

**ANALISIS SELEKSI FITUR BERBASIS *CORRELATION* PEARSON
PADA ALGORITMA *K-NEAREST NEIGHBOR* DALAM
KLASIFIKASI NILAI MAHASISWA
STUDI KASUS: UNIVERSITAS MUHAMMADIYAH
KALIMANTAN TIMUR**

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar
Sarjana Komputer

DISUSUN OLEH:

PUTRA SAYMEN

1911102441011



**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS & TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
KOTA SAMARINDA
2023**

**Analisis Seleksi Fitur berbasis *Correlation* Pearson pada
Algoritma *K-Nearest Neighbor* dalam Klasifikasi Nilai
Mahasiswa**

Studi Kasus: Universitas Muhammadiyah Kalimantan Timur

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar
Sarjana Komputer

Disusun Oleh:

Putra Saymen

1911102441011



PROGRAM STUDI S1 TEKNIK INFORMATIKA

FAKULTAS SAINS & TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

KOTA SAMARINDA

2023

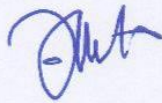
HALAMAN PENGESAHAN

Analisis Seleksi Fitur Berbasis Correlation Pearson pada Algoritma K-Nearest Neighbor dalam Klasifikasi Nilai Mahasiswa

**DISUSUN OLEH:
PUTRA SAYMEN
1911102441011**

Telah melaksanakan ujian skripsi dan dinyatakan lulus,
Pada tanggal 12 Juli 2023

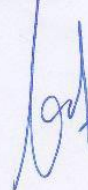
Dosen Pembimbing



Wawan Joko Pranoto, S.Kom, M.TI

NIDN. 1102057701

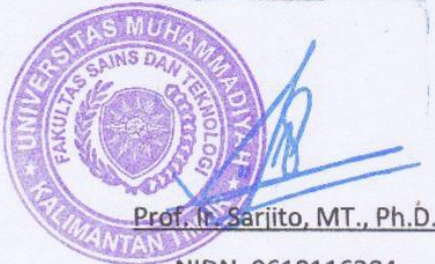
Penguji



Faldi, S.Kom., M.TI

NIDN. 1121079101

Dekan



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

Ketua Program Studi



Asslia Johar Latipah, S.Kom., M.Cs
NIDN. 1124098902

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Putra Saymen

NIM : 1911102441011

Konsentrasi : Sistem Informasi

Menyatakan bahwa skripsi dengan judul berikut:

Analisis Seleksi Fitur berbasis *Correlation Pearson* pada Algoritma *K-Nearest Neighbor* dalam Klasifikasi Nilai Mahasiswa.

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 di Perguruan 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 UMKT.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran 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 25 Juni 2023



Yang Menyatakan,

Putra Saymen

1911102441011

KATA PENGANTAR



Alhamdulillah puji syukur kepada Allah SWT berkat Rahmat dan Karunia-Nya sehingga penulis dapat menyelesaikan proposal dengan judul, “Analisis Seleksi Fitur berbasis *Correlation* Pearson pada Algoritma *k-Nearest Neighbor* dalam Klasifikasi Nilai Mahasiswa”. Adapun tujuan dari penulisan proposal ini adalah untuk memenuhi salah satu persyaratan penyelesaian studi di jurusan Teknik Informatika Fakultas Sains & Teknologi Universitas Muhammadiyah Kalimantan Timur.

Penulis menyadari betul bahwa proses penelitian dan penulisan proposal ini tidak akan berjalan dengan sebagaimana mestinya tanpa adanya dukungan dari berbagai pihak. Oleh karena itu, diucapkan rasa hormat dan terima kasih kepada:

1. Bapak Saimun dan Ibu Suntiati selaku orangtua kandung yang sudah terlalu banyak berkorban dan mendukung hingga saat ini
2. Rina Wiji Astuti, Hasan, Imron Wahyudi, S.Pd, Irawaty Jamaluddin, Elis Tiowati, S.Pd dan Hendri Aspriyadi, S.Pd yang juga turut serta memberikan dukungan sebagai saudara dan melengkapi banyak sumber daya yang dibutuhkan
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. Kamirul Hakim, Muhammad Fath Thoriq, Suci Mawaddah, Rahmat Ramadhani, Septia Intan Permata Sari dan Yuliana Dilla Evitasari 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.

Menyadari penelitian dan penulisan proposal ini masih jauh dari kata sempurna, maka dengan segala hormat serta kerendahan hati, diharapkan adanya kritik yang membangun dan saran yang bermanfaat untuk menyempurnakan penelitian ini guna mencapai hasil yang maksimal.

Samarinda, 25 Maret 2023

A handwritten signature in black ink, consisting of a large, stylized 'P' followed by a horizontal line and a vertical stroke that extends downwards.

Putra Saymen

ABSTRAK

Berdasarkan data dari Mata Kuliah Dasar Umum (MKDU) di Universitas Muhammadiyah Kalimantan Timur (UMKT) menunjukkan bahwa terjadi penurunan kualitas nilai mahasiswa di UMKT pada tahun 2020-2021. Salah satu mata kuliah yang mengalami penurunan adalah Bahasa Arab. Penelitian ini bertujuan untuk mengklasifikasi data nilai Bahasa Arab di UMKT melalui teknik *Data Mining* menggunakan algoritma *K-Nearest Neighbor* (K-NN) dengan rumus jarak *Euclidean Distance* serta pembagian data *5-Fold Cross-validation*. Kemudian, dilakukan seleksi fitur berbasis *Correlation Pearson* untuk melihat atribut data yang paling berpengaruh sekaligus melihat apakah ada perubahan akurasi pada algoritma K-NN. Hasil penelitian menunjukkan bahwa atribut yang paling berpengaruh adalah *Time_spent_on_course*, *Progress* dan *Course_completed*. Didapatkan K=5 memiliki akurasi rata-rata dari semua *fold* dengan nilai 66%, dan setelah melakukan seleksi fitur hanya naik sebesar 1%.

Kata Kunci: nilai mahasiswa; *Data Mining*; *K-Nearest Neighbor*; *Euclidean Distance*; *5-Fold Cross-validation*; *Correlation Pearson*

ABSTRACT

Based on the data from the General Basic Course (MKDU) at the University of Muhammadiyah East Kalimantan (UMKT), there has been a decrease in the quality of student grades at UMKT in the years 2020-2021. One of the courses that experienced a decline is Arabic Language. This research aims to classify the data of Arabic Language grades at UMKT through Data Mining techniques using the K-Nearest Neighbor (K-NN) algorithm with the Euclidean Distance formula and a 5-Fold Cross-validation data partitioning. Then, feature selection based on Pearson's Correlation is performed to observe the most influential data attributes and to see if there is any change in accuracy in the K-NN algorithm. The research results show that the attributes with the most influence are Time_spent_on_course, Progress, and Course_completed. It was found that K=5 has an average accuracy across all folds of 66%, and after performing feature selection, it only increased by 1%.

Keywords: *student grades; Data Mining; K-Nearest Neighbor; Euclidean Distance; 5-Fold Cross-validation; Correlation Pearson*

DAFTAR ISI

HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	2
1.4. Batasan Masalah.....	3
BAB 2 TINJAUAN PUSTAKA	4
2.1. Kinerja Mahasiswa.....	4
2.2. Pembelajaran <i>Hybrid</i>	4
2.3. <i>Data Mining</i>	4
2.3.1. Algoritma Klasifikasi	5
2.3.2. <i>k-Nearest Neighbor</i>	6
2.3.3. <i>Euclidean Distance</i>	6
2.3.4. <i>Data Preprocessing</i>	7
2.3.5. <i>K-Fold Cross-Validation</i>	7
2.3.6. <i>Confusion Matrix</i>	7
2.3.7. CRISP-DM.....	8
2.3.8. <i>Correlation Pearson</i>	10
2.4. Penelitian Terdahulu.....	11
BAB 3 METODOLOGI PENELITIAN	14
3.1. Objek Penelitian	14

3.2. Lokasi Penelitian	14
3.3. Teknik Pengumpulan Data	14
3.4. Teknik Analisis Data	15
3.5. Peralatan yang Digunakan	19
3.6. Jadwal Penelitian	20
BAB 4 HASIL DAN PEMBAHASAN	21
4.1 <i>Data Understanding</i> (Pengumpulan Data)	21
4.2 <i>Data Preparation</i>	23
4.2.1 Data Selection	23
4.2.2 Data Integration	24
4.2.3 <i>Data Transformation</i>	24
4.2.4 <i>Data Reduction</i>	27
4.3 Perhitungan Algoritma <i>k-Nearest Neighbor</i>	28
4.4. Modelling, Seleksi Fitur & Evaluasi	31
4.5 <i>Deployment</i>	36
BAB 5 KESIMPULAN DAN SARAN	37
5.1 Kesimpulan	37
5.2 Saran	37
DAFTAR PUSTAKA.....	38
LAMPIRAN	43

DAFTAR TABEL

Tabel 2. 1 Tabel Nilai Korelasi	11
Tabel 2. 2 Penelitian Terdahulu	11
Tabel 3. 1 Atribut Data MKDU	15
Tabel 3. 2 Atribut Data BAA.....	16
Tabel 3. 3 Atribut Seleksi dan Integrasi Data	17
Tabel 3. 4 Jadwal Penelitian	20
Tabel 4. 1 Data yang Didapat dari BAA.....	21
Tabel 4. 2 Data yang Didapat dari MKDU	22
Tabel 4. 3 Data dari BAA Setelah Melalui Seleksi Tahap Pertama	23
Tabel 4. 4 Data dari MKDU Setelah Melalui Seleksi Tahap Pertama	23
Tabel 4. 5 Hasil Data Integration	24
Tabel 4. 6 Atribut Jenis Kelamin Setelah Tahap Transformasi	25
Tabel 4. 7 Atribut Time Spent on Course Setelah Tahap Transformasi	25
Tabel 4. 8 Atribut Simbol Setelah Tahap Transformasi	26
Tabel 4. 9 Atribut Simbol Setelah Tahap Transformasi Menjadi Numerik.....	26
Tabel 4. 10 Hasil Data Setelah Ditransformasi	27
Tabel 4. 11 Data Setelah Melalui Tahap Reduksi.....	27
Tabel 4. 12 Sampel Data Training	28
Tabel 4. 13 Sampel Data Testing.....	28
Tabel 4. 14 Perhitungan Jarak k-Nearest Neighbor	30
Tabel 4. 15 Tabel Klasifikasi	30
Tabel 4. 16 Confusion Matrix	31
Tabel 4. 17 Total Confusion Matrix K=3.....	31
Tabel 4. 18 Total Confusion Matrix K=5.....	31
Tabel 4. 19 Total Confusion Matrix K=7	31
Tabel 4. 20 Total Confusion Matrix K=9.....	32
Tabel 4. 21 Total Confusion Matrix K=11.....	32
Tabel 4. 22 Hasil Akurasi Uji Data 5-Fold.....	32
Tabel 4. 23 Penentuan Atribut yang Digunakan	33

Tabel 4. 24 Total Confusion Matrix K=3 Setelah Seleksi Fitur	34
Tabel 4. 25 Total Confusion Matrix K=5 Setelah Seleksi Fitur	34
Tabel 4. 26 Total Confusion Matrix K=7 Setelah Seleksi Fitur	34
Tabel 4. 27 Total Confusion Matrix K=9 Setelah Seleksi Fitur	35
Tabel 4. 28 Total Confusion Matrix K=11 Setelah Seleksi Fitur	35
Tabel 4. 29 Perbandingan Hasil Akurasi Rata-rata.....	35

DAFTAR GAMBAR

Gambar 2. 1 Confusion Matrix	8
Gambar 2. 2 Alur CRISP-DM	9
Gambar 3. 1 Alur Analisis Data	15
Gambar 4. 1 Hasil Correlation Pearson	33

DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup.....	44
Lampiran 2. Data Mata Kuliah Bahasa Arab yang Diperoleh dari BAA.....	45
Lampiran 3. Data Mata Kuliah Bahasa Arab yang Diperoleh dari MKDU	47
Lampiran 4. Data BAA Setelah Melalui Tahapan Data Selection	50
Lampiran 5. Data MKDU Setelah Melalui Tahapan Data Selection	51
Lampiran 6. Seluruh Data Setelah Tahap Integrasi dan Transformasi Data	52
Lampiran 7. Seluruh Data Setelah Melalui Tahap Data Reduction.....	53
Lampiran 8. Program Python.....	54
Lampiran 9. Surat Izin Pengambilan Data MKDU	58
Lampiran 10. Surat Izin Pengambilan Data BAA	59
Lampiran 11. Lembar Bimbingan	60
Lampiran 12. Hasil Uji Turnitin	62