

**PENERAPAN *DATA MINING* UNTUK MENENTUKAN
KELAYAKAN ANGGOTA *SECURITY*
DENGAN ALGORITMA C4.5**

SKRIPSI



Oleh:
Selvina
140210060

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2018**

**PENERAPAN *DATA MINING* UNTUK MENENTUKAN
KELAYAKAN ANGGOTA *SECURITY*
DENGAN ALGORITMA C4.5**

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**



**Oleh:
Selvina
140210060**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2018**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Selvina

NPM : 140210060

Fakultas : Teknik dan Komputer

Program Studi : Teknik Informatika

Menyatakan bahwa Skripsi yang saya buat dengan judul:

“PENERAPAN *DATA MINING* UNTUK MENENTUKAN KELAYAKAN ANGGOTA *SECURITY* DENGAN ALGORITMA C4.5”

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip di dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah skripsi ini digugurkan dan gelar saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 7 Februari 2018



Selvina

140210060

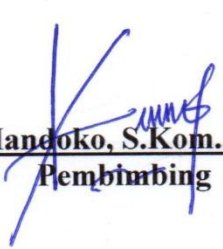
**PENERAPAN *DATA MINING* UNTUK MENENTUKAN
KELAYAKAN ANGGOTA *SECURITY*
DENGAN ALGORITMA C4.5**

**Oleh:
Selvina
140210060**

**SKRIPSI
Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 7 Februari 2018


Koko Handoko, S.Kom., M.Kom.
Pembimbing

ABSTRAK

Pentingnya jasa *security* untuk menunjang keamanan lingkungan di Indonesia masih kurang disadari oleh masyarakat. Hal tersebut dapat dilihat dari banyaknya anggota *security* yang tidak memenuhi kriteria atau syarat wajib seorang *security*. Masih banyak perekrutan anggota *security* yang dilakukan secara sembarang bahkan ada yang menggunakan unsur kekeluargaan. Penelitian ini bertujuan menentukan kelayakan anggota *security* dengan penerapan *data mining*. *Data mining* merupakan suatu proses untuk menemukan informasi-informasi yang bermanfaat dari sekumpulan *database* yang berukuran besar. Salah satu teknik yang ada pada *data mining* adalah klasifikasi. Dengan menerapkan teknik klasifikasi diharapkan dapat menghasilkan suatu pohon keputusan yang berfungsi untuk membantu direktur dan manajer dalam menentukan kelayakan anggota *security*. Adapun dalam membentuk pohon keputusan ini memerlukan suatu metode algoritma berupa algoritma C4.5. Pemodelan algoritma C4.5 dilakukan dengan perhitungan nilai entropi, *gain*, *split info* dan *gain ratio* dari data penelitian. Data penelitian bersumber dari PT Avava Duta Indonesia, data yang diambil meliputi data nama, usia, pendidikan terakhir, berat badan, tinggi badan, tes sehat, sertifikat pelatihan, dan masa kerja. Hasil penelitian adalah sebuah pohon keputusan dengan variabel-variabel yang mempengaruhi penentuan kelayakan dengan hasil akhir layak atau tidak layaknya anggota *security*. Pohon keputusan yang dihasilkan dievaluasi dengan *software RapidMiner* dan dibuktikan benar karena hasilnya sama. Berdasarkan hasil penelitian maka disimpulkan bahwa penerapan *data mining* cocok digunakan untuk menentukan kelayakan anggota *security*. Pengujian terhadap hasil penelitian juga dilakukan dengan menggunakan metode *split validation* yang menghasilkan tingkat akurasi 90,91%, *cross k-fold validation* 10 dan 3 dengan tingkat akurasi 76,32% dan 86,34%, serta *cross leave one out validation* dengan tingkat akurasi 84,21%.

Kata kunci: *Data Mining, Klasifikasi, Algoritma C4.5, Pohon Keputusan.*

ABSTRACT

The importance of security services to support environmental security in Indonesia is still poorly realized by the public. This can be seen from the number of security members who do not have the criteria or mandatory requirements of a security guard. There are still many recruitment of security members who do arbitrarily and some even use the elements of kinship. This research aims to determine the eligibility of security members with the application of data mining. Data mining is a process for finding useful information from a large set of databases. One of the techniques that exist in data mining is classification. By applying classification techniques, the research is expected to generate a decision tree that can serve to assist the director and manager in determining the eligibility of security members. As for forming a decision tree requires an algorithmic method of C4.5. Modeling by using C4.5 algorithm is done by calculation of entropy, gain, split information and gain ratio from research data. The research data is taken from PT Avava Duta Indonesia which includes data of name, age, last education, weight, height, healthy test, training certificate, and years of service. The result of research is a decision tree with the variables that influence in the determination of eligibility with the final result whether the particular member is eligible or not. The resulting decision tree is evaluated with RapidMiner and is proved true because the result is the same. Based on the results of the research, it's concluded that the application of data mining is suitable to determine the eligibility of security members. Tests on the results of research are also conducted by using split validation method with 90.91% accuracy, cross k-fold validation 10 and 3 with 76.32% and 86.34% accuracy, and cross leave one out validation with 84.21% accuracy.

Keywords: Data Mining, Classification, C4.5 Algorithm, Decision Tree.

KATA PENGANTAR

Syukur terhadap Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam.
2. Dekan Fakultas Teknik dan Komputer Universitas Putera Batam.
3. Ketua Program Studi Teknik Informatika Universitas Putera Batam.
4. Bapak Koko Handoko, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
5. Dosen dan Staff Universitas Putera Batam.
6. Bapak Ahmad Arifin selaku Direktur PT Avava Duta Indonesia yang telah memberikan izin pengumpulan data penelitian.
7. Kedua orang tua dan anggota keluarga yang selalu mendoakan dan memberi semangat.
8. Teman-teman perkuliahan yang telah membantu dan memberikan informasi yang berguna.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan hidayat serta taufikNya, Amin.

Batam, Februari 2018

Penulis

DAFTAR ISI

Halaman

HALAMAN SAMPUL DEPAN	
HALAMAN JUDUL	ii
SURAT PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
 BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah	4
1.4 Perumusan Masalah	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.6.1 Manfaat Teoritis.....	5
1.6.2 Manfaat Praktis	6
 BAB II KAJIAN PUSTAKA	
2.1 <i>Knowledge Discovery in Database (KDD)</i>	7
2.2 <i>Data Mining</i>	8
2.2.1 Tahapan <i>Data Mining</i>	10
2.2.2 Penerapan <i>Data Mining</i>	12
2.3 Metode <i>Data Mining</i>	13
2.3.1 Teknik Klasifikasi.....	15
2.3.2 Algoritma C4.5	17
2.3.3 Pohon Keputusan	19
2.3.4 <i>Rule Based</i>	20
2.4 <i>Software</i> Pendukung	21
2.5 <i>Security</i>	22
2.6 Penelitian Terdahulu	26
2.7 Kerangka Pemikiran.....	30
 BAB III METODE PENELITIAN	
3.1 Desain Penelitian	31
3.2 Teknik Pengumpulan Data.....	34
3.3 Operasional Variabel	34

3.4	Metode Analisis dan Rancangan Sistem.....	37
3.5	Lokasi dan Jadwal Penelitian.....	38
3.5.1	Lokasi Penelitian.....	38
3.5.2	Jadwal Penelitian	39

BAB IV HASIL DAN PEMBAHASAN

4.1	Analisa Data.....	40
4.1.1	Seleksi Data	40
4.1.2	<i>Cleaning</i> Data	41
4.1.3	Transformasi Data.....	42
4.1.4	Perhitungan Algoritma C4.5	45
4.2	Hasil Pengujian	64
4.2.1	Implementasi <i>RapidMiner</i>	64
4.2.2	Pengujian dengan <i>Split Validation</i>	70
4.2.3	Pengujian dengan <i>Cross Validation (K-Fold Validation)</i>	71
4.2.4	Pengujian dengan <i>Cross Validation (Leave One Out)</i>	72

BAB V SIMPULAN DAN SARAN

5.1	Simpulan	74
5.2	Saran	75

DAFTAR PUSTAKA.....	76
----------------------------	-----------

DAFTAR RIWAYAT HIDUP

SURAT KETERANGAN PENELITIAN

LAMPIRAN

DAFTAR TABEL

	Halaman
Tabel 3.1 Data Karyawan PT Avava Duta Indonesia	36
Tabel 3.2 Jadwal Penelitian.....	39
Tabel 4.1 Perhitungan Kategori Sampel.....	41
Tabel 4.2 Klasifikasi Atribut Usia.....	43
Tabel 4.3 Klasifikasi Tinggi Badan.....	43
Tabel 4.4 Klasifikasi Berat Badan.....	44
Tabel 4.5 Klasifikasi Tes Kesehatan	44
Tabel 4.6 Klasifikasi Pendidikan Terakhir.....	44
Tabel 4.7 Klasifikasi Sertifikat Pelatihan.....	45
Tabel 4.8 Klasifikasi Masa Kerja	45
Tabel 4.9 Sampel Penelitian.....	46
Tabel 4.10 Hasil Perhitungan <i>Node</i> 1.....	55
Tabel 4.11 Hasil Perhitungan <i>Node</i> 2.....	57
Tabel 4.12 Hasil Perhitungan <i>Node</i> 3.....	58
Tabel 4.13 Hasil Perhitungan <i>Node</i> 4.....	60
Tabel 4.14 Hasil Perhitungan <i>Node</i> 5.....	61
Tabel 4.15 Hasil Perhitungan <i>Node</i> 6.....	63

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Penjelasan <i>Data Mining</i>	8
Gambar 2.2 Tahapan <i>Data Mining</i> (KDD)	10
Gambar 2.3 Metode <i>Data Mining</i>	13
Gambar 2.4 Pohon Keputusan	19
Gambar 2.5 Kerangka Pemikiran	30
Gambar 3.1 Desain Penelitian	31
Gambar 3.2 Rancangan Sistem	37
Gambar 4.1 Pohon Keputusan <i>Node</i> Akar	56
Gambar 4.2 Pohon Keputusan <i>Node</i> 2	58
Gambar 4.3 Pohon Keputusan <i>Node</i> 3	59
Gambar 4.4 Pohon Keputusan <i>Node</i> 4	61
Gambar 4.5 Pohon Keputusan <i>Node</i> 5	62
Gambar 4.6 Pohon Keputusan <i>Node</i> 6	63
Gambar 4.7 Gambaran Hubungan Proses	65
Gambar 4.8 Paramater <i>Read Excel</i>	65
Gambar 4.9 <i>Import</i> Data ke <i>RapidMiner</i>	66
Gambar 4.10 Pemilihan Variabel	66
Gambar 4.11 Bagian Dalam Proses <i>Validation</i>	67
Gambar 4.12 Pohon Keputusan <i>RapidMiner</i>	67
Gambar 4.13 Hasil Pengujian <i>Split Validation</i>	70
Gambar 4.14 Hasil Pengujian <i>K-Fold Validation 10</i>	71
Gambar 4.15 Hasil Pengujian <i>K-Fold Validation 3</i>	72
Gambar 4.16 Hasil Pengujian <i>Leave One Out Validation</i>	72

DAFTAR RUMUS

	Halaman
Rumus 2.1 Perhitungan Entropi	18
Rumus 2.2 Perhitungan <i>Gain</i>	18
Rumus 2.3 Perhitungan <i>Gain Ratio</i>	19
Rumus 4.1 Perhitungan Sampel	40
Rumus 4.2 Perhitungan Sampel <i>Stratified Random</i>	41
Rumus 4.3 Perhitungan Nilai Interval	42
Rumus 4.4 Perhitungan <i>Body Mass Index</i>	43
Rumus 4.5 Perhitungan Akurasi	70

DAFTAR LAMPIRAN

- Lampiran 1. Data-data Penelitian
- Lampiran 2. Hasil Transformasi Data
- Lampiran 3. *Dataset* dalam *RapidMiner*
- Lampiran 4. Perhitungan *Node* 2
- Lampiran 5. Perhitungan *Node* 3
- Lampiran 6. Perhitungan *Node* 4
- Lampiran 7. Perhitungan *Node* 5
- Lampiran 8. Perhitungan *Node* 6