

Materi Pelatihan Data Mining

(dalam materi ini bisa di linkkan dengan mongodb dan juga interface sederhana)

1. Upload Data

Pada halaman *data awal* admin dapat mengupload data alumni dan dapat melihat daftar data alumni yang telah tersedia. Implementasi *user interface* halaman *data awal* dapat dilihat pada Gambar 4.15

Choose file

Upload

No.	No Student Registration	Name	The Place, Date of Birth	Address	Year Collese	Graduation Year	
1.	9018263	Puguh Drajat E.P	Punworejo, 4 November 1989	Tangerang Rt. 01/ 0	2009	2016	
2.	12022030	Arsyad Cahya Subrata	Bantul, 21 juni 1994	Karangpule Rt.01 , Tirtanirmolo , Kasihan Bantul	2012	2016	
3.	12020021	Yesi Novela	Tempilang, 1 Januari 1995		2012	2016	M
4.	9018269	Pandu Henwijaya	Lampung, 19 Desember 1990	Tamanlirto, Kasihan ,Bantul	2009	2016	
5.	9018109	Anjar Prasetyo	Gunungkidul, 27 April 1991	Umbulharjo , Yogyakarta	2009	2016	
6.	9020026	Moh. Dadi Kurniawan	Savanajaya, 27 Maret 1992	Saren Rt. 02 Rw. 01 Caturtunggal	2009	2016	
7.	12018060	Martin Sarlina	Lepak, 23 Maret 1994	NTB	2012	2016	
8.	11018022	Yan Adhi Permadi	Jepara, 22 juni 1993	Umbulharjo V	2011	2016	V
9.	9018212	Achicha Wahyuni	Yogyakarta, 3 juni 1991	Noyokerten, Sendangirto, Berbah	2009	2016	G
10.	9018301	Yayan Tisna Permadi	Purbalingga, 08 Oktober 1991	Tajug, Rt.04/ 05 Purbalingga	2009	2016	

Showing 1 to 10 of 108 entries

Previous

1

2

3

4

5

...

11

Next

© 2016 Foldgrowth Team

Gambar 1. Implementasi hasil load data awal

2. Pembersihan Data

Data yang dipakai dari data alumni tahun 2015-2016 yang berjumlah 102 dilakukan proses yang pertama yaitu pembersihan dan diperoleh 98 data setelah dibersihkan. Pembersihan data ini dilakukan dengan menghapus data yang tidak kosong pada atribut Umur, IPK, Toefl, lama studi, gaji pertama bekerja dan lama masa tunggu mencari kerja.

Data Item	Cleaning	Transformation	Selection	Fold-Growth
-----------	----------	----------------	-----------	-------------

Cleaning Process Alumni Data							
No.	No Student Registration	Name	The Place, Date of Birth	Address	Year Collese	Graduation Year	En
1.	9018263	Puguh Drajat E.P	Purworejo, 4 November 1969	Tangerang Rt. 01/ 0	2009	2016	
2.	9018269	Pandu Henwijaya	Lampung, 19 Desember 1990	Tamanlirto, Kasihan ,Bantul	2009	2016	
3.	9020026	Moh. Dadi Kurniawan	Savanajaya, 27 Maret 1992	Saren Rt. 02 Rw. 01 Caturtunggal	2009	2016	
4.	12018060	Martin Sarlina	Lepak, 23 Maret 1994	NTB	2012	2016	
5.	11018022	Yan Adhi Permadi	Jepara, 22 juni 1993	Umbulharjo V	2011	2016	Wi
6.	9018212	Achicha Wahyuni	Yogyakarta, 3 juni 1991	Noyokerten, Sendangirto, Berbah	2009	2016	Gt
7.	9018301	Yayan Tisna Permadi	Purbalingga, 08 Oktober 1991	Tajug, Rt.04/ 05 Purbalingga	2009	2016	
8.	12022020	Muh. Rama dhai	Pekanbaru, 29 Maret 1992	Karongan Rt/Rw. 04/11 Jogotirto, Berbah Sleman	2012	2016	
9.	12022038	Nurwahyuddi	Dili, timor-timur/ 13 April 1995	Juru Gentong J V/407 C Banguntapan	2012	2016	Mt
10.	12019039	Agus Purnomo	Klaten, 2 Agustus 1992	Dukuh Glagahwangi, Polanharjo	2012	2016	

Showing 1 to 10 of 105 entries

Previous 1 2 3 4 5 ... 11 Next

3. Integrasi Data

Pada penelitian ini, data tidak memerlukan adanya integrasi data atau penggabungan data dari beberapa database. Karena data alumni yang ada pada universitas ahmad dahlan ini berupa data excel.

4. Seleksi Data

Pemilihan data yang sesuai dengan algoritma yang akan digunakan, pemilihan data seleksi merupakan tahap awal dalam pembentukan proses data *mining* pemilihan data seleksi dalam penelitian. Pemilihan atribut dari 13 atribut menjadi 6 atribut. Hanya atribut yang dipakai yang digunakan

Data Item	Cleaning	Transformation	Selection	Fold-Growth
-----------	----------	----------------	-----------	-------------

Selection Process Alumni Data							
No.	No Student Registration	Age	Study Duration	Time to get a Job	First Salary	English Skill	Grade Point Average
1.	9018263	27	7	2.4	1475635	390	2,13
		C3	B3	A1	G2	D3	F3
2.	9018269	26	7	2.4	1475635	413	2,75
		C3	B3	A1	G2	D3	F3
3.	9020026	24	7	2.4	1475635	400	3,24
		C3	B3	A1	G2	D3	F2
4.	12018060	22	4	2.4	1475635	396	3,27
		C2	B1	A1	G2	D3	F2
5.	11018022	23	5	2.4	1475635	460	3,20
		C2	B2	A1	G2	D2	F2

Showing 1 to 10 of 210 entries

Previous 1 2 3 4 5 ... 21 Next

© 2016 Foldgrowth Team

5. Transformasi Data

Proses transformasi data yaitu proses pengkategorian atau perubahan data ke format yang sesuai untuk proses dalam data mining. Ada 3 refrensi yang diambil berdasarkan jurnal Borang Sarjana Akreditasi Program Studi Sarjana pada tahun 2012 yaitu IPK, Toefl, dan lama masa tunggu mencari kerja. hasil proses transformasi data atribut adalah sebagai berikut :

- Nilai IPK dikategorikan menjadi 3 yaitu IPK $>3.50-4.00$ = Maksimum, jika IPK $\geq 2.75-3.50$ = Rata-rata dan jika IPK <2.75 = Minimum. Hasil transformasi akan ditunjukkan pada Tabel 4.1.

Tabel 4.1. Proses Transformasi IPK

Kategori	Keterangan
IPK Minimum (Min)	Untuk IPK <2.75
IPK Rata-rata (Rat)	Untuk IPK $\geq 2.75 - 3.50$
IPK Maksimum (Max)	IPK $>3.50 - 4.00$

- b. Nilai Toefl dikategorikan menjadi 3 yaitu T1= 400-449, T2 = 450-499, dan T3 = ≥ 500 . Hasil transformasi akan ditunjukkan pada Tabel 4.2.

Tabel 4.2 Proses Tranformasi nilai Toefl

Kategori	Keterangan
T1	400 – 449
T2	450 –499
T3	≥ 500

- c. Waktu mendapat pekerjaandikategorikan menjadi 4 yaitu w1 = 0–2.9bulan, w2 = 3 – 5.9 bulan, w3 = 6 – 8.9 buln, w4 = 9 – 12 bulan. Hasil transformasi akan ditunjukkan pada Tabel 4.3.

Table 4.3 Proses tranformasi waktu mendapat kerja

Kategori	Keterangan
W1	0 -2.9 bulan
W2	3 – 5.9 bulan
W3	6 - 8.9 bulan
W4	9 -12 bulan

- d. Data atribut waktu lama studi akan dikatagorikan berdasarkan keputusan Menteri Pendidikan Nasional Nomor 232/U/2000 (2000) tentang Pedoman Penyusunan Kurikulum Pendidikan Tinggi dan Penilaian Hasil Belajar Mahasiswa. Data atribut dikategorikan tepat waktu apabila lama studi ≤ 4 tahun dan tidak tepat waktu apabila lama studi > 4 tahun. Data transformasi akan ditunjukkan pada Tabel 4.4.

Table 4.4 Proses tranformasi waktu lama studi

Kategori	Keterangan
Tepat Waktu (TW)	Jika lama studi ≤ 4 tahun
Tidak Tepat Waktu (TTW)	jika lama studi > 4 tahun

- e. Data transformasi umur akan dikategorikan berdasarkan penelitian yang dilakukan oleh (Purwaningseh, 2010) tentang Analisis pencari kerja menurut kelompok umur. Data atribut umur 15-24 tahun umumnya digolongkan produktif dalam mencari kerja dan pada umur 25-54 tahun digolongkan tidak produktif. Data transformasi umur ditunjukkan pada Tabel 4.5.

Tabel 4.5 Proses tranformasi umur

Kategori	Keterangan
Usia produktif (UP)	Untuk umur 15-24 tahun
Tidak produktif (TP)	Untuk umur 25-54 tahun

- f. Gaji dikatakan diatas UMR apabila nominalnya $\geq 1.572.200$ dan gaji dikatakan dibawah UMR apabila nominalnya $< 1.572.200$ menurut keputusan Gubernur Daerah Istimewa Yogyakarta no 235/ kep/ 2016 kota yogyakarta tentang penetapan upah minimum kabupaten/kota tahun 2017 ditunjukkan pada tabel 4.6.

Table 4.6 Proses tranformasi nominal gaji

Kategori	Keterangan
Atas UMR (A.UMR)	Gaji >=1.572.200
Bawah UMR (B.UMR)	Gaji <1.572.200

Data Item	Cleaning	Transformation	Selection	Fold-Growth
-----------	----------	----------------	-----------	-------------

Transformation Process Alumni Data									
Age	Study Duration		Time to get a Job		First Salary		English Skill		Grade Point Average
17 - 20	C1	3 - 4	B1	0 - 3	A1	< 1 Juta	G1	> 526	D1 > 3,50
21 - 23	C2	5 - 6	B2	3 - 6	A2	1 - 5 Juta	G2	420 - 525	D2 2,75 - 3,50
> 23	C3	7 - 8	B3	6 - 9	A3	> 5 Juta	G3	< 420	D3 < 2,75
				9 - 12	A4				
				> 12	A5				

No.	No Student Registration	Name	Age	Address	Study Duration	Employment	Time to get a Job	First Sa
1.	9018263	Puguh Drajat E.P	27	Tangerang Rt. 01/0	7		2.4	147563t
2.	9018269	Pandu Herwijaya	26	Tamantirto, Kasihan ,Bantul	7		2.4	147563t
3.	9020026	Moh. Dadi Kurniawan	24	Saren Rt. 02 Rw. 01 Caturtunggal	7		2.4	147563t
4.	12018060	Martin Sarlina	22	NTB	4		2.4	147563t
5.	11018022	Yan Adhi Permadi	23	Umbulharjo V	5	Wiraswasta	2.4	147563t

6. Proses Asosiasi dengan metode tertentu dan akurasinya

Metode Asosiasi : FP-Tree, Fold-Growth, Apriori, FOLDRAM , dll

Akurasi : Lift ratio dll

Data Item	Cleaning	Transformation	Selection	Fold-Growth
-----------	----------	----------------	-----------	-------------

1	2	3	4
Step 1	Step 2	Step 3	Step 4
Step 1 L1 And L2	Step 2 Frequent Item Set	Step 3 FP Tree	Step 4 Support & Confidence

Trimming Items that are not Frequent		
Frequent Item-set		
No.	Item Name	Total
1.	A1,B3,C3,F3,G2,D3	3
2.	A1,B3,C3,F2,G2,D3	2
3.	A1,B1,C2,F2,G2,D3	10
4.	A1,B2,C2,F2,G2,D2	5
5.	A1,B3,C3,F2,G1,D3	1
6.	A1,B3,C3,F2,G2,D2	4
7.	A1,B1,C3,F2,G2,D2	4
8.	A1,B1,C3,F2,G2,D3	4

Ordered Frequent Item-set		
No.	Item Name	Total
1.	A1,B1,C2,F2,G2,D2	12
2.	A1,B1,C2,F2,G2,D3	10
3.	A1,B2,C3,F2,G2,D3	7
4.	A1,B2,C3,F2,G2,D2	5
5.	A1,B2,C2,F2,G2,D2	5
6.	A1,B1,C3,F2,G2,D3	4
7.	A1,B1,C3,F2,G2,D2	4
8.	A1,B3,C3,F2,G2,D2	4

7. Metode Clustering dan akurasi

Metode clustering: K-Means, K-Medoid, AHC, SOM, Fuzzy C-Means dll

Akurasi : Confution matrik, purity tes, dll

Contoh data bisa menggunakan data alumni dengan 6 variabel : umur, IPK, lama masa studi, lama masa tunggu mencari kerja, gaji pertama bekerja dan TOEFL.

8. Metode Klasifikasi dan Akurasi

Metode Klasifikasi : Tree, C4.5, C4.8, SVM, JST, dll

Akurasi : Confution matrix dll

Contoh data bisa menggunakan data alumni dengan 6 variabel : umur, IPK, lama masa studi, lama masa tunggu mencari kerja, gaji pertama bekerja dan TOEFL.

Dengan kelas lama masa tunnggu mencari kerja : cepat, sedang dan lama

9. Metode Peramalan dan akurasi

Metode Peramalan : SVR, regresi dll

Akurasi : MAD, MSE dll