

Knowledge Discovery Data Akademik Untuk Prediksi Pengunduran Diri Calon Mahasiswa

I Gst Ayu Sri Melati¹, Linawati², I A Dwi Giriantari³

Abstract— New Students Registration at an educational institution such as STMIK STIKOM Bali is a routinely implemented every new academic year. Whereas the registration of new student candidates is always increasing from year to year, but not all prospective students continued registration to further step after they passed the selection. It would be too late to take corrective action if very few new students enrolled. By not knowing how to estimate the number of registratrations students, the institution cannot measure the time and the number of new admissions target to be achieve. In this case the use of data mining technique is proposed to provide knowledge that was previously hidden in the data warehouse. Representation of 2017 prospective student data has the lowest accuracy of 78% and student representation 2017, 2016 has 78.46% accuracy and student representation 2017,2016,2015 has 78.20% accuracy. The result of the frequent itemsets mining process shows that the emergence of advanced and backward labels is also frequent with the appearance of the emergence trend between criteria and label (advanced or backward). Task mining then is used to predict the prospective student is by classification techniques and frequent pattern. The software used for implementation is WEKA.

Keywords: *Data Mining, Classification, Decision Tree, Frequent Pattern, Knowledge Discovery.*

Intisari— Pendaftaran Mahasiswa Baru pada suatu institusi pendidikan seperti STMIK STIKOM Bali adalah sebuah kegiatan yang rutin dilaksanakan setiap tahun ajaran baru. Pendaftaran calon mahasiswa baru tersebut selalu meningkat dari tahun ketahun namun calon mahasiswa yang melakukan pendaftaran tidak semua yang melanjutkan langkah registrasi dari sejumlah calon mahasiswa yang sudah dinyatakan lulus. Hal ini akan menjadi sangat terlambat untuk mengambil tindakan apabila ternyata mahasiswa baru yang registrasi sangat sedikit jumlahnya. Dengan tidak diketahuinya mahasiswa yang registrasi, maka pihak-pihak perguruan tinggi tidak dapat mengetahui dengan pasti kapan jumlah target penerimaan mahasiswa baru tercapai. Dalam permasalahan ini penggunaan teknik data mining diharapkan dapat memberikan pengetahuan – pengetahuan yang sebelumnya tersembunyi di dalam gudang data. Representasi data calon mahasiswa 2017 memiliki akurasi paling rendah yaitu 78% dan representasi calon mahasiswa 2017, 2016 memiliki akurasi 78,46% dan representasi calon mahasiswa 2017, 2016, 2015 memiliki akurasi 78,20%. Hasil proses *frequent itemsets mining* menunjukkan bahwa kemunculan label lanjut dan mundur juga *frequent* dengan kemunculan *trend* kemunculan antara kriteria dengan label (lanjut atau mundur). *Task mining* yang digunakan untuk memprediksi calon mahasiswa yaitu dengan teknik klasifikasi dan teknik *Frequent*

Pattern. Perangkat lunak yang digunakan untuk implementasi adalah WEKA.

Kata Kunci—*Data Mining, Klasifikasi, Decision Tree, Frequent Pattern, Knowledge Discovery.*

I. PENDAHULUAN

Perkembangan teknologi informasi telah memberikan kontribusi pada cepatnya pertumbuhan jumlah data yang dikumpulkan dan disimpan dalam basis data berukuran besar (gunung data). Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode [1].

STMIK STIKOM BALI adalah salah satu sekolah tinggi manajemen informatika yang berdiri di Pulau Bali. Penerimaan mahasiswa baru pada suatu institusi pendidikan seperti STIKOM Bali adalah sebuah kegiatan yang selalu dilaksanakan setiap tahun ajaran baru. Pendaftaran calon mahasiswa baru tersebut selalu meningkat dari tahun ketahun sebesar 20-50% berdasarkan data akademik namun calon mahasiswa yang melakukan pendaftaran tidak semua yang melanjutkan langkah registrasi. Oleh karena itu dibuat sebuah prediksi dari kemungkinan pengunduran diri calon mahasiswa.

Dalam permasalahan ini klasifikasi prediksi dapat digunakan pada Penerimaan Mahasiswa Baru (PMB). Data berskala besar memunculkan fenomena data yang berjumlah besar tidak diikuti dengan peningkatan informasi yang didapat dari data tersebut [2]. Pada kasus ini model klasifikasi di buat untuk mengidentifikasi pola data untuk kelas status “lanjut” atau “mundur pendaftaran”, dari hasil penentuan pola training data [3]. Pola atau model dari training data tersebut selanjutnya diuji dengan menggunakan *test set* data.

Model klasifikasi dibuat dengan menganalisis training data. pengklasifikasian data yang bekerja relatif dengan cara yang lebih sederhana dibandingkan dengan metode pengklasifikasian data lainnya. Algoritma ini berusaha mengklasifikasikan data baru yang belum diketahui *class*-nya dengan memilih data sejumlah *k* yang letaknya terdekat dari data baru tersebut [4]

Model yang dihasilkan kemudian dapat digunakan untuk memprediksi kelas dari *unknown* data, test set data digunakan untuk pengujian dari model yang telah didapatkan pada *training* data

Dengan menerapkan beberapa metode yang telah dijabarkan di atas, diharapkan penulis dapat menganalisis pemanfaatan data calon mahasiswa untuk membangun model klasifikasi (*classifier*) dan pemanfaatan *frequent itemsets mining* pada data pendaftaran mahasiswa baru untuk memprediksi kemungkinan pengunduran diri calon mahasiswa. Perangkat lunak yang digunakan untuk implementasi adalah WEKA.

p-ISSN:1693 – 2951; e-ISSN: 2503-2372

¹ Mahasiswa, Magister Teknik Elektro, Program Pasca Sarjana Universitas Udayana, email : geamell87@yahoo.com

^{2,3} Staff Pengajar, Magister Teknik Elektro, Program Pasca Sarjana Universitas Udayana, Jl. PB. Sudirman Denpasar Bali, Telp. (0361-239599); email : linawati@unud.ac.id, dayu.giriantari@unud.ac.id

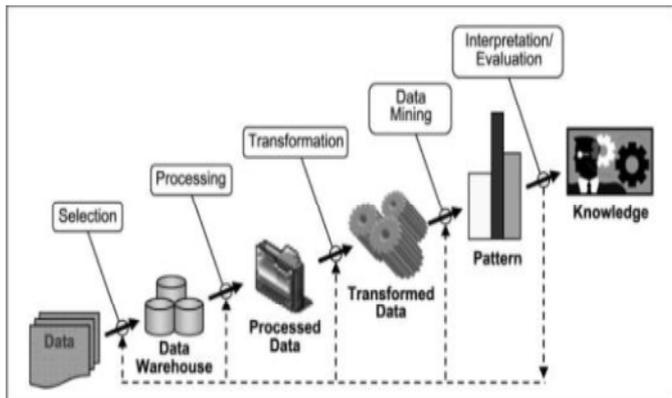


II. METODE PENELITIAN

A. Data Mining

Data mining adalah tahap analisis proses "*knowledge discovery in database*". Data mining adalah proses darimenalisis data dari perspektif yang berbeda dan meringkasnya menjadi informasi yang [5]. Data Mining merupakan salah satu teknik penting dalam mencari pengetahuan dalam sekumpulan data digital (*knowledge mining from data*) [6]. Hasil pencarian ini dapat dimanfaatkan untuk memprediksi "masa depan" dan menemukan tren berdasarkan pola dan relasi data.

Dalam penerapan data mining diperlukan berbagai perangkat lunak untuk menganalisis data guna mencari data hubungan dan pola yang dapat digunakan untuk membuat prediksi secara akurat [7].



Gambar 1. Tahapan Data Mining [6]

a. Data Selections

Memilih atau mensegmentasikan data sesuai dengan beberapa kriteria.

b. Pre-Processing dan Cleaning Data

Tahap pembersihan data dilakukan dengan menghapus informasi tertentu yang dianggap tidak perlu dan bias karena memperlambat kueri

c. Transformation

Data tidak hanya ditransfer tetapi ditransformasikan dalam lapisan dapat ditambahkan seperti lapisan demografis yang umum digunakan dalam riset pasar

d. Data Mining

Tahap ini berkaitan dengan ekstraksi pola dari data sebuah pola dapat didefinisikan sebagai seperangkat fakta (data).

e. Interpretation / Evakuasi

Pola yang didefinisikan oleh sistem ditafsirkan menjadi pengetahuan yang kemudian dapat digunakan untuk mendukung pengambilan keputusan.

B. Frequent Pattern

Frequent Patterns (FP) merupakan kumpulan pola (seperti *itemsets*, *subsequences*, atau *substructures*) yang sering muncul di dalam data set [4]. *Itemset* merupakan kumpulan *items* yang muncul bersamaan dalam sebuah *dataset*. *Subsequence* merupakan sekumpulan *itemset* yang berurutan (berdasarkan waktu atau posisi pada satu ID), jika suatu *itemset* sering berurutan dengan *itemset* yang lain maka dikatakan *frequent sequential pattern*.

C. Klasifikasi Data

Klasifikasi data (data Classification) adalah sebuah bentuk analisis data yang mengekstrak model dan menggambarkan kelas data penting. Model disebut pengklasifikasi, memprediksi kategoris (diskrit, *unordered*) label kelas [6]

Sebuah model klasifikasi atau *Classifier* dibentuk untuk memprediksi class (*categorical*) label. Jadi, dalam data yang akan dibangun model klasifikasi harus sudah ditentukan variable yang akan dijadikan *class label*. Class / label / kategori dapat diwakili oleh nilai – nilai diskrit, urutan antara nilai – nilai tidak ada artinya atau diabaikan [6].

D. Decision Tree

Decision tree adalah sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap *internal node* menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan hasil dari pengujian dan *leaf node* (simpul terminal) menyatakan kelas – kelas atau distribusi kelas [6] *Decision Tree* adalah struktur rekursif sederhana untuk mengekspresikan sebuah proses klasifikasi sekuensial di mana sebuah kasus, yang digambarkan oleh serangkaian atribut, ditugaskan ke salah satu kelas yang terpisah [8]

E. Ekstraksi Decision Tree Menjadi Aturan Klasifikasi

Ekstraksi *decision tree* menjadi aturan (*rule*) klasifikasi dilakukan setelah proses *pruning* dilakukan atau setelah *decision tree* terbentuk secara optimal. Ekstraksi *decision tree* menjadi aturan klasifikasi akan menghasilkan aturan yang berbentuk IF – THEN. Aturan IF-Then adalah ekspresi dari bentuk [4]. IF *condition* THEN *conclusion* Contoh aturan R1: R1: IF age = youth AND student = yes THEN buys computer = yes. Jika *decision tree* pada gambar (*decision tree* membeli computer) dirubah menjadi aturan IF-THEN.

F. Algoritma J48

Decision Tree J48 merupakan implementasi algoritma C4.5 (berbasis Java) pada Weka. Algoritma C4.5 digunakan untuk pemisah obyek. *Tree* atau pohon keputusan banyak dikenal sebagai bagian dari *Graph*, yang termasuk dalam irisan bidang ilmu otomata dan teori bahasa serta matematika diskrit [9]. Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan persamaan sebagai berikut:

$$Gains(S, A) = Entropy(s) - \sum^n * Entropy(st) \quad (1)$$

Keterangan:

S = himpunan kasus

A = atribut

n = jumlah partisi atribut A

|Si| = jumlah kasus pada partisi ke-i

Confusion matrix adalah tool yang dipergunakan untuk mengukur seberapa akurat model klasifikasi (*classifier*) dalam memprediksi *class label* dari tuples [4]. TP dan TN menunjukkan *classifier* melakukan prediksi secara tepat, sedangkan FP dan FN menunjukkan *classifier* melakukan salah prediksi.

TABEL I
 CONFUSION MATRIX

		Predicted class		
		yes	no	Total
Actual class	Yes	TP	FN	P
	No	FP	TN	N
Total		P ²	N ²	P + N

Dari *confusion matrix* pada TABEL I model klasifikasi dapat dihitung nilai akurasinya, *recall*, *precision* dan *F-measure*. Berikut adalah tabel formula pengukuran model klasifikasi [6] TABEL II Formula Pengukuran klasifikasi Data.

TABEL II
 FORMULA PENGUKURAN KLASIFIKASI DATA

Measure	Formula
Accuracy, recognition rate	$\frac{TP + TN}{P + N}$
Sensitivity, true positive rate, Recall	$\frac{TP}{P}$
Precision	$\frac{TP}{TP + FP}$
F, F1, F-score, harmonic mean of precision and recall	$\frac{2 * precision * recall}{Precision + recall}$

G. WEKA Data Mining Tools

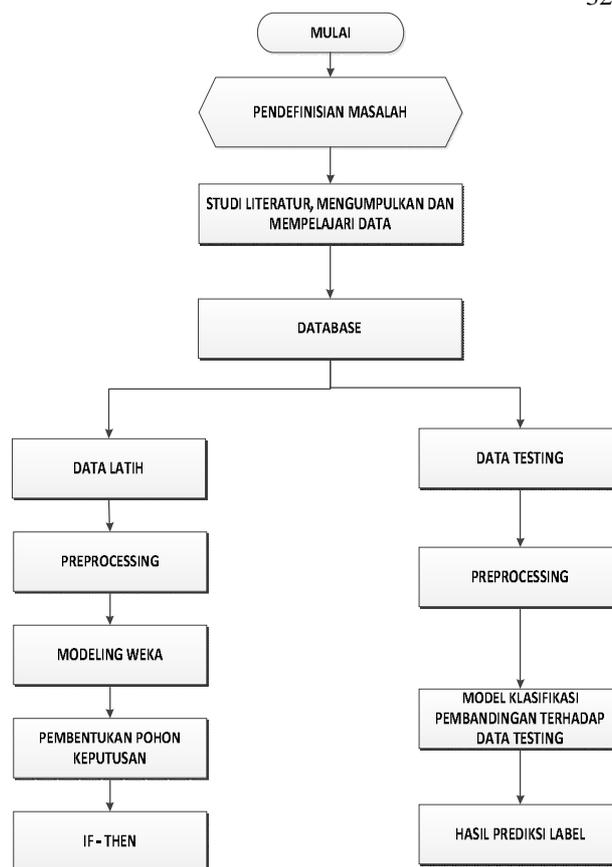
WEKA adalah sebuah paket *tools machine learning* praktis. WEKA merupakan singkatan dari *Waikato Environment for Knowledge Analysis*, yang dibuat di Universitas Waikato. *New Zealand* untuk penelitian, pendidikan dan berbagai aplikasi. Beberapa format file untuk inputnya diantaranya:

1. Format *Comma Separated values (CSV)*
 Merupakan format file teks dengan pemisah tanda koma (,) yang cukup umum digunakan. Bias membuat file ini dengan menggunakan Microsoft Excel atau membuat sendiri dengan format notepad
2. *Attribute Relation File Format (ARFF)*
 Merupakan tipe file teks yang berisi beberapa instance data yang berhubungan dengan suatu set atribut data yang dideskripsikan juga dalam file tersebut.

III. PERANCANGAN SISTEM

A. Klasifikasi

Berdasarkan permasalahan yang ada yaitu memprediksi kemungkinan pengunduran diri calon mahasiswa baru, hasil prediksinya berupa lanjut atau mengundurkan diri sebagai calon mahasiswa. Dari beberapa teknik data mining klasifikasi dan teknik *Frequent Pattern* sangat memungkinkan sekali dalam memprediksi sebuah label. Beberapa tahapan perancangan alur analisis digambarkan pada Gambar 2.



Gambar 2 Alur Analisis Klasifikasi Prediksi Pengunduran Diri Calon Mahasiswa

1. **Pendefinisian Masalah**
 Pendefinisian masalah melingkupi penjabaran masalah yang akan ditangani data akademik mengandung informasi dan pengetahuan yang bermanfaat untuk mendukung prediksi kemungkinan pengunduran diri calon mahasiswa
2. **Teknik Klasifikasi**
 Data latih yang di ambil dari data akademik PMB 3 tahun terakhir yaitu tahun 2015, 2016, 2017. Dari data latih akan dibuatkan 3 buah model klasifikasi yaitu:
 Model I : Data akademik PMB Thn 2017.
 Model II : Data akademik PMB Thn 2017, 2016
 Model III : Data akademik PMB Thn 2017, 2016, 2015.

Dalam analisis *Pembentukan Pohon Keputusan* *Pembentukan decision tree* yaitu pembentukan diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap satu atribut. Diperoleh **IF – THEN**: *decision tree* menjadi aturan klasifikasi akan menghasilkan aturan yang berbentuk **IF – THEN**.

Aturan **IF-Then** adalah ekspresi dari bentuk **IF condition THEN conclusion**.

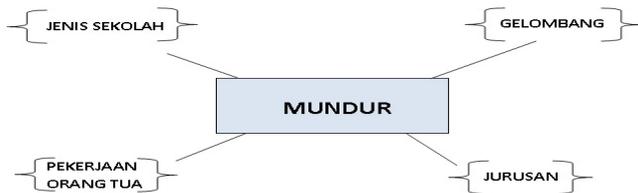
B. Analisis Pembentukan *Itemsets* Dari Data Calon Mahasiswa



Hasil Analisis menunjukkan dalam teknik *frequent pattern* yang dapat digunakan sebagai bahan pertimbangan adalah yang pertama gelombang sebagai rentan waktu pendaftaran, jenis sekolah dan jurusan menentukan kompetensi dari calon mahasiswa kemudian pekerjaan orang tua sebagai penentu finansial. Untuk membangkitkan *frequent pattern* (FP) pada data calon mahasiswa maka data gelombang, jenis sekolah, jurusan, pekerjaan orang tua sebagai sekumpulan *item*.



Gambar 3 Analisis Pembentukan Itemsets Calon mahasiswa Registrasi



Gambar 4 Analisis Pembentukan Itemsets Calon Mahasiswa Mundur

IV. HASIL DAN PEMBAHASAN

A. Data Pembentukan dan Pengujian Klasifikasi

Data pembentukan dan pengujian klasifikasi menggunakan data calon mahasiswa dalam tahun 2017, 2016, 2015. Data pembentukan *classifier* menggunakan data calon mahasiswa pada kampus STMIK STIKOM Bali, sedangkan untuk data uji menggunakan data model yang di gunakan 60% sebagai data uji dari jumlah data tahun 2017, 2016, 2015. Pada tabel berikut akan ditampilkan komposisi dari masing-masing data calon mahasiswa yang digunakan.

TABEL III
PERSENTASE DATA TRAINING DAN DATA TESTING

Area Data	Tahun	Jumlah Data	Presentase Data Training Model (60%)
STMIK Stikom Bali	2017	2706	1691
	2017,2016	5332	3393
	2017,2016,2015	5746	2591

Data pembentukan *classifier* menggunakan data calon mahasiswa pada kampus STMIK STIKOM Bali setelah dilakukan proses *preprocessing* yaitu langkah membersihkan data dengan *missing values*, *noise* data pembentukan kriteria dan pembentukan label setelah itu data calon mahasiswa akan di rubah dalam bentuk data *ARFF* yang diuji pada perangkat lunak yang digunakan untuk implementasi adalah WEKA dengan menggunakan metode C.45 jika di WEKA dinamakan dengan algoritma J.48.

B. Hasil Pengujian Klasifikasi Representasi Tahun 2017

Hasil pengujian klasifikasi menggunakan representasi data calon mahasiswa Tahun 2017 dilihat dari sisi akurasi

keseluruhan yaitu tabel calon mahasiswa 2017 dan nilai perhitungan dengan *confusion matrix*. Hasil perhitungan *confusion matrix* menghasilkan nilai *precision*, *recall* dan *F-meansure* detail dari setiap class. Dataset yang disediakan untuk memprediksi pengunduran diri calon mahasiswa jumlah data calon mahasiswa 2706 dari data calon mahasiswa tahun 2017 adalah 60% dari jumlah data digunakan sebagai data model adalah 1691 data calon mahasiswa.

TABEL IV
CORRECTLY DAN INCORRECTLY REPRESENTASI CALON MAHASISWA 2017

Area Data	Data 2017	Correctly	Incorrectly
STIKOM BALI	1691	78.06%	21.94%

TABEL V
AKURASI DETAIL KLASIFIKASI DATA CALON MAHASISWA

TF	FP	PRECISION	RECALL	F-MEASURE	CLASS
1	1	0,781	1	0.877	LANJUT
0	0	0	0	0	MUNDUR

C. Hasil Pengujian Klasifikasi Menggunakan Klasifikasi Representasi Model 2017, 2016

Hasil *confusion matrix* menghasilkan nilai *precision*, *recall* dan *F-meansure* detail dari setiap class, dan dipisahkan sesuai data model klasifikasi. Dataset yang disediakan untuk memprediksi pengunduran diri calon mahasiswa jumlah data calon mahasiswa 5332 dari data calon mahasiswa tahun 2017, 2016 adalah 60% dari jumlah data digunakan sebagai data model adalah 3357 data calon mahasiswa.

Hasil perhitungan *confusion matrix* pada data 2017, 2016 ditunjukkan pada TABEL VI

TABEL VI
CORRECTLY DAN INCORRECTLY REPRESENTASI CALON MAHASISWA 2017, 2016

Area Data	Data 2017,2016	Correctly	Incorrectly
STIKOM BALI	3357	78,46%	21,54%

Akurasi Detail menggunakan *confusion matrix* pada pengujian menggunakan data 2017, 2016.

TABEL VII
AKURASI DETAIL KLASIFIKASI DATA CALON MAHASISWA

TF	FP	PRECISION	RECALL	F-MEASURE	CLASS
1	1	0,785	1	0,879	LANJUT
0	0	0	0	0,499	MUNDUR

D. Hasil Pengujian Klasifikasi Menggunakan Klasifikasi representasi Model 2017, 2016, 2015

Hasil perhitungan *confusion matrix* menghasilkan nilai *precision*, *recall* dan *F-meansure* detail dari setiap class, dan dipisahkan sesuai data model klasifikasi. Dataset yang disediakan untuk memprediksi pengunduran diri calon mahasiswa jumlah data calon mahasiswa 5746 dari data calon mahasiswa tahun 2017, 2016, 2015 adalah 60% dari jumlah data digunakan sebagai data model adalah 4968 data calon

mahasiswa. Hasil perhitungan *confusion matrix* pada data 2017, 2016, 2015 ditunjukkan pada TABEL VIII

TABEL VIII
CORRECTLY DAN INCORRECTLY REPRESENTASI CALON MAHASISWA 2017, 2016, 2015

Area Data	Data 2017,2016,2015	Correctly	Incorrectly
STIKOM	4968	78.20%	21.80%

TABEL IX
AKURASI DETAIL KLASIFIKASI DATA CALON MAHASISWA

TF	FP	Precision	Recall	F-measure	Class
0,99	0,964	0,787	0,99	0,877	Lanjut
0,036	0,01	0,491	0,036	0,067	Mundur

E. Evaluasi Hasil Pengujian klasifikasi data Calon Mahasiswa

Hasil pengujian pertama menggunakan data testing dari data calon mahasiswa. Hasil pengujian tersebut disajikan dalam *Confusion matrix*, yang kemudian digunakan untuk menghitung rata-rata dari *precision*, *recall*, *F-measure* dan *accuracy*. Berikut adalah TABEL X perbandingan dari hasil pengujian klasifikasi menggunakan representasi data calon mahasiswa tahun 2017, 2016, 2015.

TABEL X
PERBANDINGAN AKURASI CALON MAHASISWA

Representasi	Weighted Average					Akurasi
	TF	FP	Precisi on	Recall	F- Meus ure	
Calon Mahasiswa 2017	0,781	0,781	0,609	0,781	0,684	78%
Calon Mahasiswa 2017, 2016	0,785	0,785	0,616	0,785	0,69	78,46 %
Calon Mahasiswa 2017, 2016, 2015	0,782	0,756	0,722	0,782	0,7	78,20 %

Hasil pengujian menggunakan data sama dalam kurun waktu 3 tahun menunjukkan bahwa ada peningkatan dari model 2017 ke model 2017,2016 begitu juga ada peningkatan kembali dari data 2017,2016,2015. Semakin banyak data yang digunakan untuk membuat model klasifikasi maka akurasi dan F-measure akan semakin meningkat walaupun tidak signifikan.

F. Pembentukan Frequent Pattern

Proses pembentukan *frequent pattern* menggunakan data calon mahasiswa tahun 2017, 2016, 2015. Berikut adalah pembentukan data calon mahasiswa dari data tiga tahun menggunakan item {{gelombang}, {jurusan}, {jenis sekolah}, {pekerjaan_Orangtua}, {label}} dari item yang dibentuk set itemsets pada data calon mahasiswa tiga tahun dalam pembentukan set itemsetsnya {calon_mahasiswa1}, {calon_mahasiswa2}, {calon_mahasiswa3}, {calon_mahasiswa},

{calon_mahasiswaN}}. Kemudian itemset yang dibentuk adalah {IA, IPA, SMAN Mundur}. *Support/minimum support* yang digunakan 10% dari data calon mahasiswa.

TABEL XI
MINIMUM SUPORT LANJUT ITEMSETS_2

Nama_Itemsets1	Itemsets_2	Jumlah Kemunculan
TIDAK_BEKERJA	LANJUT	2668
SMK_S	LANJUT	1128
IPA	LANJUT	892
SMA_N	LANJUT	873
KOMPUTER	LANJUT	791
SMA_S	LANJUT	462
I_A	LANJUT	440
LAIN-LAIN	LANJUT	434
IPS	LANJUT	388
SMK_N	LANJUT	315
I_B	LANJUT	271
II_C	LANJUT	252
II_B	LANJUT	244
IV_C	LANJUT	213
III_B	LANJUT	203
III_A	LANJUT	175
IV_B	LANJUT	174
I_C	LANJUT	173
SISIPAN_II	LANJUT	147
IV_A	LANJUT	141
LAIN-LAIN	LANJUT	140
SISIPAN_I	LANJUT	113
II_A	LANJUT	97
ADMINISTRASI	LANJUT	66
BAHASA	LANJUT	61
PARIWISATA	LANJUT	45
MESIN_OTOMOTIF	LANJUT	39
ELEKTRO	LANJUT	31
SEDERAJAT	LANJUT	31
KESEHATAN	LANJUT	15
MANAJAMEN	LANJUT	14
AGAMA	LANJUT	11
LISTRIK	LANJUT	7
ARSITEKTUR	LANJUT	6
BOGA	LANJUT	4

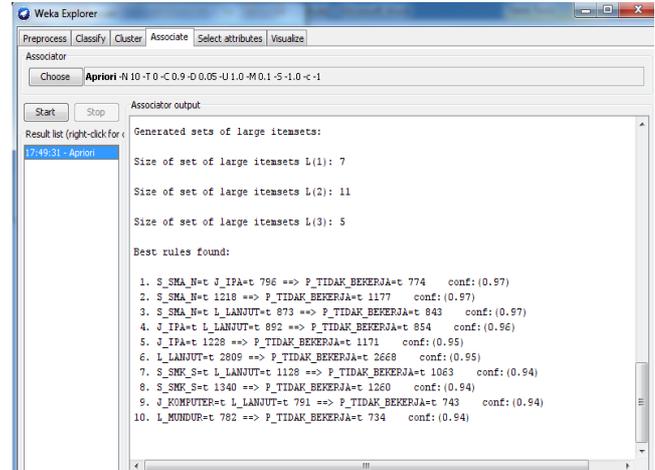


Nama_Itemsets1	Itemsets_2	Jumlah Kemunculan
BUSANA	LANJUT	2
SENI	LANJUT	2
EKONOMI	LANJUT	1
PENGACARA	LANJUT	1

Nama_Itemsets1	Itemsets_2	Jumlah Kemunculan
MANAJAMEN	MUNDUR	1
PENGACARA	MUNDUR	1

TABEL XII
MINIMUM SUPORT MUNDUR ITEMSETS_2

Nama_Itemsets1	Itemsets_2	Jumlah Kemunculan
TIDAK_BEKERJA	MUNDUR	734
SMA_N	MUNDUR	345
IPA	MUNDUR	336
SMK_S	MUNDUR	212
LAIN-LAIN	MUNDUR	139
SMA_S	MUNDUR	139
KOMPUTER	MUNDUR	135
I_A	MUNDUR	128
IPS	MUNDUR	109
I_B	MUNDUR	91
SMK_N	MUNDUR	75
I_C	MUNDUR	74
II_C	MUNDUR	67
II_B	MUNDUR	54
III_B	MUNDUR	54
IV_C	MUNDUR	53
III_C	MUNDUR	50
III_A	MUNDUR	47
LAIN-LAIN	MUNDUR	47
IV_B	MUNDUR	44
IV_A	MUNDUR	33
SISIPAN_I	MUNDUR	30
SISIPAN_II	MUNDUR	30
II_A	MUNDUR	27
ADMINISTRASI	MUNDUR	14
BAHASA	MUNDUR	12
SEDERAJAT	MUNDUR	11
PARIWISATA	MUNDUR	10
ELEKTRO	MUNDUR	7
MESIN_OTOMOTIF	MUNDUR	7
ARSITEKTUR	MUNDUR	4
KESEHATAN	MUNDUR	3
AGAMA	MUNDUR	2
LISTRİK	MUNDUR	2
BOGA	MUNDUR	1



Gambar 5 Hasil Association Rule

Frequent itemsets mining menunjukkan kemunculan **L_LANJUT** ∩ **P_Tidak Bekerja** memiliki nilai *support* paling tinggi yaitu 2668. *Frequent itemset* yang kuat juga ditunjukkan oleh *rule* 8, yaitu **S_SMK_S** ∩ **P_Tidak Bekerja** dengan nilai *support* 1260 dan *rule* 2 yang berisi **S_SMA_N** ∩ **P_Tidak Bekerja** dengan nilai *support* 1177. Dari hasil *frequent itemsets* tersebut potensi **L_LANJUT** ∩ **P_Tidak Bekerja** sering muncul bersama jenis **P_Tidak Bekerja**. Sedangkan, *association rule mining* menunjukkan bahwa alterasi **L_MUNDUR** dan **L_MUNDUR** menjadi *confident*, ini artinya jika items mendukung telah terpenuhi maka kemungkinan **L_MUNDUR** dan **L_MUNDUR**.

V. KESIMPULAN

Hasil pengujian menunjukkan, analisis *knowledge discovery* data akademik untuk prediksi pengunduran diri calon mahasiswa menggunakan teknik klasifikasi dan teknik *frequent pattern* dapat membantu program pendaftaran mahasiswa baru dalam mengurangi risiko kekurangan dalam mendapatkan mahasiswa. Klasifikasi (*classifier*) menggunakan data calon mahasiswa dapat mendukung keputusan pada kampus STMIK STIKOM Bali menjadi lebih kuantitatif dalam menentukan para pendaftar di STMIK STIKOM Bali. Hasil proses *frequent itemsets mining* menunjukkan bahwa kemunculan label lanjut dan mundur juga *frequent* dengan kemunculan pekerjaan ortu. Dari hasil *frequent itemsets mining* dan *association rule* dapat disimpulkan *frequent itemset* dapat menampilkan trend kemunculan antara kriteria dengan label (lanjut atau mundur). Sedangkan *association* justru tidak menampilkan asosiasi *rool mining* antar label dengan kriteria.

Dari Teknik klasifikasi dengan hasil pengujian pada data calon mahasiswa pada kampus STMIK STIKOM Bali Representasi data calon mahasiswa 2017 memiliki akurasi paling rendah yaitu 78% dan representasi calon mahasiswa 2017, 2016 memiliki akurasi 78,46% dan representasi calon

mahasiswa 2017, 2016, 2015 memiliki akurasi 78,20%. Pada saat pengujian menggunakan data model klasifikasi nilai akurasi meningkat serta *precision*, *recall* dan *F-measure* juga meningkat. Semakin banyak data yang digunakan untuk membuat model klasifikasi maka akurasi dan F-measure akan semakin meningkat walaupun tidak signifikan. Teknik *frequent itemsets mining* menunjukkan kemunculan memiliki nilai *support* paling tinggi yaitu 2668. *Frequent itemset* yang kuat juga ditunjukkan oleh $S_SMK_S \cap P_Tidak\ Bekerja$ dengan nilai *support* 1260 dan $S_SMA_N \cap P_Tidak\ Bekerja$ dengan nilai *support* 1177. Calon mahasiswa yang berasal dari SMA_N dan (atau) jurusan IPA berpotensi mundur. Sedangkan item tidak bekerja memiliki nilai *frequent* dengan label mundur juga tinggi yaitu 734 hal ini dapat menjadi pengetahuan bahwa orang tua calon mahasiswa yang memiliki pekerjaan tidak tetap juga berpotensi mundur.

REFERENSI

- [1] Han Jiwei, Kamber, Pei, 2012, "*Data Mining Concepts and Techniques Third Edition*". Morgan Kaufman Publishers
- [2] Adie Wahyudi Oktavia G, I Ketut Gede Darma Putra, I Putu Agung Bayupati, "Implementasi Algoritma Apriori Untuk Menemukan *Frequent Itemset* Dalam Keranjang Belanja" *Majalah Ilmiah Teknologi Elektro Universitas Udayana*, Vol. 15, No.2, Juli –Desember 2016
- [3] Anddri, Yesi Novaria Kunang, Sri Murniati, 2013 "Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universal Bina Darma Palembang". Nasional Informatika 2013.
- [4] I Gede Harsemadi, Made Sudarma, Nyoman Pramaita "Implementasi Algoritma K-Nearest Neighbor Pda Perangkat Lunak Pengelompokan Musil Untuk Menentukan Suasana Hati", *Majalah Ilmiah Teknologi Elektro Universitas Udayana* Vol. 16, No.1, Januari –April 2017
- [5] Ankita A Nichat, Dr. Anjali B Raut. 2017. "*Predicting and Analysis of Student Performance Using Decision Tree Technique*". *International Journal* Vol 5 2017
- [6] Han Jiawai, Pei Jian, Yin Yuwen, Mao Runying. "*Mining Frequent Pattern Candidate Generation: A Frequent Pattern Tree Aproach*". *Data Mining and Knowledge Discovery*, 2014
- [7] Wahyudin, Ari Wijaya, Swamardika. "*Data Mning For Clustering Revenue Plain Expense Area (APBD) By Using K-Means Algorithm*". *International Journal Of Engineering And Emerging Technology*, Vol. 2, No 1 Januari-June 2017
- [8] Jiawei Han, Micheline Kamber, Jian Pei. "*Data Mining Concepts and Techniques Third Edition*". Simon Fraser University Elsvier 2012
- [9] Cristina Oprrea, Della Miora Popescu, Anca Gabriela Petrescu, Irina Barb. "*Data Mining based Model To Improve University Management*" *Journal Of Science And Arts*, 2017.



{ Halaman ini sengaja dikosongkan }