

**PENERAPAN ALGORITMA C4.5 DENGAN *FEATURE FORWARD SELECTION* UNTUK ANALISIS CAPAIAN INDIKATOR KINERJA UTAMA BERDASARKAN *TRACER STUDY* (STUDI KASUS: FASILKOM UNSIKA)**

**Fenny Rahmayani, Betha Nurina Sari, Iqbal Maulana, Garno, Rini Mayasari**

Program Studi Informatika S1, Fakultas Ilmu Komputer  
Universitas Singaperbangsa Karawang Jl. H.S. Ronggo Waluyo  
Telukjambe Timur, Karawang, Jawa Barat, Indonesia 41361  
1910631170020@student.unsika.ac.id

**ABSTRAK**

Perguruan tinggi, sebagai pusat ilmu, penelitian, dan pengabdian masyarakat, dituntut untuk fokus pada capaian Indikator Kinerja Utama Perguruan Tinggi Negeri (IKU-PTN). Salah satu kunci dalam mengelola kinerja perguruan tinggi adalah melalui IKU-PTN yang mendukung relevansi dengan industri dan dunia kerja terdapat pada IKU1. Namun, data *tracer study* Fasilkom Unsika menunjukkan capaian IKU1 pada tahun 2020 dan 2021 di bawah target, menandakan perlunya analisis pola data alumni. Penelitian ini menggunakan algoritma C4.5 untuk mengevaluasi kelayakan alumni dalam mencapai IKU1. Setelah *feature selection*, ditemukan atribut paling berpengaruh seperti "Status Bekerja," "Gaji," "Mendapat Pekerjaan Kurang dari 6 Bulan," "1.2 x UMP," dan "Jenis Tempat Bekerja." Model terbaik, yaitu model 5, menghasilkan akurasi 98.77% dengan *Weighted Precision* rata-rata 100%, *Recall* 96%, dan *f1-Score* 89%. Penerapan *Decision Tree* C4.5 dan metodologi *Knowledge Discovery in Databases* (KDD) menghasilkan 5 aturan penting yang membantu menilai pencapaian IKU1 berdasarkan fitur-fitur signifikan. Penelitian ini memberikan panduan bagi lembaga pendidikan untuk memahami faktor-faktor yang memengaruhi kesuksesan penempatan kerja alumni serta memberikan prediksi akurat untuk meningkatkan capaian IKU1. Kontribusi ini berpotensi mendorong evaluasi dan pengembangan program pendidikan yang lebih efektif.

**Kata kunci:** Algoritma C4.5, *Knowledge Discovery in Database*, *Data Mining*, *Feature Selection*.

## 1. PENDAHULUAN

Perguruan tinggi sebagai lembaga ilmu, pengetahuan, penelitian, serta pengabdian kepada masyarakat, dituntut untuk dapat lebih fokus dalam merealisasikan target kinerjanya. Salah satu kunci dalam mengatur kinerja perguruan tinggi ialah melalui Indikator Kinerja Utama Perguruan Tinggi Negeri (IKU-PTN) yang ditetapkan melalui Keputusan Menteri Pendidikan dan Kebudayaan [3]. Salah satu tujuan IKU adalah untuk meningkatkan relevansi perguruan tinggi dengan kebutuhan industri, dunia usaha, dan dunia kerja. Hal ini tercantum pada penilaian IKU 1 yaitu lulusan mendapatkan pekerjaan yang layak. Indikator yang digunakan pada IKU 1 terdiri dari pekerjaan, studi lanjut dan kewirausahaan. Berdasarkan data hasil monitoring *tracer study* akademik Fasilkom Unsika tahun 2021 target persentase capaian IKU 1 untuk lulusan tahun 2020 sebesar 55%, namun pada realisasinya mendapat persentase sebanyak 26,75%. Sedangkan tahun 2022, target persentase capaian IKU 1 untuk lulusan tahun 2021 sebesar 50%, realisasinya per tanggal 28 November 2022 mencapai 33,12%. Hal ini menunjukkan bahwa selama dua tahun IKU 1 Fasilkom tidak mencapai target. Berdasarkan data tersebut, analisis pada data *tracer study* perlu dilakukan guna mencapai target indikator kinerja utama. Pihak perguruan tinggi membutuhkan analisis pola data alumni yang dapat digunakan untuk menyusun strategi agar mencapai IKU 1 pada tahun berikutnya. Salah satu cara untuk mendapatkan

informasi, pengetahuan, dan pola dari *tracer study* adalah *data mining*.

Salah satu metode analisis dalam data mining adalah klasifikasi. Klasifikasi adalah teknik pembelajaran mesin yang mengkategorikan data ke dalam sejumlah kelas tertentu dengan memprediksi kelas label atau kategori untuk data baru [6]. Untuk mengetahui pola atau standar pemilihan subjek, maka digunakan suatu algoritma untuk melakukan proses klasifikasi menggunakan algoritma C4.5 dengan pohon keputusan. Algoritma C4.5 merupakan algoritma untuk membentuk pohon keputusan dengan menghitung nilai *gain* di mana nilai terbesar akan digunakan sebagai *node* atau *node root* [11].

Berdasarkan penelitian sebelumnya yang dilakukan oleh [8] berupa pembuatan model prediksi kelulusan siswa yang diperoleh dari proses klasifikasi data menggunakan *decision tree* dengan algoritma C4.5. Dari hasil klasifikasi data siswa dapat dilakukan analisis faktor-faktor apa saja yang dapat mempengaruhi kelulusan siswa. Hasil klasifikasi adalah model prediksi yang memiliki akurasi sebesar 82,24% dan menyatakan bahwa faktor yang paling berpengaruh dalam memprediksi kelulusan siswa adalah IPK pada tahun kedua. Penelitian selanjutnya dilakukan oleh [1] yang berjudul "Implementasi *Naive Bayes* Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan" penelitian tersebut menggunakan metode *Naive Bayes* dengan proses data *Selection* dan *preprocessing* pada datanya. Dari 1240 data uji hanya 603 data yang berhasil diklasifikasi

dengan benar. Evaluasi dihasilkan nilai akurasi sebesar 48,629% dengan parameter pengujian 10 *Folds Cross-Validation*. Penelitian lainnya oleh [9] mengenai prediksi waktu tunggu alumni mendapatkan pekerjaannya dengan menggunakan algoritma *decision tree* (C4.5) tunggal dan dibandingkan algoritma *decision tree* (C4.5) dengan fitur *Forward Selection*. Hasil menunjukkan bahwa algoritma *decision tree* (C4.5) dengan fitur *forward selection* meraih performa terbaik dengan nilai *Accuracy* 80,37%, *Precision* 79,56%, *Recall* 81,34%, *F1-Score* 80,40% dan *AUC* 0.914 yang termasuk ke dalam kategori *excellent Classification*.

Dari hasil penelitian sebelumnya, algoritma C4.5 dengan fitur *forward selection* direkomendasikan untuk digunakan. Penelitian ini menerapkan algoritma C4.5 dengan fitur *forward selection* dalam menganalisis capaian indikator kinerja utama berdasarkan data *tracer study* Fasilkom Unsika. Penelitian ini diharapkan dapat menjadi informasi yang dapat membantu fakultas dalam menyusun strategi agar mencapai IKU pada tahun berikutnya.

## 2. TINJAUAN PUSTAKA

### 2.1. Indikator Kinerja Utama

Indikator Kinerja Utama dikeluarkan oleh Menteri Pendidikan dan Kebudayaan melalui Keputusan Menteri Pendidikan dan Kebudayaan Nomor 754/P/2020 merupakan ukuran kinerja baru bagi perguruan tinggi untuk mewujudkan perguruan tinggi yang adaptif dengan berbasis luaran lebih konkret [3].

### 2.2. Algoritma C4.5

*Decision tree* atau pohon keputusan merupakan bagian dari *data mining*. Pohon keputusan sangat bermanfaat dalam pengambilan keputusan dengan kemampuannya menyederhanakan sehingga memberikan solusi dari sebuah masalah [10].

### 2.3. K-Fold Cross Validation

*K-Fold Cross Validation* merupakan teknik verifikasi yang membagi data menjadi k bagian, dan setiap bagian dimasukkan dalam proses klasifikasi. Dengan menggunakan benarasi *K-Fold Cross*, maka uji k akan dilakukan. Setiap pengujian menggunakan satu data pengujian, dan bagian k-1 tersebut akan menjadi data latih, kemudian data pengujian tersebut ditukar dengan satu data latih sehingga diperoleh data yang berbeda untuk setiap pengujian [2].

### 2.4. Forward Selection

Tahapan *Forward Selection* diawali dengan fitur himpunan kosong lalu menambahkan fitur yang terpakai, kemudian seluruh fitur dievaluasi. Salah satu fitur ditambahkan pada fitur himpunan yang merupakan bagian dari fitur sebelumnya dan juga fitur yang baru dibuat, lalu dievaluasi kembali. Untuk mengurangi jumlah evaluasi, hanya subset fitur terbaik yang disimpan [2].

### 2.5. Heatmap

*Heatmap* adalah matriks korelasi yang memvisualisasikan hubungan 2 dimensi antara fitur - fitur dengan variabel target. Korelasi bisa bernilai positif, yaitu dua variabel bergerak dalam arah yang sama (keduanya meningkat) atau negatif ketika variabel bergerak berlawanan arah (satu meningkat, dan satu menurun). *Heatmap* yang terkait dengan matriks korelasi membantu mengidentifikasi fitur yang paling terkait satu sama lain dengan variabel target [7].

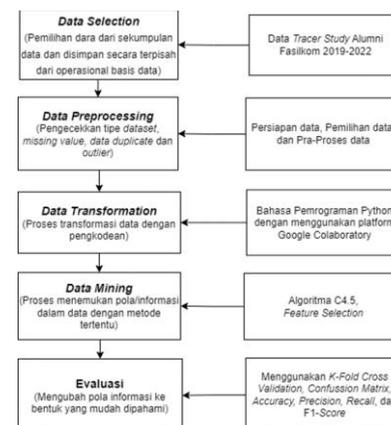
### 2.6. Python

Bahasa Python diciptakan oleh Guido van Rossum di Desember 1989 di Centrum Wiskunde & Informatica di Belanda untuk dijadikan sebagai penerus program ABC bahasa, yang menangani dan menghubungkan sistem operasi Amoeba. Nama bahasa pemrograman diturunkan dari acara *BBC favorit van Rossum "Monty Python's Flying Circus"*. Fitur-fitur baru telah ditambahkan secara konsisten tahun 90-an yang menyebabkan rilis Python 2.0 pada 16 Oktober, 2000 dan Python 3.0 pada 3 Desember 2008 [4].

### 2.7. Google Colaboratory

Google menawarkan Colaboratory sebagai cara untuk menjalankan kode di browser dengan menghubungkan pengguna dengan *runtime* yang dihosting pada server *cloud* Google. Ini memungkinkan penggunaan *machine learning* secara independen dari batasan perangkat keras pribadi dengan menggunakan perangkat keras yang disediakan oleh Google [5].

## 3. METODE PENELITIAN



Gambar 1. Rancangan penelitian metodologi KDD

Pada Penelitian ini, analisis data *tracer study* dengan capaian Indikator Kinerja Utama 1 menggunakan objek penelitian data alumni Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang tahun 2019-2022. Metode atau tahapan untuk mencapai tujuan dari penelitian ini melalui metodologi *Knowledge Discovery in Database (KDD)* yang terdiri dari lima tahapan, yaitu *Data Selection*,

Data Preprocessing, Data Transformation, Data Mining, dan Evaluasi. Rancangan penelitian dapat dilihat pada gambar 1.

A. Data Selection

Tahap pertama yaitu menyeleksi data dari kumpulan data *tracer study* alumni Universitas Singaperbangsa Karawang. Data dipilih dan dipisahkan dari basis data utama dengan mengambil beberapa variabel yang akan digunakan. Data yang dibutuhkan berupa data alumni Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang tahun 2019-2022.

B. Data Preprocessing

Tahap kedua adalah *data preprocessing* yang dilakukan dengan pembersihan data seperti menghapus duplikasi data alumni, data yang tidak sesuai, data yang hilang, dan memperbaiki data yang berbeda. Pada tahap ini dapat dilakukan dengan menggunakan bahasa pemrograman python dengan bantuan platform Google Colaboratory.

C. Data Transformation

Tahap selanjutnya adalah *data transformation* yaitu proses pengkodean data agar dapat diubah sesuai dengan data yang sudah dipilih sebelumnya. Proses pengkodean ini nantinya menggunakan bahasa pemrograman *python* dengan bantuan platform Google Colaboratory.

D. Data Mining

Tahap selanjutnya adalah *data mining* yaitu proses menemukan pola atau menarik informasi dalam data terpilih dengan menggunakan algoritma C4.5. Proses *data mining* ini nantinya menggunakan bahasa pemrograman *python* dengan bantuan platform Google Colaboratory. Pada fase ini juga ditambahkan metode *feature selection* yaitu *forward selection* untuk dapat meningkatkan hasil evaluasi. Adapun 3 Skenario yang akan dilakukan untuk mendapatkan hasil terbaik yang dapat dilihat pada tabel 1.

Tabel 1. Skenario Pembagian Dataset

No.	Skenario	Algoritma C4.5	Algoritma C4.5 + Feature Selection
1.	A (90:10)	Model 1	Model 4
2.	B (80:20)	Model 2	Model 5
3.	C (70:20)	Model 3	Model 6

Pembagian dataset dipartisi menjadi set *data training* dan *data testing* dengan perbandingan tertentu, di mana dalam skenario A menggunakan perbandingan 10% *data training* dan 90% *data testing*, skenario B menggunakan perbandingan 20% *data training* dan 80% *data testing*, dan skenario C menggunakan perbandingan 30% *data training* dan 70% *data testing*. Dengan menentukan 3 skenario dapat menghemat waktu komputasi yang dimana semakin banyak skenario yang diuji maka semakin

lama waktu yang dibutuhkan untuk melatih dan menguji model.

E. Evaluasi

Tahap terakhir pada metodologi KDD adalah evaluasi. Pada tahap ini dilakukan perubahan pola menjadi informasi yang mudah dipahami. Tahapan ini meliputi pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan tujuan sebelumnya untuk menganalisis target indikator kinerja utama berdasarkan data *tracer study* dengan melakukan validasi menggunakan *k-Folds cross validation* dan evaluasi menggunakan *confusion matrix* dengan parameter nilai *accuracy*, *precision*, *recall*, dan *F1-score*

4. HASIL DAN PEMBAHASAN

Hasil penelitian data mining yang telah dilakukan adalah bagaimana menerapkan algoritma C4.5 dalam menganalisis capaian IKU 1 Fasilkom Unsika. Penelitian ini menggunakan metode KDD dengan tahapan-tahapan yakni: *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan evaluasi.

A. Data Selection

Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional. Data diperoleh dari hasil kuisioner *tracer study* tahun kelulusan 2019-2022 dengan total sebanyak 815 *record* dengan 24 atribut. Pendeskripsian data dapat dilihat lebih rinci pada tabel 2.

Tabel 2. Skenario Deskripsi Dataset

No	Atribut	Keterangan
1	Npm	Nomor Pokok Mahasiswa
2	Tahun Lulus	Tahun kelulusan alumni
3	Nama	Nama lengkap alumni
4	Status Bekerja	Status pekerjaan alumni
5	Gaji	Pendapatan dalam satu bulan bekerja
6	Provinsi Tempat Bekerja	Provinsi tempat alumni bekerja
7	Mendapat Pekerjaan Kurang dari 6 Bulan	Lama waktu tunggu alumni setelah lulus sampai mendapatkan pekerjaan
8	1.2 x UMP	Upah Minimum Provinsi dikali 1.2 (sesuai syarat IKU1)
9	IKU1	Indikator Kinerja Utama 1 : Lulusan Mendapat Pekerjaan yang Layak
10	Keterkaitan Bidang Studi dengan Pekerjaan	Tingkat keterkaitan bidang pekerjaan alumni dengan bidang studi
11	Kesesuaian Tingkat Pendidikan dengan Pekerjaan	Kesesuaian tingkat pendidikan dengan bidang pekerjaan alumni
12	Jenis Tempat Bekerja	Jenis usaha/tempat bekerja alumni
13	Nama Tempat Bekerja	Nama tempat alumni bekerja

No	Atribut	Keterangan
14	Sumber Dana Kuliah	Sumber dana kuliah alumni
15	a1	Keahlian berdasarkan bidang ilmu
16	a2	Bahasa inggris
17	a3	Manajemen waktu
18	a4	Kepemimpinan
19	a5	Etika
20	b1	Keahlian berdasarkan bidang ilmu
21	b2	Bahasa inggris
22	b3	Manajemen waktu
23	b4	Kepemimpinan
24	b5	Etika

Pada tahap *data selection*, beberapa atribut digunakan sebagai acuan dan perhitungan untuk mencapai Indikator Kinerja Utama (IKU1). Dalam rangka memperbaiki dan menyederhanakan format data, maka diputuskan untuk menghapus kolom 'Nama' karena kolom 'NPM' sudah mencakup informasi identitas yang diperlukan. Hal ini dilakukan untuk menghindari duplikasi data dan memastikan konsistensi. Selain itu, penghapusan kolom 'Provinsi Tempat Bekerja' karena informasinya sudah termasuk dalam kolom '1.2 UMP'. Penghapusan kolom 'Nama Tempat Bekerja' juga akan membantu mencapai keseragaman dan konsistensi data yang digunakan dalam penelitian ini.

Dengan demikian atribut yang akan digunakan dalam penelitian ini dapat dilihat pada Tabel 3.

Tabel 3. Atribut Terpilih

No	Atribut Terpilih	Keterangan
1	Npm	ID
2	Tahun Lulus	Atribut Utama
3	Status Bekerja	Atribut Utama
4	Gaji	Atribut Utama
5	Mendapat Pekerjaan Kurang dari 6 Bulan	Atribut Utama
6	1.2 x UMP	Atribut Utama
7	IKU1	Kelas Target
8	Keterkaitan Bidang Studi dengan Pekerjaan	Atribut Utama
9	Kesesuaian Tingkat Pendidikan dengan Pekerjaan	Atribut Utama
10	Jenis Tempat Bekerja	Atribut Utama
11	Sumber Dana Kuliah	Atribut Utama
12	a1	Atribut Utama
13	a2	Atribut Utama
14	a3	Atribut Utama
15	a4	Atribut Utama
16	a5	Atribut Utama
17	b1	Atribut Utama
18	b2	Atribut Utama
19	b3	Atribut Utama
20	b4	Atribut Utama
21	b5	Atribut Utama

B. Data Preprocessing

Tahapan selanjutnya adalah *data preprocessing*, yang melibatkan beberapa langkah penting dalam

menangani kualitas data sebelum proses *data mining*. Berikut adalah urutan tahapan yang diperlukan:

- a. Pengecekan Tipe Data pada Dataset
- b. Pengecekan *Missing Value*
- c. Pengecekan *Data Outlier*
- d. Pengecekan *Data Duplicate*

C. Data Transformation

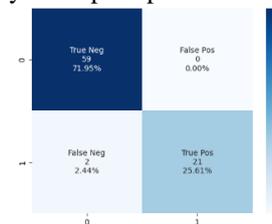
Pada tahapan ini, melakukan proses transformasi/permodelan data dengan tujuan untuk memudahkan data yang digunakan. Pemodelan hanya bisa jika data bertipe numerik, maka perlu ditransformasi lagi agar data diubah menjadi numerik. Oleh karena itu data yang bertipe kategori harus diubah terlebih dahulu menggunakan bantuan *LabelEncoder*.

D. Data Mining

1. Model Comparison

a. Skenario A - Model 1

Dari 82 entri data yang merupakan *data testing* didapat hasil yaitu seperti pada Gambar 2.



Gambar 2. Confussion matrix skenario A model 1

Berdasarkan *Confussion Matrix* tersebut, dapat dihitung beberapa metrik evaluasi performa model klasifikasi, sebagaimana pada Gambar 3.

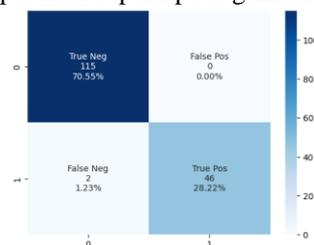
Classification Report 1:

	precision	recall	f1-score	support
0	1.00	0.91	0.95	23
1	0.97	1.00	0.98	59
accuracy			0.98	82
macro avg	0.98	0.96	0.97	82
weighted avg	0.98	0.98	0.98	82

Gambar 3. Classification report model 1

b. Skenario B – Model 2

Dari 163 entri data yang merupakan *data testing* didapat hasil seperti pada gambar 4.



Gambar 4. Confussion matrix model 1

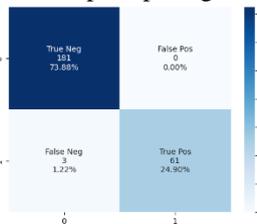
Berdasarkan *Confusion Matrix* tersebut, dapat dihitung beberapa metrik evaluasi performa model klasifikasi, sebagaimana terlihat pada gambar 5.

Classification Report 2:				
	precision	recall	f1-score	support
0	1.00	0.96	0.98	48
1	0.98	1.00	0.99	115
accuracy			0.99	163
macro avg	0.99	0.98	0.99	163
weighted avg	0.99	0.99	0.99	163

Gambar 5. Classification report Model 2

c. Skenario C – Model 3

Dari 245 entri data yang merupakan data testing didapat hasil seperti pada gambar 6.



Gambar 6. Confusion matrix skenario C model 3

Berdasarkan Confusion Matrix tersebut, dapat dihitung beberapa metrik evaluasi performa model klasifikasi, sebagaimana terlihat pada gambar 7.

Classification Report 3:				
	precision	recall	f1-score	support
0	1.00	0.95	0.98	64
1	0.98	1.00	0.99	181
accuracy			0.99	245
macro avg	0.99	0.98	0.98	245
weighted avg	0.99	0.99	0.99	245

Gambar 7. Classification report model 3

2. Feature Forward Selection Comparison

Tahapan ini dilakukan untuk menentukan atribut dengan akurasi terbaik sehingga membuat model baru yang lebih akurat dengan menggunakan atribut yang lebih sedikit dan mempengaruhi IKU1.

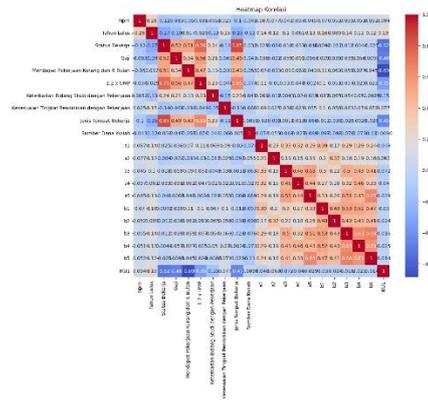
Penerapan Forward Selection menghasilkan fitur terbaik yang didapat dari dataset. Dari 815 entri data dan 20 atribut yang merupakan data testing didapat hasil yaitu seperti pada gambar 8.

```

Fitur terbaik:
Npm
Tahun Lulus
Status Bekerja
Gaji
Mendapat Pekerjaan Kurang dari 6 Bulan
1.2 x UMP
Akurasi terbaik: 0.9877300613496932
    
```

Gambar 8. Hasil forward selection

Berdasarkan gambar 9 fitur terpilih berdasarkan hasil Forward Selection adalah “Npm”, “Tahun Lulus”, “Status Bekerja”, “Gaji”, “Mendapat Pekerjaan Kurang dari 6 Bulan”, dan “1.2 UMP”. Namun jika dilihat dari penggunaannya, “Npm” belum dapat digunakan sebagai fitur terbaik karena Npm merupakan atribut ID. Oleh karena itu, untuk mendukung hasil forward selection dapat diperkuat dengan menggunakan visualisasi korelasi fitur menggunakan heatmap pada gambar 9.

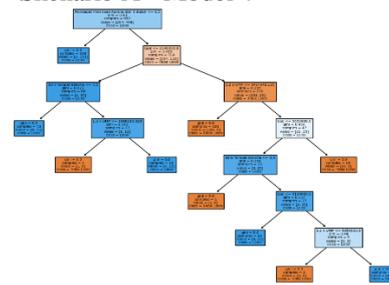


Gambar 9. Heatmap korelasi fitur dengan kelas target

Dengan melihat hasil dari feature Selection dimana atribut terpilih adalah “Npm”, “Tahun Lulus”, “Status Bekerja”, “Gaji”, “Mendapat Pekerjaan Kurang dari 6 Bulan”, dan “1.2 UMP” yang berada pada nilai koefisien terbesar -0.1. Maka atribut terpilih yang didapat dari heatmap berada pada nilai koefisien dibawah -0.1 yang memiliki warna lebih pekat. Atribut terpilih berdasarkan heatmap membentuk model baru dengan dataset yang berisi 5 atribut diantaranya: “Status Bekerja”, “Gaji”, “Mendapat Pekerjaan Kurang dari 6 Bulan”, “1.2 UMP” dan “Jenis Tempat Bekerja”. Dataset ini kemudian digunakan sebagai data latih baru yang di deskripsikan dengan X2.

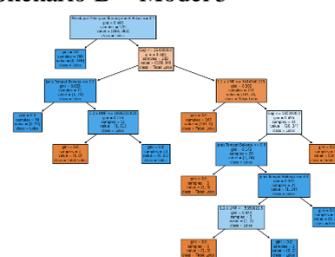
Pembuatan model dengan feature selection tidak jauh berbeda dengan tanpa feature selection, model dibuat dengan 3 buah skenario pembagian dataset baru kemudian algoritma Decision Tree C4.5 diterapkan ke dalam data tersebut dan disimpan ke dalam variabel “model\_4” untuk skenario A, “model\_5” untuk skenario B, dan “model\_6” untuk skenario C. Hasilnya berupa akurasi dari kinerja model pada setiap skenario yang telah dibuat.

a. Skenario A - Model 4



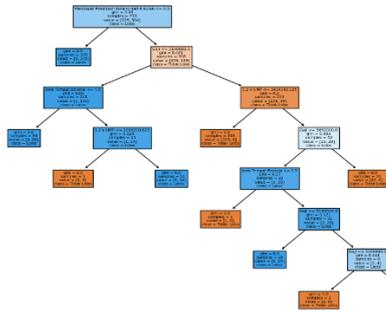
Gambar 10. Decision tree C4.5 pada model 4

b. Skenario B – Model 5



Gambar 11. Decision tree C4.5 pada model 5

c. Skenario C – Model 6

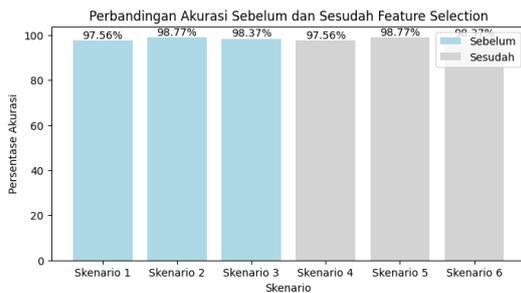


Gambar 12. Decision tree C4.5 pada model 6

E. Evaluasi

Pada tahap evaluasi ini, dilakukan penilaian terhadap semua model yang telah dibangun untuk mengidentifikasi model yang paling cocok atau optimal. Evaluasi dilakukan dengan menggunakan metrik atau kriteria tertentu yang relevan. Proses evaluasi bertujuan untuk mengevaluasi performa dan kualitas model dalam memprediksi atau mengklasifikasikan data baru.

a. Akurasi Model



Gambar 13. Perbandingan akurasi model

Grafik pada gambar 13 menunjukkan akurasi model dari setiap kondisi dan skenario. Dapat diketahui bahwa model yang dibangun sebelum menggunakan *feature selection* sudah menunjukkan akurasi yang sangat tinggi. Jika dibandingkan dengan sesudah menggunakan *feature selection* hasilnya tidak mengalami perubahan meskipun atribut yang digunakan mengalami penyusutan hingga tersisa 5 atribut terbaik yang dipilih dari *feature selection*.

Sebelum menggunakan *feature selection*, model 2 menunjukkan akurasi tertinggi. Sedangkan sesudah penggunaan *feature selection*, model 5 menunjukkan akurasi tertinggi dengan hanya menggunakan 5 atribut terpilih. Model tersebut memiliki akurasi 98.77% sehingga dapat menghasilkan *rules* dari permodelan pohon keputusan yang baik. Maka dari hasil tersebut, pohon keputusan yang diambil adalah model 5.

b. KFold Cross Validation

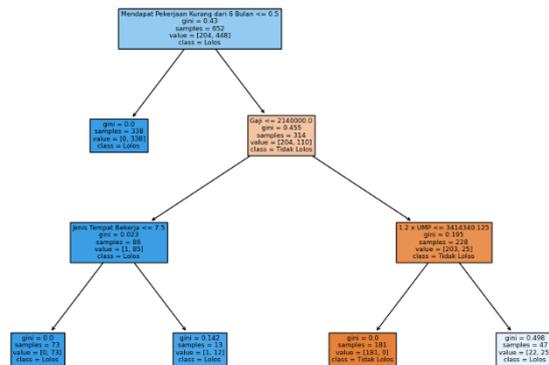
Data pelatihan yang diperoleh dari *Cross Validation* memiliki performa yang baik. Hal ini dapat dibuktikan melalui tabel 4.

Tabel 4. Perbandingan hasil evaluasi 10-fold cross validation model

Percobaan	Sebelum Feature Selection		Sesudah Feature Selection	
	Skor	Rata-Rata Skor	Skor	Rata-rata Skor
1	1.0	0.986	1.0	0.986
2	1.0		1.0	
3	1.0		1.0	
4	0.987		0.987	
5	1.0		1.0	
6	0.987		0.987	
7	0.975		0.975	
8	0.975		0.975	
9	0.962		0.962	
10	1.0		1.0	

Pada tabel 3 hasil dari *cross validation* setelah dilakukan *feature selection* memiliki performa yang baik dalam data pengujian sehingga model dapat memprediksi data baru dengan benar. Skor tersebut menunjukkan bahwa model dapat memprediksi dengan benar sekitar 98% dari seluruh *data testing* yang diberikan. Meskipun rata-rata skor menunjukkan hasil yang sama dengan sebelum menggunakan *feature selection*, rata-rata tersebut tidak mengalami penurunan sehingga penggunaan *feature selection* dapat memperkecil dataset namun tidak mengurangi akurasi. Dapat disimpulkan juga bahwa model mampu melakukan prediksi dengan akurasi yang tinggi pada *data testing*.

Melalui penggunaan model 5 dengan *feature selection* yang merupakan model dengan performa terbaik, terbentuklah sebuah pohon keputusan yang menghasilkan cabang-cabang dengan keputusan yang serupa atau tidak signifikan dalam memprediksi kelas target pada gambar 11. Pohon keputusan ini memiliki tingkat kompleksitas yang tinggi dan sulit untuk dipahami, sehingga membuat perumusan aturan (*rules*) menjadi sulit. Untuk mengatasi hal ini, diperlukan proses *pruning* atau pemangkasan yang bertujuan untuk menyederhanakan struktur pohon keputusan yang kompleks tersebut. Hasil dapat dilihat pada gambar 14.



Gambar 14. Decision tree setelah di-pruning

## 5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah bahwa penerapan algoritma C4.5 dengan metode *feature selection* pada capaian IKU1 telah menghasilkan model keputusan yang signifikan. Pemilihan fitur melalui *forward selection* dan validasi dengan *heatmap* telah membantu mengidentifikasi fitur-fitur kritis seperti "Status Bekerja", "Gaji", "Mendapat Pekerjaan Kurang dari 6 Bulan", "1.2 x UMP", dan "Jenis Tempat Bekerja" yang berpengaruh terhadap IKU1. Dalam implementasi *decision tree* C4.5 dengan metode KDD, model terbaik dari 3 model yang dihasilkan setelah fitur dipilih adalah model 5 (skenario B), mencapai akurasi 98.77% dengan *metrics Precision, Recall*, dan *f1-Score* yang tinggi. Saran untuk penelitian mendatang meliputi eksplorasi analisis lebih komprehensif terhadap seluruh IKU, penggunaan metodologi seperti CRISP-DM, eksperimen dengan algoritma klasifikasi berbeda, dan pertimbangan metode seleksi fitur lainnya untuk memperoleh hasil yang lebih optimal.

## DAFTAR PUSTAKA

- [1] Adnyana (2020) *PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK PREDIKSI WAKTU TUNGGU ALUMNI MENDAPATKAN PEKERJAAN (STUDI KASUS: FASILKOM UNSIKA)*.
- [2] Bode, A. (2018) 'K-NEAREST NEIGHBOR DENGAN FEATURE SELECTION MENGGUNAKAN BACKWARD ELIMINATION UNTUK PREDIKSI HARGA KOMODITI KOPI ARABIKA', *Ilkom Jurnal Ilmiah*, 9, pp. 188–195. Available at: <https://doi.org/https://doi.org/10.33096/ilkom.v9i2.139.188-195>.
- [3] Dirjendikti (2020) 'Buku-Panduan-Indikator-Kinerja-Utama-PTN (1)', in P. Nurwadani (ed.). Jakarta: Direktorat Jendral Pendidikan Tinggi - Jakarta. Available at: <http://dikti.kemdikbud.go.id/wp-content/uploads/2021/06/Buku-Panduan-IKU-2021-28062021.pdf>.
- [4] Kadiyala, A. and Kumar, A. (2018) 'Applications of Python to evaluate environmental data science problems', *Environmental Progress and Sustainable Energy*. John Wiley and Sons Inc., pp. 1580–1586. Available at: <https://doi.org/10.1002/ep.12786>.
- [5] Kempten, W. (2020) 'Overview of Current State of the Art Lane Detection Models and Methods and Testing and Training a SCNN on Google Colaboratory Felix Pius Umscheid', *Research Gate* [Preprint], (December). Available at: <https://doi.org/10.13140/RG.2.2.25239.44962>.
- [6] Miranda, A. and Muslim Lhaksana, K. (2022) 'Classification Analysis of Waiting Period for Telkom University Alumni to Get Jobs Using Decision Tree and Support Vector Machine', *Technology and Science (BITS)*, 4(2). Available at: <https://doi.org/10.47065/bits.v4i2.1963>.
- [7] Mondal, Z. (2022) 'Application of Corporate Governance indices to predict Bankruptcy of Companies using Machine Learning', in *Master Thesis*, pp. 1–104.
- [8] Putri, D.Y., Andreswari, R. and Hasibuan, M.A. (2018) 'Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study: Information Systems Students of Telkom University', in *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*. Medan: International Conference on Cyber and IT Service Management (CITSM). Available at: <https://doi.org/https://doi.org/10.1109/CITSM.2018.8674366>.
- [9] Rezkika, F. et al. (2021) 'KLASIFIKASI MASA TUNGGU ALUMNI UNTUK MENDAPATKAN PEKERJAAN MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus: Fasilkom Unsika)', *Progresif: Jurnal Ilmiah Komputer*, 17, pp. 95–106. Available at: <https://doi.org/10.35889/progresif.v17i2.652>.
- [10] Rizmayanti, A.I. et al. (2021) '9–18 Diterima Februari 10', *JURNAL SWABUMI*, 9(1), p. 2021.
- [11] Sudrajat, R., Irianingsih, I. and Krisnawan, D. (2018) 'Analysis of data mining classification by comparison of C4.5 and ID algorithms', in *IOP Conference Series: Materials Science and Engineering*. Institute of Physics Publishing. Available at: <https://doi.org/10.1088/1757-899X/166/1/012031>.