

**ANALISIS KINERJA DALAM MENDETEKSI STUDENT LOSES
BERDASARKAN NILAI *GAIN* DENGAN *SPLITE FEATURE REDUCTION*
MODEL PADA ALGORITMA C4,5**

TESIS

Oleh:

MOHAMMAD YUSUP

NIM : 184211013



**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

MIKROSKIL

MEDAN

2019

LEMBAR PENGESAHAN

ANALISIS KINERJA DALAM MENDETEKSI STUDENT LOSER
BERDASARKAN NILAI *GAIN* DENGAN *SPLIT FEATURE REDUCTION*
MODEL PADA ALGORITMA C4,5

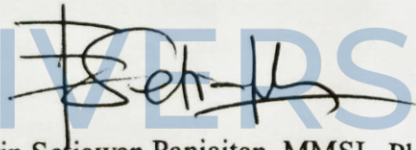
TESIS

Diajukan Untuk Melengkapi Persyaratan
Guna Mendapatkan Gelar Magister Setrata Dua
Program Studi Magister Teknologi Informasi

Oleh:

MOHAMMAD YUSUP
NIM : 184211013


Disetujui Oleh
Dosen Pembimbing I



Ir. Erwin Setiawan Panjaitan, MMSI., Ph.D.

Medan, 30 Januari 2020
Diketahui dan disahkan Oleh

Ketua Program Studi
Magister Teknologi Informasi



Dr. Ronsen Purba, M.Sc.

Saya yang membuat pernyataan ini adalah mahasiswa Jurusan/Program studi S2 Magister Teknologi Informasi STIMIK Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Mohammad Yusup
Nim : 184211013
Peminatan : Teknologi Informasi

Saya telah melaksanakan penelitian dan penulisan tesis dengan judul:

“ANALISIS KINERJA DALAM MENDETEKSI STUDENT LOSES BERDASARKAN NILAI *GAIN* DENGAN *SPLITE FEATURE REDUCTION MODEL* PADA ALGORITMA C4,5”.

Dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya), dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. bila dikemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh stimik mikroskil medan, yakni mencabut ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah. selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada STIMIK Mikroskil Medan, hak bebas royalti non eksklusif (*non-exclusive royalty free right*) atas tesis saya beserta perangkat yang ada (jika diperlukan). dengan hak ini, STIMIK Mikroskil Medan berhak menyimpan mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak atau elektronik, selama mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi tesis saya guna pengembangan karya dimasa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak.

Demikian pernyataan ini saya perbuat dengan sungguh sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan 30 Januari 2020

Saya yang membuat pernyataan



Mohammad Yusup

KATA PENGANTAR

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tesis yang berjudul “Analisis Kinerja Dalam Mendeteksi Student Loses Berdasarkan Nilai *Gain* Dengan *Split Feature Reduction Model* Pada Algoritma C4,5”.

Tesis ini dibuat untuk melengkapi persyaratan kurikulum pada Program Studi Magister Teknologi Informasi Strata Dua, STMIK Mikroskil Medan. Penulis menyampaikan banyak terima kasih kepada:

1. Bapak Dr. Pahala Sirait, S.T., M.Kom., selaku Ketua STMIK Mikroskil Medan.
2. Bapak Dr. Ronsen Purba, M.Sc. selaku Ketua Program Studi Magister Teknologi Informasi.
3. Bapak Ir. Erwin Setiawan Panjaitan MMSI., Ph.D. selaku pembimbing I dalam proses penulisan tesis ini.
4. Bapak Roni yunis, S.Kom., M.T. selaku Dosen Pembimbing yang telah membimbing penulis dalam proses penyelesaian tesis ini.
5. Bapak Djoni, S.Kom., M.T.I., selaku Wakil Ketua I STMIK Mikroskil Medan.
6. Bapak atau Ibu Dosen Penguji atas masukan/saran yang diberikan. Dan Dosen STMIK Mikroskil Medan yang telah membantu proses penulisan tesis ini.
7. 6. Anggota keluarga, teman, serta saudara-saudari yang terus memberikan dukungan penuh kepada penulis selama proses penulisan tesis ini.
8. Semua pihak yang telah membantu dalam penulisan tesis ini. Penulis menyadari bahwa masih terdapat banyak kekurangan yang ada. Oleh sebab itu, kritik dan saran yang bersifat membangun akan diterima. Akhir kata, semoga Tesis ini dapat bermanfaat, bagi lembaga dan masyarakat. Terima kasih.

Medan, 26 Januari 2020

Penulis

DAFTAR ISI

ABSTRAK	
KATA PENGANTAR	4
DAFTAR ISI.....	ii
DAFTAR GAMBAR.....	v
DAFTAR TABEL.....	vi
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.1.1. Masalah Penelitian.....	4
1.2.1. Identifikasi Masalah.....	4
1.3. Tujuan dan Manfaat Penelitian.....	5
1.3.1. Tujuan Penelitian.....	5
1.3.2. Manfaat Penelitian.....	5
1.4. Ruang Lingkup.....	6
1.5. Metodologi Penelitian.....	6
1.6. Sistematika Penulisan.....	7
BAB II.....	9
2.1. Tinjauan Pustaka.....	9
2.1.1. Perguruan Tinggi.....	9
2.2. <i>Data Mining</i>	11
2.2.1. <i>Clasification</i>	12
2.2.2. <i>Clustering</i>	13
2.2.3. <i>Regression</i>	13
2.2.4. <i>Assosiation</i>	13
2.2.5. <i>Forecasting</i>	14
2.2.6. <i>Sequence Analysis</i>	15

2.2.2. <i>Deviation Analysis</i>	15
2.3. <i>Class Imbalance</i> Kelas tidak Seimbang.....	17
2.3.1. Pengaruh ketidakseimbangan.....	18
2.3.2. Tantangan Pada Kelas tidak seimbang.....	18
2.3.3. Distribusi kelas tidak Seimbang.....	19
2.3.4. Kompleksitas Kelas Atau Kelas <i>Overlapping</i>	20
2.3.5. <i>Small Disjunct</i>	20
2.4. Penanganan kelas Tidak Seimbang.....	21
2.4.1. Pendekatan Level Data dengan <i>Feature Selection</i>	21
2.4.2. <i>Split Feature Reduction Model</i>	22
2.5. Algoritma C4.5.....	22
2.5.1. Pembuatan pohon Keputusan.....	23
2.5.2. Evaluasi dan Pemangkasan Pohon (<i>Opsional</i>).....	23
2.5.2. Membangun Aturan dari dari Pohon keputusan (<i>Opsional</i>)	24
2.6. Validasi dan Evaluasi Pada Klasifikasi.....	25
2.6.1. <i>Cross Validation</i>	27
2.6.2. <i>Confusion Matrix</i>	28
2.6.1. <i>Receiver Operating Character Curve (ROC)</i>	29
2.7. Penelitian terkait	33
2.8. Kerangka Konsep Pemecahan Masalah	36
2.9. Hipotesis.....	37
BAB III METODOLOGI PENELITIAN	38
3.1. Analisis Masalah	38
3.1.1. Metode Pengumpulan Data	39
3.1.2. Pengolahan Data Awal (<i>Preprocessing</i>).....	40
3.1.3. Proses Pemberian Label	43

3.2.1. Kerangka Model Penelitian.....	43
3.3.1. Validasi dan Evaluasi.....	44
3.3.2. Alat Analisis Data Penelitian	44
3.4. Teknik Analysis dan Pengujian Model	46
3.4.1. Teknik Analysis	47
3.4.2. Pengujian Model	47
3.4.3. Kontribusi Penelitian.....	48
BAB IV	50
4.1. Hasil	50
4.1.1. Hasil Pengumpulan Data dan Pencarian Nilai Gain	50
4.1.2. Penentuan Peringkat Berdasarkan Nilai Gain.....	55
4.2. Pembahasan.....	56
4.2.1. Analisis klasifikasi algoritma C4.5	56
4.2.2. Analisis kinerja algoritma C4.5 dengan <i>Confusion Matrix</i>	57
4.2.3. Hasil Analisis Kinerja klasifikasi C4.5 dengan <i>Confusion Matrix</i>	61
4.2.4. Hasil Kinerja proses pengujian dengan <i>cross validation</i>	62
4.2.5. Hasil estimasi Kinerja ROC (<i>Receiver Operating Characteristic</i>)	64
BAB V	68
5.1. Kesimpulan	68
5.2. Saran.....	69
DAFTAR PUSTAKA	70
LAMPIRAN.....	73

DAFTAR GAMBAR

Gambar 2.1. Pertumbuhan Mahasiswa PTS di Indonesia 2018.....	9
Gambar 2.2. Pertumbuhan Mahasiswa UNPAB beberapa tahun terakhir.....	10
Gambar 2.3. Persentase lulus, aktif dan tidak aktif 2015 s/d 2019.....	11
Gambar 2.4. Klasifikasi - <i>Decision Tree</i>	14
Gambar 2.5. Tahap Data Mining.	Error! Bookmark not defined. 5
Gambar 2.6. Dampak ukuran sampel kecil (a dan b).....	19
Gambar 2.7. Contoh tidak seimbang antara Kelas (a dan b).....	21
Gambar 2.8. Proses klasifikasi data.....	24
Gambar 2.9. Tahap pengujian dan penggunaan aturan.....	24
Gambar 2.10. Contoh pohon keputusan.....	25
Gambar 2.11. <i>Cross Validation</i>	28
Gambar 2.12. Konsep kerangka pemecahan masalah.....	36
Gambar 3.1. Model Tahapan Penelitian.....	38
Gambar 3.2. Kerangka Model yang diusulkan.....	40
Gambar 3.3. Flowchart Model alur proses.....	42
Gambar 4.1. Kelas overlapping pada setiap fitur.....	55
Gambar 4.2. grafik nilai gain ratio sebelum di urutkan.....	56
Gambar 4.3. grafik nilai gain ratio sesudah di urutkan.....	56
Gambar 4.4. Hasil akurasi dan peringkat masing masing split.....	63
Gambar 4.5. Peningkatan kinerja AUC pada pengujian <i>cross validation</i>	65

Gambar 4.6. kurva ROC kinerja model algoritma C4.5 Split 167

Gambar 4.7. kurva ROC kinerja model algoritma C4.5 Split 267

Gambar 4.8. kurva ROC kinerja model algoritma C4.5 Split 368

Gambar 4.9. kurva ROC kinerja model algoritma C4.5 Split 469



UNIVERSITAS MIKROSKIL

DAFTAR TABEL

Tabel 2.1. Confusion matrix	25
Tabel 2.2. Nilai AUC dan keterangan.....	27
Tabel 2.3. Penelitian Terkait	31
Tabel 3.1. Atribut Dataset Mahasiswa	37
Tabel 3.2. Spesifikasi dan Atribut Dataset Mahasiswa.....	38
Tabel 3.3. Nilai AUC	43
Tabel 3.4. Spesifikasi hardware dan software.....	44
Tabel 4.1. Perhitungan awal penentuan nilai Gain rasio.....	50
Tabel 4.2. Perhitungan penentuan peringkat berdasarkan nilai gain rasio.....	56
Tabel 4.3. Hasil Kinerja dengan proses SFRM dengan klasifikasi C4.5	60
Tabel 4.4. Pengujian dengan <i>cross validation</i> dengan SFRM	62



UNIVERSITAS
MIKROSKIL