
Analisis Pola Pemilihan Konsentrasi Ilmu Jurusan Sistem Informasi Di STMIC TIME

Herman
Program Studi Teknik Informatika
Sekolah Tinggi Manajemen Informatika dan Komputer – TIME
Jalan Merbabu No. 32 AA-BB
e-mail : hrman_ang@yahoo.com

Abstrak

Penelitian ini berisi tentang pola atau karakteristik pemilihan konsentrasi ilmu Jurusan Sistem Informasi di STMIC TIME. Dengan adanya penelitian ini diharapkan dapat membantu mahasiswa yang masih bingung dalam memilih konsentrasi ilmu yang akan diambil nantinya. Database yang dipergunakan dalam penelitian ini adalah mata kuliah beserta nilainya dengan total 34 mata kuliah, data diambil dari data akademik STMIC TIME tahun 2008 sampai tahun 2014. Pengolahan data menggunakan teknik data mining dengan algoritma decision tree. Perangkat lunak yang digunakan untuk membantu menyajikan hasil pohon adalah perangkat lunak WEKA versi 3.6.11. Diharapkan dengan adanya penelitian sebagai acuan mahasiswa dalam menentukan konsentrasi ilmu Jurusan Sistem Informasi dapat membantu mahasiswa memilih konsentrasi ilmu yang tepat, sehingga minat belajar mahasiswa dapat ditingkatkan, bakat mahasiswa dapat berkembang, dan membantu STMIC TIME dalam menyediakan sarana dan prasarana dalam mendukung proses pembelajaran.

Kata Kunci : Sistem Informasi, Database, Akademik

1. Latar Belakang

STMIC TIME memiliki dua Jurusan yaitu Jurusan Teknik Informatika dan Jurusan Sistem Informasi, pada Jurusan Sistem Informasi terbagi lagi dua konsentrasi ilmu, yaitu Sistem Informasi Akuntansi dan Sistem Informasi Bisnis.

Masih Ada mahasiswa yang masih bingung, tidak mengetahui minat dan bakat sendiri akan memilih konsentrasi ilmu yang mana setelah memasuki semester lima. Telah banyak dijumpai kasus terjadi kesalahan pemilihan konsentrasi oleh mahasiswa karena tidak mengetahui kemampuan dan minat sendiri. Kesalahan pemilihan konsentrasi ilmu dan jurusan akan mengakibatkan menurunnya prestasi mahasiswa, keengganan minat belajar, sampai berhenti atau putusnya perkuliahan di tengah jalan.

Selain mahasiswa, STMIC TIME juga kesulitan dalam memprediksi pemilihan konsentrasi ilmu pada mahasiswa Jurusan Sistem Informasi, hal tersebut menyebabkan STMIC TIME kesulitan menyediakan sarana dan prasarana yang diperlukan untuk mendukung proses pembelajaran. Pada akhirnya kesalahan pemilihan konsentrasi ilmu selain mempengaruhi prestasi akademik mahasiswa, keengganan belajar mahasiswa juga dapat menurunkan kinerja kerja STMIC TIME dan kenyamanan belajar mahasiswa.

Data berikut adalah data mahasiswa Jurusan Sistem Informasi yang telah memilih konsentrasi ilmu yaitu angkatan tahun 2008 sampai dengan tahun 2014:

Tabel 1. Data Mahasiswa Sistem Informasi

ANGKATAN	AKUNTANSI	BISNIS
2008	33	20
2009	36	29
2010	26	30
2011	32	42
2012	27	37
2013	19	29
2014	16	42

(Sumber: Data Akademik STMIC TIME)

Untuk menghindari hal-hal yang dapat mengganggu kenyamanan pembelajaran, prestasi akademik, keengganan belajar, dan menurunnya kinerja STMIC TIME diperlukan sebuah penelitian yang dapat mengetahui kemampuan, bakat, dan minat mahasiswa berdasarkan nilai mata kuliah pada semester awal, sebelum mahasiswa memilih konsentrasi ilmu pada Jurusan Sistem Informasi di STMIC TIME. Penelitian dilakukan dengan menggunakan data mining, dimana penerapan teknik data mining dilakukan pada data akademik mahasiswa. Sedangkan teknik data mining yang dipergunakan adalah metode decision tree. Pemilihan metode decision tree karena metode ini mudah dipahami, dapat menangani nilai nominal dan nilai diskrit, serta dapat menangani data dengan mempunyai value

yang hilang. Untuk perangkat yang dipergunakan untuk mengolah data data mining adalah perangkat lunak yang berbasis open source yaitu WEKA 3.6.11.

2. Landasan Teori

Data Mining

Beberapa definisi mengenai *data mining* antara lain :

1. *Data mining is the tool which exactly enables discovering of emerging patterns and important business information.* (Ruso & Stojanovic, 2012)
2. *Data mining is the core step of knowledge discovery in database (KDD) and inter disciplinary field includes database management system, machine learning, statistics, neural network, fuzzy logic, etc.* (Kumar, Kapil, & Bhatia, 2012)
3. *Data mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. (Mujjasih, 2011)
4. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar. (Pudjiantoro, Renaldi, & Teogunadi, 2011)

Adanya suatu pemahaman bahwa *data mining* bukan hanya suatu kumpulan alat untuk mengisolasi masing-masing data yang sama sekali berbeda dari yang satu dengan yang lainnya dan mencocokkannya dengan masalah. *Data mining* juga mencocokkan permasalahan data dengan teknik penyelesaian data yang tepat. Namun, yang terjadi dalam prakteknya adalah *data mining* menjadi proses yang berulang-ulang. (Kantardzic, 2011)

Para analisis sering kali salah mempelajari data dan meneliti dengan menggunakan beberapa teknik analisis data. Pada akhirnya memutuskan untuk melihat teknik analisis data yang lainnya dan kemudian kembali ke proses awal lagi, sehingga terkadang memperoleh hasil yang lebih baik atau hasil yang berbeda-beda. Oleh sebab itu, sangatlah penting untuk menemukan permasalahan yang ada atau memperkirakan dependensi dari data atau menemukan data agar dapat menerapkan langkah-langkah standar untuk menggambar kesimpulan dari data (Kantardzic, 2011).

Decision Tree

Sebuah model pohon keputusan terdiri dari satu *set* aturan untuk membagi suatu populasi heterogen besar menjadi lebih kecil, kelompok yang lebih homogen dengan memperhatikan suatu variabel target tertentu.

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root*.

Pada *decision tree* terdapat 3 jenis *node*, yaitu:

1. *Root Node*, merupakan node paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal Node*, merupakan node percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf Node* atau terminal *node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Adapun langkah-langkah dalam proses pengolahan dan analisis data dengan metode *decision tree* adalah sebagai berikut :

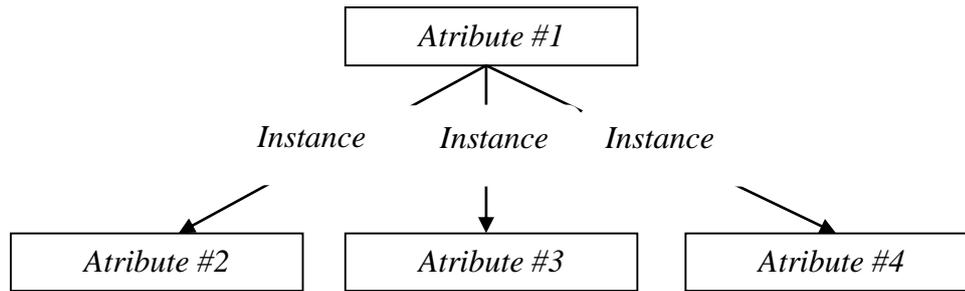
1. Mengubah data menjadi *tree*

Identity Atribut	Atribut #1	Atribut #2	Atribut #3	Atribut #n	Target Atribut

Dari data *sample* kemudian ditentukan *node* terpilih dengan menghitung nilai *entropy* masing-masing atribut.

- a. Untuk menentukan *node* terpilih, gunakan nilai *entropy* dari setiap atribut dengan *sample* yang ditentukan.
- b. Atribut dengan *entropy* yang paling kecil adalah *node* terpilih.
- c. Persamaan untuk menghitung *entropy*.

$$-P_{(+)}\text{Log}_2P_{(+)} - P_{(-)}\text{Log}_2P_{(-)}$$



Gambar 1. Data Menjadi Tree

2. Mengubah *tree* menjadi *rule*

Selanjutnya dari *decision tree*, kemudian diubah menjadi bentuk *rule* dan kemudian *rule-rule* tersebut dianalisis untuk mendapatkan pola terhadap data yang diteliti.

Ada 2 jenis penyusunan *rule* yaitu :

Konjunction[^]
 If *Attribute#1=Instance*[^]*Attribute#2=Instance* THEN Target *Attribute=Instance*

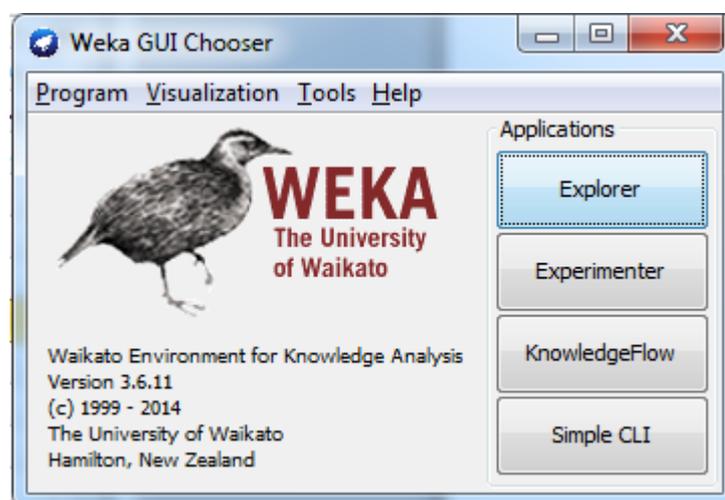
Disjunction[∨]
 If *Attribute#1=Instance*[∨]*Attribute#1=Instance* THEN Target *Attribute=Instance*

WEKA

WEKA merupakan aplikasi *data mining* yang berbasis *open source* (GPL) dan berengine JAVA. WEKA dikembangkan oleh Universitas Waikato di Selandia Baru.

Menurut situs resmi WEKA (*Waikato Environment for Knowledge Analysis*), “WEKA is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.”

The Waikato Environment for Knowledge Analysis (Weka) adalah rangkaian lengkap perpustakaan kelas Java yang mengimplementasikan banyak *state-of-the-art* pembelajaran mesin dan algoritma *data mining*. Weka tersedia secara bebas di *World Wide Web* dan menyertai teks baru pada dokumen *data mining* dan sepenuhnya menjelaskan semua algoritma yang dikandungnya. Aplikasi yang ditulis menggunakan *library class* pada Weka yang dapat dijalankan pada komputer manapun dengan kemampuan *browsing Web*, ini memungkinkan pengguna untuk menerapkan teknik pembelajaran mesin untuk data mereka sendiri terlepas dari *platform* komputer. (Witten, Frank & Hall, 2011).



Gambar 2. Tampilan Awal WEKA

(Sumber: <http://www.cs.waikato.ac.nz/ml/weka/>)

WEKA mulai dikembangkan sejak tahun 1994 dan telah menjadi *software data mining open source* yang populer. WEKA mempunyai kelebihan seperti mempunyai banyak algoritma *data mining* dan *machine learning*, kemudahan dalam penggunaannya, selalu *up-to-date* dengan algoritma-algoritma yang baru. *Software* WEKA tidak hanya digunakan untuk akademik saja namun cukup banyak dipakai oleh perusahaan untuk meramalkan bisnis dari suatu perusahaan. Ian H. Witten merupakan latar belakang dibalik kesuksesan WEKA. Beliau merupakan profesor di *Universitas of Waikato, New Zealand*, yang menekuni *Digital Library, Text Mining, Machine Learning* dan *Information Retrieval*. Pada Weka ada beberapa metode pemilihan *variable* dari suatu *dataset*, diantaranya *BestFirst, ExhaustiveSearch, FCBFSearch, GeneticSearch, GreedyStepwise, RaceSearch, RandomSearch, Rankerdan, RankerSearch*. Metode atau Teknik yang digunakan Weka adalah *Predictive* dan *Descriptive* karena Weka mendukung teknik-teknik *data preprocessing, clustering, classification, regression, visualization, dan feature Reduction*. Semua teknik Weka adalah didasarkan pada asumsi bahwa data tersedia sebagai *flat file* tunggal atau hubungan, dimana setiap titik data digambarkan oleh sejumlah tetap atribut (biasanya, atribut numeric atau nominal, tetapi beberapa jenis atribut lain juga didukung).

3. Metodologi Penelitian

Tujuan penelitian ini adalah mencari pola pemilihan konsentrasi ilmu bagi mahasiswa jurusan sistem informasi dan mengetahui mata kuliah yang berpengaruh dalam pemilihan tersebut. Untuk mengetahui karakteristik pemilihan konsentrasi ilmu mahasiswa sistem informasi diperlukan suatu kemampuan dalam mencari pola tersebut, hal inilah yang dapat dilakukan oleh *data mining*. *Data mining* mampu menemukan informasi penting berupa pola dengan mempelajari informasi mahasiswa dan nilai akademik. Dengan menemukan pola tersebut dalam setiap karakteristik, *data mining* dapat mengetahui mata kuliah mana saja yang paling berpengaruh dalam pemilihan konsentrasi ilmu jurusan sistem informasi.

1. Pengumpulan Data

Data yang dikumpulkan berasal dari nilai akademik STMIK TIME yaitu dari tahun 2008 sampai tahun 2014. Data yang dikumpulkan terdiri dari data mahasiswa, pemilihan konsentrasi ilmu, dan nilai mata kuliah. Objek penelitian pada penelitian ini adalah mahasiswa/mahasiswi sistem informasi di STMIK TIME.

2. Pembersihan Data (Data Cleaning)

Pembersihan data berfungsi agar data lebih mudah ditransformasikan ke dalam suatu format yang lebih mudah dan efektif dalam proses *data mining*. Proses pembersihan yaitu menghilangkan noise, mengisi data yang tidak konsisten, data yang tidak relevan, dan data yang hilang sehingga sesuai dengan penelitian.

3. Penggabungan Data (Data Integration)

Setelah pembersihan data, baru dilakukan penggabungan data dari berbagai *database* ke dalam satu *database* baru yaitu menggabungkan *database* mahasiswa, mata kuliah, dan nilai sehingga mendapat satu *database* yang sesuai untuk *data mining*.

4. Seleksi Data (Data Selection)

Tidak semua data yang telah digabungkan dapat dipergunakan untuk kebutuhan data mining, untuk mendapatkan database yang dapat dipakai dalam proses data mining dilakukan seleksi data. Seleksi dilakukan pada data agar nantinya dapat ditentukan variabel mana saja yang dapat menjadi variabel penelitian. Setelah seleksi data, didapat bahwa mata kuliah yang mempengaruhi mahasiswa memilih konsentrasi ilmu jurusan sistem informasi adalah dari semester satu sampai dengan semester lima, pada semester enam mahasiswa telah melakukan pemilihan konsentrasi ilmu mana yang akan diambil, apakah Sistem Informasi Akuntansi atau Sistem Informasi Bisnis.

Tabel 2. Database Setelah Seleksi Data

No	Atribut	Instance
1	Pengantar Teknologi Informatika	A, B, C
2	Pengantar Akuntansi	A, B, C
3	Pengantar Manajemen	A, B, C
4	Matematika Diskrit	A, B, C
5	Struktur Data	A, B, C
6	Pengantar Bisnis	A, B, C
7	Pengantar Sistem Informasi	A, B, C
8	Sistem Berkas	A, B, C
9	Data Mining	A, B, C
10	Analisa dan Perancangan Sistem Informasi	A, B, C
11	Manajemen Proyek Sistem Informasi	A, B, C
12	Sistem Pendukung Keputusan	A, B, C
13	Pengantar Web	A, B, C
14	Praktikum Pengantar Web	A, B, C

15	Algoritma dan Pemrograman	A, B, C
16	Praktikum Algoritma dan Pemrograman	A, B, C
17	Pengantar Basis Data	A, B, C
88	Praktikum Pengantar Basis Data	A, B, C
19	Pemrograman Visual I	A, B, C
20	Praktikum Pemrograman Visual I	A, B, C
21	Pemrograman Visual II	A, B, C
22	Praktikum Pemrograman Visual II	A, B, C
23	Manajemen Sistem Basis Data	A, B, C
24	Praktikum Manajemen Sistem Basis Data	A, B, C
25	Desain Web	A, B, C
26	Praktikum Desain Web	A, B, C
27	Statistik	A, B, C
28	Praktikum Statistik	A, B, C
29	Pemrograman Orientasi Objek	A, B, C
30	Praktikum Pemrograman Orientasi Objek	A, B, C
31	Rekayasa Sistem Informasi	A, B, C
32	Jaringan Komputer	A, B, C
33	Praktikum Jaringan Komputer	A, B, C
34	Sistem Pakar	A, B, C
35	Jurusan	SIA, SIB

5. Data Mining

Data yang telah disiapkan, akan dianalisis dengan menggunakan perangkat lunak WEKA algoritma decision tree. Secara umum untuk membangun pohon keputusan adalah sebagai berikut:

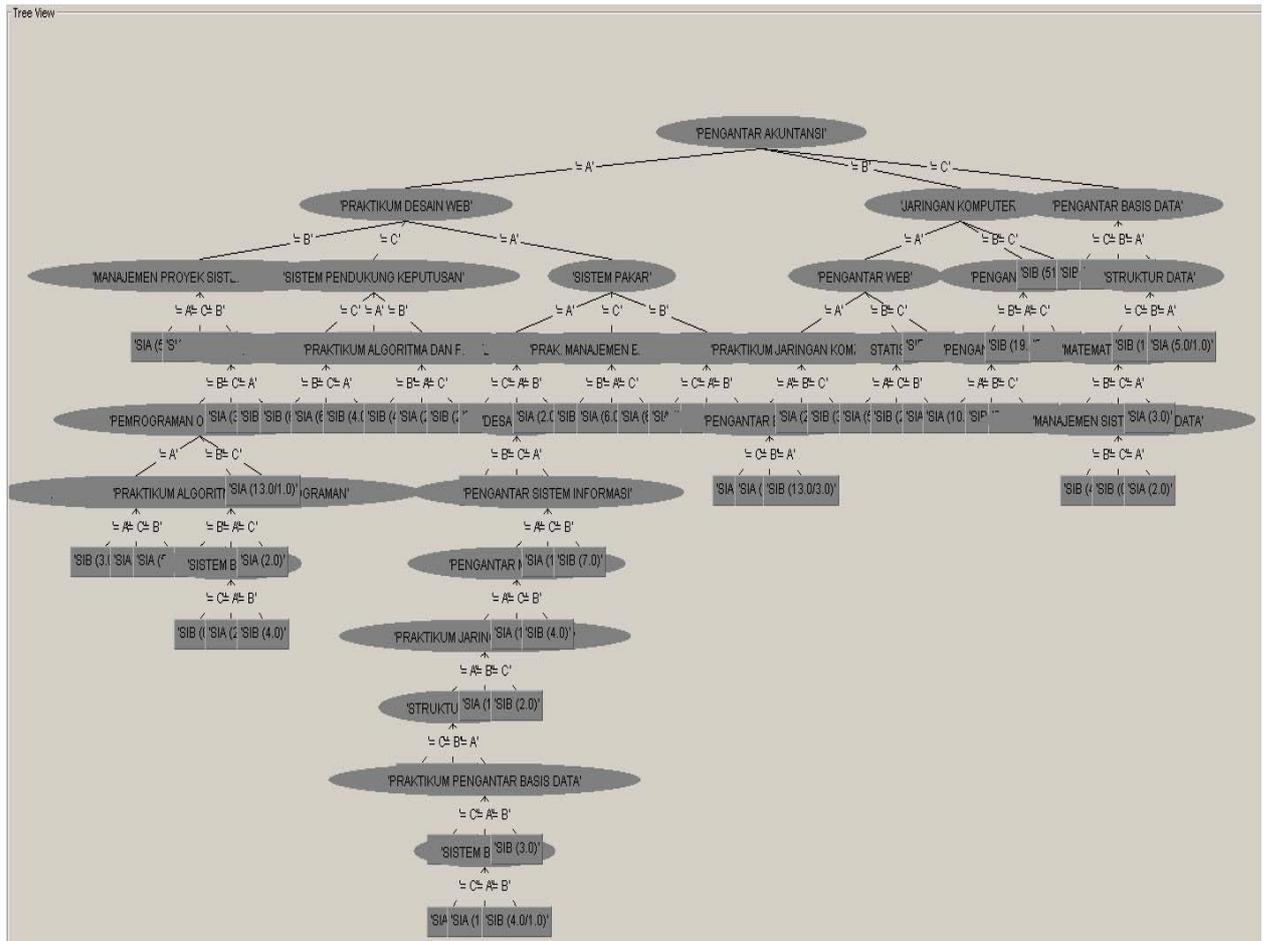
- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk masing-masing nilai.
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus selesai.

6. Pattern Evaluation

Tujuan pattern evaluation adalah untuk mengidentifikasi pola tertentu yang ditemukan. Hasil *output* dari WEKA berupa *tree* menggambarkan informasi dari data yang telah *dimining*. Evaluasi terhadap hasil *data mining* ini bertujuan untuk memberikan pengetahuan mata kuliah apa saja yang menjadi penentu pemilihan konsentrasi ilmu jurusan sistem informasi.

7. Knowledge Presentation

Knowledge Presentation dilakukan dengan menyajikan pengetahuan kepada pengguna setelah metode *data mining* yang digunakan memperoleh pengetahuan.



Gambar 3. Tree Hasil Data Mining

4. Kesimpulan

Dari hasil penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Pada penelitian yang menggunakan decision tree, dengan klasifikasi yang dilakukan mengacu pada 34 mata kuliah, dan setelah dilakukan pengolahan mata kuliah yang berpengaruh pada penentuan pola karakteristik pemilihan konsentrasi ilmu jurusan sistem informasi, diperoleh pola karakteristik berupa pohon keputusan atas mata kuliah yang mempengaruhi pemilihan program studi.
2. Berdasarkan perhitungan data mining dengan metode decision tree atas data mahasiswa STMIK TIME diperoleh kesimpulan bahwa mata kuliah yang paling berpengaruh terhadap pemilihan konsentrasi ilmu jurusan sistem informasi adalah Pengantar Akuntansi.

Untuk hasil yang lebih baik diperlukan hal-hal sebagai berikut:

1. Jumlah data yang dipergunakan perlu ditambah, karena semakin banyak data akan tinggi keakuratan hasil *data mining*.
2. Dapat dilakukan dengan menggunakan metode *data mining* lainnya sebagai perbandingan.

Daftar Pustaka

- [1] Ruso, J. & Stojanovic, V. (2012). *Occupational Health And Safety Using Data Mining*. International Journal for Quality Research. Vol. 6 No. 4 pp. 355-364. University of Belgrade, Serbia
- [2] Kumar, R., Kapil, A. K., & Bhatia, A. (2012). *Modified Tree Classification In Data Mining*. Global Journal Of Computer Science and Techonology. Vol. 12 Issue 2 Vers. 1.0 pp. 59-62. Global Journals Inc., USA
- [3] Mujiasih, S. (2011). *Pemanfaatan Data Mining Untuk Prakiraan Cuaca*. Jurnal Meteorologi dan Geofisika. Vol. 12 No. 2 pp. 189-195. Pusat Meteorologi Penerbangan dan Maritim BMKG, Jakarta
- [4] Pudjiantoro, T. H., Renaldi, F. & Teogunadi, A. (2011). *Penerapan Data Mining Untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru*. Konferensi Nasional Sistem dan Informatika. Bali
- [5] Kantardzic, M. (2011). *Data Mining Concepts, Models, Methods, and Algorithms*. Second Edition. USA : A John Willey & Sons, Inc

- [6] Baradwaj, B. K. & Pal, S. (2011). *Mining Educational Data to Analyze Students Performance*. International Journal of Advanced Computer Science and Applications (IJACSA). Vol. 2 No. 6 pp. 63-69
- [7] Borkar, S. & Rajeswari, K. (2013). *Predicting Students Academic Performance Using Education Data Mining*. IJCSMC, Vol.2, Issue. 7, July 2013, pg 273-279
- [8] Kusrini, & Hartati, S. (2007). *Implementation of C4.5 Algorithm to Evaluate The Cancellation Possibility of New Student Applicants at STMIK AMIKOM Yogyakarta*. The International Conference on Electronic Engineering and Informatics (ICEEI)
- [9] Pandey, U.K., & Pal, S. (2011). *Data Mining: A Prediction of Performer or Underperformer Using Classification*. International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 2 (2)
- [10] Kumar, R., Kapil, A. K., & Bhatia, A. (2012). *Modified Tree Classification In Data Mining*. Global Journal Of Computer Science and Techonology. Vol. 12 Issue 2 Vers. 1.0 pp. 59-62. Global Journals Inc., USA
- [11] Kumar, S. A. & Vijayalakshmi, M. N. (2011). *Efficiency Of Decision Trees In Predicting Student's Academic Performance*. Research Scholar. CS & IT 02 pp. 335–343