

**OPTIMASI ALGORITMA *DECISION TREE C4.5* MENGGUNAKAN  
METODE *CORRELATION-BASED FEATURE SELECTION* PADA  
KLASIFIKASI NILAI MAHASISWA**

**STUDI KASUS : UNIVERSITAS MUHAMMADIYAH  
KALIMANTAN TIMUR**

**SKRIPSI**

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar  
sarjana komputer

**DISUSUN OLEH :**

**KAMIRUL HAKIM**

**1911102441124**



**PROGRAM STUDI S1 TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR  
SAMARINDA**

**2023**

**Optimasi Algoritma *Decision Tree C4.5* menggunakan Metode  
*Correlation-Based Feature Selection* pada Klasifikasi Nilai**

**Mahasiswa**

**Studi Kasus : Universitas Muhammadiyah  
Kalimantan Timur**

**SKRIPSI**

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar  
sarjana komputer

**Disusun Oleh :**

**Kamirul Hakim**

**1911102441124**



**PROGRAM STUDI S1 TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR  
SAMARINDA**

**2023**

## HALAMAN PENGESAHAN

### OPTIMASI ALGORITMA *DECISION TREE* C4.5 MENGGUNAKAN METODE *CORRELATION-BASED FEATURE SELECTION* PADA KLASIFIKASI NILAI MAHASISWA

DISUSUN OLEH :

**KAMIRUL HAKIM**

**1911102441124**

Telah melaksanakan ujian skripsi dan dinyatakan lulus,

Pada tanggal 13 Juli 2023

Dosen Pembimbing

Wawan Joko Pranoto, S.Kom., M.Ti  
NIDN. 1102057701

Penguji

Faldi, S.Kom., M.Ti  
NIDN. 1121079101

Dekan

Prof. Ir. Sarjito, MT., Ph.D  
NIDN. 0610116204

Ketua Program Studi

Assijalohar Latipah, M.Cs  
NIDN. 1124098902

## PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan dibawah ini,

Nama Mahasiswa : Kamirul Hakim

Nim : 1911102441124

Konsentrasi : Sistem Informasi

Menyatakan skripsi dengan judul berikut :

**Optimasi Algoritma *Decision Tree* C4.5 Menggunakan Metode *Correlation-Based Feature Selection* Pada Klasifikasi Nilai Mahasiswa Studi Kasus Universitas Muhammadiyah Kalimantan Timur**

Dosen Pembimbing : Wawan Joko Pranoto, S.Kom, M.TI

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar Sarjana Komputer, baik di Universitas Muhammadiyah Kalimantan Timur (UMKT) maupun diperguruan tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas Muhammadiyah Kalimantan Timur (UMKT)
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi muhammadiyah.

Samarinda , 13 Juli 2023

Yang Mengesahkan,



Kamirul Hakim

## PRAKATA



Alhamdulillah, puji syukur saya panjatkan kehadiran Allah SWT atas rahmat dan kasih sayang-Nya sehingga penulis dapat menyelesaikan proposal yang berjudul “Optimasi Algoritma *Decision Tree* C4.5 Menggunakan Metode *Correlation-Based Feature Selection* Klasifikasi Nilai Mahasiswa”. Tujuan penulisan proposal ini adalah untuk memenuhi salah satu syarat untuk memperoleh gelar sarjana dari Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur. Penulis menyadari sepenuhnya bahwa penelitian dan penulisan proposal ini tidak akan berjalan sebagaimana mestinya tanpa dukungan dari berbagai pihak. Banyak rasa hormat dan terima kasih kepada:

1. Bapak Ramli dan Ibu Sahari Malam selaku orangtua kandung yang sudah terlalu banyak berkorban dan mendukung hingga saat ini
2. Saudara yang saya banggakan Karman S.Kom, Kartini, Kartina, Kartika Yana Dan Kahar Lutviansyah selaku Saudara Kandung saya yang memberikan dukungan kepada saya.
3. Bapak Wawan Joko Pranoto, S.Kom., M.TI selaku dosen pembimbing skripsi yang sudah banyak meluangkan waktu, tenaga dan juga memberikan banyak saran
4. Ibu Asslia Johar Latipah, M.Cs sebagai Ketua Program Studi Teknik Informatika di Universitas Muhammadiyah Kalimantan Timur
5. Bapak Prof. Ir. Sarjito, M.T.,Ph.D., selaku Dekan Fakultas Sains dan Teknologi Universitas Muhammadiyah Kalimantan Timur
6. Bapak Prof. Dr. H. Bambang Setiaji, selaku Rektor Universitas Muhammadiyah Kalimantan Timur
7. Seluruh dosen baik dari Program Studi Teknik Informatika atau pun program studi lain di Universitas Muhammadiyah Kalimantan Timur yang sudah memberikan warisan istimewa berupa ilmu.

8. Putra Saymen, Muhammad Fath Thoriq, Suci Mawaddah, Rahmat Ramadhani, Septia Intan Permatasari dan Dilla Nur Fadillah sebagai rekan mahasiswa satu bimbingan yang selalu menemani serta menjadi sistem pendukung
9. Seluruh rekan mahasiswa Program Studi Teknik Informatika angkatan tahun 2019 dan kepada semua pihak yang tidak bisa disebutkan satu persatu.

Mengingat penelitian dan penulisan Skripsi ini masih jauh dari kesempurnaan, maka dengan segala hormat dan kerendahan hati, kritik dan saran yang bersifat membangun sangat diharapkan untuk menyempurnakan penelitian ini agar mendapatkan hasil yang maksimal.

Samarinda, 13 Juli 2023

Kamirul Hakim

## ABSTRAK

Universitas Muhammadiyah Kalimantan Timur (UMKT) menerapkan sistem pembelajaran hybrid (daring dan luring) karena keterbatasan ruangan dan jumlah mahasiswa yang besar. Namun, pembelajaran jarak jauh memiliki dampak negatif terhadap hasil belajar mahasiswa. Data juga menunjukkan penurunan nilai rata-rata pada Mata Kuliah Dasar Umum Bahasa Arab di UMKT, yang mungkin disebabkan oleh pembelajaran hybrid yang tidak efektif dan tantangan lainnya. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang berpengaruh dalam klasifikasi nilai Bahasa Arab mahasiswa. Hasil penelitian menunjukkan bahwa Fitur-fitur yang berpengaruh signifikan dalam pengambilan keputusan adalah "Comments", "Tugas\_2", "UTS", dan "Tugas\_1", sementara jenis kelamin tidak memiliki pengaruh signifikan. Evaluasi model menggunakan validasi silang dengan rata-rata k-fold menunjukkan bahwa sebelum dilakukan seleksi fitur menggunakan metode CFS, akurasi mencapai 88%. Namun, setelah seleksi fitur menggunakan CFS, akurasi meningkat menjadi 92%. Hasil ini menunjukkan bahwa seleksi fitur dengan menggunakan metode CFS dapat signifikan meningkatkan kinerja klasifikasi dalam penilaian nilai Bahasa Arab mahasiswa.

**Kata Kunci :** Akurasi, Optimasi, Algoritma C4.5, CFS.

## **ABSTRACT**

*East Kalimantan Muhammadiyah University (UMKT) implements a hybrid learning system (online and offline) due to limited space and a large number of students. However, distance learning has negative impacts on students' learning outcomes. Data also indicates a decrease in the average grades for the General Arabic Language course at UMKT, possibly due to ineffective hybrid learning and other challenges. This research aims to analyze the influential factors in classifying students' Arabic Language grades. The research results show that the significant features in decision-making are "Comments," "Assignment\_2," "Midterm Exam," and "Assignment\_1," while gender does not have a significant influence. Model evaluation using cross-validation with the average k-fold demonstrates that before feature selection using the CFS method, the accuracy reaches 88%. However, after feature selection using CFS, the accuracy increases to 92%. These results indicate that feature selection using the CFS method can significantly improve the classification performance in assessing students' Arabic Language grades.*

**Keywords:** Accuracy, Optimization, C4.5 Algorithm, CFS.



# DAFTAR ISI

<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>PERNYATAAN KEASLIAN SKRIPSI.....</b>	<b>iii</b>
<b>PRAKATA .....</b>	<b>iv</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRACT .....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR GAMBAR .....</b>	<b>xii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiii</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
<b>BAB 2 TINJAUAN PUSTAKA.....</b>	<b>5</b>
2.1 Kinerja Mahasiswa .....	5
2.2 Data Mining.....	5
2.3 Klasifikasi .....	6
2.4 Python .....	6
2.5 Algoritma Decision Tree C 4.5 .....	6
2.6 Correlation-Based Feature Selection .....	8
2.7 Data Preprocessing .....	9
2.8 Cross Validation.....	10
2.9 Confusion Matrix.....	10
2.10 CRISP-DM .....	11

2.11 Penelitian Terdahulu .....	12
<b>BAB 3 METODOLOGI PENELITIAN .....</b>	<b>14</b>
3.1 Objek Penelitian .....	14
3.2 Teknik Pengumpulan Data.....	14
3.3 Alur Penelitian.....	15
3.4 Peralatan Yang Digunakan.....	19
3.5.1. Hardware (Perangkat Keras) .....	19
3.5.2. Software (Perangkat Lunak) .....	19
<b>BAB 4 HASIL DAN PEMBAHASAN .....</b>	<b>21</b>
4.1. Business Understanding (Pemahaman Bisnis).....	21
4.2. Data Understanding (Pemahaman Data) .....	21
4.3. Data Preparation ( Pengolahan Data) .....	25
4.3.1. Data Selection Dan Data Integration.....	25
4.3.2. Data Transformation .....	25
4.3.3. Cross Validation 10 K-Fold .....	27
4.4. Modeling (Pemodelan) .....	35
4.4.1. Perhitungan Algoritma C4.5.....	35
4.4.2. Pemodelan Algoritma C4.5 Dengan Python .....	38
4.4.3. Pemodelan Algoritma C4.5 Dan Metode CFS .....	42
4.5. Pembahasan.....	46
<b>BAB 5 PENUTUP .....</b>	<b>48</b>
<b>DAFTAR PUSTAKA .....</b>	<b>49</b>
<b>LAMPIRAN .....</b>	<b>52</b>

## DAFTAR TABEL

Table 3. 1 Keterangan Atribut Data Openlarning .....	15
Table 3. 2 Keterangan Atribut Data MKDU.....	16
Table 3. 3 Proses Seleksi Dan Integrasi Data .....	17
Tabel 4. 1 Data Nilai Bahasa Arab Dari BAA .....	21
Tabel 4. 2 Data Openlearning Dari MKDU.....	22
Tabel 4. 3 Data Data Seleksi Dan Data Integrasi Tahap Awal .....	25
Tabel 4. 4 Hasil Transformasi Data Atribut Time Spent on Course.....	26
Tabel 4. 5 Transformasi Atribut Jenis Kelamin.....	26
Tabel 4. 6 Hasil Transformasi Data Atribut Simbol .....	27
Tabel 4. 7 Data Latih Fold 1 .....	28
Tabel 4. 8 Data Uji Fold 1.....	28
Tabel 4. 9 Data Latih Fold 2 .....	28
Tabel 4. 10 Data Uji Fold 2.....	29
Tabel 4. 11 Data Latih Fold 3 .....	29
Tabel 4. 12 Data Uji Fold 3.....	30
Tabel 4. 13 Data Latih Fold 4 .....	30
Tabel 4. 14 Data Uji Fold 4.....	30
Tabel 4. 15 Data Latih Fold 5 .....	31
Tabel 4. 16 Data Uji Fold 5.....	31
Tabel 4. 17 Data Latih Fold 6 .....	32
Tabel 4. 18 Data Uji Fold 6.....	32
Tabel 4. 19 Data Latih Fold 7 .....	32
Tabel 4. 20 Data Uji Fold 7.....	33
Tabel 4. 21 Data Latih Fold 8 .....	33
Tabel 4. 22 Data Uji Fold 8.....	33
Tabel 4. 23 Data Latih Fold 9 .....	34
Tabel 4. 24 Data Uji Fold 9.....	34
Tabel 4. 25 Data Latih Fold 10 .....	35

Tabel 4. 26 Data Uji Fold 10.....	35
Tabel 4. 27 Entropy Atribut Jenis_Kelamin.....	36
Tabel 4. 28 Confusion Matrix Fold 1.....	37
Tabel 4. 29 Confusion Matrix Fold 2.....	37
Tabel 4. 30 Hasil Evaluasi 10 K-Fold.....	38
Tabel 4. 31 Dataset.....	38
Tabel 4. 32 Rata-Rata Confusion Matrix.....	40
Tabel 4. 34 Confusion Matrix .....	44

## DAFTAR GAMBAR

Gambar 2. 1 CRISP-DM.....	11
Gambar 3. 1 Tahapan Alur Penelitian.....	15
Gambar 4. 1 Confusion Matrix.....	39
Gambar 4. 2 Confusion Matrix Lanjutan.....	39
Gambar 4. 3 Rata-Rata Akurasi Confusion Matrix.....	39
Gambar 4. 4 Pohon Keputusan.....	41
Gambar 4. 5 Hasil Seleksi Fitur CFS.....	42
Gambar 4. 6 Hasil Confusion Matrix CFS.....	43
Gambar 4. 7 Hasil Confusion Matrix Lanjutan CFS.....	43
Gambar 4. 8 Rata-Rata Confusion Matrix dengan CFS.....	43
Gambar 4. 9 Pohon Keputusan Seleksi Fitur CFS.....	45

## DAFTAR LAMPIRAN

Lampiran 1 Riwayat Hidup .....	53
Lampiran 2 Data Gabungan MKDU dan BAA Nilai Bahasa Arab.....	54
Lampiran 3 Data Selection Dan Data Integration .....	55
Lampiran 4 Data Seleksi Fitur CFS .....	56
Lampiran 5 Source Code .....	57
Lampiran 6 Lembar Konsultasi .....	60
Lampiran 7 Surat Izin Pengambilan Data .....	62
Lampiran 8 Lembar Tindak Lanjut Revisi.....	64
Lampiran 9 Surat Keterangan Tidak Menggunakan Uji Validitas .....	65
Lampiran 10 Hasil Uji Turnitin .....	66