

**PENERAPAN TEKNIK *DATA MINING* ALGORITMA C4.5 UNTUK
PREDIKSI KESESUAIAN TINGKAT PENDIDIKAN DENGAN BIDANG
PEKERJAAN PADA ALUMNI UNIVERSITAS SRIWIJAYA**

SKRIPSI

**Sebagai Salah Satu Syarat Memperoleh Gelar
Sarjana Sains Bidang Studi Matematika**



Oleh

**BUDI SANTOSO
NIM 08011281520084**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
JANUARI 2020**

LEMBAR PENGESAHAN

**PENERAPAN TEKNIK *DATA MINING* ALGORITMA C4.5 UNTUK
PREDIKSI KESESUAIAN TINGKAT PENDIDIKAN DENGAN BIDANG
PEKERJAAN PADA ALUMNI UNIVERSITAS SRIWIJAYA**

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh Gelar

Sarjana Matematika


Oleh

BUDI SANTOSO

NIM. 08011281520084

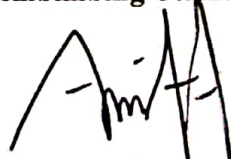
Indralaya, Januari 2020

Pembimbing Pembantu



Drs. Ali Amran, M.T
NIP. 19661213 199402 1 001

Pembimbing Utama



Anita Desiani, M.Kom
NIP. 19771211 200312 2 002

Mengetahui

Ketua Jurusan Matematika



Drs. Sugandi Yuhdin, M.M
NIP. 19590727 198603 1 003

LEMBAR PERSEMBAHAN

MOTTO :

“Dan Aku tidak menciptakan jin dan manusia melainkan agar mereka beribadah kepada-Ku”.

(Q.S. Al-Dzariat : 56)

“Apa yang melewatkanmu tidak akan pernah menjadi takdirmu, dan apa yang di takdirkan untukmu tidak akan pernah melewatkanmu”.

(Umar ibn Khattab)

Skripsi ini Kupersembahkan kepada:

- ❖ **Allah *Subhanahu wa ta'ala***
- ❖ **Agamaku**
- ❖ **Ayah dan Ibuku**
- ❖ **Semua Keluargaku**
- ❖ **Semua Guru dan Dosen**
- ❖ **Teman-temanku**
- ❖ **Almamaterku**

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah penulis ucapkan syukur kepada Allah SWT karena berkat rahmat, karunia, kasih sayang dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Penerapan teknik *Data mining* algoritma C4.5 untuk prediksi kesesuaian tingkat pendidikan dengan bidang pekerjaan pada alumni universitas sriwijaya"**. Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Bidang Studi Matematika di Fakultas Matematika dan Ilmu Pengeratahuan Alam Universitas Sriwijaya.

Dengan segala hormat dan kerendahan hati penulis mengucapkan terima kasih yang sebesar-besarnya kepada Ayahanda Ahmat Doro Siregar dan Ibunda Khoiria Ritonga serta Saudara-saudariku Ainan Rawani Siregar, Akmal Syarif Siregar, Milfa Aisyaroh Siregar, dan Fajar Romadhonsyah Siregar atas segala kasih sayang, perhatian, dukungan, serta doanya. Penulisan skripsi ini tidak lepas bantuan sebagai pihak baik secara langsung maupun tidak langsung, untuk itu penulis menyampaikan terimakasih sebesar-besarnya kepada :

1. Bapak Drs. Sugandi Yahdin, M.M, selaku Ketua Jurusan Matematika dan Ibu Des Alwine Zayanti, M.Si selaku Sekertaris Jurusan Matematika atas bimbingan yang telah diberikan selama penulis belajar di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

2. Ibu Anita Desiani, M.Kom selaku Pembimbing Utama dan Bapak Drs. Ali Amran, M.T selaku Pembimbing Kedua. Terima kasih telah bersedia menyediakan waktu, pikiran, nasehat, motivasi dan saran serta kesabaran memberikan arahan dan bimbingan terbaik penulis dalam penyusunan skripsi ini.
3. Bapak Drs. Sugandi Yahdin, M.M, Ibu Des Alwine Zayanty, M.Si, dan Ibu Novi Rustiana Dewi, M.Si selaku Dosen Penguji yang telah bersedia meluangkan waktu dalam memberikan tanggapan, kritik dan saran yang bermanfaat dalam perbaikan dan penyelesaian skripsi ini, serta Ibu Sri Indra Maiyanti, M.Si selaku Dosen Pembimbing Akademik yang telah memberikan pengarahan akademik kepada penulis di Jurusan Matematika FMIPA Universitas Sriwijaya
4. Seluruh Dosen di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya yang telah memberkan ilmu yang bermanfaat, nasehat, serta bimbingan selama penulis menjalani perkuliahan.
5. Bapak Irwansyah selaku admin dan Ibu Hamidah selaku pegawai tata usaha jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah membantu penulis selama masa perkuliahan.
6. Pihak *Career Development Center* (CDC) Universitas Sriwijaya terima kasih telah memberikan kesempatan kepada penulis memberikan data kuisinoer penelitian.
7. Sahabat-sahabat penulis di bangku perkuliahan dan juga lembaga organisasi baik skala Jurusan, Fakultas, dan Universitas.
8. Semua Pihak yang tidak bisa disebutkan satu persatu yang telah memberikan bantuan, do'a, serta semangat dalam menyelesaikan skripsi ini.

Penulis mengharapkan semoga skripsi ini dapat menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Indralaya, Januari 2020

Penulis

THE IMPLEMENTATION OF C4.5 DATA MINING TECHNIQUES FOR PREDICTION OF ACCORDANCE OF EDUCATION LEVELS WITH THE FIELD OF WORK AT THE ALUMNI OF THE SRIWIJAYA UNIVERSITY

By:

Budi Santoso
08011281520084

ABSTRACT

Universitas Sriwijaya has a talent and career development center (CDC) unit for alumni. The CDC website provided a questionnaire form for alumni to fill out. The features on the questionnaire had relevance regarding the level of education of alumni with their job such as lectures, research projects, internships, practicum, field work, discussions, time to get a job, fields of science, outside the field of science, English, internet, computers, and the relationship between study program and work. Based on data on the CDC from 2013 to 2015, there were 1143 alumni of Sriwijaya University who filled out the questionnaire form completely, but only 1019 alumni who had jobs. The selection of attributes was done by calculating using the C4.5 algorithm . This study used the k-fold cross-validation confusion matrix method to see the success rate of the C4.5 algorithm. The result of accuracy in the testing process using confusion matrix was 92% for the decision tree method using the C4.5 algorithm. The best features selected were 9 features, namely the level of learning in the discussion, the level of competence outside the field of science, the level of learning in field work, the level of competency in English skills, the level of learning in internships, the level of learning in practicum, the level of competence in the field of science, the level of learning in research projects, as well as the time to get a job. Based on the accuracy obtained, it can be concluded that the C4.5 algorithm worked very well in the feature selection process and improved accuracy.

Keywords: Data mining, Classification, Decession tree, C4.5 Algorithm, education level.

PENERAPAN TEKNIK *DATA MINING* ALGORITMA C4.5 UNTUK PREDIKSI KESESUAIAN TINGKAT PENDIDIKAN DENGAN BIDANG PEKERJAAN PADA ALUMNI UNIVERSITAS SRIWIJAYA

Oleh :

Budi Santoso
08011281520084

ABSTRAK

Universitas Sriwijaya memiliki pusat pengembangan bakat dan karir bagi alumni yang disebut dengan CDC (*Career Development Centre*). CDC menyediakan kuisisioner bagi alumni, di mana fitur-fitur pada formulir kuisisioner tersebut memiliki relevansi tentang tingkat pendidikan alumni dengan pekerjaan mereka seperti perkuliahan, proyek riset, magang, praktikum, kerja lapangan, diskusi, waktu mendapatkan pekerjaan, bidang ilmu, diluar bidang ilmu, bahasa inggris, internet, komputer, dan hubungan program studi dengan pekerjaan. Berdasarkan data pada CDC tahun 2013 sampai 2015 terdapat 1143 alumni Universitas Sriwijaya yang mengisi formulir kuisisioner dengan lengkap, tetapi alumni yang memiliki pekerjaan sebanyak 1019 alumni. Pemilihan atribut dilakukan dengan melakukan penghitungan menggunakan algoritma C4.5 sehingga mendapatkan atribut dengan nilai *gain* tertinggi untuk dijadikan akar pada pembentukan pohon keputusan. Hasil dari pencarian nilai-nilai *gain* tersebut membentuk pohon keputusan untuk memperoleh atribut yang berpengaruh pada penelitian ini. Penelitian ini menggunakan *k-fold cross-validation* dengan jumlah 10 *fold* metode *confusion matrix* untuk melihat tingkat keberhasilan dari algoritma C4.5. Hasil akurasi pada proses uji menggunakan *confusion matrix* ialah 92%. Fitur terbaik yang terpilih sebanyak 9 fitur yaitu tingkat pembelajaran pada diskusi, tingkat kompetensi luar bidang ilmu, tingkat pembelajaran pada kerja lapangan, tingkat kompetensi kemampuan bahasa inggris, tingkat pembelajaran pada magang, tingkat pembelajaran pada praktikum, tingkat kompetensi bidang ilmu, tingkat pembelajaran dalam proyek riset, serta waktu mendapatkan pekerjaan setelah lulus. Berdasarkan akurasi yang diperoleh dapat disimpulkan bahwa algoritma C4.5 bekerja sangat baik dalam proses pemilihan fitur dan meningkatkan akurasi.

Kata Kunci: *Data mining*, Klasifikasi, Pohon Keputusan, Algoritma C4.5, tingkat Pendidikan

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Pembatasan Masalah.....	3
1.4 Tujuan.....	4
1.5 Manfaat.....	4
BAB II.....	6
TINJAUAN PUSTAKA.....	6

2.1 Data Mining.....	6
2.2 Klasifikasi	8
2.3 Pohon Keputusan	9
2.4 Algoritma C4.5	9
2.5 Teknik Pengujian <i>K-fold cross validation</i>	10
2.6 Confusion matrix	10
BAB III.....	13
METODOLOGI PENELITIAN.....	13
3.1 Tempat.....	13
3.2 Waktu.....	13
3.3 Alat.....	13
3.4 Metode Penelitian.....	13
BAB IV.....	18
HASIL DAN PEMBAHASAN.....	18
BAB V.....	83
KESIMPULAN DAN SARAN.....	83
5.1 Kesimpulan.....	83
5.2 Saran.....	83
DAFTAR PUSTAKA.....	84

DAFTAR TABEL

	Halaman
Tabel 2.1. Model Confusion matrix	10
Tabel 2.2. Kategori nilai akurasi	11
Tabel 4.1. Data atribut yang berhubungan dengan kuisisioner.....	19
Tabel 4.2. Data nilai dari atribut yang telah diisi oleh alumni.....	21
Tabel 4.3. Data contoh untuk perhitungan manual.....	22
Tabel 4.4. Nilai entropy dan gain dari seluruh atribut.....	24
Tabel 4.5. Nilai entropy dan gain dari A1 bernilai 1.....	28
Tabel 4.6. Nilai entropy dan gain dari A1 bernilai 2	31
Tabel 4.7. Nilai entropy dan gain dari A2 bernilai 1	35
Tabel 4.8. Nilai entropy dan gain dari A2 bernilai 3.....	38
Tabel 4.9. Prediksi atribut berpengaruh.....	49
Tabel 4.10. Perbandingan hasil prediksi dengan atribut asli.....	50
Tabel 4.11. Confusion matrix prediksi atribut.....	51
Tabel 4.12. Model confusion matrix data penelitian.....	83
Tabel 4.13. Hasil kinerja model confusion matrix	86

DAFTAR GAMBAR

	Halaman
Gambar 3.1. Flowchart perhitungan Algoritma C4.5	16
Gambar 4.1. Akar pohon keputusan perhitungan manual	28
Gambar 4.2. cabang pohon keputusan perhitungan manual	34
Gambar 4.3. Pohon keputusan Algoritma C4.5 perhitungan manual.....	42

BAB I PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan salah satu dari lima negara dengan jumlah penduduk terbanyak di dunia. Berdasarkan data yang dipublikasikan oleh Badan Pusat Statistik (BPS, 2016) jumlah penduduk Indonesia sebanyak 255.461.700 jiwa pada tahun 2015. Tingginya jumlah penduduk ini tidak dibarengi dengan kondisi lapangan pekerjaan yang masih kurang di Indonesia sehingga kebanyakan alumni lulusan perguruan tinggi bekerja tidak sesuai dengan tingkat pendidikannya. Peran pendidikan sangatlah penting bagi peningkatan kualitas masyarakat Indonesia karena akan membantu menumbuhkan produktivitas kerja bagi bangsa Indonesia.

Pada tahun 2013 Universitas Sriwijaya membentuk lembaga *Career Development Center* (CDC) yang berfungsi sebagai pusat pengembangan karir Universitas Sriwijaya dalam menyikapi pencapaian kelulusan alumni Universitas Sriwijaya yang diisi dalam bentuk boring (Gofar & Susmanto, 2018). Menurut (Gofar & Susmanto, 2018) *Tracer Study* yang dilakukan oleh CDC Universitas Sriwijaya pada tahun 2017 terdapat 5839 alumni yang terlacak secara online dan hanya 577 alumni yang mengisi form tracer yang disiapkan oleh CDC Universitas Sriwijaya, dimana terdapat 119 alumni yang menyatakan bahwa tidak ada kesesuaian bidang ilmu saat pendidikan dengan pekerjaan saat ini. Data ini menunjukkan bahwa masih banyak alumni dari Universitas Sriwijaya yang bekerja akan tetapi tidak sesuai

dengan bidang ilmu saat pendidikan di perguruan tinggi, sehingga perlu dilakukan program pengembangan kualitas untuk calon alumni agar terserap di pasar kerja secara maksimal.

Ada banyak cara yang dapat dilakukan untuk menentukan kesesuaian pendidikan dengan bidang pekerjaan pada Alumni Perguruan Tinggi, salah satunya yaitu dengan *data mining*. *Data mining* ialah suatu proses untuk mencari pola atau informasi data dalam cakupan besar dengan menggunakan teknik, metode, atau algoritma tertentu. Beberapa teknik dalam data mining antara lain yaitu *association rule mining*, *clustering*, klasifikasi, *neural network*, dan *genetic algorithm*. Salah satu teknik dalam data mining ialah pengklasifikasian. Klasifikasi adalah proses menemukan konsep untuk memperkirakan data dari suatu objek yang belum diketahui labelnya (Febie Elfaladonna, 2019). Salah satu metode dalam menyelesaikan kasus klasifikasi ialah pohon keputusan. Pohon keputusan merupakan struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi kumpulan yang kecil dalam sebuah aturan keputusan, dimana data dalam pohon keputusan akan berbentuk atribut yang dinyatakan dalam suatu parameter sebagai kriteria dalam pembentukan pohon (Mardi, 2017). Salah satu algoritma yang dapat dilakukan dalam membuat pohon keputusan ialah Algoritma C4.5, dimana algoritma C4.5 ini memiliki kelebihan dalam memproses generalisasi data dari setiap aturan dengan cara membuang/mengabaikan kondisi awal, sehingga akurasi klasifikasi akan meningkat (Budiman & Parandani, 2018). Salah satu contoh penggunaan algoritma C4.5 ialah penentuan pola masa studi mahasiswa Jurusan Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya dengan kategori cukup baik

menghasilkan akurasi keberhasilan prediksi sebesar 84,2857% untuk proses latih *K-fold cross-validation* (Yahdin et al., 2019).

Penelitian ini dilakukan dengan menggunakan penerapan Algoritma C4.5 untuk memprediksi kesesuaian tingkat pendidikan dengan bidang pekerjaan alumni Universitas Sriwijaya tahun 2013, 2014, 2015 melalui metode pengujian *K-fold cross-validation* dengan mencari akurasi, presisi, dan *recall* yang akan mengetahui tingkat akurasi pada atribut yang berpengaruh sehingga nantinya akan dijadikan saran dan masukan untuk Universitas Sriwijaya menyiapkan fasilitas atau materi dalam meningkatkan mutu kualitas mahasiswa Universitas Sriwijaya yang telah dinyatakan lulus dari Universitas Sriwijaya.

1.2 Perumusan Masalah

Bagaimana memprediksi kesesuaian tingkat pendidikan dengan bidang pekerjaan pada alumni Universitas Sriwijaya.

1.3 Pembatasan Masalah

Penelitian ini dibatasi oleh beberapa atribut yang tersedia pada data *tracer* alumni yang berhubungan dengan bidang pekerjaannya saat ini dan hanya mengambil data untuk alumni Universitas Sriwijaya tahun 2013, 2014, dan 2015. Adapun atribut yang digunakan ialah :

- a. Tingkat pembelajaran pada perkuliahan

- b. Tingkat kompetensi dalam proyek riset
- c. Tingkat pembelajaran pada magang
- d. Tingkat pembelajaran pada praktikum
- e. Tingkat pembelajaran pada kerja lapangan
- f. Tingkat pembelajaran pada diskusi
- g. Waktu mendapat kerja
- h. Tingkat kompetensi bidang ilmu
- i. Tingkat kompetensi luar bidang ilmu
- j. Tingkat kompetensi bahasa inggris
- k. Tingkat kompetensi kemampuan internet
- l. Tingkat kompetensi komputer
- m. Hubungan program studi dengan pekerjaan
- n. Tingkat pendidikan yang sesuai dengan bidang pekerjaan

1.4 Tujuan

Memperoleh atribut yang berpengaruh pada kesesuaian tingkat pendidikan dengan bidang pekerjaan alumni Universitas Sriwijaya dengan melihat nilai akurasi prediksi yang diperoleh menggunakan Algoritma C4.5

1.5 Manfaat

Manfaat dari hasil penelitian ini adalah :

1. Mengetahui atribut mana yang bakal mempengaruhi pekerjaan alumni Universitas Sriwijaya dengan menghubungkan bidang minat pendidikannya.
2. Menyiapkan program pendidikan kampus dengan menyesuaikan atribut yang berpengaruh pada pekerjaan agar alumni Universitas Sriwijaya memperoleh pekerjaan sesuai dengan jenjang dan bidang pendidikannya.

DAFTAR PUSTAKA

- Budiman, A. S., & Parandani, X. A. (2018). Uji Akurasi Klasifikasi Dan Validasi Data Pada Penggunaan Metode Membership Function Dan Algoritma C4.5 Dalam Penilaian Penerima Beasiswa. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 565–578. <https://doi.org/10.24176/simet.v9i1.2021>
- Febie Elfaladonna, A. R. (2019). Analisa Metode Classification-Decission Tree Dan Algoritma. *Science And Information Technology*, 2(1), 10–17.
- Gofar, N., & Susmanto, P. (2018). Tracer Study 2018 (Lulusan Tahun 2016). *Noer Fikri*, KUSNO HARIANTO, HENY PRATIWI, Y. S. (2019). Sistem.
- Han, J., & Kamber, M. (2012). *DataMiningConceptsAndTechniques*.
- Mahmood Ali, M., S. Qaseem, M., Rajamani, L., & A, G. (2013). Extracting Useful Rules Through Improved Decision Tree Induction Using Information Entropy. *International Journal of Information Sciences and Techniques*, 3(1), 27–41. <https://doi.org/10.5121/ijist.2013.3103>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan. *Seminar Nasional Sistem Informasi Indonesia, November*, 2–3.
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>

- Rismayanti. (2016). Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan. *Media Infotama*, 12(2), 116–120.
- Rosandy, T. (2016). PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus: KSPPS / BMT AL-FADHILA. *Jurnal Teknologi Informasi Magister Darmajaya*, 2(01), 52–62.
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Sugara, B., Widyatmoko, D., Prakoso, B. S., & Saputro, D. M. (2018). Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak. *Seminar Nasional Teknologi Informasi Dan Komunikasi (SENTIKA)*, 2018(Sentika), 87–96.
- Yahdin, S., Desiani, A., Amran, A., Rodiah, D., & Solehan. (2019). Pattern recognition for study period of student in Mathematics Department with C4.5 algorithm data mining technique at the Faculty of Mathematics and Natural Science Universitas Sriwijaya. *Journal of Physics: Conference Series*, 1282(1), 0–6. <https://doi.org/10.1088/1742-6596/1282/1/012014>