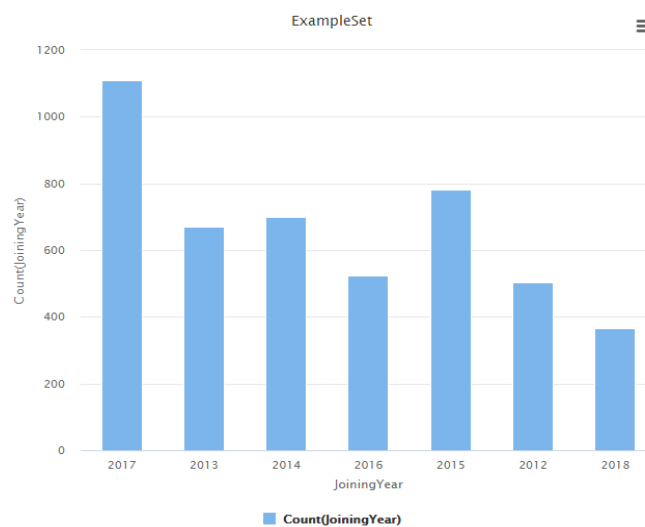
Gambar grafik dibawah merupakan gambaran karakteristik data *education* dari keseluruhan dataset sebanyak 4.653 record, digambarkan terdapat 179 *PHD*, 3601 *Bachelors* dan 873 *Masters*.



Gambar 4. 7 Data *Education*

c. *Joining Year*

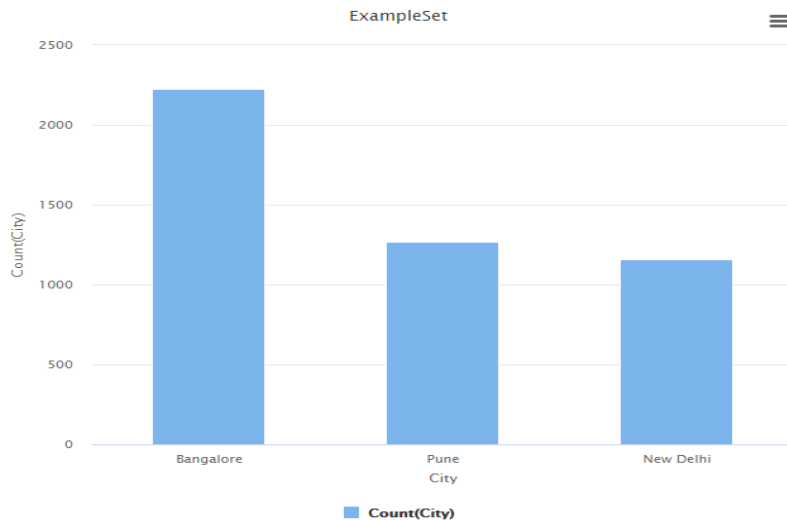
Gambar grafik dibawah merupakan gambaran karakteristik data *joining year* diambil dari data tahun 2012 sampai 2018 dari keseluruhan dataset sebanyak 4.653 record.



Gambar 4. 8 Data *Joining Year*

d. *City*

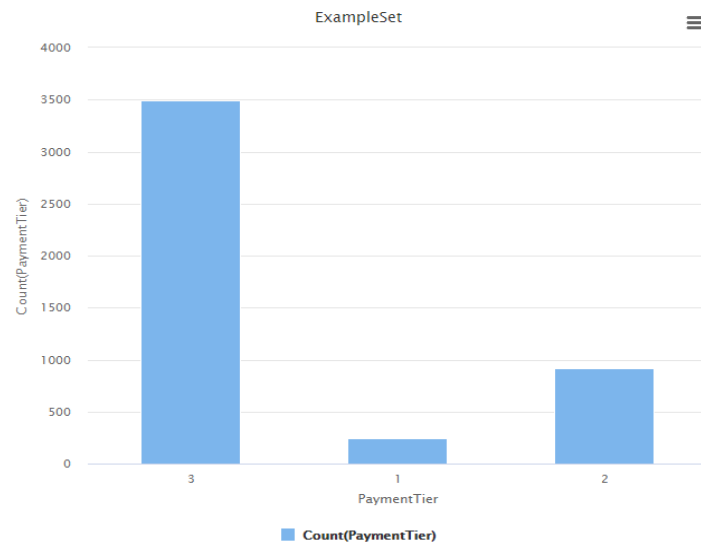
Gambar grafik dibawah merupakan gambaran karakteristik data *city* terdiri dari kota Bangalore sebanyak 2228, Pune 1268 dan kota New Delhi 1157 dari keseluruhan dataset sebanyak 4.653 *record*.



Gambar 4. 9 Data *City*

e. *Payment Tier*

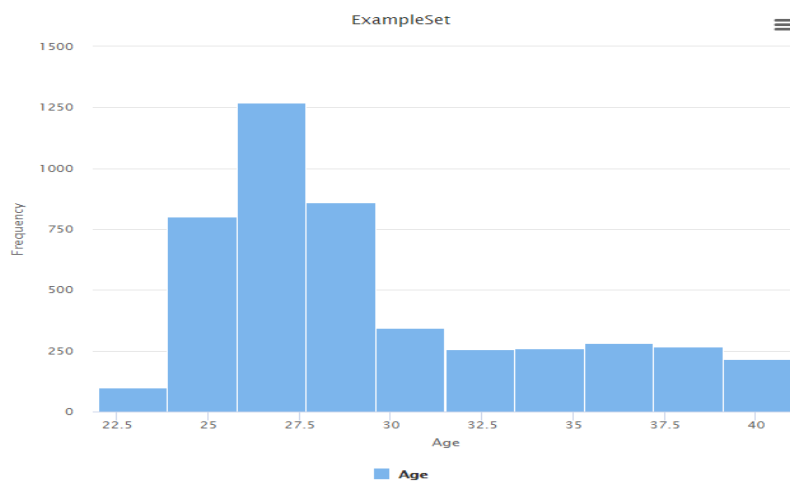
Gambar grafik dibawah merupakan gambaran karakteristik data *payment tier* dari keseluruhan dataset sebanyak 4.653 *record*, digambarkan terdapat 3492 dengan *tier* = 3, 918 *tier* = 2 dan 243 *tier* = 1.



Gambar 4. 10 Data *Payment Tier*

f. *Age*

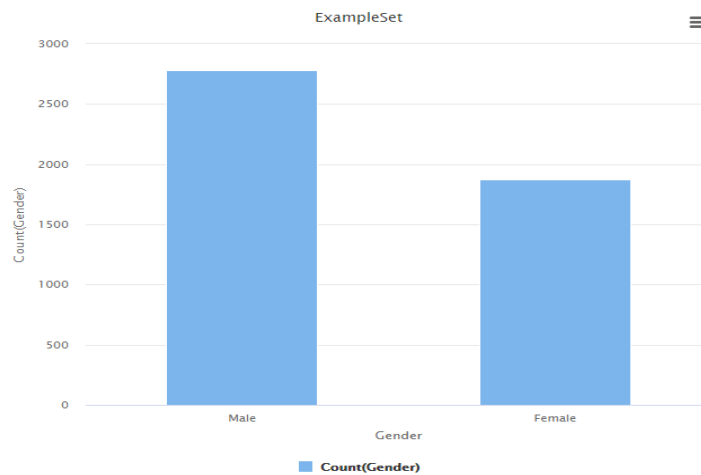
Gambar grafik dibawah merupakan gambaran karakteristik data *age* terdiri dari umur 22 sampai 41 tahun dari keseluruhan dataset sebanyak 4.653 *record*, dan didapatkan rata-rata sebesar 29,393.



Gambar 4. 11 Data *Age*

g. *Gender*

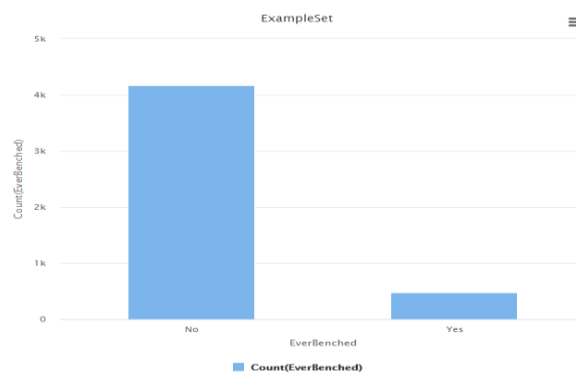
Gambar grafik dibawah merupakan gambaran karakteristik data *gender* terdiri dari *male* dan *female* dari keseluruhan dataset sebanyak 4.653 *record*, digambarkan terdapat 2778 data *male*, 1875 data *female*.



Gambar 4. 12 Data *Gender*

h. *Ever Benched*

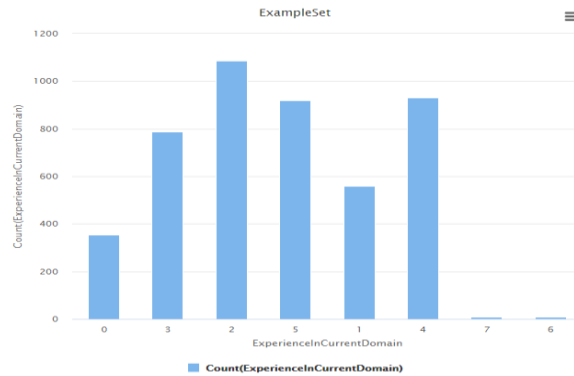
Gambar grafik dibawah merupakan gambaran karakteristik data *ever benched* terdiri dari data *yes* dan *no*. Dari keseluruhan *dataset* sebanyak 4.653 *record*, digambarkan terdapat 4175 data *yes* dan 478 data *no*.



Gambar 4. 13 Data *Ever Benched*

i. *Experience In Current Domain*

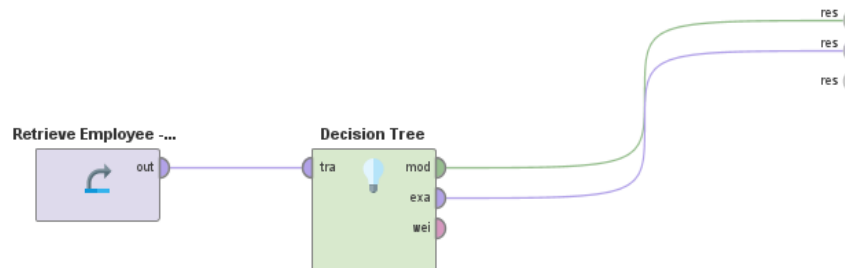
Gambar grafik dibawah merupakan gambaran karakteristik data *experience in current domain* terdiri dari range angka 0 sampai 7 dari keseluruhan dataset sebanyak 4.653 *record*, digambarkan terdapat 0 = 355, 1 = 558, 2=1087, 3=786, 4=931, 5=919, 6=8, 7=9 data.



Gambar 4. 14 Data *Experience In Current Domain*

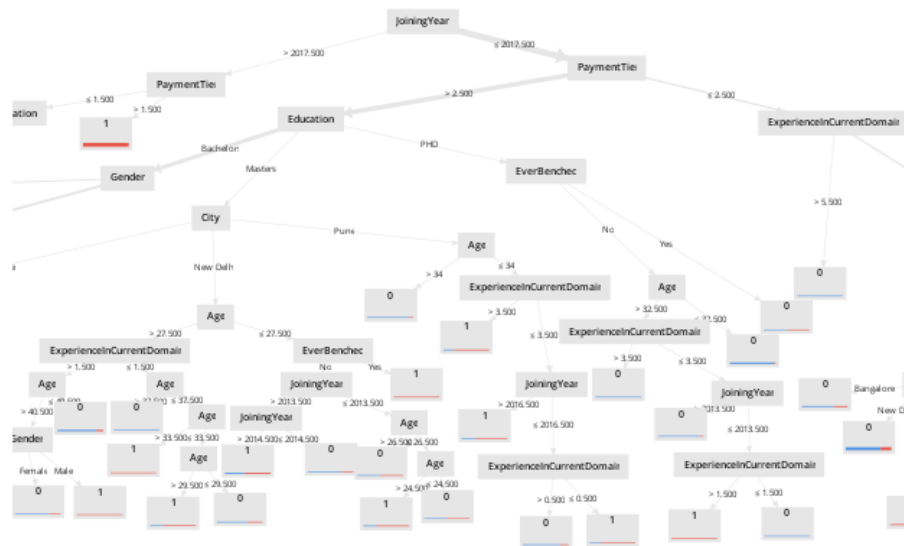
4.1.3 Algoritma *Decision Tree* C4.5

Pada tahap ini yaitu dataset yang sudah di *import* kedalam rapidminer akan dilakukan klasifikasi dengan menggunakan model algoritma *Decision Tree* C4.5 untuk menentukan pola dan hasil dari pohon keputusan.



Gambar 4. 15 Tampilan Proses *Decision Tree* C4.5

Berikut ini adalah tampilan pohon keputusan pada algoritma *Decision Tree C4.5* dimana *node* *root* awal yaitu *joining year*.



Gambar 4. 16 Tampilan Pohon Keputusan *Decision Tree C4.5*

Selanjutnya kita dapat melihat deskripsi pohon keputusan diatas, dapat kita lihat seperti gambar dibawah ini.

Tree

```
JoiningYear > 2017.500
|   PaymentTier > 1.500: 1 {0=0, 1=359}
|   PaymentTier ≤ 1.500
|   |   Education = Bachelors: 0 {0=4, 1=0}
|   |   Education = Masters
|   |   |   ExperienceInCurrentDomain > 3: 1 {0=0, 1=2}
|   |   |   ExperienceInCurrentDomain ≤ 3: 0 {0=1, 1=1}
JoiningYear ≤ 2017.500
|   PaymentTier > 2.500
|   |   Education = Bachelors
|   |   |   Gender = Female
|   |   |   |   City = Bangalore: 0 {0=401, 1=88}
|   |   |   |   City = New Delhi: 0 {0=214, 1=25}
|   |   |   |   City = Pune: 1 {0=27, 1=106}
|   |   |   Gender = Male
|   |   |   |   ExperienceInCurrentDomain > 6.500
|   |   |   |   |   JoiningYear > 2014.500: 0 {0=2, 1=0}
|   |   |   |   |   JoiningYear ≤ 2014.500: 1 {0=0, 1=2}
|   |   |   |   ExperienceInCurrentDomain ≤ 6.500: 0 {0=1570, 1=207}
|   |   Education = Masters
|   |   |   City = Bangalore
|   |   |   |   JoiningYear > 2016.500
|   |   |   |   |   ExperienceInCurrentDomain > 4.500: 1 {0=0, 1=3}
|   |   |   |   |   ExperienceInCurrentDomain ≤ 4.500
|   |   |   |   |   |   ExperienceInCurrentDomain > 1.500
|   |   |   |   |   |   |   Gender = Female
|   |   |   |   |   |   |   |   Age > 27: 1 {0=2, 1=5}
|   |   |   |   |   |   |   |   Age ≤ 27: 0 {0=3, 1=0}
|   |   |   |   |   |   |   Gender = Male
|   |   |   |   |   |   |   |   ExperienceInCurrentDomain > 3.500: 1 {0=2, 1=3}
|   |   |   |   |   |   |   |   ExperienceInCurrentDomain ≤ 3.500: 0 {0=4, 1=0}
|   |   |   |   |   |   |   |   ExperienceInCurrentDomain ≤ 1.500: 1 {0=0, 1=2}
|   |   |   |   JoiningYear ≤ 2016.500
|   |   |   |   |   Age > 28.500
|   |   |   |   |   |   Age > 29.500
|   |   |   |   |   |   |   JoiningYear > 2012.500: 1 {0=3, 1=13}
|   |   |   |   |   |   |   JoiningYear ≤ 2012.500: 0 {0=2, 1=1}
|   |   |   |   |   |   Age ≤ 29.500: 0 {0=2, 1=0}
|   |   |   |   |   Age ≤ 28.500
|   |   |   |   |   |   Age > 24.500
|   |   |   |   |   |   |   Gender = Female
|   |   |   |   |   |   |   |   ExperienceInCurrentDomain > 4.500: 0 {0=1, 1=1}
|   |   |   |   |   |   |   |   ExperienceInCurrentDomain ≤ 4.500: 1 {0=0, 1=7}
|   |   |   |   |   |   |   Gender = Male: 1 {0=0, 1=15}
|   |   |   |   |   |   |   |   Age ≤ 24.500: 0 {0=1, 1=1}
|   |   |   City = New Delhi
|   |   |   |   Age > 27.500
|   |   |   |   |   ExperienceInCurrentDomain > 1.500
|   |   |   |   |   |   Age > 40.500
|   |   |   |   |   |   |   Gender = Female: 0 {0=3, 1=1}
|   |   |   |   |   |   |   Gender = Male: 1 {0=0, 1=2}
|   |   |   |   |   |   Age ≤ 40.500: 0 {0=80, 1=14}
|   |   |   |   |   ExperienceInCurrentDomain ≤ 1.500
|   |   |   |   |   |   Age > 37.500: 0 {0=7, 1=0}
|   |   |   |   |   |   Age ≤ 37.500
|   |   |   |   |   |   |   Age > 33.500: 1 {0=0, 1=5}
|   |   |   |   |   |   |   Age ≤ 33.500
|   |   |   |   |   |   |   |   Age > 29.500: 1 {0=2, 1=5}
|   |   |   |   |   |   |   |   Age ≤ 29.500: 0 {0=8, 1=4}
```

```

| | | | Age ≤ 27.500
| | | | | EverBenchd = No
| | | | | | JoiningYear > 2013.500
| | | | | | | JoiningYear > 2014.500: 1 {0=24, 1=31}
| | | | | | | JoiningYear ≤ 2014.500: 0 {0=11, 1=4}
| | | | | | | JoiningYear ≤ 2013.500
| | | | | | | | Age > 26.500: 0 {0=5, 1=5}
| | | | | | | | Age ≤ 26.500
| | | | | | | | | Age > 24.500: 1 {0=7, 1=17}
| | | | | | | | | Age ≤ 24.500: 0 {0=1, 1=1}
| | | | | | | | EverBenchd = Yes: 1 {0=0, 1=7}
| | | | | City = Pune
| | | | | | Age > 34: 0 {0=11, 1=1}
| | | | | | Age ≤ 34
| | | | | | | ExperienceInCurrentDomain > 3.500: 1 {0=6, 1=20}
| | | | | | | ExperienceInCurrentDomain ≤ 3.500
| | | | | | | | JoiningYear > 2016.500: 1 {0=4, 1=9}
| | | | | | | | JoiningYear ≤ 2016.500
| | | | | | | | | ExperienceInCurrentDomain > 0.500: 0 {0=17, 1=4}
| | | | | | | | | ExperienceInCurrentDomain ≤ 0.500: 1 {0=1, 1=3}
| | | | | Education = PHD
| | | | | | EverBenchd = No
| | | | | | | Age > 32.500
| | | | | | | | ExperienceInCurrentDomain > 3.500: 0 {0=5, 1=0}
| | | | | | | | ExperienceInCurrentDomain ≤ 3.500
| | | | | | | | | JoiningYear > 2013.500: 0 {0=11, 1=1}
| | | | | | | | | JoiningYear ≤ 2013.500
| | | | | | | | | | ExperienceInCurrentDomain > 1.500: 1 {0=0, 1=3}
| | | | | | | | | | ExperienceInCurrentDomain ≤ 1.500: 0 {0=2, 1=0}
| | | | | | | | | Age ≤ 32.500: 0 {0=91, 1=3}
| | | | | | | | EverBenchd = Yes: 0 {0=1, 1=1}
| | | | | PaymentTier ≤ 2.500
| | | | | | ExperienceInCurrentDomain > 5.500: 0 {0=4, 1=0}
| | | | | | ExperienceInCurrentDomain ≤ 5.500
| | | | | | | JoiningYear > 2016.500
| | | | | | | | EverBenchd = No
| | | | | | | | | Age > 23.500
| | | | | | | | | | City = Bangalore: 0 {0=33, 1=14}
| | | | | | | | | | City = New Delhi: 0 {0=211, 1=64}
| | | | | | | | | | City = Pune
| | | | | | | | | | | Gender = Female
| | | | | | | | | | | | Education = Bachelors: 1 {0=1, 1=37}
| | | | | | | | | | | | Education = Masters: 0 {0=24, 1=11}
| | | | | | | | | | | Gender = Male: 0 {0=54, 1=20}
| | | | | | | | | Age ≤ 23.500
| | | | | | | | | | Education = Bachelors
| | | | | | | | | | | Gender = Female: 1 {0=1, 1=2}
| | | | | | | | | | | Gender = Male: 0 {0=2, 1=1}
| | | | | | | | | | | Education = Masters: 1 {0=2, 1=6}
| | | | | | | | | EverBenchd = Yes
| | | | | | | | | | Age > 29.500: 0 {0=13, 1=2}
| | | | | | | | | | Age ≤ 29.500
| | | | | | | | | | | Age > 27.500
| | | | | | | | | | | | Gender = Female: 1 {0=0, 1=2}
| | | | | | | | | | | | Gender = Male: 0 {0=1, 1=1}
| | | | | | | | | | | Age ≤ 27.500: 1 {0=0, 1=18}
| | | | | | JoiningYear ≤ 2016.500
| | | | | | | PaymentTier > 1.500
| | | | | | | | Education = Bachelors
| | | | | | | | | Gender = Female: 1 {0=0, 1=289}
| | | | | | | | | Gender = Male
| | | | | | | | | | ExperienceInCurrentDomain > 3.500: 0 {0=18, 1=1}
| | | | | | | | | | ExperienceInCurrentDomain ≤ 3.500: 1 {0=0, 1=27}
| | | | | | | | Education = Masters
| | | | | | | | | JoiningYear > 2013.500
| | | | | | | | | | EverBenchd = No
| | | | | | | | | | | City = Bangalore: 1 {0=0, 1=2}
| | | | | | | | | | | City = New Delhi: 0 {0=2, 1=1}
| | | | | | | | | | | City = Pune: 0 {0=27, 1=8}
| | | | | | | | | | | EverBenchd = Yes: 1 {0=0, 1=2}
| | | | | | | | | | | JoiningYear ≤ 2013.500: 1 {0=1, 1=26}
| | | | | | | | | Education = PHD
| | | | | | | | | | Gender = Female: 0 {0=13, 1=0}
| | | | | | | | | | Gender = Male: 1 {0=0, 1=5}
| | | | | | | PaymentTier ≤ 1.500
| | | | | | | | Age > 22.500
| | | | | | | | | Age > 40.500: 0 {0=2, 1=0}
| | | | | | | | | Age ≤ 40.500
| | | | | | | | | | City = Bangalore
| | | | | | | | | | | Education = Bachelors: 0 {0=64, 1=15}
| | | | | | | | | | | Education = Masters: 1 {0=0, 1=3}
| | | | | | | | | | | Education = PHD: 0 {0=4, 1=0}
| | | | | | | | | | | City = New Delhi: 0 {0=19, 1=13}
| | | | | | | | | | | City = Pune
| | | | | | | | | | | | Gender = Female: 1 {0=8, 1=42}
| | | | | | | | | | | | Gender = Male: 0 {0=13, 1=2}
| | | | | | | | Age ≤ 22.500: 1 {0=0, 1=4}

```

Gambar 4. 17 Deskripsi Pohon Keputusan

Mencari *entropy* dan *gain* untuk menentukan *node root* awal, diketahui dari proses diatas bahwa dengan menggunakan tools rapidminer node root awal adalah age. Disini kita akan melakukan pengukuran nilai *entropy* dan *gain* dengan cara menghitung manual. Proses ini adalah untuk pembuktian apakah tools rapidminer berjan dengan baik.

Menghitung nilai *entropy* total pada dataset

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
Total		4653	1600	3053	0,928469	

Rumus :

$$\text{Entropy (S)} \sum_{i=1}^n -p_i \log_2 p_i$$

$$= - \left(\left(\frac{1600}{4653} \right) \times \log_2 \left(\frac{1600}{4653} \right) \right) + \left(\left(\frac{3053}{4653} \right) \times \log_2 \left(\frac{3053}{4653} \right) \right) = 0,928469$$

Setelah diketahui *entropy* total pada dataset, maka kita selanjutnya akan menghitung *entropy* dan *gain* pada setiap atribut.

Menghitung *entropy* dan *gain* pada atribut *education*

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
Total		4653	1600	3053	0,928469	
Education						0,015280318
	Bachelors	3601	1129	2472	0,8972	
	Masters	873	426	447	0,447	
	PHD	179	45	134	0,813484	

E (semua (Bachelors))

$$= - \left(\left(\frac{1129}{3601} \right) \times \log_2 \left(\frac{1129}{3601} \right) \right) + \left(\left(\frac{2472}{3601} \right) \times \log_2 \left(\frac{2472}{3601} \right) \right) = 0,8972$$

E (semua (Masters))

$$= - \left(\left(\frac{426}{873} \right) x \log_2 \left(\frac{426}{873} \right) \right) + \left(\left(\frac{447}{873} \right) x \log_2 \left(\frac{447}{873} \right) \right) = 0,447$$

E (semua (PHD))

$$= - \left(\left(\frac{45}{179} \right) x \log_2 \left(\frac{45}{179} \right) \right) + \left(\left(\frac{134}{179} \right) x \log_2 \left(\frac{134}{179} \right) \right) = 0,813484$$

G (semua (Education))

$$= - \left(\left(\frac{3601}{4653} \right) x 0,8972 \right) + \left(\left(\frac{873}{4653} \right) x 0,447 \right) + \left(\left(\frac{179}{4653} \right) x 0,813484 \right) = 0,015280318$$

Sinfo (semua (Gender))

$$= \left(- \left(\frac{3601}{4653} \right) x \log_2 \left(\frac{3601}{4653} \right) - \left(\frac{873}{4653} \right) x \log_2 \left(\frac{873}{4653} \right) - \left(\frac{179}{4653} \right) x \log_2 \left(\frac{179}{4653} \right) \right) = 0,919914$$

Gratio (semua (Gender))

$$= \frac{0,015280318}{0,919914} = 0,016611$$

Lakukan perhitungan diatas ke seluruh atribut, maka setelah dilakukan perhitungan diatas didapatkan hasil entropy dan gain setiap atribut seperti dibawah ini.

Algoritma DTC4.5								
		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain	Split Info	Gain Ratio
Total		4653	1600	3053	0,928469			
Education						0,0152	0,9199	0,0166
	Bachelors	3601	1129	2472	0,8972			
	Masters	873	426	447	0,999583			

	PHD	179	45	134	0,813484			
Joining Year						0,13420	2,7298	0,0659
	2012	504	109	395	0,753298			
	2013	669	224	445	0,919783			
	2014	699	173	526	0,807286			
	2015	781	320	461	0,976359			
	2016	525	118	407	0,768768			
	2017	1108	299	809	0,841257			
	2018	367	362	5	0,103958			
City						0,01996	1,5190	0,0131
	Bangalore	2228	595	1633	0,837208			
	Pune	1268	477	791	0,955302			
	New Delhi	1157	528	629	0,994496			
Payment Tier						0,05040	0,9951	0,0506
	1	243	89	154	0,947753			
	2	918	551	367	0,970823			
	3	3492	961	2531	0,848828			
Age						0,00067	0,9329	0,0007
	<=29	3030	1074	1956	0,937984			
	>29	1623	526	1097	0,908775			
Gender						0,03490	0,9726	0,0358
	Male	2778	716	2062	0,823316			
	Female	1875	884	991	0,99765			
Ever Benched						0,0042	0,4775	0,0089
	Yes	478	217	261	0,993879			
	No	4175	1383	2792	0,916206			
Experience InCurrent Domain						0,0021	2,5492	0,0008
	0	355	124	231	0,933438			
	1	558	188	370	0,921838			
	2	1087	399	688	0,948392			
	3	786	299	487	0,958329			
	4	931	297	634	0,903304			
	5	919	288	631	0,897041			
	6	8	2	6	0,811278			
	7	9	3	6	0,918296			
Nilai Gain Maksimal								0,0659

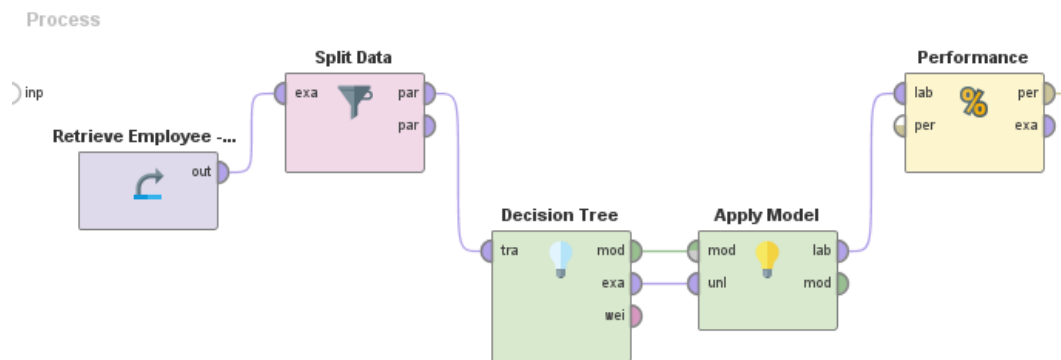
Maka

didapatkan hasil *gain ratio* terbesar yaitu 0,065946 dimana *attribute* tersebut adalah *joining year*,

atribut ini akan menjadi *node root* awal sebagai pohon keputusan, selanjutnya untuk mencari cabang node yang lain lakukan Kembali pencarian *entropy* dan *gain* pada seluruh *attribute*.

4.1.4 Algoritma *Decision Tree* C4.5 Dengan Split Validation

Penerapan algoritma *Decision Tree* pada rapidminer dengan menggunakan split validation dengan nilai akurasi, *precision*, *confusion matrix* atau nilai *recall* dan nilai AUC dengan pembagian data *training* dan data *testing* sebesar 80:20, alur pada rapidminer dapat dilihat pada gambar berikut.



Gambar 4. 18 Proses *Decision Tree* C4.5 dan *Split Validation*

Atribut yang digunakan pada algoritma *Decision Tree* C4.5 ini melibatkan semua atribut yang berjumlah 8, Adapun atribut dan bobot dapat dilihat dari tabel berikut.

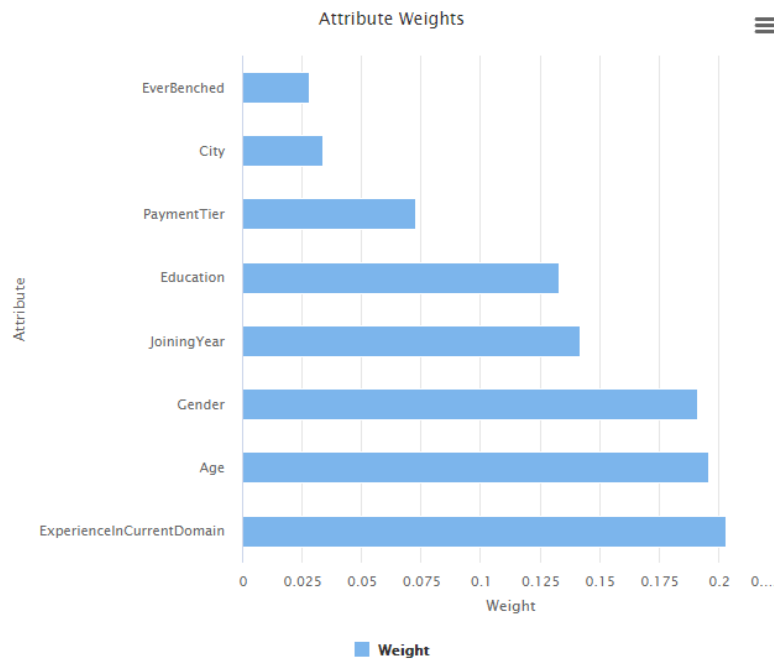
Tabel 4. 2 Tabel Atribut Yang Digunakan Sebelum PSO

NO	Attribute	Weight
1	Education	0.133
2	Joining Year	0.142
3	City	0.034
4	Payment Tier	0.073

5	Age	0.196
6	Gender	0.191
7	Ever Benched	0.028
8	Experience In Current Domain	0.203

Dari keterangan tabel diatas bahwa setiap atribut mempunyai pembobotan masing masing untuk digunakan sebagai model klasifikasi prediksi loyalitas karyawan dengan menggunakan algoritma *Decision Tree C4.5*.

Dari pembobotan atribut diatas maka dapat dilihat juga visualisasi grafik perbandingan antar atribut seperti gambar dibawah ini.



Gambar 4. 19 Gambar Pembobotan Atribut Dengan Algoritma Decision Tree C4.5

Hasil Akurasi *Decision Tree* C4.5 dan *Split Validation*, pada pembagian data sampling 80% dan data testing 20% maka didapatkan akurasi sebesar 87,37%

accuracy: 87.37%

	true 0	true 1	class precision
pred. 0	2357	385	85.96%
pred. 1	85	895	91.33%
class recall	96.52%	69.92%	

Gambar 4. 20 Nilai Akurasi Dengan Split Validation

Menghitung akurasi :

Rumus :
$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{2357 + 895}{2357 + 895 + 385 + 85} = \frac{3252}{3722} = 87,37\%$$

Berikut adalah hasil *precision* yaitu mendapatkan 91,33% seperti gambar dibawah ini.

precision: 91.33% (positive class: 1)

	true 0	true 1	class precision
pred. 0	2357	385	85.96%
pred. 1	85	895	91.33%
class recall	96.52%	69.92%	

Gambar 4. 21 Nilai Precision Dengan Split Validation

Menghitung *Precision*, Rumus :
$$\text{Precision} = \frac{TP}{TP+FP}$$

$$P(1) = \frac{2357}{2357+385} = \frac{2357}{2742} = 85.95\%$$

$$P(0) = \frac{895}{895+85} = \frac{895}{980} = 91.33\%$$

Berikut adalah hasil *confusion matrix* yaitu mendapatkan 62,92% seperti gambar dibawah ini.

recall: 69.92% (positive class: 1)

	true 0	true 1	class precision
pred. 0	2357	385	85.96%
pred. 1	85	895	91.33%
class recall	96.52%	69.92%	

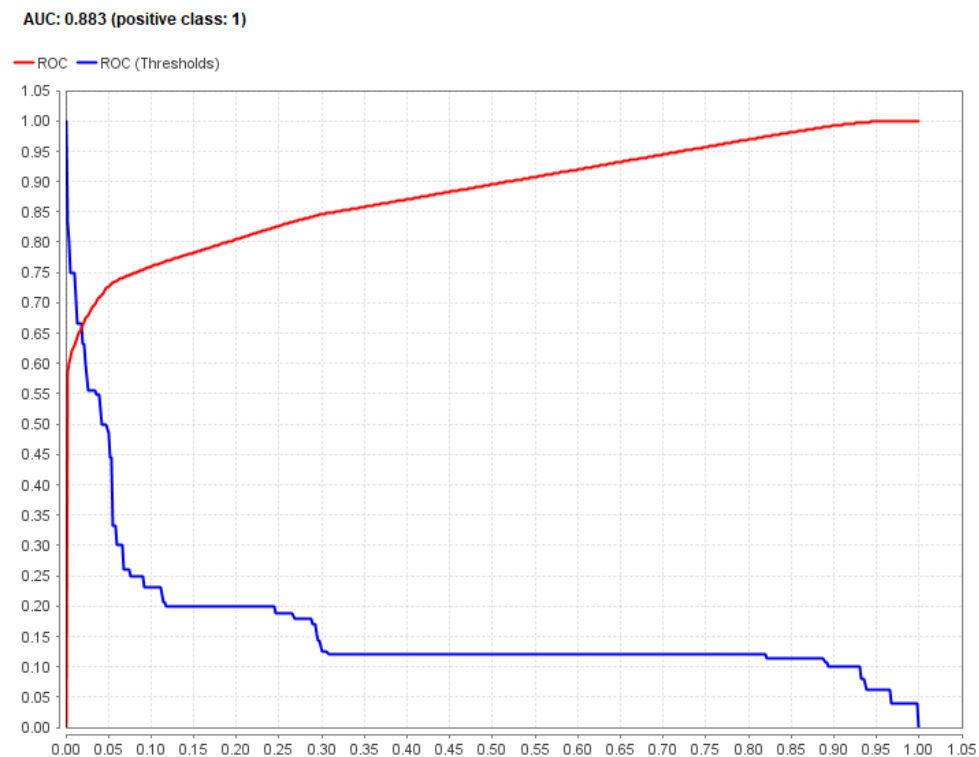
Gambar 4. 22 Nilai Confusion Matrix Dengan Split Validation

Menghitung *Recall*, Rumus : $Recall = \frac{TP}{TP+FN}$

$$R(1) = \frac{2357}{2357+85} = \frac{2357}{2442} = 96,51\%$$

$$R(0) = \frac{895}{895+385} = \frac{895}{1280} = 69,92\%$$

Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,883 seperti gambar dibawah ini.



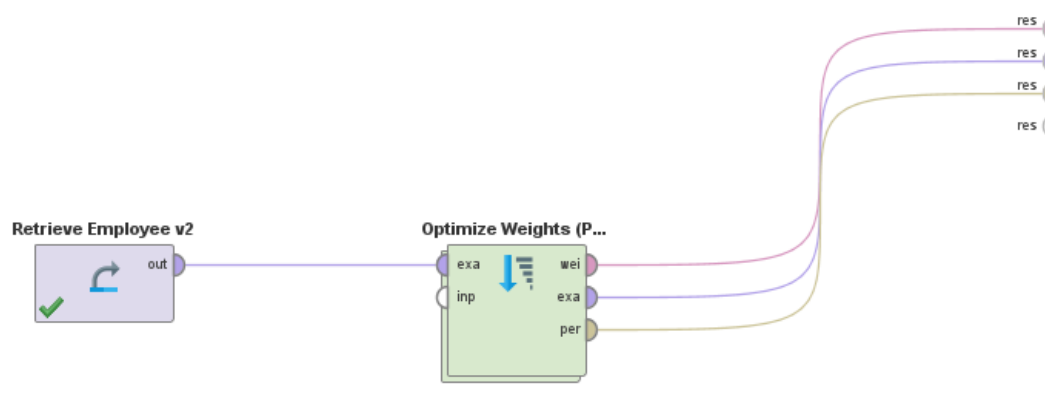
Gambar 4. 23 Kurva AUC Dengan Split Validation

Tabel 4. 3 Hasil dari *Decision Tree* dan *Split Validation* (80:20)

No	DTC4.5 (80:20)	Hasil Nilai
1	<i>Accuracy</i>	87.37%
2	<i>Precision</i>	91.33%
3	<i>Recall / Confusion Matrix</i>	69.92%
4	AUC	0.883

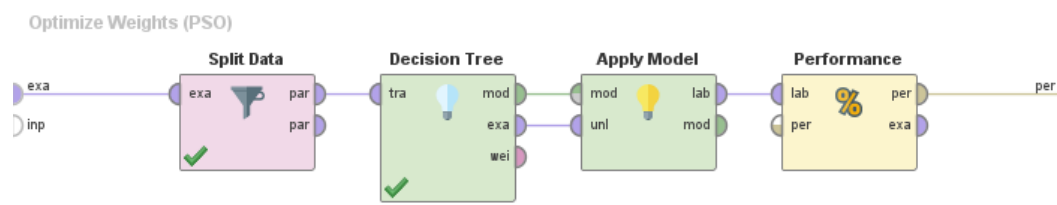
4.1.5 *Particle Swarm Optimization* (PSO) dan *Decision Tree* C4.5 Dengan *Split Data* (80:20)

Metode ini adalah untuk meningkatkan akurasi dalam klasifikasi dengan menggunakan Teknik seleksi fitur pembobotan atribut dengan menggunakan PSO dengan *Decision Tree* C4.5. penerapan metode ini dapat dilihat pada gambar dibawah ini.



Gambar 4. 24 Proses PSO dan *Decision Tree* C4.5

Setelah seperti gambar diatas kita masukkan algoritma *Decision Tree* C4.5 kedalam PSO dan tambahkan split data 80:20 seperti gambar dibawah ini.



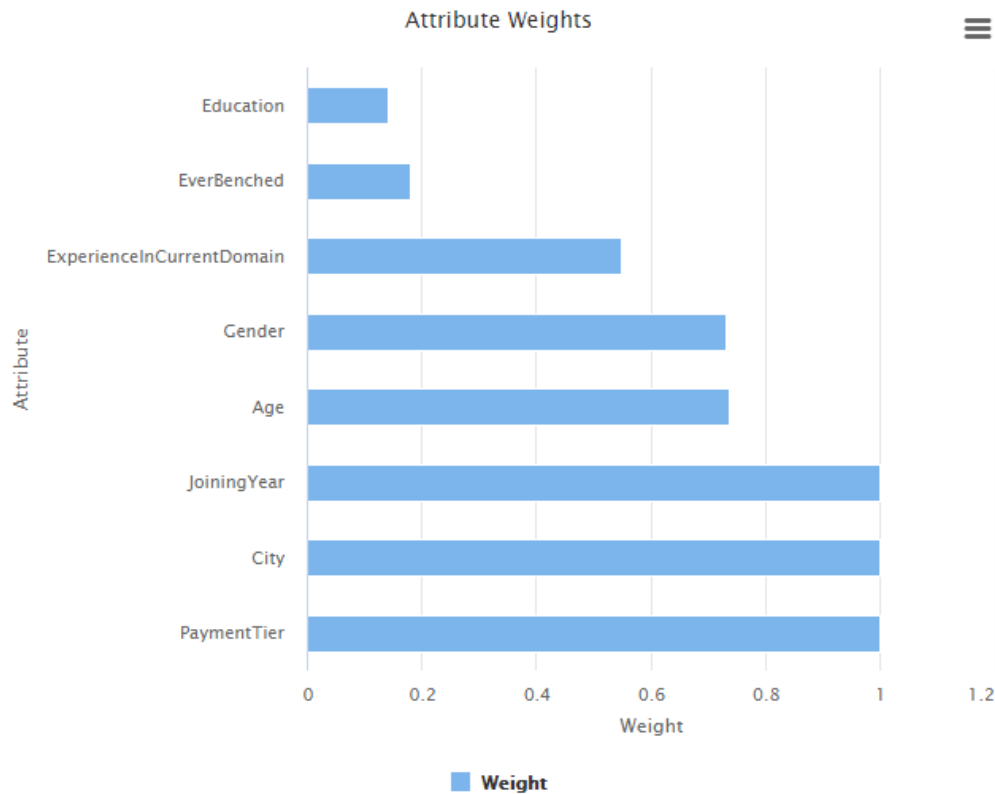
Gambar 4. 25 Sub Proses PSO dan *Decision Tree* C4.5

Atribut yang digunakan pada algoritma *Decision Tree* C4.5 ini melibatkan semua atribut yang berjumlah 8, Adapun atribut dan bobot dapat dilihat dari tabel berikut.

Tabel 4. 4 Tabel Atribut Yang Digunakan Sesudah PSO

NO	Attribute	Weight
1	Education	0.141
2	Joining Year	1
3	City	1
4	Payment Tier	1
5	Age	0.732
6	Gender	0.732
7	Ever Benched	0.181
8	Experience In Current Domain	0.548

Dari keterangan tabel diatas bahwa terdapat atribut yang memiliki nilai 0 yaitu gender yang artinya pada proses PSO atribut tersebut tidak berkontribusi dalam proses klasifikasi prediksi loyalitas karyawan dengan menggunakan algoritma *Decision Tree* C4.5.



Gambar 4. 26 Gambar Pembobotan Atribut Dengan PSO

Dari pengaplikasian PSO dan *Decision Tree* C4.5 pada rapidminer maka dihasilkan peningkatan akurasi, dimana dapat dilihat perbandingan akurasi, precision dan recall serta AUC sesudah dan sebelum di tambahkan metode PSO.

Berikut adalah hasil *accuracy* yaitu mendapatkan 88,80% seperti gambar dibawah ini.

accuracy: 88.80%

	true 0	true 1	class precision
pred. 0	2379	354	87.05%
pred. 1	63	926	93.63%
class recall	97.42%	72.34%	

Gambar 4. 27 Nilai Akurasi PSO dan *Decision Tree* C4.5

Berikut adalah hasil *precision* yaitu mendapatkan 93,63% seperti gambar dibawah ini.

precision: 93.63% (positive class: 1)

	true 0	true 1	class precision
pred. 0	2379	354	87.05%
pred. 1	63	926	93.63%
class recall	97.42%	72.34%	

Gambar 4. 28 Nilai *Precision* PSO dan *Decision Tree* C4.5

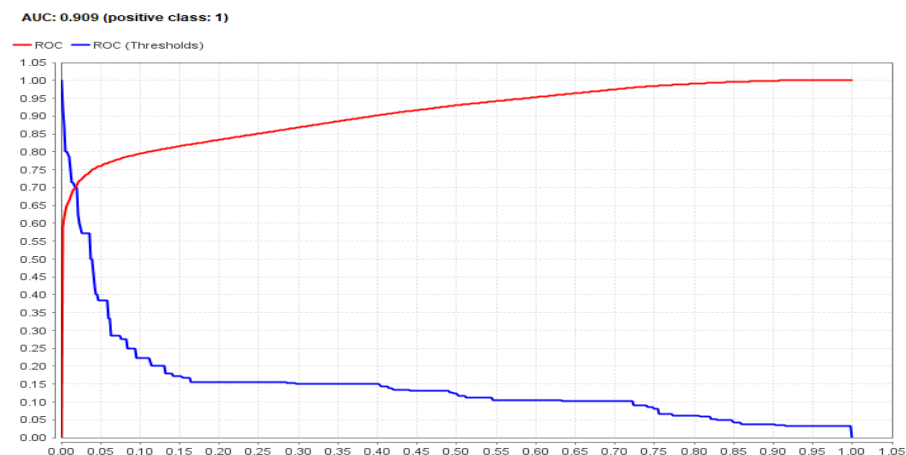
Berikut adalah hasil *confusion matrix* yaitu mendapatkan 72,34% seperti gambar dibawah ini.

recall: 72.34% (positive class: 1)

	true 0	true 1	class precision
pred. 0	2379	354	87.05%
pred. 1	63	926	93.63%
class recall	97.42%	72.34%	

Gambar 4. 29 Nilai *Confusion Matrix* PSO dan *Decision Tree* C4.5

Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,909 seperti gambar dibawah ini.



Gambar 4. 30 Kurva AUC PSO dan *Decision Tree* C4.5

Tabel 4. 5 Hasil dari PSO dan *Decision Tree* (80:20)

No	PSO + DTC4.5 (80:20)	Hasil Nilai
----	----------------------	-------------

1	<i>Accuracy</i>	88.80%
2	<i>Precision</i>	93.63%
3	<i>Recall / Confusion Matrix</i>	72.34%
4	AUC	0.909

4.2 Pembahasan Hasil Penelitian

Pembahasan hasil penelitian yaitu membahas hasil penelitian yang telah dilakukan. Penelitian yang dilakukan yang pertama yaitu mengklasifikasi untuk prediksi loyalitas karyawan dengan menggunakan algoritma *decision tree* C4.5 dan penelitian ke dua yaitu untuk menabahi nilai akurasi sehingga ditambahkan metode seleksi fitur PSO dan Algoritma *Decision tree* C4.5 serta dataset yang digunakan diterapkan split data dimana perbandingan data training dan data testing sebesar 80:20.

Untuk melihat perbandingan akurasi sebelum dan sesudah ditambahkan PSO dapat kita lihat pada tabel dibawah ini.

1. Peningkatan Akurasi

Tabel 4. 6 Perbandingan DTC4.5 dan PSO+DCT5.4

No	Keterangan	DTC4.5 (80:20)	PSO + DTC4.5 (80:20)
1	<i>Accuracy</i>	87.37%	88.80%
2	<i>Precision</i>	91.33%	93.63%
3	<i>Recall / Confusion Matrix</i>	69.92%	72.34%
4	AUC	0.883	0.909

Pada tabel diatas dapat dilihat bahwa dari percobaan menggunakan data *training* dan *testing* dengan perbandingan 80:20 memakai algoritma DTC4.5 dan PSO + DTC4.5 menghasilkan akurasi yang berbeda, sudah jelas bahwa nilai akurasi setelah ditambahkan metode PSO kedalam algoritma DTC4,5 memiliki peningkatan..

2. Prediksi Atribut terpenting dalam loyalitas karyawan

Dapat dilihat pada gambar dibawah ini setelah dilakukan PSO dan DTC4.5 maka dapat diketahui atribut yang paling berpengaruh ada 3, yaitu *joining year*, *city* dan *payment tier*. Dapat dilihat dari tabel pembobotan dibawah ini.

Tabel 4. 7 Prediksi Atribut terpenting dalam loyalitas karyawan

NO	Attribute	Weight
1	Education	0.141
2	Joining Year	1
3	City	1
4	Payment Tier	1
5	Age	0.732
6	Gender	0.732
7	Ever Benched	0.181
8	Experience In Current Domain	0.548

