PERBANDINGAN ALGORITMA C4.5 DAN CART DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

Skripsi

oleh
DEA ALVERINA
71130036

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2017

PERBANDINGAN ALGORITMA C4.5 DAN CART DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

Skripsi

Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

DEA ALVERINA
71130036

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2017

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PERBANDINGAN ALGORITMA C4.5 DAN CART DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 20 Oktober 2017

DEA ALVERINA
71130036

HALAMAN PERSETUJUAN

Judul Skripsi	:	PERBANDINGAN ALGORITMA C4.5 DAN CART DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA
Nama Mahasiswa	:	DEA ALVERINA
N I M	:	71130036
Matakuliah	:	Skripsi (Tugas Akhir)
Kode	:	TIW276
Semester	:	Gasal
Tahun Akademik	:	2017/2018

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 20 Oktober 2017

Dosen Pembimbing I	Dosen Pembimbing II
Antonius Rachmat C., S.Kom.,M.Cs.	R. Gunawan Santosa, Drs. M.Si.

HALAMAN PENGESAHAN

PERBANDINGAN ALGORITMA C4.5 DAN CART DALAM MEMPREDIKSI KATEGORI INDEKS PRESTASI MAHASISWA

Oleh: DEA ALVERINA / 71130036

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 16 Oktober 2017

Yogyakarta, 20 Oktober 2017
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., S.Kom.,M.Cs.
2. R. Gunawan Santosa, Drs. M.Si.
3. Rosa Delima, S.Kom., M.Kom.
4. Nugroho Agus Haryono, M.Si

Dekan	Ketua Program Studi
(Budi Susanto, S.Kom., M.T.)	(Gloria Virginia, Ph.D.)

UCAPAN TERIMAKASIH

Dalam proses penulisan tugas akhir ini, penulis mendapatkan banyak bantuan baik saran, kritik, serta bimbingan dari berbagai pihak. Oleh karena itu, sudah sepantasnya penulis mengantarkan ucapan terima kasih kepada:

1. Bapak Budi Susanto, S.Kom, M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
3. Bapak Antonius Rachmat C., S.Kom.,M.Cs. selaku pembimbing I yang telah membimbing dan mengarahkan penulis dalam penyusunan tugas akhir ini.
4. Bapak R. Gunawan Santosa, Drs. M.Si. selaku pembimbing II yang telah membimbing dan mengarahkan penulis dalam penyusunan tugas akhir ini.
5. Orang tua yang selalu mendoakan, memberikan motivasi dan pengorbanan baik dari segi moril dan materi kepada penulis sehingga dapat menyelesaikan tugas akhir ini dengan baik.
6. Teman-teman mahasiswa/i Program Studi Teknik Informatika 2013 Universitas Kristen Duta Wacana yang telah memberikan motivasi untuk menyelesaikan tugas akhir ini.
7. Semua pihak yang tidak dapat disebutkan satu per satu yang telah ikut memberikan dukungan baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa masih banyak kekurangan, baik dalam penelitian ini maupun dalam penulisan laporan penelitian. Akhir kata penulis mengucapkan terima kasih kepada semua pihak yang telah membantu. Semoga tugas akhir ini dapat bermanfaat bagi kita semua dan menjadi bahan masukan bagi dunia pendidikan.

Penulis

KATA PENGANTAR

Puji syukur dan terima kasih kepada Tuhan Yang Maha Esa karena atas berkat rahmat-Nya penulis dapat menyelesaikan tugas akhir yang berjudul "Perbandingan Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa" dengan lancar.

Laporan tugas akhir ini diajukan guna melengkapi sebagai syarat dalam mencapai gelar sarjana strata satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta. Penulis menyadari meskipun telah berusaha untuk menyajikan pembahasan sebaik mungkin, namun masih terdapat kekurangan dalam tugas akhir ini. Hal ini terjadi dikarenakan masih terbatasnya kemampuan dan pengetahuan penulis. Penulis mengharapkan kritik dan saran yang membangun untuk menyempurnakan tugas akhir ini.

Dalam proses penyusunan tugas akhir ini, penulis banyak mengalami kendala, namun berkat bantuan, bimbingan, dan kerjasama dari berbagai pihak serta berkah dari Tuhan Yang Maha Esa sehingga kendala-kendala yang dihadapi dapat teratasi. Oleh karena itu, penulis mengucapkan terima kasih dan penghargaan kepada Bapak Antonius Rachmat C., S.Kom.,M.Cs. selaku pembimbing I dan Bapak R. Gunawan Santosa, Drs. M.Si. selaku pembimbing II yang telah bersedia membimbing dengan sabar dan bersedia meluangkan waktu, tenaga dan pikiran dalam memberikan bimbingan, motivasi dan arahan serta saran-saran yang sangat berharga bagi penulis dalam menyusun tugas akhir ini.

Penulis

INTISARI

Perbandingan Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa

Beberapa faktor, internal maupun external, mempengaruhi tinggi rendahnya prestasi akademis mahasiswa baru. Faktor external meliputi asal SMA, kategori SMA dan status SMA. Faktor internal meliputi kemampuan Spasial, Verbal, Numerik, dan Analogi (Santosa & Rachmat, 2016). Dengan memprediksi indeks prestasi (IP) mahasiswa semester 1, fakultas dapat menyaring calon mahasiswa baru.

Penelitian ini mencoba melakukan prediksi berbagai macam kategori IP semester satu mahasiswa Fakultas Teknologi Informasi Universitas Kristen Duta Wacana (FTI UKDW), menggunakan algoritma *decision tree* C4.5 dan CART. Penelitian ini juga mengeksplorasi berbagai parameter seperti pengkategorian atribut numerik, keseimbangan data, jumlah kategori IP, dan ketersediaan atribut yang berbeda karena perbedaan ketersediaan data antara jalur prestasi dan jalur non-prestasi.

Hasil penelitian menunjukkan bahwa algoritma CART dan C4.5 memiliki akurasi yang sama dalam memprediksi data jalur prestasi, dengan akurasi tertinggi sebesar 86.86%. Namun rata-rata akurasi dari semua skenario yang memiliki parameter yang berbeda-beda untuk C4.5 sebesar 41.48% dan CART sebesar 42.65%. Algoritma C4.5 dan CART lebih cocok untuk memprediksi IP menggunakan data jalur prestasi. Parameter untuk skenario prediksi yang paling akurat adalah ketika atribut numerik untuk data jalur non-prestasi di *threshold*, data latih tidak diseimbangkan dan kategori indeks prestasi sebanyak 2 kategori.

Kata Kunci: *Decision Tree*, C4.5, CART, Prediksi, Tabel *Crosstab*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Performa akademis mahasiswa baru di semester pertama merupakan hal yang penting untuk diperhatikan. Performa akademis yang di bawah rata-rata dapat menimbulkan berbagai masalah, di antaranya adalah efek berantai performa rendah untuk semester-semester berikutnya.

Terdapat beberapa faktor yang mempengaruhi tinggi rendahnya prestasi akademis mahasiswa baru. Faktor-faktor tersebut meliputi faktor eksternal, yang meliputi kategori asal SMA (pulau Jawa atau luar pulau Jawa), kategori SMA (SMA atau SMK), dan status SMA (Negeri atau Swasta). Selain faktor eksternal, terdapat juga faktor internal, yang meliputi kemampuan Spatial, kemampuan Verbal, kemampuan Numerik, dan kemampuan Analogi (Santosa & Rachmat, 2016). Data-data dari faktor tersebut diperoleh pada saat pendaftaran mahasiswa baru, namun tidak semua data dapat diperoleh. Hal tersebut disebabkan oleh adanya dua jalur penerimaan mahasiswa baru, yaitu jalur prestasi dan jalur non-prestasi. Penerimaan mahasiswa baru yang melalui jalur prestasi akan dilihat dari faktor eksternal dan yang melalui jalur non-prestasi akan dilihat dari faktor eksternal maupun faktor internal. Mahasiswa baru yang mendaftar melalui jalur prestasi maupun jalur non-prestasi, akan diuji kemampuan bahasa Inggrisnya yang dipetakan menjadi beberapa level, yaitu level 1, level 2, level 3, atau level ESP.

Prediksi indeks prestasi (IP) mahasiswa semester satu yang dilakukan secara dini bisa digunakan untuk menanggulangi masalah-masalah yang mungkin akan ditimbulkan oleh mahasiswa di kemudian hari. Fakultas bisa melacak mahasiswa-mahasiswa yang berpotensi memiliki IP semester satu yang rendah untuk kemudian dilakukan pembimbingan lebih lanjut terhadap mahasiswa

tersebut. Selain itu, prediksi juga bisa manfaatkan untuk melakukan penyaringan calon mahasiswa baru.

IP merupakan alat ukur prestasi mahasiswa di bidang akademis yang didapat dari nilai rata-rata semua matakuliah yang diambil mahasiswa selama satu semester. Sedangkan masukan dari prediksi tersebut adalah faktor-faktor internal dan eksternal mahasiswa.

Pada penelitian ini akan dilakukan prediksi berbagai kategori IP mahasiswa baru Fakultas Teknologi Informasi Universitas Kristen Duta Wacana (FTI UKDW) yang akan menggunakan data mahasiswa dari tahun 2008 sampai 2015 menggunakan algoritma C4.5 dan algoritma CART. Data tersebut diperoleh dari penelitian yang telah dilaksanakan oleh Santosa & Rachmat (2016). Penelitian ini juga mengeksplorasi berbagai parameter seperti pengkategorian atribut numerik, keseimbangan data, jumlah kategori IP, dan ketersediaan atribut yang berbeda, yang disebabkan oleh perbedaan ketersediaan data antara jalur prestasi dan jalur non-prestasi.

1.2. Perumusan Masalah

Berikut adalah permasalahan yang akan diselesaikan dalam penelitian ini:

(a) Apakah algoritma C4.5 dapat diterapkan untuk memprediksi kategori IP semester satu untuk mahasiswa baru FTI UKDW tahun 2015?
(b) Apakah algoritma CART dapat diterapkan untuk memprediksi kategori IP semester satu untuk mahasiswa baru FTI UKDW tahun 2015?
(c) Seberapa akurat algoritma C4.5 dan CART dengan berbagai jumlah kategori IP yang berbeda, keseimbangan data, dan pengkategorian atribut numerik dalam memprediksi kategori IP semester satu mahasiswa baru FTI UKDW tahun 2015?

1.3. Batasan Penelitian

Beberapa batasan masalah dalam penelitian ini sebagai berikut:

(a) IP yang diteliti adalah indeks prestasi semester satu untuk mahasiswa baru FTI UKDW tahun 2015.

(b) Data yang akan dijadikan sebagai sampel data latih adalah data mahasiswa baru FTI UKDW tahun 2008 sampai 2014.

(c) Data yang akan dijadikan untuk pengujian adalah data mahasiswa baru FTI UKDW tahun 2015.

(d) Penelitian ini akan dilakukan pada 2 tipe kelompok data mahasiswa baru FTI UKDW yang berbeda, yaitu data jalur prestasi dan jalur non-prestasi. Jalur prestasi memiliki 4 atribut, yaitu kategori (SMA atau SMK), status SMA (Negeri atau Swasta), kategori asal SMA (Jawa atau Luar Jawa), dan level bahasa Inggris (Level 1, 2, 3, atau ESP). Jalur non-prestasi memiliki 8 atribut, yaitu kategori (SMA atau SMK), status SMA (Negeri atau Swasta), kategori asal SMA (Jawa atau Luar Jawa), level bahasa Inggris (Level 1, 2, 3, atau ESP), kemampuan Spasial, kemampuan Verbal, kemampuan Numerik, dan kemampuan Analogi.

(e) Pada jalur non-prestasi akan dilakukan 2 pengujian yaitu menggunakan data atribut numerik yang di-*binning* dan data atribut numerik yang di-*threshold*. Atribut numerik yang di-*binning* adalah nilai numerik, nilai verbal, nilai spasial, dan nilai analogi dengan jumlah kategori sebanyak 5 kategori.

(f) Perbandingan akurasi algoritma C4.5 dan CART dengan berbagai jumlah kategori IP dihitung dari 2-5 kategori yang berbeda dalam memprediksi kategori IP.

(g) Data yang tidak dapat diprediksi karena memiliki *missing value*, tidak dihitung sebagai pembagi dalam perhitungan akurasi algoritma.

(h) Algoritma C4.5 akan digunakan di dalam sistem untuk mengkategorikan IP semester satu mahasiswa baru FTI UKDW tahun 2015.

(i) Algoritma CART akan digunakan di dalam sistem untuk mengkategorikan IP semester satu mahasiswa baru FTI UKDW tahun 2015.

(j) *Software* yang digunakan untuk membuat program pada penelitian ini adalah **Node.js**, **Electron** untuk tampilan antarmuka, dan **NoSQL** sebagai *database*.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menerapkan algoritma C4.5 dan CART dalam memprediksi kategori IP semester satu mahasiswa baru FTI UKDW tahun 2015 dan mengetahui perbandingan akurasi algoritma C4.5 dan CART dengan berbagai jumlah kategori IP yang berbeda, keseimbangan data, dan pengkategorian atribut numerik dalam memprediksi kategori IP semester satu mahasiswa baru FTI UKDW tahun 2015.

1.5. Metode Penelitian

Metodologi yang dilakukan dalam penelitian ini adalah:

(a) Studi Pustaka

Studi pustaka dilakukan dengan mencari dan mempelajari sumber-sumber pustaka yang berkaitan dengan teori *Data Mining*, algoritma C4.5, algoritma CART, dan metode tabulasi silang (*crosstab*).

(b) Pengumpulan Data

Penelitian ini dilakukan dengan mengambil data dari penelitian yang telah dilakukan oleh Santosa & Rachmat (2016). Jumlah data berdasarkan jumlah mahasiswa FTI dari tahun 2008 sampai dengan tahun 2015, yaitu 2.017 data mahasiswa. Data mahasiswa tersebut akan dikategorikan berdasarkan jalur penerimaan, yaitu jalur prestasi dan jalur non-prestasi. Jalur prestasi memiliki atribut kategori (SMU atau SMK), status (negeri

atau swasta), lokasi (jawa atau luar jawa), dan level (1, 2, 3, atau ESP). Sedangkan jalur non-prestasi memiliki atribut kategori (SMU atau SMK), status (negeri atau swasta), lokasi (jawa atau luar jawa), level (1, 2, 3, atau ESP), nilai numerik, nilai verbal, nilai spasial, dan nilai analogi.

(c) Perancangan Sistem

Pada tahap ini akan dirancang program untuk melakukan prediksi kategori IP dengan menggunakan algoritma C4.5 dan CART.

(d) Pembangunan Sistem

Pembangunan sistem prediksi kategori IP semester satu mahasiswa baru FTI UKDW dibuat dengan menggunakan *software* **Node.js**, **Electron**, dan **NoSQL**.

(e) Pengujian dan Analisis

Metode yang digunakan untuk mengukur keakuratan hasil prediksi adalah tabel tabulasi silang (*crosstab*). Tabel *crosstab* akan digunakan untuk menghitung akurasi dari algoritma C4.5 dan algoritma CART.

1.6. Sistematika Penulisan

Pada bab 1, berisi pendahuluan yang terdiri dari lima bagian. Bagian pertama berisi latar belakang masalah dari penelitian tentang prediksi IP mahasiswa baru berdasarkan data-data mahasiswa. Pada bagian kedua berisi perumusan masalah yang membahas tentang pertanyaan untuk menyelesaikan masalah penelitian. Bagian ketiga berisikan batasan masalah yang berkaitan dengan jangkauan dan memperjelas batasan di dalam penelitian ini. Pada bagian keempat berisikan tujuan penelitian yang dilakukan. Kemudian bagian kelima berisikan tentang metode penelitian yang digunakan yaitu mengambil data mahasiswa baru, perancangan sistem, pembuatan sistem, dan membandingkan akurasi algoritma

C4.5 dan CART menggunakan tabel *crosstab*. Bagian terakhir berisikan tentang sistematika penulisan yang berisikan penjelasan pada tiap bab di dalam skripsi.

Pada bab 2, ditulis tinjauan pustaka dan landasan teori. Tinjauan pustaka berisi dari beberapa teori yang didapatkan dari berbagai sumber pustaka yang dapat mendukung penelitian. Landasan teori memuat penjelasan tentang konsep-konsep dan prinsip utama yang diperlukan untuk memecahkan masalah riset.

Pada bab 3, ditulis analisis dan perancangan sistem. Pada bab ini, ditulis analisis dari teori-teori yang digunakan dan bagaimana menerapkan teori-teori tersebut pada penelitian terkait. Bab ini memuat bahan/materi dari penelitian, variabel yang digunakan, data yang dikumpulkan, dan cara perancangan.

Pada bab 4, ditulis bab implementasi dan analisis sistem. Bab ini memuat hasil riset dan pembahasan dari riset yang ditulis secara terpadu. Hasil riset dan pembahasan meliputi penjelasan mengenai sistem yang dihasilkan dan hasil dari perbandingan akurasi algoritma C4.5 dan CART dengan menggunakan metode *crosstab* yang dilakukan terhadap sistem tersebut.

Pada bab 5, ditulis bab kesimpulan dan saran. Bab ini memuat pernyataan singkat dan tepat untuk menyimpulkan hasil analisis pada bab 4 dan juga saran-saran untuk kegiatan riset lain yang akan dilakukan. Saran-saran ini memuat aktifitas-aktifitas atau metode dan teknik pengembangan yang dapat memperbaiki kekurangan dari penelitian ini.

BAB 5
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian yang telah dilakukan oleh penulis, terdapat beberapa kesimpulan yang diperoleh yaitu sebagai berikut:

a) Algoritma C4.5 dan CART dapat memberikan prediksi kategori indeks prestasi semester 1 mahasiswa baru FTI UKDW.
b) Prediksi tertinggi jalur prestasi yang menggunakan data latih sebanyak 613 data dan data uji sebanyak 193 data dicapai oleh kedua algoritma, yaitu C4.5 dan CART dengan akurasi sebesar 86.86%. Namun, rata-rata akurasi dari semua skenario yang memiliki parameter yang berbeda-beda untuk C4.5 sebesar 50.11% dan CART sebesar 50.35%.
c) Prediksi tertinggi jalur non-prestasi yang menggunakan data latih sebanyak 1.150 data dan data uji sebanyak 61 data, algoritma CART, dengan akurasi sebesar 63.16%, lebih baik daripada algoritma C4.5, dengan akurasi sebesar 61.54%. Namun, rata-rata akurasi dari semua skenario yang memiliki parameter yang berbeda-beda untuk C4.5 sebesar 36.78% dan CART sebesar 38.04%.
d) Dengan adanya dua macam jalur penerimaan mahasiswa baru FTI UKDW, algoritma C4.5 dan CART lebih cocok digunakan untuk jenis jalur penerimaan melalui jalur prestasi daripada jalur non-prestasi.
e) Algoritma C4.5 dimana atribut numerik di *binning* memiliki rata-rata akurasi yang lebih baik, yaitu sebesar 35.83% dibandingkan dengan atribut numerik yang di *threshold*, yaitu sebesar 34.87%. Sedangkan algoritma CART dimana atribut numerik di *threshold* memiliki rata-rata akurasi yang lebih baik, yaitu sebesar 39.66% dibandingkan dengan atribut numerik yang di *binning*, yaitu sebesar 35.09%

f) Data yang tidak seimbang memiliki rata-rata akurasi yang baik, yaitu C4.5 memiliki rata-rata akurasi sebesar 47.23% dan CART memiliki rata-rata akurasi sebesar 50.32%, dibandingkan dengan data yang seimbang.

g) Berdasarkan banyaknya kategori indeks prestasi mahasiswa, kategori indeks prestasi sebanyak 2 kategori memiliki rata-rata akurasi yang lebih tinggi, yaitu rata-rata akurasi algoritma C4.5 sebesar 60.29% dan algoritma CART sebesar 62.98%.

h) Ketika nilai akurasi algoritma C4.5 dan CART dibandingkan dengan nilai akurasi metode Regresi Logistik, algoritma K-Nearest Neighbor, dan algorima Naïve Bayes, algoritma yang paling baik untuk jalur prestasi adalah algoritma C4.5 dan CART dengan nilai akurasi kedua algoritma sebesar 86.86% dan algoritma yang paling baik untuk jalur non-prestasi adalah algoritma CART dengan nilai akurasi sebesar 63.16%.

5.2 Saran

Permasalahan pada penelitian ini mengenai *missing value* pada saat pengujian dapat diatasi dengan cara *discard*. Dari penelitian milik Gavankar & Sawarkar (2015), selain mengatasi *missing value* dengan cara *discard*, ada beberapa pendekatan lain untuk mengatasi permasalahan *missing value* tersebut, yaitu *Imputation* (prediksi menggunakan estimasi *missing value* atau distribusinya), C4.5 *Strategy* (prediksi menggunakan distribusi *missing value* dan kombinasi model prediksi secara probablisitik), *Null Strategy* (*null* dianggap sebagai *value*), dan *Lazy Decision Tree* (*tree* dibuat hanya menggunakan *value* yang diketahui). Saran untuk perkembangan penelitian dapat meneliti akurasi dari berbagai pendekatan tersebut.

DAFTAR PUSTAKA

Andrian, Y., & Wayahdi, M. R. (2014). Analisis Kinerja Data Mining Algoritma C4.5 Dalam Menentukan Tingkat Minat Siswa yang Mendaftar di Kampus ABC. *https://www.academia.edu/9203748/Analisis_Kinerja_Data_Mining_Algoritma_C4.5_Dalam_Menentukan_Tingkat_Minat_Siswa_yang_Mendaftar_di_Kampus_ABC*.

Andriani, A. (2012). Penerapan Algoritma C4.5 pada Program Klasifikasi Mahasiswa Droupout. *Seminar Nasional Matematika 2012*, 139-147.

Alpaydın, E. (2014). *Introduction to Machine Learning* (2nd ed.). London: The MIT Press.

Gavankar, S., & Sawarkar, S. (2015). Decision Tree: Review of Techniques for Missing Values at Training, Testing and Compatibility. *IEEE Computer Society*, 122-126. Doi:10.1109/AIMS.2015.29.

Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann.

Indratno, I., & Irwinsyah, R. (1998). Aplikasi Analisis Tabulasi Silang (Crosstab) Dalam Perencanaan Wilayah dan Kota. *Jurnal PWK, 9*, 48-59.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. A John Wiley & Sons, Inc.

Lakshmi, T., Martin, A., Begum, R., & Venkatesan, V. (2013). An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data. *International Journal of Modern Education and Computer Science, 5*(5), 18-27. doi:10.5815/ijmecs.2013.05.03.

Margasari, A. (2014). Penerapan Metode CART (Classification and Regression Trees) dan Analisis Regresi Logistik Biner Pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya. *Jurnal Matematika, F.MIPA, Universitas Brawijaya*, 257-260.

Rahmayuni, I. (2014). Perbandingan Performansi Algoritma C4.5 dan CART Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Informatika Politeknik Negeri Padang. *Jurnal TEKNOIF, 2*(1), 40-46.

Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis* (1st ed.). Yogyakarta: Graha Ilmu.

Santosa, R. G., & Rachmat, A. (2016). *Regresi Logistik Untuk Prediksi Kategori IP Mahasiswa Fakultas Teknologi Informasi UKDW*. Laporan tidak dipublikasi.

Sari, V. H. A., Santosa, R. G., & Rachmat, A. (2016). *Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier dalam Memprediksi Kategori Indeks Prestasi Mahasiswa*. Makalah tidak dipublikasi.

Untari, D. (2014). Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5. *http://eprints.dinus.ac.id/13181/*.