

**SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN
PROGRAM STUDI DI UNIVERSITAS
DENGAN ALGORITMA C4.5
(STUDI KASUS: UNIVERSITAS MULTIMEDIA NUSANTARA)**

***DECISION SUPPORT SYSTEM IN CHOOSING STUDY
PROGRAM AT UNIVERSITY USING C4.5 ALGORITHM
(A CASE STUDY AT UNIVERSITAS MULTIMEDIA
NUSANTARA)***

Gilbert Sirait¹, Seng Hansun²

Program Studi Teknik Informatika
Universitas Multimedia Nusantara, Tangerang, Indonesia
¹gilbertsirait@gmail.com, ²hansun@umn.ac.id

Abstrak

Lebih dari setengah dari enam puluh persen mahasiswa baru yang telah memilih jurusan beralih dari jurusan dipilih sebelumnya ke jurusan yang baru. Hal ini terjadi karena kurangnya informasi mengenai jurusan yang tersedia di universitas dan ketidaksesuaian antara jurusan sebelumnya dengan minat dan bakat mahasiswa tersebut. Keputusan seorang mahasiswa dalam memilih jurusan sering dipengaruhi oleh orang tua, kerabat, dan teman-teman. Akibatnya, mahasiswa akan memilih jurusan yang tidak sesuai dengan minat atau bakat mereka. Hal ini akan mengakibatkan lingkungan belajar yang tidak kondusif yang pada gilirannya akan mengakibatkan penurunan kualitas sumber daya manusia yang dihasilkan karena lingkungan belajar yang tidak mendukung. Berdasarkan permasalahan tersebut, sistem pendukung keputusan berbasis *desktop* dibangun untuk membantu mahasiswa memilih jurusan yang relevan dengan minat atau bakat mereka. Sistem ini akan membantu mahasiswa memilih jurusan dengan menerapkan algoritma pohon keputusan, yang disebut C4.5. Faktor yang sedang dipertimbangkan untuk sistem dalam membuat keputusan adalah nilai studi calon mahasiswa. Setelah beberapa percobaan yang dirancang untuk menguji akurasi sistem menggunakan *Confusion Matrix* dan *Cross Validation*, akurasi maksimum 95% tercatat sebagai hasil terbaik dari sistem dengan menggunakan *Confusion Matrix*.

Kata kunci: sistem pendukung keputusan, jurusan perguruan tinggi, algoritma C4.5, *Confusion Matrix*, *Cross Validation*

Abstract

Over half of sixty percent of new college students who had chosen their major switched their major. This happened due to information about the available major at the university was not clearly informed and the student's previous major was irrelevant to their interest and talent. A college student's decision in choosing a major was often influenced by parents, relatives, and friends. As a result, students would choose a major that was irrelevant to their interest or talent. This would result in an uncondusive learning environment which would lead to degradation of the human resource quality. Based on the problem aforementioned, a desktop-based decision support system was built to help students choose a major relevant to their interest or talent. The system would help student choose their major by implementing a decision tree algorithm called C4.5. The factor considered for the system to decide was the student's school grades. After several experiments designed to test the system's accuracy using Confusion Matrix and Cross Validation, a maximum accuracy of 95% was recorded as the best result of the system by using Confusion Matrix.

Keywords: *decision support system, college's major selection, C4.5 algorithms, confusion matrix, cross validation*

Tanggal Terima Naskah : 07 Oktober 2016
Tanggal Persetujuan Naskah : 21 Desember 2016

1. PENDAHULUAN

Dalam situs www.internetworldstats.com [1], Indonesia merupakan negara yang memiliki jumlah penduduk terpadat nomor empat di dunia pada tahun 2015. Reputasi tersebut menjadikan Indonesia memiliki jumlah Sumber Daya Manusia (SDM) yang sangat berlimpah. Namun, memiliki SDM yang berlimpah tidak menjadi tolok ukur kemajuan suatu bangsa dalam berbagai bidang. Menurut data dari *Human Development Index* yang dikutip dari artikel Kompasiana berjudul “Kualitas SDM Indonesia di Dunia”, Indonesia berada pada peringkat 108 di dunia dari segi kualitas SDM [2]. Dengan fakta tersebut, perlu adanya pembenahan dalam hal pendidikan nasional dengan cara melakukan pengawasan dan penilaian terhadap lembaga pendidikan untuk mengetahui kekurangan yang ada dalam mengevaluasi pendidikan yang sedang berjalan agar memperoleh SDM yang mampu beradaptasi terhadap perubahan dan bersaing pada tingkat nasional dan internasional [3].

Menurut penelitian Dr. Fritz Grupe [4], terdapat lima puluh persen dari enam puluh persen mahasiswa baru yang telah menentukan jurusan, pindah dari jurusan yang telah dipilihnya. Fakta tersebut disebabkan kurangnya informasi mengenai jurusan yang terdapat di universitas dan bakat minat yang dimiliki para calon mahasiswa itu sendiri. Akibatnya, para mahasiswa akan membuat keputusan yang bertolak belakang dengan minat dan bakat yang dimiliki dan menciptakan suasana yang tidak kondusif dalam proses belajar selama menjalani perkuliahan, yang pada akhirnya mempengaruhi kualitas SDM karena kurang memaksimalkan bakat dan minat yang dimiliki.

Selain dilatarbelakangi permasalahan tersebut, penelitian ini juga merupakan lanjutan dari penelitian Joko Haryanto [5] yang mendapatkan hasil 63% dari saran aplikasi yang telah dibuat, tetapi dengan menggunakan algoritma C4.5. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang digunakan karena memiliki kelebihan dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, dan dapat menangani atribut bertipe diskret dan numerik [6].

2. SISTEM PENDUKUNG KEPUTUSAN

Michael S. Scoot Morton pada awal tahun 1970-an memperkenalkan konsep Sistem Pendukung Keputusan (SPK). Konsep SPK ditandai dengan sistem interaktif yang berbasis komputer, yang membantu pengambilan keputusan dengan memanfaatkan data dan model dengan tujuan untuk menyelesaikan suatu masalah yang bersifat tidak terstruktur dan semi terstruktur [3]. SPK sebatas alat bantu untuk pembuatan suatu keputusan.

Proses pengambilan keputusan dibagi menjadi empat tahapan proses. Proses-proses tersebut adalah sebagai berikut [3]:

a. Tahap penelurusan (*Intelligence*)

Tahap tempat dilakukannya pendefinisian masalah serta indentifikasi informasi yang dibutuhkan, yang berkaitan dengan persoalan yang dihadapi.

b. Perancangan (*Design*)

Tahap tempat dilakukannya analisis dalam kaitan untuk mencari atau merumuskan alternatif pemecahan masalah. Di dalam tahapan ini juga terdapat proses merancang atau

membangun model pemecahan masalah dan disusun menjadi berbagai alternatif pemecahan masalah.

c. *Pemilihan (Choice)*

Setelah perancangan alternatif solusi, maka dipilih kembali alternatif solusi yang paling sesuai. Kemudahan tahapan ini dipengaruhi oleh suatu alternatif yang memiliki nilai kuantitas tertentu dan terukur.

d. *Implementasi (Implementation)*

Merupakan tahapan terakhir dari keputusan yang telah diambil. Pada tahapan ini juga diperlukan susunan serangkaian tindakan yang terencana agar hasil keputusan dapat dipantau dan disesuaikan bila terjadi suatu perbaikan.

3. **DATA MINING**

Data mining memiliki beberapa karakteristik sebagai berikut [7].

a. *Data mining* merupakan penemuan sesuatu yang tersembunyi dan mempunyai pola data tertentu yang tidak diketahui sebelumnya.

b. *Data mining* menggunakan data yang besar untuk mendapatkan hasil yang dapat dipercaya.

c. *Data mining* bertujuan untuk membuat keputusan yang kritis, terutama dalam hal bisnis.

Data mining terdiri atas beberapa metode pengolahan, yaitu [8]:

a. *Predictive modelling*

Predictive modelling merupakan pengolahan *data mining* dengan melakukan prediksi atau peramalan. Metode ini bertujuan untuk membangun suatu model prediksi nilai yang mempunyai ciri khas tertentu. Algoritma yang menggunakan metode ini adalah *Linear Regression*, *Neural Network*, *Support Vector Machine*, dan lain-lain.

b. *Association*

Association merupakan teknik *data mining* yang mempelajari hubungan antardata. Contoh penggunaan teknik ini adalah untuk menganalisis perilaku mahasiswa yang terlambat mengumpulkan tugas. Contoh algoritma yang menggunakan teknik ini adalah FP-Growth dan Apriori.

c. *Clustering*

Clustering merupakan teknik pengelompokan data ke dalam kelompok tertentu. Contoh algoritma yang menggunakan teknik *clustering* adalah K-Means dan Fuzzy C-Means. Contoh kasus untuk *clustering* adalah terdapat empat fakultas di dalam Universitas Multimedia Nusantara, di dalam masing-masing fakultas dikelompokkan kembali menjadi beberapa program studi.

d. *Classification*

Classification merupakan teknik mengklasifikasikan data. Perbedaan dengan *clustering* adalah terletak pada data yang digunakan. Pada *clustering* tidak terdapat variabel *dependent*, sedangkan pada *classification* diwajibkan terdapat variabel *dependent*. Contoh algoritma yang menggunakan metode *classification* adalah ID3 dan K Nearest Neighbors.

4. **DECISION TREE**

Decision tree atau pohon keputusan merupakan metode yang mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan tersebut dapat dengan mudah dipahami dan dapat diekspresikan dalam bentuk bahasa basis data, seperti *Structured Query Language* untuk mencari *record* pada suatu kategori tertentu. Selain itu, pohon keputusan bermanfaat untuk mengeksplorasi data dan menemukan

hubungan yang tersembunyi antara sejumlah calon variabel *input* dengan sejumlah variabel target [9].

Pohon keputusan dipilih karena proses pembangunannya lebih cepat dan hasil dari model yang dibangun lebih mudah untuk dipahami sehingga *decision tree* merupakan metode klasifikasi yang paling populer untuk digunakan [10].

5. ALGORITMA C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser*) yang dikembangkan oleh J. Ross Quinlan, seorang peneliti di bidang mesin pembelajaran. Algoritma C4.5 mengalami perkembangan dengan berbasis *supervised learning* [6]. Perbaikan C4.5 mengalami puncaknya karena adanya perbaikan yang meliputi metode untuk menangani *numeric attributes*, *missing values*, *noisy data*, dan aturan yang menghasilkan *rules* dan *trees* [11].

Beberapa tahapan dalam membuat sebuah *decision tree* dalam algoritma C4.5, yaitu [8]:

1. Mempersiapkan *data training*. *Data training* didapat dari data *history* yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dengan cara menghitung nilai *gain* dari masing-masing atribut. Nilai *gain* yang paling tinggi akan menjadi akar yang pertama. Sebelum menghitung nilai *gain* dari atribut, terlebih dahulu hitung nilai *entropy*.

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

S = himpunan kasus
 n = jumlah partisi S
 P_i = proporsi S_i terhadap S

3. Kemudian hitung nilai *gain* dengan menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| Entropy(S_i) \dots\dots\dots (2)$$

S = himpunan kasus
 A = fitur
 n = jumlah partisi atribut A
 $|S_i|$ = proporsi S_i terhadap S
 $|S|$ = jumlah kasus dalam S

4. Ulangi langkah kedua dan langkah ketiga sampai *record* terpartisi.
5. Proses *decision tree* akan berhenti ketika:
 - a. Semua *record* dalam simpul n mendapat kelas yang sama
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi
 - c. Tidak ada *record* di dalam cabang yang kosong

Pada algoritma C4.5 untuk memperbaiki informasi dari *gain* dapat menggunakan *gain ratio* [12].

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \dots\dots\dots (3)$$

dimana S merupakan ruang (*sample*) data yang digunakan untuk *training* dan A adalah atribut. $Gain(S, A)$ merupakan informasi *gain* pada atribut A , $SplitInfo$ adalah nilai *split information* pada atribut yang didapat dari rumus sebagai berikut:

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots (4)$$

6. METODOLOGI PENELITIAN DAN HASIL IMPLEMENTASI

6.1 Metodologi Penelitian

Dalam melakukan penelitian ini, terdapat langkah-langkah penelitian yang digunakan sebagai berikut:

a. Studi Literatur

Tahap pertama dalam penelitian ini adalah melakukan studi literatur mengenai sistem pendukung keputusan, *data mining*, dan algoritma C4.5 menggunakan referensi buku, jurnal, artikel, skripsi, dan berbagai sumber untuk mendukung perancangan dan pembangunan sistem.

b. Pengumpulan Data

Data diperoleh dengan cara meminta data dari Divisi Teknologi Informasi (TI) Universitas Multimedia Nusantara (UMN) dan pihak psikolog UMN. Data yang didapat berupa nilai rapor mahasiswa semester 1 dan 2 kelas 12 pada saat SMA, program studi, minat, jurusan, dan nomor induk mahasiswa (NIM) dari data mahasiswa angkatan 2015/2016. Sebelum data tersebut dapat digunakan sebagai *data training*, terdapat beberapa proses dalam melakukan pengolahan data agar data tersebut dapat digunakan sesuai dengan target penelitian. Proses yang pertama adalah melakukan pemilihan data, data yang digunakan adalah data mahasiswa yang menggunakan jalur masuk universitas melalui jalur akademik. Proses yang kedua adalah menjumlahkan data nilai semester 1 dan semester 2 untuk mendapatkan nilai rata-rata. Setelah didapatkan nilai rata-rata, nilai tersebut dikonversikan dengan konversi nilai sebagai berikut: nilai di atas 85 digolongkan sebagai A, nilai antara 80 dan 85 digolongkan sebagai A-, nilai antara 75 dan 80 digolongkan sebagai B+, nilai diantara 70 dan 75 digolongkan sebagai B, dan nilai di bawah 70 digolongkan sebagai C. Jika terdapat data yang kosong atau *missing values*, dalam kasus ini adalah tidak terdapatnya nilai Fisika bagi penjurusan SMA IPS, maka dilakukan pengisian data dengan cara memilih data yang terendah dari seluruh nilai Fisika yang ada. Setelah semua proses tersebut selesai, maka data tersebut dapat digunakan sebagai *data training*.

c. Perancangan Desain Sistem

Setelah mendapatkan semua data yang dibutuhkan, dilakukan perancangan desain sistem yang akan digunakan untuk melakukan pendukung keputusan bagi para calon mahasiswa dengan menggunakan algoritma C4.5.

d. Pembangunan Sistem dan Penerapan Algoritma

Melakukan pembangunan sistem berbasis *desktop* dengan menggunakan bahasa pemrograman berbasis C#. Penerapan algoritma terlebih dahulu dilakukan dengan cara melakukan perhitungan secara manual, kemudian diterapkan ke dalam sistem. Setelah berhasil diterapkan terhadap sistem, maka algoritma perhitungan manual dan perhitungan melalui sistem akan dibandingkan. Jika perbandingan tersebut menghasilkan hasil yang sama, maka dapat dikatakan penerapan algoritma berhasil.

e. Uji Coba dan Analisis

Sistem yang telah selesai dibangun digunakan untuk mengolah *data training* maupun *data testing*. Metode yang digunakan untuk uji coba keakuratan sistem yang telah dibangun adalah dengan menggunakan metode *Confusion Matrix* dan *Cross Validation*. Dari hasil pengujian berdasarkan metode-metode tersebut didapatkan hasil tingkat keakuratan sistem dalam bentuk persentase.

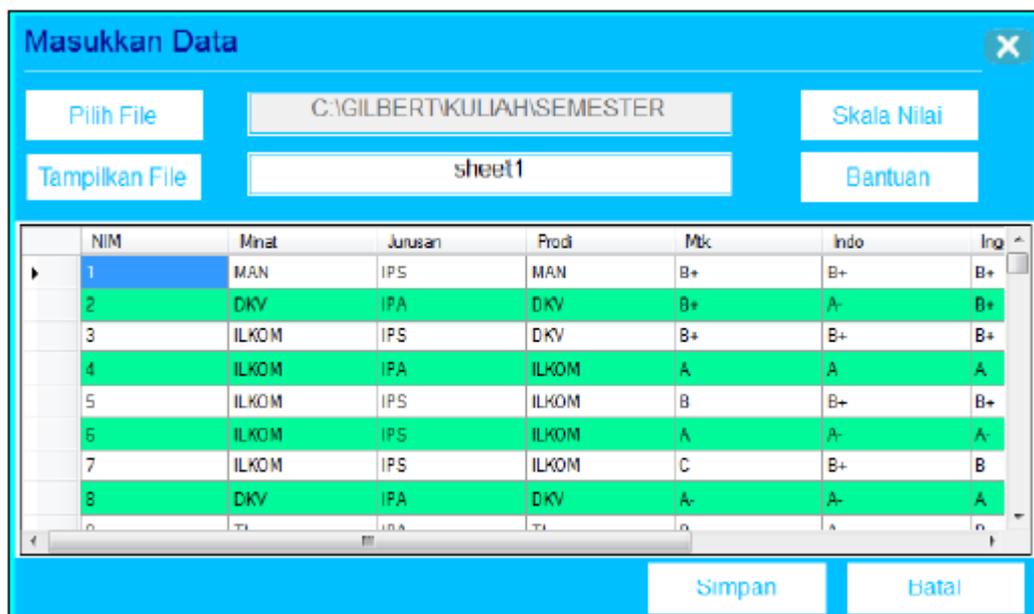
6.2 Implementasi Sistem

Gambar 1 merupakan tampilan halaman menu utama yang di dalamnya terdapat menu-menu lainnya, seperti menu *Input Data Training*, *Input Data Testing*, Lihat Hasil Prediksi, Lihat *Database*, dan Profil. Pada halaman ini juga terdapat *button* 'Info' untuk melihat cara menggunakan aplikasi secara singkat.



Gambar 1. Halaman *Home*

Gambar 2 merupakan tampilan halaman dari Menu *Input Data Training Excel* yang bertujuan untuk memasukkan *data training* ke dalam *database*. Terdapat *button* ‘Pilih File’ yang berfungsi untuk memilih *file* yang akan dimasukkan, *button* ‘Tampilkan File’ yang berfungsi untuk menampilkan *file* pada *datagridview*, *button* ‘Skala Nilai’ untuk melihat skala nilai, *button* ‘Simpan’ untuk menyimpan data ke dalam *database*, *button* ‘Batal’ untuk membatalkan aksi atau menutup halaman, dan *button* ‘X’ untuk keluar dari aplikasi.



Gambar 2. Tampilan Halaman *Input Data Training Excel*

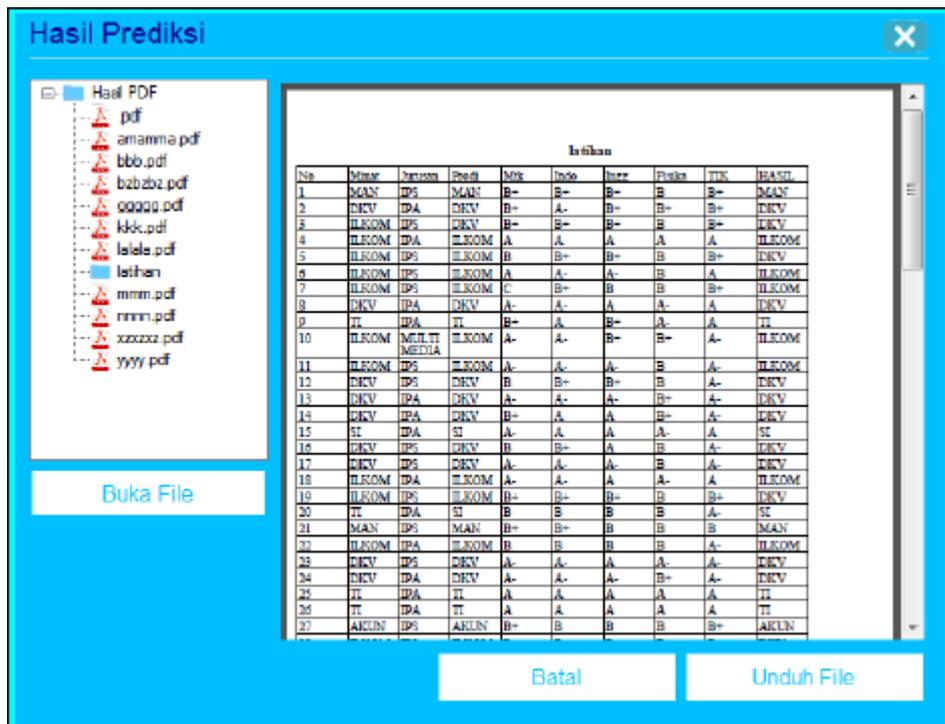
Gambar 3 merupakan tampilan halaman dari *Input Data Testing Excel* yang bertujuan untuk melakukan uji coba terhadap sistem dengan menggunakan data berformat

excel. Terdapat beberapa *button*, seperti *button* ‘Proses Data’ untuk memproses *data training* dari *database* dan membentuk *treeview* dari data tersebut, *button* ‘Tampilkan File’ untuk menampilkan *file* yang telah dipilih, *button* ‘Edit Data’ untuk melakukan *edit* data yang terdapat pada *file*, *button* ‘OK’ untuk memasukkan data ke dalam sistem, *button* ‘Proses’ untuk melakukan proses data dan menampilkan hasil pada data *gridview*, *button* ‘Simpan’ untuk menyimpan hasil ke dalam *folder*.



Gambar 3. Tampilan Halaman *Input Data Testing* Excel

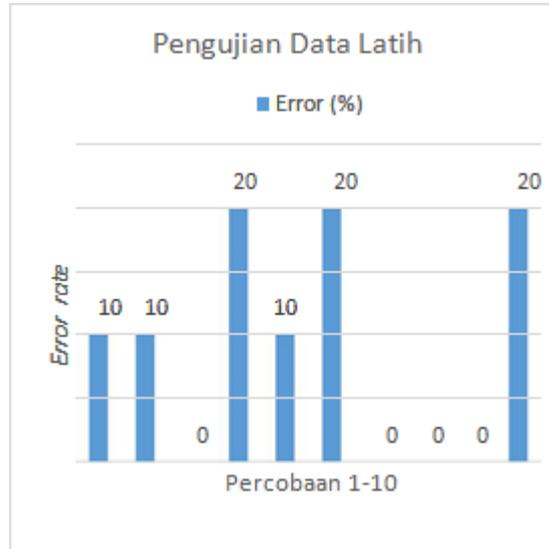
Gambar 4 merupakan tampilan halaman dari Hasil Prediksi yang bertujuan untuk menampilkan *file* yang telah disimpan pada saat melakukan uji coba dengan menggunakan *data testing*. Terdapat beberapa *button*, seperti *button* ‘Buka File’ untuk membuka *file* yang telah dipilih, *button* ‘Batal’ untuk membatalkan aksi melihat hasil prediksi, *button* ‘Unduh File’ untuk melakukan unduh *file* yang telah terpilih, dan terdapat *acrobat pdf reader* untuk membaca *file-file* berformat pdf.



Gambar 4. Tampilan Halaman Hasil Prediksi

7. UJI COBA DAN PEMBAHASAN

Pengujian pertama dilakukan dengan menggunakan metode *cross validation* dengan nilai K sebesar 10.



Gambar 5. Diagram *error* hasil pengujian data latih

Persentase *error* = $(\sum \text{error setiap partisi})/K\text{-Fold Cross validation}$. Dari percobaan dengan menggunakan metode *cross validation* dengan nilai K sebesar 10, didapatkan persentase *error* sebesar 9% dan persentase keberhasilan sebesar 91%.

Pengujian kedua dilakukan dengan menggunakan metode *confusion matrix* untuk mencari tingkat akurasi dari sistem dengan menggunakan algoritma C4.5 dan didapatkan persentase keberhasilan sebesar 95% dengan menggunakan 100 *data training*.

Tabel 1. Penguujian Data Latih

	PREDICTED						
	T I	S K	S I	DK V	ILKO M	MA N	AKU N
TI	5	0	0	0	0	0	0
SK	0	2	0	0	0	0	0
SI	1	0	6	0	0	0	0
DKV	0	0	1	27	1	0	0
ILKO M	0	0	0	0	36	0	0
MAN	0	0	0	0	0	16	0
AKUN	0	0	0	0	0	1	3

Tabel 1 merupakan tabel yang berisikan hasil pengujian data latih menggunakan metode *Confusion Matrix*. Terdapat empat buah hasil yang tidak sesuai dengan kondisi aktual. Data pertama mempunyai data aktual berupa SI, tetapi mempunyai data prediksi berupa TI. Data kedua mempunyai data aktual berupa DKV, tetapi mempunyai data prediksi berupa SI. Data ketiga mempunyai data aktual berupa DKV, tetapi mempunyai hasil prediksi berupa ILKOM, dan data keempat mempunyai data aktual berupa AKUN, tetapi mempunyai data prediksi berupa MAN dan terdapat satu buah data yang belum dapat diprediksikan. Dari hasil Tabel 1 dapat ditentukan nilai keberhasilan sistem sebagai berikut.

$$\text{Persentase Accuracy} = ((5 + 2 + 6 + 27 + 36 + 3 + 16) / 100) * 100\% = 95\%$$

$$\text{Persentase Error} = 100\% - 95\% = 5\%$$

Dari hasil perhitungan di atas, didapatkan hasil *accuracy* dari sistem sebesar 95% dengan menggunakan metode Confusion Matrix dan persentase *error* sebesar 5%.

8. KESIMPULAN

Rancang bangun sistem berbasis *desktop* untuk mendukung keputusan pemilihan program studi dengan menggunakan algoritma C4.5 telah berhasil dilakukan di Universitas Multimedia Nusantara. Hasil akurasi sistem yang didapatkan dari pengujian hasil rekomendasi dengan menggunakan metode *Cross Validation* adalah sebesar 91%, sedangkan untuk pengujian aplikasi berdasarkan metode *Confusion Matrix* adalah sebesar 95%.

REFERENSI

- [1]. Populasi Penduduk Dunia. 2016. Sumber: <http://www.internetworldstats.com/stats8.htm> (diakses pada 21 Februari 2016).
- [2]. Kualitas SDM Indonesia. 2016. Sumber: http://www.kompasiana.com/hesabhirawa/kualitas-sdm-indonesia/didunia_550095bea333111773511436 (diakses pada 21 Februari 2016).
- [3]. Meliala, F.P. 2012. Sistem Pendukung Keputusan Pemilihan Program Studi dengan Metode Fuzzy Multi Criteria Decision Making (FMCDM) [Tesis] <http://repository.usu.ac.id/handle/123456789/35206> (diakses 25 Februari 2016).
- [4]. Gayle, B.R. (2005). Sumber: http://www.nbcnews.com/id/10154383/ns/business-your_retirement/%20-%20.Tqvit1aJznA#.VtAZ__197IU (diakses pada 26 Februari 2016).
- [5]. Haryanto, J. 2014. Aplikasi Pendukung Keputusan Pemilihan Program Studi dengan Metode Fuzzy Logic (Studi Kasus: Universitas Multimedia Nusantara) [Tesis]. Universitas Multimedia Nusantara.
- [6]. Han, J., Kamber, M. 2001. Data Mining Concepts and Techniques. Morgan Kaufman Pub: USA.
- [7]. Davies dan Beynon, P. 2004. Database Systems. Third Edition. Palgrave Macmillan: New York.
- [8]. Larose. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.
- [9]. Kusrini. 2008. Computing for Humanity Algoritma C4.5. Yogyakarta.
- [10]. Ginting, R. 2014. Sistem Pendukung Keputusan. USU Press: Medan.
- [11]. Witten, I.H. dan Frank, E. 2005. Data Mining: Practical Machine Learning Tools and Techniques. Second Edition. Morgan Kauffman: San Fransisco.
- [12]. Bramer, M. 2007. Principles of Data Mining. Springer: London.