

PENENTUAN UKT MAHASISWA UNS DENGAN ALGORITMA *ITERATIVE DICHOTOMISER THREE* DAN *CLASSIFICATION VERSION 4.5*

¹Laviga Azizah Larytasari, ²Yuliana Susanti, ³Respatiwan

^{1,2,3}Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
e-mail: lavigaziza@gmail.com

Abstrak

Uang Kuliah Tunggal (UKT) merupakan biaya yang ditanggung oleh setiap mahasiswa berdasarkan kemampuan ekonomi orang tua/wali mahasiswa. UKT bertujuan untuk meringankan beban pembayaran bagi mahasiswa khususnya mahasiswa kurang mampu yaitu dengan menerapkan sistem subsidi silang. Besaran UKT tiap mahasiswa tidak sama. Pengklasifikasian berdasarkan pendapatan orang tua, *voltase* listrik, rekening listrik, Pajak Bumi dan Bangunan (PBB), jumlah tanggungan, dan kepemilikan rumah untuk menentukan besaran UKT. Penelitian ini menerapkan metode klasifikasi dengan algoritma *Iterative Dichotomiser Three* (ID3) dan *Classification Version 4.5* (C4.5) serta menghitung nilai akurasi. Hasil penerapan dengan algoritma ID3 menghasilkan 184 klasifikasi dan menunjukkan nilai akurasi sebesar 80.39%, sedangkan pada algoritma C4.5 menghasilkan 75 klasifikasi dan menunjukkan nilai akurasi sebesar 77.95%.

Kata Kunci: UKT, klasifikasi, ID3, C4.5, nilai akurasi

Abstract

UKT is the cost borne by each student based on the economic ability of the student's parents / guardians. UKT aims to ease the burden of payment for students, especially underprivileged students by implementing a cross subsidy system. The amount of UKT for each student is not similar. The classification is based on parent's income, electricity voltage, electricity account, Land and Building Tax (PBB), number of dependents, and home ownership to determine the amount of UKT. This study applies classification methods, namely the *Iterative Dichotomiser Three* (ID3) algorithm and *Classification Version 4.5* (C4.5) and to calculate the accuracy. The results of the application with the ID3 algorithm produces 184 classification and show an accuracy value of 80.39%, while the C4.5 algorithm produces 75 classification and shows an accuracy value of 77.95%.

Keywords: UKT, classification, ID3, C4.5, accuracy value

PENDAHULUAN

UKT merupakan bagian dari biaya kuliah tunggal dimana biaya ini ditanggung oleh setiap mahasiswa yang didasarkan pada kemampuan ekonomi dari masing-masing mahasiswa (Permenristekdikti No. 39, 2016). Sistem UKT telah diterapkan mulai tahun akademik 2013/2014 bagi mahasiswa baru program S1 di seluruh PTN di Indonesia. Sistem ini meniadakan uang pangkal yang harus dibayar oleh setiap mahasiswa dengan semua biaya lain menjadi sebuah biaya tunggal. Tujuan dari penerapan sistem UKT yaitu mampu meringankan beban pembayaran pendidikan mahasiswa khususnya mahasiswa kurang mampu. UKT menerapkan sistem subsidi silang yang didasarkan pada kondisi ekonomi dan sosial orang tua/wali mahasiswa, sehingga pembayaran UKT setiap mahasiswa tidak sama. Hal ini menimbulkan berbagai tanggapan dari berbagai kalangan, karena kebijakan ini dianggap tidak adil.

Besarnya UKT yang harus di bayar oleh masing-masing mahasiswa mengalami kesulitan karena harus memperhatikan pertimbangan kondisi perekonomian orang tua/ wali mahasiswa dalam pengelompokan. Oleh karena itu, diperlukan teknik pengambilan keputusan. Teknik tersebut berguna untuk mengelompokkan besarnya UKT yang sesuai dengan kemampuan ekonomi masing-masing mahasiswa. Teknik klasifikasi pada *data mining* dapat digunakan untuk memberikan pengetahuan penting yang menjadi informasi dari suatu kumpulan data. Teknik klasifikasi berguna untuk membantu dalam mengelompokkan besarnya UKT berdasarkan kemampuan ekonomi orang tua/ wali mahasiswa. Klasifikasi merupakan suatu pengelompokan yang dilakukan ke dalam kelas tertentu berdasarkan ciri-ciri yang sama (Hamandoko dan Tairas, 1999). Teknik klasifikasi dapat disajikan dalam bentuk pohon keputusan. Pohon keputusan merupakan salah satu metode pada data mining yang menggunakan struktur berupa pohon yaitu dengan mengubah bentuk data dalam tabel menjadi model pohon keputusan, lalu mengubahnya lagi menjadi aturan yang mudah dimengerti (Jayanti dkk, 2008). Pohon

keputusan yang digunakan dalam penelitian ini yaitu pohon keputusan dengan algoritma *Iterative Dichotomiser Three* (ID3) dan *Classification version 4.5* (C4.5). Pemilihan variabel dengan algoritma ID3 diproses dengan perhitungan *gain* sedangkan algoritma C4.5 diproses dengan perhitungan *gain ratio*.

Penelitian sebelumnya dilakukan oleh Aradea dkk (2011), yang menggunakan algoritma ID3 dalam penentuan pola data penerimaan mahasiswa baru dan didapatkan nilai akurasi sebesar 100%. Penelitian Adhatrao *et al.* (2013), yang membandingkan algoritma ID3 dengan C4.5 untuk memprediksi kinerja siswa, didapatkan hasil akurasi yang sama. Swastina (2013) melakukan penelitian dengan menerapkan algoritma C4.5 untuk penentuan jurusan mahasiswa dan didapatkan nilai akurasi yang tinggi. Menurut penelitian yang dilakukan oleh Archana *et al.* (2014), algoritma ID3 adalah algoritma yang lebih akurat dibandingkan dengan algoritma C4.5. Pada penelitian ini, diterapkan algoritma ID3 dan C4.5 untuk menentukan penggolongan besarnya UKT mahasiswa UNS serta menghitung nilai akurasi kedua metode.

Algoritma ID3

Algoritma ID3 merupakan algoritma matematika yang berguna untuk menghasilkan suatu pohon keputusan yang mampu mengklasifikasikan suatu obyek. Perhitungan yang digunakan oleh algoritma ID3 adalah dengan *entropy* dan *informasi gain*. *Entropy* dan *informasi gain* dapat dihitung dengan rumus :

$$Entropy(S) = - \sum_{j=1}^n p_j \log_2 p_j \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

dengan S adalah himpunan kasus, n adalah jumlah kelas, p_j adalah proporsi dari S_j terhadap S , A adalah variabel, k adalah banyaknya nilai yang mungkin untuk atribut A , S_i adalah jumlah sampel untuk nilai i , S adalah jumlah seluruh sampel data, dan $Entropy(S_i)$ adalah entropy untuk sampel-sampel yang memiliki nilai i .

Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 banyak digunakan karena memiliki beberapa kelebihan yaitu menghasilkan pohon keputusan yang mudah diinterpretasikan, dapat digunakan untuk menangani variabel bertipe kategorikal atau diskrit dan numerik, mempunyai tingkat akurasi yang dapat diterima, dan dapat menangani *missing value* (Witten *et al.*, 2011). Perhitungan algoritma C4.5 dimulai dari menghitung *entropy*, *gain*, *split info*, dan *gain ratio*. *Entropy* dan *gain* memiliki rumus yang sama dengan (1) dan (2), sedangkan *split info* dan *gain ratio* dapat dihitung dengan rumus :

$$SplitInfo(S, A) = - \sum_{i=1}^k \frac{|S_i|}{|S|} \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (3)$$

$$Gain Ratio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (4)$$

Nilai *gain ratio* terbesar digunakan sebagai variabel akar. Selanjutnya yaitu mengulangi proses sampai semua cabang memiliki kelas yang sama.

Confussion Matrix

Confusion matrix merupakan tabel yang digunakan untuk mencatat hasil klasifikasi dan menghitung akurasi (Han dan Kamber, 2006). Akurasi yaitu presentase dari seberapa baik model mengkorelasi antara hasil dengan variabel-variabel yang ada dalam data.

Tabel 1. Confusion Matrix

Observasi	Prediksi	
	<i>True (+)</i>	<i>False (-)</i>
<i>True (+)</i>	<i>True positive (TP)</i>	<i>False negative (FN)</i>
<i>False (-)</i>	<i>False positive (FP)</i>	<i>True negative (TN)</i>

Rumus yang digunakan untuk perhitungan akurasi yaitu sebagai berikut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang diperoleh dari kantor Seleksi Penerimaan Mahasiswa Baru (SPMB) UNS. Data sekunder yang digunakan yaitu jumlah UKT mahasiswa SI UNS sebagai variabel dependen dan pendapatan orang tua per bulan, PLN (*voltase* listrik), PLN (pajak listrik), PBB, jumlah tanggungan, dan kepemilikan rumah sebagai variabel independen. Adapun langkah-langkah dalam penelitian ini, sebagai berikut:

1. Mendiskripsikan dan mengkategorikan data baik variabel dependen maupun independen
2. Melakukan perhitungan klasifikasi dengan algoritma ID3
3. Melakukan perhitungan klasifikasi dengan algoritma C4.5
4. Mengulangi proses perhitungan sampai semua variabel mempunyai kelas
5. Menghitung nilai akurasi lalu membandingkan nilai akurasi algoritma ID3 dan C4.5.

HASIL DAN PEMBAHASAN

1. Deskripsi Data

Data yang digunakan yaitu data jumlah UKT Mahasiswa S1 UNS. Jumlah total data sebanyak 984 data. Rincian kategori dari masing-masing variabel yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Kategori variabel

Variabel	Kode	Kategori
UKT	1	500000
	2	1000000
	3	>1000000
	4	>4500000
	5	>6000000
Pendapatan	P1	<1500000
	P2	1500000-2500000
	P3	2500000-3500000
	P4	>3500000
<i>Voltase</i> listrik	V1	450 W
	V2	900 W
	V3	≥ 1300 W
Pajak	J1	<100000
	J2	100000-200000
	J3	>200000
PBB	B1	<50000
	B2	50000-100000
	B3	>100000
Jumlah tanggungan	T1	1-2
	T2	3-4
	T3	>4
Kepemilikan rumah	R1	Mempunyai rumah
	R2	Tidak mempunyai rumah

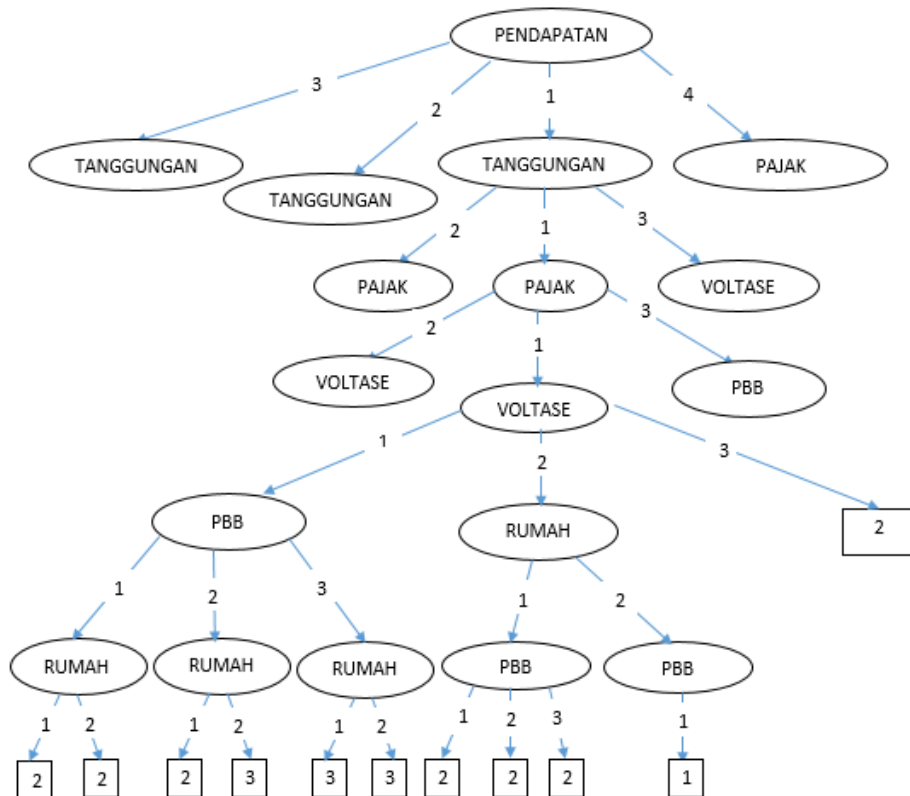
2. Analisis Algoritma ID3

Melakukan analisis algoritma ID3 dengan perhitungan *entropy* dan *information gain* dari tujuh variabel data. Hasil perhitungan ditunjukkan pada tabel 3.

Tabel 3. Perhitungan node akar

Variabel	Kode	Entropy	Information gain
Pendapatan	P1	1.3013	0.9583
	P2	1.1745	
	P3	1.1756	
	P4	0.7558	
Voltase listrik	V1	1.9903	0.1996
	V2	1.907	
	V3	1.2306	
Pajak	J1	2.0246	0.2936
	J2	1.863	
	J3	1.1501	
PBB	B1	2.092	0.1224
	B2	1.7894	
	B3	1.4017	
Jumlah tanggungan	T1	1.8885	0.0269
	T2	1.9067	
	T3	2.1331	
Kepemilikan rumah	R1	1.874	0.0393
	R2	2.0151	

Didapatkan bahwa variabel pendapatan merupakan variabel yang paling berpengaruh karena mempunyai nilai *information gain* tertinggi. Ada 4 kategori pada variabel pendapatan yaitu P1, P2, P3, dan P4. Dari keempat kategori itu dilakukan perhitungan lebih lanjut. Diperoleh pohon keputusan sebagai berikut.



Gambar 1. Pohon keputusan ID3

Pembentukan aturan klasifikasi keseluruhan didapatkan 184 klasifikasi. Pembentukan aturan klasifikasi untuk pendapatan dibawah Rp 1.500.000 (P1), tanggungan keluarga kurang dari tiga orang (T1), dan Pajak kurang dari Rp 100.000 (J1) adalah sebagai berikut.

1. Jika P1, T1, J1, V1, B1, dan R1, maka masuk kelompok UKT golongan 2.
2. Jika P1, T1, J1, V1, B1, dan R2, maka masuk kelompok UKT golongan 2.
3. Jika P1, T1, J1, V1, B2, dan R1, maka masuk kelompok UKT golongan 2.
4. Jika P1, T1, J1, V1, B2, dan R2, maka masuk kelompok UKT golongan 3.
5. Jika P1, T1, J1, V1, B3, dan R1, maka masuk kelompok UKT golongan 3.
6. Jika P1, T1, J1, V1, B3, dan R2, maka masuk kelompok UKT golongan 3.
7. Jika P1, T1, J1, V2, R1, dan B1, maka masuk kelompok UKT golongan 2.
8. Jika P1, T1, J1, V2, R1, dan B2, maka masuk kelompok UKT golongan 2.
9. Jika P1, T1, J1, V2, R1, dan B3, maka masuk kelompok UKT golongan 2.
10. Jika P1, T1, J1, V2, R2, dan B1, maka masuk kelompok UKT golongan 1.
11. Jika P1, T1, J1, dan V3, maka masuk kelompok UKT golongan 2.

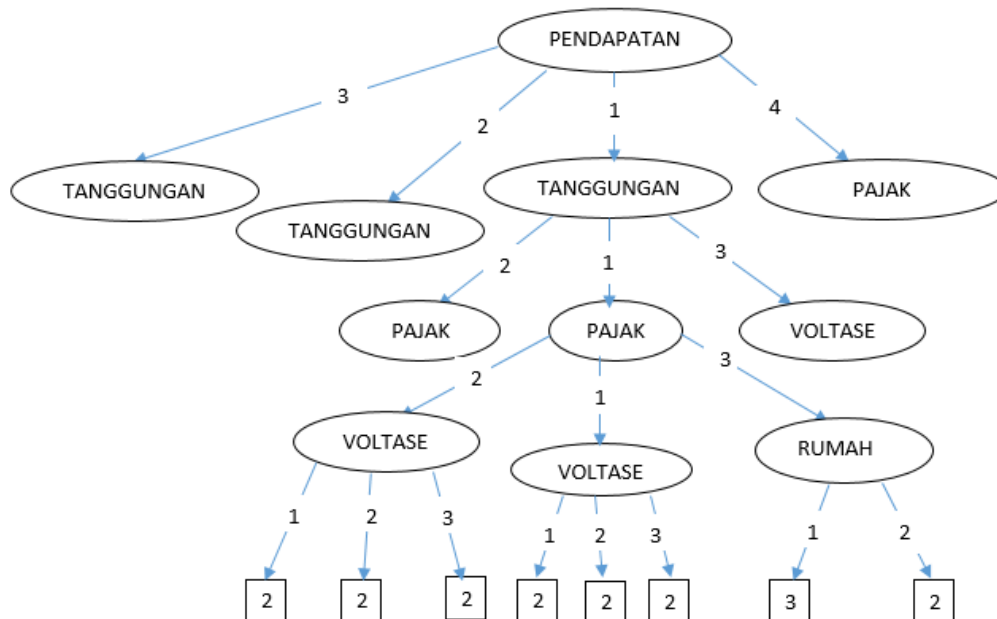
3. Analisis Algoritma C4.5

Analisis dengan algoritma C4.5 menggunakan perhitungan *gain ratio*. Nilai *gain ratio* tertinggi digunakan untuk mengetahui variabel yang paling berpengaruh. Hasil perhitungan ditunjukkan pada tabel 4.

Tabel 4. Perhitungan node akar

Variabel	Gain	Split info	Gain ratio
Pendapatan	0.9583	1.7442	0.5494
Voltase listrik	0.1996	1.5302	0.1304
Pajak	0.2936	1.5762	0.1863
PBB	0.1224	1.5146	0.0808
Jumlah tanggungan	0.0269	1.1458	0.0234
Kepemilikan rumah	0.0393	0.6058	0.0649

Didapatkan bahwa variabel pendapatan mempunyai nilai *gain ratio* tertinggi. Selanjutnya, dari 4 kategori pendapatan tersebut, dilakukan perhitungan lebih lanjut. Diperoleh pohon keputusan dengan pemangkasan sebagai berikut.



Gambar 2. Pohon keputusan C4.5

Pembentukan aturan klasifikasi keseluruhan didapatkan 75 klasifikasi. Pembentukan aturan klasifikasi untuk pendapatan dibawah Rp 1.500.000 (P1), dan tanggungan keluarga kurang dari tiga orang (T1) adalah sebagai berikut.

1. Jika P1, T1, J1, dan V1, maka masuk kelompok UKT golongan 2.
2. Jika P1, T1, J1, dan V2, maka masuk kelompok UKT golongan 2.
3. Jika P1, T1, J1, dan V3, maka masuk kelompok UKT golongan 2.
4. Jika P1, T1, J2, dan V1, maka masuk kelompok UKT golongan 2.
5. Jika P1, T1, J2, dan V2, maka masuk kelompok UKT golongan 2.
6. Jika P1, T1, J2, dan V3, maka masuk kelompok UKT golongan 2.
7. Jika P1, T1, J3, dan R1, maka masuk kelompok UKT golongan 3.
8. Jika P1, T1, J3, dan R2, maka masuk kelompok UKT golongan 2.

4. Analisis Akurasi

Metode ID3 dan C4.5 mempunyai tingkat ketepatan dalam memprediksi besarnya UKT tiap mahasiswa. Hasil tingkat ketepatan dengan menggunakan algoritma ID3 dan C4.5 dapat dilihat dengan menggunakan *confussion matrix* seperti Tabel.

Tabel 5. Perhitungan nilai akurasi ID3

Observasi	Prediksi				
	Kel. 1	Kel. 2	Kel. 3	Kel. 4	Kel. 5
Kel. 1	3	1	1	0	0
Kel. 2	8	98	28	1	0
Kel. 3	0	15	177	32	1
Kel. 4	0	0	15	131	24
Kel. 5	0	0	0	67	382

$$Akurasi\ total = \frac{3 + 98 + 177 + 131 + 382}{5 + 135 + 225 + 170 + 449} \times 100\% = 80.39\%$$

Berdasarkan Tabel 3, menunjukkan bahwa dari 984 data secara keseluruhan terdapat 791 data dengan klasifikasi benar, sehingga diperoleh presentase akurasi untuk memprediksi penggolongan besarnya UKT secara tepat yaitu 80.39% dengan tingkat kesalahan 19.61%.

Tabel 6. Perhitungan nilai akurasi C4.5

Observasi	Prediksi				
	Kel. 1	Kel. 2	Kel. 3	Kel. 4	Kel. 5
Kel. 1	2	1	0	0	0
Kel. 2	9	97	32	3	0
Kel. 3	0	16	167	31	1
Kel. 4	0	0	22	102	7
Kel. 5	0	0	0	95	399

$$Akurasi\ total = \frac{2 + 97 + 167 + 102 + 399}{3 + 141 + 215 + 131 + 494} \times 100\% = 77.95\%$$

Berdasarkan Tabel 4, menunjukkan bahwa dari 984 data secara keseluruhan terdapat 767 data dengan klasifikasi benar, sehingga diperoleh presentase akurasi untuk memprediksi penggolongan besarnya UKT secara tepat yaitu 77.95% dengan tingkat kesalahan 22.05%.

SIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan bahwa dengan algoritma ID3 didapatkan 184 klasifikasi dengan nilai presentase akurasi untuk memprediksi penggolongan UKT secara tepat yaitu 80.39%. Sedangkan dengan algoritma C4.5 didapatkan 75 klasifikasi dengan nilai presentase akurasi yaitu 77.95%. Hal ini menunjukkan bahwa algoritma ID3 mempunyai aturan klasifikasi yang lebih banyak dan nilai akurasi yang lebih tinggi daripada algoritma C4.5. Pada penelitian selanjutnya, dapat dilakukan penerapan dengan metode klasifikasi data mining yang lain untuk menganalisis data UKT.

DAFTAR PUSTAKA

- Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R. and Honrao, V. 2013. Predicting Students Performance Using ID3 and C4.5 Classification Algorithms. *International Journal of Data Mining & Knowledge Management Process*, Vol. 3, No. 5.
- Aradea, Satriyo, A. Aruyan, Z., dan Yuliana, A. 2011. Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru. *Jurnal Penelitian Sitotika*, Vol. 7, No. 1, ISSN:1693-9670.
- Arcana, S., and Dr. K. Elangovan. 2014. Survey of Classification Techniques in Data Mining. *International journal of Computer Science and Mobile Applications*, Vol. 2 Issue. 2, pg. 65-71.
- Hamandoko dan J. Tairas. 1999. *Pengantar Klasifikasi Persepuluhan Dewey*. Jakarta: BPK Gunung Mulia.
- Han, J., and Kamber, M. 2006. *Data Mining Concept Tehniques*. Morgan Kauffman Publisher, San Fransisco.
- Jayanti, N., Puspitodjati, S., dan Elida, T. 2008. Teknik Klasifikasi Pohon Keputusan Untuk Memprediksi Kebangkrutan Bank Berdasarkan Rasio Keuangan Bank. *Proceeding, Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen*, ISSN: 1411-6286.
- Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 39 Tahun 2016 Tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri di Lingkungan Kementerian Riset, Teknologi, dan Pendidikan.
- Swastina, L. (2013). Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. *Jurnal GEMA AKTUALITA*, Vol. 2, No. 1.
- Utama, T. D., Sihwi, S. W. dan Doewes, A. 2014. Implementasi Algoritma Iterative Dichotomizer 3 Pada Penyeleksian Program Mahasiswa Wirausaha UNS. *Jurnal ITSMART*, Vol 3, No. 2.
- Witten, I. F., H.E., and Hall, M. A. 2011. *Data Mining: Practical Machine Learning Tools and Technique*. Third Edition. Elsevier, Inc. Massachusetts.