

ANALISIS TEKNIK BOOTSTRAP AGGREGATING DENGAN ALGORITMA DECISION TREE UNTUK MEMPREDIKSI DATA TRACER LULUSAN

Bani Nurhakim ^{1*}, Mulyawan ², Cep Lukman Rohmat ³

¹ Manajemen Informatika, STMIK IKMI Cirebon, Cirebon, Indonesia

² Sistem Informasi, STMIK IKMI Cirebon, Cirebon, Indonesia

³ Rekayasa Perangkat Lunak, STMIK IKMI Cirebon, Cirebon, Indonesia

baninurhakim@gmail.com

ABSTRAK

Ketidaksesuaian antara bidang kerja lulusan dengan kompetensi mereka di bidang pekerjaan masing-masing, yang tidak sesuai dengan latar belakang pendidikan dan keahlian mereka, memiliki dampak signifikan terhadap fleksibilitas pekerjaan dan dapat mempengaruhi kinerja kerja. Langkah antisipatif yang dapat digunakan untuk pengambilan keputusan terkait penelitian yang dilakukan, dan menurut hasil wawancara dengan manajemen tracer study STMIK IKMI Cirebon, hingga saat ini, informasi mengenai prediksi kesesuaian pekerjaan lulusan belum diperoleh, meskipun prediksi tersebut diperlukan untuk mempersiapkan sumber daya manusia agar bekerja sesuai dengan kompetensi mereka. Penelitian ini merupakan penelitian eksperimental dan studi kasus menggunakan metode Knowledge Discovery in Database (KDD). Algoritma yang digunakan adalah pohon keputusan C4.5 dengan teknik bootstrap aggregating. Dengan menganalisis, mengimplementasikan, dan mengeksperimentkan model yang akan menghasilkan evaluasi akurasi model pohon keputusan C4.5 terbaik dengan teknik bootstrap aggregating. Kesesuaian bidang kerja lulusan dipengaruhi oleh atribut yang paling dominan, yaitu IPK lulusan (100%) dengan IPK tertinggi maksimal 3,71 untuk 12 orang, usia lulusan (97,4%) dengan 776 lulusan yang termuda berusia 21 tahun, dan durasi studi (53,85%). Sementara itu, atribut yang paling sedikit mempengaruhi kesesuaian bidang kerja lulusan adalah tahun kelulusan (10,4%). Model optimasi klasifikasi terbaik diperoleh melalui pengaturan parameter gain ratio, menghasilkan model prediksi yang sangat baik. Berdasarkan evaluasi akurasi model terbaik, skor akurasi adalah 95,27% dan skor AUC adalah 98,60%.

Kata Kunci: Data Mining, Tracer Study Lulusan, Decision Tree, Bootstrap Aggregating

1. PENDAHULUAN

Abad ke-21 menuntut kesesuaian antara keterampilan dan jenis pekerjaan. Perubahan cepat di dunia kerja akibat globalisasi dan revolusi teknologi memerlukan antisipasi dan evaluasi terhadap kompetensi yang diperlukan. Permintaan akan lulusan berkualitas tinggi menjadi perhatian penting bagi pemerintah, institusi pendidikan, dan masyarakat. Berdasarkan data modal manusia, hanya 11,01% dari tenaga kerja Indonesia yang memiliki kualifikasi pendidikan tinggi menurut sumber yang relevan. Hal ini menunjukkan perlunya melakukan pemetaan dan peningkatan kualifikasi pendidikan tinggi sebagai bagian dari upaya untuk memastikan kualitas pendidikan.

Ketidaksesuaian antara bidang kerja lulusan [1] dan kompetensi mereka yang ditempatkan pada pekerjaan yang tidak sesuai dengan latar belakang pendidikan [2] dan keahlian mereka dapat berdampak signifikan terhadap fleksibilitas kerja dan kinerja pekerjaan. Hal ini dinyatakan dalam Peraturan Menteri No. 49/2014 tentang standar nasional pendidikan tinggi, yang menekankan bahwa pencapaian lulusan berkualitas baik dipengaruhi oleh kesesuaian bidang studi mereka dengan pekerjaan mereka.

Berdasarkan profil lulusan dari berbagai program studi di STMIK IKMI Cirebon, langkah antisipatif untuk mengumpulkan informasi potensial tentang lulusan dapat digunakan untuk pengambilan keputusan dalam penelitian. Menurut wawancara dengan manajer

tracer study di STMIK IKMI Cirebon, universitas tersebut telah meluluskan lebih dari 1.000 mahasiswa dari program sarjana Ilmu Komputer, program diploma Sistem Informasi Manajemen, dan Akuntansi Terkomputerisasi sejak tahun 2016 hingga 2022. Namun, meskipun memiliki database lulusan yang bekerja, informasi mengenai prediksi kesesuaian di bidang mereka masing-masing masih belum tersedia. Informasi ini diperlukan untuk mempersiapkan sumber daya manusia agar bekerja sesuai dengan kompetensi mereka. Penelitian sebelumnya telah dilakukan tentang prediksi kesesuaian pekerjaan untuk lulusan.

Dalam artikel jurnal tahun 2018 berjudul "Relevansi Lulusan Perguruan Tinggi dalam Penempatan Kerja" [3] membahas tingkat relevansi lulusan dalam penempatan kerja. Variabel yang diteliti termasuk tingkat relevansi lulusan terhadap pekerjaan yang mereka peroleh, dengan lulusan program Administrasi Bisnis mencapai persentase 67,2% di bidang administrasi dan manajemen, dan 32,8% di bidang non-administrasi dan manajemen. Sementara itu, faktor-faktor lain seperti keterampilan dalam bekerja juga diperhitungkan.

2. TINJAUAN PUSTAKA

Artikel penelitian berjudul [1] "Prediksi Bidang Pekerjaan untuk Lulusan Program Vokasi dalam Studi Kesekretariatan Menggunakan Teknik Klasifikasi Data Mining," yang diterbitkan pada tahun 2018,

membahas hasil pemodelan menggunakan teknik bootstrap aggregating melalui algoritma pohon keputusan, yang memiliki kinerja terbaik dengan tingkat akurasi 95,8%, sedangkan algoritma Naïve Bayes memiliki tingkat akurasi 81,9%.

Artikel berjudul [2] "Implementasi Algoritma Pohon Keputusan C4.5 untuk Prediksi Penyakit Diabetes" membahas pembuatan model prediksi untuk diabetes menggunakan algoritma C4.5 dan mengevaluasi akurasi. Model tersebut dikembangkan menggunakan data dari Pima Indians Diabetes Databases (PPID) yang berasal dari UCI Machine Learning Repository. Model berbasis algoritma pohon keputusan C4.5 mencapai akurasi 70,32% dengan 9 aturan, di mana 4 dan 5 aturan mengklasifikasikan "ya" untuk memprediksi penyakit diabetes.

"Pengklasifikasian [3] Waktu Tunggu Alumni untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5" bertujuan untuk memprediksi waktu tunggu alumni untuk mendapatkan pekerjaan menggunakan algoritma pohon keputusan tunggal (C4.5) [4] dan membandingkannya dengan algoritma C4.5 dengan fitur seleksi maju. Studi ini menemukan bahwa algoritma C4.5 dengan fitur seleksi maju memberikan kinerja terbaik dengan akurasi 80,37%, presisi 79,56%, recall 81,34%, f-measure 80,40%, dan AUC 0,914, yang masuk dalam kategori klasifikasi yang sangat baik.

Implementasi algoritma Pohon Keputusan [5] dalam mengklasifikasikan kompetensi siswa [6], dengan tujuan mendapatkan pola rekomendasi kompeten dan tidak kompeten. Studi ini menemukan akurasi sebesar 76,96%, yang menunjukkan bahwa siswa memiliki kompetensi yang relevan dengan bidang masing-masing.

Perbedaan utama studi ini dibandingkan dengan lima studi sebelumnya adalah dataset yang digunakan, yang diperoleh dari tracer study lulusan STMIK IKMI Cirebon. Selain itu, perbedaannya juga terletak pada algoritma yang diterapkan, yaitu pohon keputusan C4.5 menggunakan teknik bootstrap aggregating.

3. METODE PENELITIAN

Data dikumpulkan melalui wawancara dan observasi langsung dengan pengelola tracer study lulusan program S1 Teknik Informatika, D3 Manajemen Informatika, dan D3 Komputerisasi Akuntansi dari tahun 2016 hingga 2022 di STMIK IKMI Cirebon.

Penelitian ini adalah studi kuantitatif yang mengumpulkan data tracer study lulusan program S1 Teknik Informatika, D3 Manajemen Informatika, dan D3 Akuntansi Terkomputerisasi di STMIK IKMI Cirebon dari tahun 2016 hingga 2022.

Proses KDD [10][11] mencakup beberapa tahap: pemilihan data, pra-pemrosesan data, integrasi data, transformasi data, data mining [12][13], evaluasi pola, dan representasi pengetahuan. Tahap pemilihan data melibatkan pemilihan dataset yang diperoleh dari

pengumpulan data primer tracer study [14][15]. Tahap pra-pemrosesan data mempersiapkan data menggunakan operator seperti pemilihan atribut, penyaringan contoh, penghapusan duplikat, dan penggantian. Tahap integrasi data melibatkan penggabungan dataset dari tracer study program S1 Teknik Informatika, D3 Manajemen Informatika, dan D3 Akuntansi Terkomputerisasi dari tahun 2016 hingga 2022 di STMIK IKMI Cirebon ke dalam tabel Microsoft Excel untuk atribut yang dipilih. Tahap transformasi data melibatkan transformasi dan konsolidasi dataset agar sesuai untuk tujuan data mining. Tahap data mining menggunakan algoritma Pohon Keputusan dan alat RapidMiner untuk mengekstrak pola dari data dalam proses intelijen. Tahap evaluasi pola mengidentifikasi pola data yang mewakili basis pengetahuan untuk menghitung akurasi. Akhirnya, tahap representasi pengetahuan menggunakan model yang dibuat untuk membuat prediksi tentang kecocokan lulusan dengan bidang pekerjaan.

4. HASIL DAN PEMBAHASAN

4.1. Data Collection

Wawancara dengan kepala tracer study mengungkapkan data lulusan dari tracer study yang mencakup tahun 2016-2022, dengan total 1.692 catatan. Dataset contoh tracer study mencakup atribut-atribut seperti NIM (integer & ID), Kecocokan (label binomial), Program Studi (integer), Tahun Masuk/Tahun Lulus (integer, menunjukkan tahun masuk pertama & tahun kelulusan), IPK (real), Durasi Studi (integer), dan Usia saat lulus (integer). Aplikasi RapidMiner menampilkan jenis data dari atribut tracer study.

4.2. Data Preprocessing

Pada tahap pra-pemrosesan, data tracer study dipersiapkan untuk analisis kualitas dengan mengelola, mentransformasi, atau mengeliminasi data yang tidak diperlukan. Data yang diproses hanya mencakup data yang Relevan dan Tidak Relevan dengan bidang kerja lulusan, terdiri dari 1.184 catatan Relevan dan 508 catatan Tidak Relevan, dengan total 1.692 catatan. Data ini diimpor ke dalam RapidMiner menggunakan operator Read Excel beserta parameter yang sesuai.

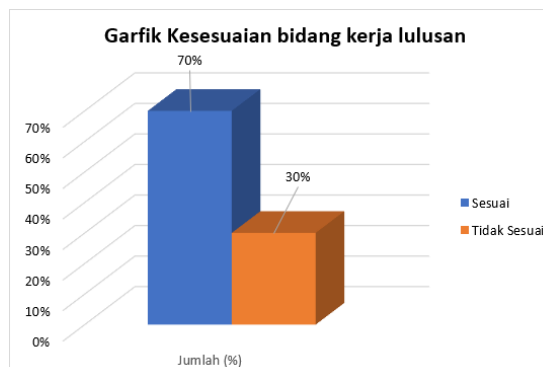
4.3. Data Transformation

Data lulusan tracer study yang telah diimpor dan dibersihkan digunakan sebagai input untuk model klasifikasi prediktif. Tujuan dari transformasi data ini adalah untuk memprediksi atau mengklasifikasikan lulusan berdasarkan atribut relevan seperti bidang pekerjaan atau tingkat kecocokan pekerjaan. Proses klasifikasi melibatkan penerapan algoritma dan teknik machine learning untuk mengenali pola dan hubungan dalam data. Berdasarkan kecocokan bidang pekerjaan lulusan, dataset dalam studi ini dapat disajikan dalam

bentuk Tabel 1 dan divisualisasikan dalam bentuk grafik seperti yang ditunjukkan pada Gambar 1.

Table 1. Jumlah Kesesuaian Bidang Pekerjaan Lulusan

No	Kesesuaian	Jumlah (%)
1	Sesuai	70%
2	Tidak Sesuai	30%

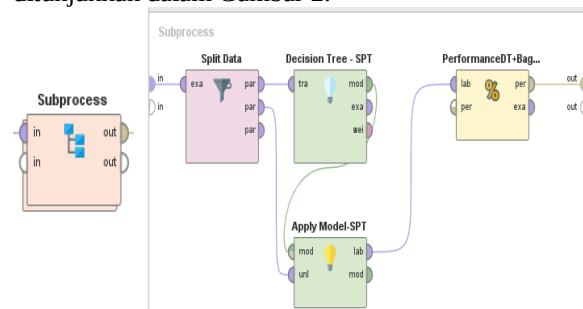


Gambar 1. Grafik Kesesuaian Bidang Kerja Lulusan

4.4. Modeling

Pemanfaatan algoritma pohon keputusan C4.5 untuk model prediksi kecocokan bidang pekerjaan lulusan dari dataset tracer study STMIK IKMI Cirebon. Tujuannya adalah untuk menghasilkan model implementasi untuk memperkirakan kecocokan bidang pekerjaan. Sebelum ini, operator

penggabungan menggunakan Subproses, seperti yang ditunjukkan dalam Gambar 2.



Gambar 2. Model Subprocess Operator Decision Tree

Operator subprocess digunakan untuk mengatur alur pemrosesan data, termasuk operator pembagian data yang membagi dataset menjadi dua bagian: 90% untuk pelatihan (sehingga menghasilkan 1.523 data) dan 10% untuk pengujian (sehingga menghasilkan 169 data) dari total 1.692 data.

4.5. Evaluation

Evaluasi penerapan dua algoritma, yaitu pohon keputusan C4.5 dan pohon keputusan C4.5 dengan metode bagging (bootstrap aggregating), dilakukan menggunakan Cross Validation dengan teknik 10-fold nFold dan teknik pengambilan sampel stratifikasi. Teknik ini memastikan distribusi kelas (didefinisikan oleh label atau atribut target) dalam subset dataset sama dengan di seluruh dataset.

Tabel 2. Dataset

NO	NIM	NAMA	PROGRAM STUDI	THN MSK	THN LLS	IPK	MASA STUDI	USIA	PEKERJ AAN	KESESU AIAN
1	41101477	Yogi Nurohman	1	10	16	33,31	6	27	6	Sesuai
2	41101394	Elan Wisesa Amirullah	1	10	17	3,00	7	30	20	Tidak Sesuai
3	41101378	Wisnu Dodi Wijaya	1	10	16	3,19	6	34	9	Sesuai
4	41101274	Stefanus Nugroho	1	10	16	3,35	6	26	9	Sesuai
...										
1.692	41101373	Taufik R Firdaus	1	10	16	3,31	6	23	8	Sesuai

Studi ini mengevaluasi penerapan dua algoritma, yaitu pohon keputusan C4.5 dan pohon keputusan C4.5 dengan teknik bagging, menggunakan 10-fold cross-validation dan pengambilan sampel stratifikasi. Hasilnya menunjukkan bahwa akurasi algoritma pohon keputusan C4.5 adalah 94,92% berdasarkan matriks kebingungan yang diperoleh. Matriks Kebingungan yang ditunjukkan dalam Gambar 3 adalah teknik yang dapat digunakan untuk melakukan perhitungan akurasi dalam konteks data mining.

accuracy: 94.91% +/- 1.66% (micro average: 94.92%)			
	true TIDAK SESUAI	true SESUAI	class precision
pred. TIDAK SESUAI	457	35	92.89%
pred. SESUAI	51	1149	95.75%
class recall	89.96%	97.04%	

Gambar 3. Confusion Matrix untuk Algoritma Decision Tree C4.5

accuracy: 95.27% +/- 1.87% (micro average: 95.27%)			
	true TIDAK SESUAI	true SESUAI	class precision
pred. TIDAK SESUAI	455	27	94.40%
pred. SESUAI	53	1157	95.62%
class recall	89.57%	97.72%	

Gambar 4. Confusion Matrix untuk Algoritma Bootstrap Aggregating

Eksperimen menunjukkan bahwa algoritma Pohon Keputusan C4.5 dengan kriteria gain ratio memiliki akurasi tertinggi sebesar 95,27%. Sementara itu, akurasi algoritma dengan kriteria indeks gini dan gain informasi terendah sebesar 94,80%.

Studi ini menghitung nilai AUC dari penerapan algoritma pohon keputusan dan pohon keputusan C4.5 dengan teknik bootstrap aggregating. Ditemukan bahwa nilai AUC dari penerapan pohon keputusan C4.5 dengan teknik bagging lebih tinggi (98,60%) dibandingkan dengan pohon keputusan C4.5 saja

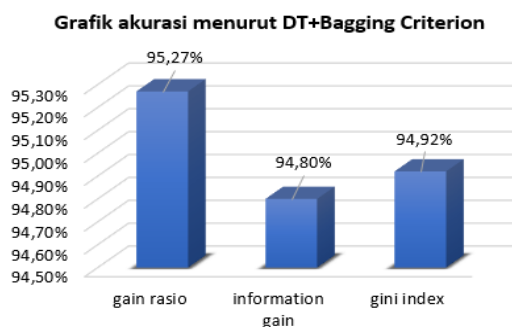
(95,90%). Hal ini menunjukkan bahwa penerapan teknik bagging pada model menghasilkan model prediksi klasifikasi yang sangat baik berdasarkan teori AUC.

Table 3. Comparison of Accuracy and AUC

No	Model	Performance Accuracy	AUC
1	Decision Tree C4.5 + Bagging	95,27 %	98,60 %
2	Decision Tree	94,74 %	95,90 %

4.6. Pembahasan

Penelitian ini membedakan dirinya dari lima studi terkait lainnya yang membahas penerapan algoritma untuk menentukan kesiapan kerja siswa, memprediksi kecocokan pekerjaan bagi lulusan, dan menganalisis kompetensi lulusan. Studi ini berfokus pada bidang kecocokan pekerjaan bagi lulusan STMIK IKMI Cirebon dari tahun 2016 hingga 2022, dengan menerapkan algoritma pohon keputusan C4.5 dan teknik Bootstrap Aggregating. Penerapan teknik Bootstrap Aggregating pada algoritma pohon keputusan C4.5 meningkatkan akurasi dan nilai AUC model prediksi, yang menjadi sorotan utama dari penelitian ini.



Gambar 5. Grafik Akurasi DT+Bagging dan DT Criterion

Penelitian ini menggunakan metode KDD yang berbeda dari studi terkait sebelumnya dan menghitung nilai AUC menggunakan teknik Bootstrap Aggregating. Temuan menunjukkan tingkat akurasi sebesar 95,27%, yang lebih tinggi dibandingkan dengan studi terkait sebelumnya dengan tingkat akurasi 85%.

Penelitian ini merekomendasikan teknik Bootstrap Aggregating dengan algoritma Pohon Keputusan sebagai metode yang efektif untuk memprediksi kecocokan bidang pekerjaan lulusan. Selain itu, hasil prediksi dengan akurasi 95,27%, lebih tinggi daripada akurasi studi sebelumnya sebesar 85%, menunjukkan bahwa model yang dikembangkan memiliki potensi yang baik untuk diterapkan dalam kasus serupa. Menurut Gorunescu [7], akurasi 95,27% dengan nilai AUC yang baik termasuk dalam kategori klasifikasi yang sangat baik. Sebuah studi oleh Yoga Pristyanto menunjukkan bahwa penggunaan pohon

keputusan + bagging menghasilkan akurasi yang lebih tinggi dibandingkan dengan pohon keputusan saja, yaitu 97,8% dibandingkan dengan 96,7%. Hasil perbandingan kinerja menunjukkan bahwa penerapan metode pohon keputusan + bagging dapat meningkatkan kemampuan model dan memberikan kinerja terbaik pada semua dataset yang digunakan.

5. KESIMPULAN DAN SARAN

Berdasarkan analisis teknik Bootstrap Aggregating dengan algoritma Pohon Keputusan C4.5 untuk memprediksi data Tracer Study lulusan menggunakan Rapidminer, ditemukan bahwa 70% lulusan memiliki kecocokan bidang pekerjaan dan faktor yang paling dominan mempengaruhi kecocokan bidang pekerjaan adalah IPK, usia lulusan, dan lama studi. Implementasi model menggunakan teknik ini menunjukkan bahwa ada parameter optimal, dan model yang menghasilkan parameter ini dipilih sebagai model optimasi klasifikasi terbaik. Model terbaik memiliki akurasi sebesar 95,27% dan nilai AUC sebesar 98,60%, yang menunjukkan bahwa teknik Bootstrap Aggregating dengan kriteria gain ratio menghasilkan prediksi yang termasuk dalam klasifikasi yang sangat baik untuk kecocokan bidang pekerjaan lulusan.

DAFTAR PUSTAKA

- [1] D. R. Utari, "PREDIKSI BIDANG KERJA BAGI LULUSAN PROGRAM STUDI VOKASI SEKRETARIS MENGGUNAKAN TEKNIK KLASIFIKASI DATA MINING," pp. 1–23, 2018.
- [2] N. Noviandi, "Implementasi Algoritma Decision Tree C4.5 Untuk Prediksi Penyakit Diabetes," *Indones. Heal. Inf. Manag. J.*, vol. 6, no. 1, pp. 1–5, 2018, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/142>
- [3] F. Rezkika, B. N. Sari, and A. Susilo, "Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5," *Progresif J. Ilm. Komput.*, vol. Vol. 17, pp. 95–106, 2021.
- [4] A. Asroni, B. Masajeng Respati, and S. Riyadi, "Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta," *Semesta Tek.*, vol. 21, no. 2, pp. 158–165, 2020, doi: 10.18196/st.212222.
- [5] N. W. Wardani and N. K. Ariasih, "Analisa Komparasi Algoritma Decision Tree C4.5 dan Naïve Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail," *Int. J. Nat. Sci. Eng.*, vol. 3, no. 3, p. 103, 2019, doi: 10.23887/ijnse.v3i3.23113.
- [6] H. Rifa'i, Ryan Hamonangan, Dian Ade Kurnia, Kaslani, and Mulyawan, "Implementasi Algoritma Decision Tree Dalam Klasifikasi Kompetensi Siswa," *KOPERTIP J. Ilm. Manaj.*

- Inform. dan Komput.*, vol. 6, no. 1, pp. 15–20, 2022, doi: 10.32485/kopertip.v6i1.131.
- [7] F. Gorunescu, *Data mining concepts models and techniques*, vol. 4, no. 1. 2011.