

**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI
DATA KANDIDAT *OUTSOURCING* MENGGUNAKAN
METODE DECISION TREE C4.5**

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

Ahmad Ardhy Ansyah

20083010008

**PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
SURABAYA
2024**

LEMBAR PENGESAHAN

IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI DATA KANDIDAT *OUTSOURCING* MENGGUNAKAN METODE DECISION TREE C4.5

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : Selasa, 14 Mei 2024

Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya

Oleh :

AHMAD ARDHY ANSYAH

NPM. 20083010008

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

Dr. I Gede Susrama Mas Diyasa, ST., MT.
NIP. 197006192021211009

Pembimbing 1

Tresna Maulana Fahrudin, S.ST., MT
NIP. 199305012022031007

Fakultas Ilmu Komputer
Dekan,

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 196811261994032001

Penguji 2

Aviolla Terza Damalliana, S.SI., M.Stat.
NIP. 199408022022032015

Pembimbing 2

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.
NIP. 198012052005011002

Mengetahui,

Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.
NIP. 198012052005011002

Surabaya, Mei, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Ahmad Ardhy Ansyah
NPM : 20083010008
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI DATA KANDIDAT *OUTSOURCING* MENGGUNAKAN METODE DECISION TREE C4.5

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 03 Mei 2024

Hormat Saya



Ahmad Ardhy Ansyah
NPM. 20083010008

ABSTRAK

IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI DATA KANDIDAT *OUTSOURCING* MENGGUNAKAN METODE DECISION TREE C4.5

Nama Mahasiswa / NPM : Ahmad Ardhy Ansyah / 20083010008
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Tresna Maulana Fahrudin, S.ST., MT
Dosen Pembimbing 2 : Dr. Eng. Ir. Dwi Arman Prasetya.,ST., MT., IPU.

TOG Indonesia adalah perusahaan yang bergerak sebagai penyedia jasa tenaga kerja IT profesional. Untuk menyediakan tenaga kerja, perusahaan terus membuka lowongan agar memenuhi kebutuhan mitra. Saat ini perusahaan masih melakukan seleksi kandidat dengan metode konvensional, sehingga mengakibatkan pemilihan kandidat yang memakan waktu lama dan keputusan yang tidak konsisten, dan data riwayat kandidat tidak tersimpan secara terstruktur karena berkas tidak disimpan dalam penyimpanan. Oleh karena itu, penelitian ini akan membuat sebuah model yang dapat digunakan untuk klasifikasi data kandidat, model klasifikasi yang dipilih yaitu Machine Learning seperti Decision Tree C4.5, yang secara otomatis mengklasifikasikan berdasarkan profil dan kompetensi kandidat. Setelah diimplementasikan hasil model ini memiliki akurasi dan presisi tinggi, terutama pada validasi menggunakan *holdout* atau *percentage split* dengan proporsi data *train* dan *test* 70:30 yang mencapai akurasi terbaik 99% dan presisi 90%, lalu variabel yang paling menentukan proses prediksi rekomendasi kandidat yaitu variabel pendidikan kandidat, karena menjadi *root node* pohon klasifikasi. Model tersebut digunakan untuk *website* aplikasi klasifikasi kandidat tenaga kerja di TOG Indonesia. Aplikasi telah berhasil dibuat dan dapat memprediksi data kandidat sesuai dengan ketentuan yang sudah ditetapkan oleh HR perusahaan.

Kata kunci: *Klasifikasi, Outsourcing, Decision Tree, Algoritma C4.5.*

ABSTRACT

IMPLEMENTATION OF DATA MINING FOR OUTSOURCING CANDIDATE DATA CLASSIFICATION USING C4.5 DECISION TREE METHOD

Student Name / NPM : Ahmad Ardhy Ansyah / 20083010008
Study Program : Sains Data, FASILKOM, UPN Veteran Jatim
Advisor 1 : Tresna Maulana Fahrudin, S.ST., MT
Advisor 2 : Dr. Eng. Ir. Dwi Arman Prasetya.,ST., MT., IPU.

TOG Indonesia is a company engaged in the provision of professional IT labor services. In order to fulfill labor requirements, the company continues to create vacancies in order to meet the needs of its partners. Currently, the company employs conventional methods for candidate selection, which results in the identification of candidates that require a significant amount of time and produces inconsistent decisions. Furthermore, the candidate history data is not stored in a structured manner due to the lack of storage for the files. Consequently, this research will develop a model for candidate data classification. The chosen classification model is machine learning, specifically the decision tree C4.5, which automatically classifies based on candidate profiles and competencies. Once implemented, the results of this model demonstrate high accuracy and precision, particularly in validation using a holdout or percentage split with a 70:30 proportion of train and test data. This achieves the highest accuracy of 99% and precision of 90%. The candidate education variable emerges as the most influential factor in the candidate recommendation prediction process, as it serves as the root node of the classification tree. The model has been implemented on the website of the labor candidate classification application at TOG Indonesia. The application has been successfully created and can predict candidate data in accordance with the conditions set by the company's HR.

Keyword: *Classification, Outsourcing, Decision Tree, C4.5 Algorithm.*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat dan karunian-Nya penulis dapat menyelesaikan skripsi ini dengan judul “Implementasi Data Mining Untuk Klasifikasi Data Kandidat *Outsourcing* Menggunakan Metode Decision Tree C4.5”. Penulisan skripsi ini ditujukan untuk memenuhi syarat untuk memperoleh gelar Sarjana Sains Data pada program studi S1 Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Veteran Jawa Timur. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada

1. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU sebagai Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Prof. Dr. Ir. Novirina Hendrasarie, M.T. sebagai Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
3. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Tresna Maulana Fahrudin, S.ST., M.T. dan Dr. Eng. Ir. Dwi Arman Prasetya.,ST., MT., IPU sebagai Dosen Pembimbing 1 dan 2.
5. TOG Indonesia sebagai perusahaan untuk menjadi objek penelitian yang memberikan banyak informasi selama masa penelitian
6. Kedua orang tua dan keluarga yang senantiasa memberikan dorongan dan doa.
7. Teman – teman yang sudah menjadi *support system* ketika sedang suntuk dalam penyusunan laporan ini.

Serta pihak-pihak lain yang tidak dapat disebutkan satu-persatu. Semoga laporan ini dapat dipergunakan dengan sebaik-baiknya

Surabaya, Mei 2024
Penulis

Ahmad Ardhy Ansyah

DAFTAR ISI

HALAMAN SAMBUL	i
LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
BAB I. PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
BAB II .TINJAUAN PUSTAKA.....	5
2.1. Dasar Teori.....	5
2.1.1. <i>Outsourcing</i>	5
2.1.2. Klasifikasi pada Data Mining.....	6
2.1.3. Analisis Data Eksploratif	7
2.1.4. <i>Validation Model</i>	8
2.1.5. <i>Decision Tree</i>	9
2.1.6. Algoritma C4.5.....	10
2.1.7. <i>Confusion Matrix</i>	14
2.2. Penelitian Terdahulu	16
BAB III. METODOLOGI PENELITIAN.....	20
3.1. Variabel Penelitian dan Sumber Data	20
3.2. Langkah Analisis.....	24
3.2.1. Studi Literatur	24

3.2.2.	Persiapan Data.....	24
3.2.3.	Praproses Data.....	28
3.2.4.	Pembuatan Model Decision Tree C4.5	30
3.2.5.	Evaluasi Model	39
3.2.6.	<i>Deployment</i> Aplikasi.....	39
3.2.7.	Skenario Pengujian	39
3.3.	Diagram Alir Penelitian	40
3.4.	Jadwal Penelitian.....	41
BAB IV. HASIL DAN PEMBAHASAN		42
4.1.	Persiapan Data.....	42
4.2.	Pengolahan Data.....	42
4.2.1.	<i>Collecting Data</i>	42
4.2.2.	<i>Encode Data</i>	44
4.2.3.	<i>Feature Engineering</i>	46
4.3.	Pembuatan Model Decision Tree C4.5	48
4.3.1.	Proses <i>Split Data</i>	48
4.3.2.	Pembuatan Model Decision Tree C4.5	49
4.4.	Evaluasi	49
4.4.1.	Evaluasi Model Decision Tree C4.5	50
4.4.2.	Evaluasi <i>Training</i> Model Klasifikasi	50
4.4.3.	Evaluasi Fase <i>Validation</i>	53
4.5.	<i>Deployment</i> Aplikasi	60
4.6.	Uji Coba	62
BAB V. PENUTUP.....		66
5.1	Kesimpulan	66
5.2	Saran.....	67
DAFTAR PUSTAKA		68
LAMPIRAN.....		71
BIODATA PENULIS		83

DAFTAR GAMBAR

Gambar 2. 1. Pohon Decision Tree (Setio dkk., 2020)	10
Gambar 3. 1. Diagram Alir Persiapan Data	25
Gambar 3. 2. Gambar Diagram Alir Pengolahan Data	28
Gambar 3. 3. Gambar Diagram Alir Pembuatan Model	31
Gambar 3. 4. Hasil Pohon Decision Tree C4.5	37
Gambar 3. 5. Diagram Alir Penelitian	41
Gambar 4. 1. Banyaknya Jumlah Kategori Usia	45
Gambar 4. 2. Hasil Akhir Praproses Data	46
Gambar 4. 3. Kelas Data Sebelum <i>Oversampling</i>	47
Gambar 4. 4. Jumlah Data X Setelah <i>Oversampling</i>	47
Gambar 4. 5. Jumlah Data y Setelah <i>Oversampling</i>	48
Gambar 4. 6. <i>Classification Report</i> Data Tanpa <i>Oversampling</i> 70:30	51
Gambar 4. 7. <i>Classification Report</i> Data Tanpa <i>Oversampling</i> 80:20	51
Gambar 4. 8. <i>Classification Report</i> Data Tanpa <i>Oversampling</i> 90:10	52
Gambar 4. 9. <i>Classification Report</i> Data <i>Oversampling</i> 70:30	52
Gambar 4. 10. <i>Classification Report</i> Data <i>Oversampling</i> 80:20	53
Gambar 4. 11. <i>Classification Report</i> Data <i>Oversampling</i> 90:10	53
Gambar 4. 12. <i>Classification Report</i> Skenario 1	54
Gambar 4. 13. <i>Classification Report</i> Skenario 2	55
Gambar 4. 14. <i>Classification Report</i> Skenario 3	56
Gambar 4. 15. <i>Classification Report</i> Skenario 4	57
Gambar 4. 16. <i>Classification Report</i> Skenario 5	58
Gambar 4. 17. <i>Classification Report</i> Skenario 6	59
Gambar 4. 18. Perbandingan Akurasi Hasil Pengujian	60
Gambar 4. 19. Menu Utama <i>Website</i> Aplikasi	61
Gambar 4. 20. Tampilan <i>Form</i> Input Data Klasifikasi	61
Gambar 4. 21. Tampilan Halaman Riwayat Data Kandidat	62
Gambar 4. 22 Tampilan Riwayat Kandidat Hasil Pengujian	64

DAFTAR TABEL

Tabel 2. 1. Tabel <i>Confusion Matrix</i>	14
Tabel 3. 1. Tabel Struktur Data Penelitian	20
Tabel 3. 2. Tabel Lanjutan Kolom Struktur Data Penelitian.....	21
Tabel 3. 3. Tabel Lanjutan Kolom Struktur Data Penelitian.....	22
Tabel 3. 4. Tabel Rangkuman Nama Fitur	22
Tabel 3. 5. Tabel Informasi Data	26
Tabel 3. 6. Tabel Rekap Perhitungan Manual.....	34
Tabel 3. 7. Skenario Pengujian Penelitian	40
Tabel 3. 8. Jadwal Kegiatan Penelitian	41
Tabel 4. 1. Tabel Data Awal	43
Tabel 4. 2. Rentang Nilai Kategori Usia	44
Tabel 4. 3. Tabel Data Hasil Kategorisasi Usia	45
Tabel 4. 4. Jumlah Proporsi Data	49
Tabel 4. 5. <i>Confusion Matrix</i> Skenario 1	54
Tabel 4. 6. <i>Confusion Matrix</i> Skenario 2	55
Tabel 4. 7. <i>Confusion Matrix</i> Skenario 3	56
Tabel 4. 8. <i>Confusion Matrix</i> Skenario 4	57
Tabel 4. 9. <i>Confusion Matrix</i> Skenario 5	58
Tabel 4. 10. <i>Confusion Matrix</i> Skenario 6	59
Tabel 4. 11. Tabel Uji Data Kandidat	63
Tabel 4. 12. Perbandingan Pengujian Label Klasifikasi	65

DAFTAR LAMPIRAN

Lampiran 1. Hasil Uji Plagiasi	71
Lampiran 2. Data Penelitian.....	72
Lampiran 3. <i>Source Code</i> yang digunakan untuk analisis	73