

**NASKAH PUBLIKASI (*MANUSCRIPT*)**

**ANALISIS SELEKSI FITUR BERBASIS *CORRELATION PEARSON* PADA  
ALGORITMA *K-NEAREST NEIGHBOR* DALAM KLASIFIKASI NILAI  
MAHASISWA**

***FEATURE SELECTION ANALYSIS BASED ON PEARSON CORRELATION  
IN K-NEAREST NEIGHBOR ALGORITHM FOR STUDENT GRADE  
CLASSIFICATION***

Putra Saymen, Wawan Joko Pranoto, Faldi



**DISUSUN OLEH:**

**PUTRA SAYMEN**

**1911102441011**

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

**2023**

**Naskah Publikasi (*Manuscript*)**

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

***Feature Selection Analysis Based on Pearson Correlation in K-Nearest Neighbor Algorithm for Student Grade Classification***

Putra Saymen, Wawan Joko Pranoto, Faldi



**Disusun Oleh:**

**Putra Saymen**

**1911102441011**

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

**HALAMAN PENGESAHAN**

**ANALISIS SELEKSI FITUR BERBASIS *CORRELATION PEARSON* PADA  
ALGORITMA *K-NEAREST NEIGHBOR* DALAM KLASIFIKASI NILAI  
MAHASISWA**

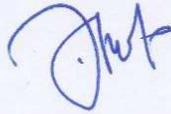
NASKAH PUBLIKASI

DISUSUN OLEH:

**PUTRA SAYMEN**

**1911102441011**

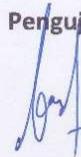
Dosen Pembimbing



**Wawan Joko Pranoto, S.Kom., M.T.I**

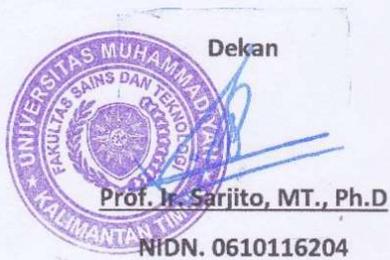
**NIDN. 1102057701**

Penguji



**Faldi, S.Kom., M.T.I**

**NIDN. 1121079101**



Dekan

**Prof. Ir. Sarjito, MT., Ph.D**

**NIDN. 0610116204**



Ketua Program Studi

**Asllia Johar Latipah, S.Kom., M.Cs**

**NIDN. 1124098902**

# ANALISIS SELEKSI FITUR BERBASIS *CORRELATION PEARSON* PADA ALGORITMA *K-NEAREST NEIGHBOR* DALAM KLASIFIKASI NILAI MAHASISWA

Putra Saymen<sup>1\*</sup>, Wawan Joko Pranoto<sup>2</sup>, Faldi<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

email: 1911102441011@umkt.ac.id\*

**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* (KNN) dengan rumus jarak *Euclidean Distance* serta pembagian data *5-Fold Cross-validation* dan evaluasi *Confusion Matrix*. Kemudian, dilakukan seleksi fitur berbasis *Correlation Pearson* untuk melihat atribut data yang paling berpengaruh sekaligus melihat apakah ada perubahan akurasi pada algoritma KNN. 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 1,52%,  $K=7$  dengan nilai akurasi rata-rata 65% mengalami penurunan 3,08%, dan sisanya tidak mengalami perubahan atau tetap.

**Kata Kunci :** Nilai mahasiswa, *data mining*, *K-Nearest Neighbor*, *5-fold Cross-Validation*, *Correlation Pearson*, *Confusion Matrix*.

**Abstract:** Based on the data from the General Basic Course (MKDU) at the University of Muhammadiyah in East Kalimantan (UMKT), it is indicated that there was a decline in the quality of student grades at UMKT in the years 2020-2021. One of the courses that experienced a decline is Arabic Language. The purpose of this study is to classify the data of Arabic Language grades at UMKT using Data Mining techniques, utilizing the K-Nearest Neighbor (KNN) algorithm with the Euclidean Distance formula, 5-Fold Cross-validation for data division, and Confusion Matrix for evaluation. Furthermore, a feature selection based on Pearson Correlation is conducted to identify the most influential data attributes and to observe any changes in accuracy in the KNN algorithm. The research results show that the most influential attributes are *Time\_spent\_on\_course*, *Progress*, and *Course\_completed*. It was found that  $K=5$  had an average accuracy of 66% across all folds, and after the feature selection, the accuracy increased by only 1.52%. However, with  $K=7$ , there was a decrease of 3.08% in average accuracy, while the rest did not experience any changes and remained constant.

**Keywords :** Student performance, *data mining*, *K-Nearest Neighbor*, *5-fold Cross-Validation*, *Correlation Pearson*, *Confusion Matrix*.

## PENDAHULUAN

Institusi pendidikan di Indonesia tentunya memiliki sistem pembelajaran yang beragam. Ada yang menerapkan pembelajaran daring ataupun luring secara konsisten, namun ada pula yang menggabungkan keduanya, hal ini biasa disebut dengan pembelajaran *hybrid*. Salah satu institusi pendidikan yang sedang menerapkan sistem pembelajaran *hybrid* adalah Universitas Muhammadiyah Kalimantan Timur (UMKT). Namun, sistem seperti ini akan mempengaruhi kegiatan belajar-mengajar, khususnya pada faktor penilaian.

Berdasarkan survei dari Pemerintah Indonesia yang dilakukan pada tahun 2020, sebanyak 33% pelajar merasa kesulitan untuk belajar dari rumah karena tidak terbiasa belajar di luar kelas, 28% karena sarana prasarana terbatas, 22% karena keterbatasan pembiayaan, 14% lainnya dan 3% karena terbatasnya dukungan orangtua [1]. Hal ini diperkuat berdasarkan data dari Mata Kuliah Dasar Umum (MKDU) UMKT yang menunjukkan kualitas nilai mahasiswa UMKT dari angkatan tahun 2020 hingga 2021 mengalami

penurunan. Salah satu mata kuliah yang mengalami penurunan nilai adalah Bahasa Arab. Nilai rata-rata mata kuliah Bahasa Arab di UMKT pada angkatan tahun 2017 adalah 57,77. Kemudian, pada angkatan tahun 2018 mengalami peningkatan yaitu 76,61. Lalu di tahun 2019 dan 2020 menjadi lebih tinggi yaitu 82,40 dan 84,80. Tapi pada tahun 2021, nilai Bahasa Arab mengalami penurunan yang cukup dramatis, bahkan lebih rendah dari 3 tahun sebelumnya, yakni 74,18 [2].

Permasalahan yang telah diuraikan di atas dikhawatirkan akan mempengaruhi UMKT sebagai institusi pendidikan untuk menjaga kualitas pembelajaran guna menghasilkan lulusan yang berkualitas. Oleh karena itu, perlu diadakan evaluasi dan monitoring secara dini terhadap kinerja mahasiswa dalam kegiatan pembelajaran. Manfaat yang utama dari evaluasi adalah meningkatkan kualitas pembelajaran dan selanjutnya akan terjadi peningkatan kualitas pendidikan [3]. Demi menjaga kualitas sistem pembelajaran *hybrid*, maka perlu dilakukan klasifikasi terhadap data nilai mahasiswa

Bahasa Arab dalam perkuliahan *hybrid* khususnya di UMKT.

Ada banyak cara yang bisa digunakan untuk mengklasifikasi nilai mahasiswa, salah satunya dengan menerapkan konsep Data Mining. Data Mining sendiri memiliki beberapa algoritma. Salah satu algoritma yang bisa digunakan untuk mengklasifikasi nilai mahasiswa adalah *K-Nearest Neighbor* (KNN). Algoritma KNN dipilih karena merupakan salah satu algoritma yang paling sederhana guna memecahkan masalah klasifikasi, dan sering memberikan hasil yang kompetitif dan signifikan [4]. KNN juga efektif pada *data training* yang besar dan dapat menghasilkan data yang lebih akurat [5]. Berdasarkan studi literatur yang telah dilakukan, dapat dilihat bahwa algoritma KNN memiliki berbagai macam kelebihan. Namun masih terdapat kekurangan dalam algoritma KNN, yaitu memiliki masalah pada pemilihan nilai K [6]. Penentuan nilai K pada algoritma KNN merupakan sesuatu yang sangat penting, karena nilai K adalah jumlah tetangga yang akan diukur berdasarkan metrik jarak yang akan mempengaruhi klasifikasi [7]. Klasifikasi tentunya juga sangat berpengaruh pada nilai akurasi yang dihasilkan. Ada berbagai macam metode yang bisa digunakan untuk melakukan optimasi atau seleksi fitur guna meningkatkan akurasi pada model algoritma, salah satu metode yang bisa digunakan adalah *Correlation Pearson*. *Correlation Pearson* sendiri dipilih karena mampu mengatasi fitur-fitur dengan tipe data campuran [8]. Selain mampu meningkatkan akurasi, *Correlation Pearson* juga efektif dan cepat dalam mengatasi fitur yang sangat banyak [9]. Oleh karena itu, pada penelitian ini, dilakukan analisis seleksi fitur berbasis *Correlation Pearson* pada Algoritma KNN dalam klasifikasi nilai mahasiswa di UMKT.

## TINJAUAN PUSTAKA

### Penelitian Terdahulu

Terdapat beberapa referensi terkait penelitian ini, di antaranya adalah:

Penelitian [10] yang berjudul Penerapan Algoritma *K-Nearest Neighbor* Untuk Memprediksi Kelulusan Mahasiswa. Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa metode KNN dapat digunakan untuk memprediksi kelulusan mahasiswa di STMIK Pelita Nusantara.

Penelitian [11] dengan judul Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Metode *Modified K-Nearest Neighbor* (MK-NN). Di mana pada penelitian tersebut, untuk setiap pengujian yang dilakukan, akurasi dari MK-NN berada di atas akurasi KNN. Sehingga dapat disimpulkan bahwa MK-NN menghasilkan akurasi lebih baik dibanding KNN.

Penelitian [12] berjudul Optimalisasi dalam Mengidentifikasi Mahasiswa Jalur Cepat (*Fast-track*) menggunakan Metode *K-Nearest Neighbor*. Hasil yang diuji untuk memperoleh ketepatan metode *K-*

*Nearest Neighbor* pada saat penyeleksian penerimaan mahasiswa *fast-track* Fakultas MIPA di Universitas Andalas berdasarkan IPK semester 1 sampai dengan 6 dan jumlah SKS sampai dengan semester 6 menggunakan  $k = 3$  adalah 95,79%.

Penelitian [13] berjudul Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Topik Skripsi Mahasiswa di Fakultas Ilmu Komputer. Hasil penelitian menunjukkan optimasi nilai K menggunakan *k-fold cross validation* menghasilkan tingkat akurasi yaitu 56,67% dengan nilai *k-fold cross validation* = 2 dan nilai  $K=5$ .

Penelitian [14] berjudul *Implementation of K-Nearest Neighbour (Knn) Algorithm to Predict Student'S Performance*. Didapatkan nilai K terbaik adalah 3, 6, dan 9 untuk mendapatkan prediksi terbaik. Hasil ini diperoleh dengan mencoba nilai K adalah 3 sampai 60. Nilai prediksi tersebut kemudian dibandingkan, hasil prediksi salah yang persentase terkecil adalah yang terbaik.

Penelitian [15] berjudul Perbandingan Performa Antara Algoritma *Naive Bayes* Dan *K-Nearest Neighbour* Pada Klasifikasi Kanker Payudara. Pada penelitian tersebut, pengujian model KNN menghasilkan akurasi, presisi sehat, *recall* sakit dan AUC lebih tinggi dibandingkan *Naive Bayes*. Sehingga dapat ditarik kesimpulan berdasarkan penelitian bahwa KNN lebih baik dibandingkan *Naive Bayes*.

Penelitian [16] berjudul Prediksi Tingkat Kelulusan Tepat Waktu dengan Metode *Naive Bayes* dan *K-Nearest Neighbor* dengan hasil akurasi tertinggi diraih oleh algoritma KNN yaitu 98.7%.

Penelitian [17] berjudul Optimasi Nilai K pada Algoritma *k-Nearest Neighbor* untuk Prediksi Akademik Mahasiswa yang Bekerja. Hasil penelitian menunjukkan bahwa optimasi nilai menggunakan *5-fold cross-validation* menghasilkan tingkat akurasi sebesar 0,55 dan nilai  $K=3$ . Kemudian, hasil uji *performance* dengan *Confusion Matrix* menunjukkan nilai sebesar 85,71%.

Terakhir, penelitian [18] dengan judul Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa menggunakan Algoritma *K-Nearest Neighbor*. Hasil dari penelitian menunjukkan algoritma KNN mendapatkan akurasi terbaik sebesar 75,95%.

Perbedaan semua penelitian di atas pada penelitian ini adalah melakukan seleksi fitur menggunakan *Correlation Pearson* untuk mengetahui pengaruh dari metode tersebut pada akurasi dari algoritma KNN. Kemudian, dilakukan perbandingan nilai akurasi sebelum dan sesudah menerapkan seleksi fitur menggunakan *Correlation Pearson*. Lalu, indikator yang digunakan pada penelitian ini adalah data riwayat perkuliahan mahasiswa pada mata kuliah Bahasa Arab angkatan tahun 2020-2021.

## Data Mining

*Data mining* adalah suatu proses yang dilakukan guna menciptakan koleksi data besar menjadi sebuah pengetahuan baru [19]. Data Mining juga bisa disebut sebagai proses penggalian dan analisis menggunakan perangkat otomatis atau semi-otomatis untuk mengolah sejumlah besar data dengan tujuan mengidentifikasi pola yang signifikan [20]. Algoritma Data Mining dibagi menjadi beberapa kelompok berdasarkan fungsi dan tujuannya masing-masing, di antaranya [21]:

1. Algoritma deskripsi. Algoritma ini digunakan untuk mengidentifikasi pola yang sering terjadi dan berulang. Pola tersebut kemudian diubah menjadi sebuah aturan yang dapat diterapkan.
2. Algoritma klasifikasi. Cara kerja algoritma ini adalah melakukan pengelompokan data dari hubungan antara variabel kriteria dan variabel target.
3. Algoritma prediksi. Algoritma ini memiliki prinsip kerja yang hampir serupa seperti algoritma klasifikasi, yang menjadi pembeda adalah algoritma prediksi tidak hanya mengidentifikasi target dari data yang ada, namun juga menghasilkan output yang bisa digunakan untuk memprediksi masa depan.
4. Algoritma estimasi. Didefinisikan juga sebagai perkiraan atau prediksi. Perbedaannya terletak pada fokus pengelompokannya, di mana pengelompokan pada estimasi lebih berfokus pada nilai numerik dan daripada kategori.
5. Pengklasteran. Algoritma ini melakukan pengelompokan data dari kemiripan nilai (homogen). Data yang dapat dikelompokkan merupakan hasil pengamatan *record* data, kelas dan objek-objek yang memiliki kesamaan. Jika dibandingkan dengan klasifikasi, pengklasteran tidak melibatkan penggunaan variabel keputusan atau target.
6. Asosiasi. Algoritma ini melakukan suatu kelompok, kumpulan, atau persekutuan. Proses asosiasi melakukan pencarian atribut yang muncul atau selalu muncul secara bersamaan.

### Klasifikasi

Klasifikasi merupakan salah satu peran dalam *data mining* yang menerapkan metode pendekatan prediktif, di mana sebuah kumpulan data uji digunakan untuk mengevaluasi keakuratan suatu model, yang pada umumnya *dataset* dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membentuk model, sedangkan data uji digunakan untuk menguji model tersebut [20]. Beberapa algoritma klasifikasi termasuk *Cart*, *ID3*, *C4.5*, *J48*, *C5.0*, *nearest neighbor*, *naive bayes*, dan masih banyak lagi [21].

### Algoritma K-Nearest Neighbor

Sesuai dengan namanya, “*nearest neighbor*”, KNN mengklasifikasikan data berdasarkan prinsip “kedekatan” dengan tetangga terdekatnya [22]. Dalam konteks ini, jumlah data/tetangga terdekat dalam algoritma KNN ditentukan oleh pengguna yang disebut dengan “K”. Secara sederhana, KNN dapat dijelaskan sebagai berikut [22]:

1. Tentukan k buah tetangga terdekat (*nearest neighbor*) dengan cara acak.
2. Ubahlah himpunan data ke dalam bentuk ruang vektor.
3. Bagi himpunan data menjadi data pelatihan dan data pengujian.
4. Lakukan perhitungan jarak (*d*) antara data pengujian dengan data pelatihan.
5. Urutkan nilai *d* dari yang terkecil hingga yang terbesar.
6. Ambil dan pisahkan data dari hasil penyortiran (*sorting*) sejumlah K.
7. Perhatikan kelas (*class*) mayoritas.
8. Lakukan klasifikasi data pengujian berdasarkan mayoritas.

Untuk menemukan kedekatan data uji dengan tetangga terdekat, dapat dicari menggunakan rumus jarak *Euclidean distance* [22].

### Euclidean Distance

Menurut Banjarsari dkk (2015), fungsi dari *Euclidean distance* yaitu untuk mengukur kedekatan antara dua buah objek dengan menggambarkannya sebagai garis lurus yang teratur dan langsung [23]. Rumus persamaan *Euclidean distance* dapat dituliskan sebagai berikut [24]:

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^p (a_k - b_k)^2} \quad (1)$$

- $a_k$  = Sampel data
- $b_k$  = Data *testing*
- $p$  = Dimensi data
- $i$  = Variabel data

### Data Preprocessing

*Data preprocessing* melibatkan tahapan dalam proses Data Mining yang bertujuan untuk mendapatkan hasil analisis yang lebih akurat, serta bermanfaat untuk mengurangi waktu komputasi dan mengompresi nilai data tanpa mengubah informasinya [25]. *Data preprocessing* terdiri dari empat tahap utama, yaitu seleksi data, pemrosesan data atau pembersihan data, transformasi data dan reduksi data [26].

### K-Fold Cross-Validation

*Cross-Validation* adalah metode statistik yang digunakan untuk mengevaluasi kinerja suatu model

algoritma dengan melibatkan pembagian data ke dalam dua *subset*, yaitu data pelatihan (*training*) dan data pengujian (*testing*) [27]. Model dilatih menggunakan *subset* pelatihan dan divalidasi oleh *subset* pengujian, kemudian pemilihan jenis *Cross-Validation* bergantung pada ukuran kumpulan data yang digunakan [27]. Metode *K-fold* sendiri merupakan salah satu dari beberapa metode dalam *Cross-Validation*, di mana metode ini *dataset* dibagi menjadi k kelompok, dan setiap kelompok terdiri dari data *training* dan data *testing* dengan jumlah yang sama [27].

### Confusion Matrix

Matriks kebingungan, juga dikenal sebagai *Confusion Matrix*, digunakan untuk menggambarkan hasil perbandingan klasifikasi yang dilakukan oleh suatu model [28]. Biasanya, matriks kebingungan divisualisasikan dalam bentuk tabel yang menunjukkan kinerja model klasifikasi pada rangkaian data pengujian dengan nilai sebenarnya yang diketahui. Gambar 1 menunjukkan contoh tabel tersebut untuk mempermudah pemahaman.

Sumber: [28]

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP (True Positive)	FP (False Positive)
	Negative (0)	FN (False Negative)	TN (True Negative)

Gambar 1. *Confusion Matrix*

TP adalah jumlah data *point* berlabel *yes* dengan nilai yang diidentifikasi benar, TN adalah jumlah data *point* berlabel *no* dengan nilai yang diidentifikasi salah, FP adalah jumlah data *point* berlabel *yes* dengan nilai sebenarnya yang diidentifikasi salah, sedangkan FN adalah jumlah data *point* berlabel *no* dengan nilai sebenarnya yang diidentifikasi benar.

Nilai akurasi pada model dapat dihitung dengan menggunakan rumus di bawah [29]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

### Correlation Pearson

Analisis korelasi digunakan untuk mengukur kekuatan hubungan antara dua variabel, di mana variabel lainnya dianggap dikendalikan atau tetap (sebagai variabel kontrol), tujuan dari analisis korelasi adalah untuk mengetahui sejauh mana hubungan tersebut berkorelasi satu sama lain [30]. *Koefisien Pearson* memiliki jangkauan dari angka -1 hingga +1, korelasi 1 menunjukkan setiap kenaikan positif dalam sebuah variabel maka terjadi

peningkatan positif dalam variabel lain, -1 berarti jika ada peningkatan positif dalam satu variabel maka terjadi penurunan negatif pada variabel lainnya, dan nol untuk setiap variabel yang tidak berhubungan [31]. Rumus *Correlation Pearson* dapat ditulis sebagai berikut [31]:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2] [n \sum y^2 - (\sum y)^2]}} \quad (3)$$

- $n$  = banyak data (*record*)
- $x$  = sum dari nilai atribut  $x$
- $y$  = sum dari nilai atribut  $y$
- $xy$  = sum dari nilai nilai atribut  $x, y$
- $x^2$  = sum dari nilai atribut  $x^2$
- $y^2$  = sum dari nilai atribut  $y^2$

Menurut Jonathan Sarwono, berikut adalah tabel interval kekuatan hubungan korelasi dari *Correlation Pearson* [32]:

Tabel 1. Nilai Korelasi

No.	Nilai r	Interpretasi
1	0,00	Tidak ada hubungan
2	0,01 - 0,09	Hubungan kurang berarti
3	0,10 - 0,29	Hubungan moderat
4	0,30 - 0,49	Hubungan kuat
5	0,50 - 0,69	Hubungan sangat kuat
6	0,70 - 0,89	Hubungan mendekati sempurna
7	>0,90	sempurna

## METODE

### Objek Penelitian

Penelitian ini menjadikan kinerja mahasiswa di UMKT yang mengalami penurunan nilai mata kuliah Bahasa Arab dari angkatan tahun 2020 sampai 2021 sebagai objek penelitian.

Penelitian ini berlokasi di Universitas Muhammadiyah Kalimantan Timur (UMKT) yang bertempat di Jl. Ir. H. Juanda No. 15 Samarinda, Kalimantan Timur.

### Teknik Pengumpulan Data

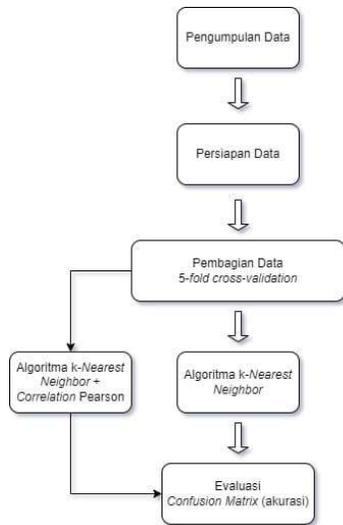
Ada dua teknik pengumpulan data yang digunakan pada penelitian ini, di antaranya:

1. Observasi  
Dilakukan pengamatan langsung pada struktur data yang akan diambil dari riwayat perkuliahan di MKDU dan BAA UMKT. Data yang akan diambil dari MKDU dan BAA adalah data mata kuliah Bahasa Arab tahun angkatan 2020 sampai 2021. Tahapan ini menghasilkan kumpulan *dataset* riwayat perkuliahan mahasiswa UMKT pada mata kuliah Bahasa Arab dari Unit MKDU UMKT. *Dataset* tersebut akan digunakan pada proses pemodelan.
2. Studi Dokumen

Studi dokumen melakukan pengumpulan data pada dokumen-dokumen yang bersangkutan dengan penelitian yang sedang dilaksanakan saat ini. Tahapan ini dilakukan dengan mempelajari jurnal, buku, dan referensi lainnya guna mendukung penelitian ini.

### Alur Analisis Data

Adapun alur analisis data dapat dilihat pada Gambar 2.



Gambar 2. Alur Analisis Data

Penjelasan alur analisis data pada Gambar 2 adalah sebagai berikut:

#### 1. Pengumpulan Data

Pengumpulan data dilakukan dengan melakukan observasi ke bagian MKDU dan BAA UMKT. Data dari MKDU yang didapatkan berasal dari riwayat perkuliahan pada platform *Openlearning* dengan jumlah 983 *record*. Sedangkan data yang didapat dari BAA berjumlah 4055 *record*. Data yang telah diperoleh dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Atribut Data MKDU

No.	Atribut	Keterangan
1	<i>Profile name</i>	<i>Id</i> mahasiswa pada sistem <i>OpenLearning</i>
2	<i>Learner name</i>	Nama mahasiswa
3	<i>Learner email</i>	Email mahasiswa
4	<i>Enrolment ID</i>	<i>Id</i> pendaftaran <i>OpenLearning</i>
5	<i>Institution Membershi p ID</i>	<i>Id</i> anggota institusi
6	<i>Enrolment date</i>	Tanggal daftar

7	<i>Completion date</i>	Tanggal menyelesaikan mata kuliah
8	<i>Time spent on course</i>	Lama waktu mahasiswa mengakses mata kuliah
9	<i>Progress</i>	Persentase kemajuan mahasiswa
10	<i>% Course completed</i>	Persentase kemajuan mahasiswa menyelesaikan mata kuliah
11	<i>Certificate ID</i>	<i>Id</i> sertifikat
12	<i>Comments</i>	Jumlah komentar mahasiswa selama perkuliahan
13	Tugas 1	Nilai tugas 1
14	Tugas 2	Nilai tugas 2
15	UTS	Nilai ujian tengah semester

Tabel 3. Atribut Data BAA

No.	Atribut	Keterangan
1	NIM	Nomor induk mahasiswa
2	Nama mahasiswa	Nama mahasiswa
3	Nilai akhir	Nilai akhir mahasiswa
4	Bobot	Bobot nilai berdasarkan standar penilaian
5	Simbol	Skala penilaian

#### 2. Persiapan Data (*Data Preparation*)

Setelah dikumpulkan, kemudian dilakukan identifikasi, pemilihan, pembersihan dan perubahan ke dalam format yang diperlukan. *Data preparation* juga bisa disebut sebagai *data preprocessing*. *Data training* adalah data mahasiswa semua program studi UMKT yang mengikuti mata kuliah Bahasa Arab pada tahun angkatan 2020-2021. Data dari MKDU dan BAA UMKT meliputi atribut yang dapat dilihat pada Tabel 2 dan Tabel 3.

##### a) *Data Selection dan Integration*

Seleksi dan integrasi data diperlukan untuk meningkatkan akurasi serta efisiensi dari model algoritma yang dipilih. Proses seleksi dan integrasi dilakukan melalui pendekatan Excel. Hasilnya dapat dilihat pada Tabel 4.

Tabel 4. Atribut seleksi dan Integrasi Data

No.	Atribut	Tipe Data
1	Jenis Kelamin	<i>Binominal</i>

2	<i>Time spent on course</i>	Numerik
3	<i>Progress</i>	Numerik
4	<i>% Course Completed</i>	Numerik
5	<i>Comments</i>	Numerik
6	Tugas 1	Numerik
7	Tugas 2	Numerik
8	UTS	Nominal
9	Simbol	<i>Binominal</i>

b) *Data Transformation*

Dilakukan proses data transformasi guna mengubah format pada nilai data sesuai yang dibutuhkan. Data transformasi dilakukan melalui pendekatan Excel.

c) *Data Reduction*

Tahap data reduction dilakukan untuk menyeimbangkan kelas dalam data. Teknik *undersampling* dipilih untuk melakukan *data reduction*. *Undersampling* adalah teknik yang digunakan untuk mengurangi kelas mayoritas hingga seimbang dengan kelas minoritas. Tahapan ini dilakukan dengan bahasa pemrograman *Python*.

3. *Modelling*

Pemodelan akan diimplementasikan menggunakan *Python*. Pertama-tama, data melalui proses pembagian menjadi *data training* dan data testing dengan metode *5-fold Cross-validation*. *Data training* tersebut yang nantinya akan digunakan oleh model algoritma untuk pelatihan, sementara data testing digunakan sebagai proses evaluasi model algoritma yang sudah dilatih. Model algoritma yang digunakan untuk mengklasifikasi nilai mahasiswa adalah KNN. Untuk menghitung rumus jarak antar data, bisa menggunakan rumus persamaan *Euclidean distance*. Jika ada dua buah objek yang akan diukur kedekatannya dengan menggunakan rumus *Euclidean distance*, contoh kasusnya adalah sebagai berikut:

$$X = [9,5,5,3]$$

$$Y = [2,2,4,8]$$

Maka jarak nilai antara objek X dan Y akan dihitung seperti di bawah:

$$d_{(x,y)}:$$

$$\sqrt{(9-2)^2 + (5-2)^2 + (5-4)^2 + (3-8)^2}$$

Berdasarkan perhitungan di atas, maka didapatkan nilai jarak antar data ada 9,17. Setelah melakukan pemodelan, selanjutnya adalah tahap seleksi fitur. Seleksi fitur dilakukan menggunakan metode *Correlation Pearson*. *Correlation Pearson* bekerja dengan cara mencari hubungan korelasi antar dua buah variabel atau atribut. Jika ada sebuah atribut yang dibandingkan

dengan atribut target dan tidak memiliki hubungan korelasi, maka atribut tersebut bisa dibuang. Contoh perhitungan *Correlation Pearson* dapat dilihat di bawah ini:

$$r = \frac{6(\sum 20485) - (\sum 247)(\sum 486)}{\sqrt{[6\sum 11409 - (\sum 247)^2][6\sum 40022 - (\sum 486)^2]}}$$

Berdasarkan perhitungan di atas, maka didapatkan hasil 0.5298, yang menunjukkan bahwa dua buah variabel memiliki korelasi positif sedang, karena angkanya mendekati positif 1.

4. Evaluasi

Melakukan perbandingan dengan mempertimbangkan nilai komparasi *Confusion Matrix*. Teknik *Confusion Matrix* yang digunakan adalah *accuracy*. Evaluasi dilakukan sebelum dan sesudah menerapkan seleksi fitur untuk melihat perbedaan akurasi pada model algoritma.

5. *Deployment*

Tahap *deployment* akan memberikan rekomendasi dari hasil algoritma KNN dan hasil seleksi fitur untuk meningkatkan akurasi serta mengetahui fitur paling berpengaruh pada tahap klasifikasi.

## HASIL DAN PEMBAHASAN

### *Data Preparation*

Sebelum melakukan pemodelan, dilakukan proses *data preparation* dengan 4 tahapan sebagai berikut:

1. *Data Selection*

Tahap data selection dilakukan setelah melakukan pengumpulan data. Pada tahapan ini, atribut data yang telah dikumpulkan akan diseleksi berdasarkan dua tahapan, yaitu menyeleksi atribut yang tidak diperlukan dan melakukan seleksi fitur berbasis *Correlation Pearson*. Namun, proses seleksi fitur dengan *Correlation Pearson* dilakukan setelah melakukan implementasi pada model algoritma KNN. Hal tersebut bertujuan untuk membandingkan nilai akurasi sebelum dan sesudah melakukan seleksi fitur. Tahap pertama dilakukan dengan menghapus atribut Nama Lengkap, Nilai Akhir dan Bobot pada data yang diperoleh dari BAA. Sedangkan untuk data dari MKDU, atribut yang dihapus adalah *Profile name*, *Learner name*, *Learner email*, *Enrolment ID*, *Institution of Membership ID*, *Enrolment date*, dan *Certificate ID*.

2. *Data Integration*

Data integration dilakukan untuk sinkronisasi sekaligus menggabungkan data dari BAA dan MKDU UMKT yang telah

melalui tahap pertama dari data selection. Data yang digabungkan adalah atribut Jenis Kelamin, Simbol, Time Spent on Course, Progress, % Course Complete, Comments, Tugas 1, Tugas 2 dan UTS.

3. *Data Transformation*

Pada tahapan ini, dilakukan perubahan tipe data pada nilai atribut sebelum masuk ke tahap pemodelan. Ada tiga atribut yang ditransformasikan, yaitu Jenis Kelamin, *Time Spent on Course* dan Simbol. Atribut Jenis Kelamin merupakan keterangan jenis kelamin tiap individu mahasiswa dengan tipe data awal adalah “L” untuk laki-laki dan “P” untuk perempuan. Tipe data tersebut akan diubah menjadi numerik. Kemudian, Atribut selanjutnya yang akan ditransformasi adalah *Time Spent on Course*. *Time Spent on Course* adalah lama waktu mahasiswa menyelesaikan kursus. Waktu yang tertera berupa jam dan menit, pada tahapan ini akan ditransformasi menjadi menit secara keseluruhan. Setelah itu, atribut ke tiga yang akan ditransformasi selanjutnya adalah Simbol. Atribut Simbol juga akan digunakan sebagai target saat pemodelan. Target yang awalnya memiliki 8 kriteria yaitu A, AB, B, BC, C, D, E dan T akan diubah menjadi hanya dua kriteria. Nilai A, AB dan B akan diubah menjadi “1”, sedangkan nilai BC dan seterusnya akan diubah menjadi “0”. Angka 1 mewakili “LULUS”, dan angka 0 mewakili “TIDAK LULUS”.

4. *Data Reduction*

Tahap *data reduction* dilakukan guna menangani ketidakseimbangan kelas pada himpunan data. Pada tahap ini, teknik *data reduction* yang digunakan adalah *undersampling*. Cara kerja teknik *undersampling* adalah mengurangi jumlah *record* dalam target mayoritas agar menciptakan keseimbangan dengan target minoritas. Target yang menjadi mayoritas merupakan target “1” dengan jumlah 877 *record* dan akan dikurangi agar menjadi seimbang dengan data minoritas yaitu “0” yang berjumlah 66 *record*. Penghapusan data mayoritas dengan teknik *undersampling* dilakukan secara acak.

**Modelling, Seleksi Fitur & Evaluasi**

Setelah dilakukan tahapan *data preparation*, kemudian dilakukan *modelling* algoritma KNN dengan membagi data menggunakan teknik *5-Fold Cross-Validation* pada seluruh atribut. Kemudian dilakukan tahapan seleksi fitur menggunakan metode *Correlation Pearson* dengan bahasa pemrograman *Python*. Nilai korelasi dari hasil perhitungan *Correlation Pearson* dapat dilihat pada Tabel 5.

**Tabel 5. Penentuan Atribut yang Digunakan**

Atribut	Nilai Korelasi	Hasil
Jenis_Kelamin	-0.094	Tidak digunakan
<i>Time_spent_on_course</i>	0.32	Digunakan
<i>Progress</i>	0.42	Digunakan
<i>Course_compl eted</i>	0.42	Digunakan
<i>Comments</i>	0.14	Tidak digunakan
Tugas 1	-0.17	Tidak digunakan
Tugas 2	0.049	Tidak digunakan
UTS	-0.0017	Tidak digunakan
Simbol	1	Digunakan

Pada Tabel 5, terlihat bahwa atribut yang dipilih adalah *Time\_spent\_on\_course*, *Progress* dan *Course\_completed*. Setelah melalui tahap seleksi fitur, dilakukan lagi pemodelan ulang pada algoritma KNN untuk melihat pengaruh *Correlation Pearson* pada akurasi dari KNN. Perbandingan hasil akurasi rata-rata yang didapat dari keseluruhan *fold* sebelum dan sesudah seleksi fitur dapat dilihat pada Tabel 6.

**Tabel 6. Perbandingan Hasil Akurasi Rata-Rata**

Nilai K	Mean Accuracy (Sebelum Seleksi Fitur)	Mean Accuracy (Setelah Seleksi Fitur)	Status
K=3	0.63	0.64	+1,59%
K=5	0.66	0.66	0%
K=7	0.65	0.63	-3,08%
K=9	0.64	0.64	0%
K=11	0.62	0.62	0%

Bisa dilihat pada Tabel 6, bahwa terdapat perubahan dalam hal akurasi pada algoritma KNN setelah menerapkan seleksi fitur *Correlation Pearson*. Nilai K=3 mengalami kenaikan, sedangkan K=7 mengalami penurunan, dan yang lainnya tidak mengalami perubahan atau tetap.

## Deployment

Berdasarkan hasil penelitian, akurasi rata-rata dari keseluruhan *fold* yang didapatkan pada algoritma KNN yang paling tinggi adalah 66%, sedangkan yang paling rendah adalah 62%. Peningkatan akurasi KNN menggunakan seleksi fitur berbasis *Correlation Pearson* pada data nilai Bahasa Arab mahasiswa di UMKT juga tidak begitu efisien dan tidak direkomendasikan. Kemudian, fitur yang dianggap paling berpengaruh menurut perhitungan *Correlation Pearson* pada data nilai Bahasa Arab mahasiswa UMKT adalah *Time Spent on Course*, *Progress* dan *Course Completed*.

## KESIMPULAN DAN SARAN

### Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka kesimpulan dari penelitian adalah sebagai berikut:

1. Berdasarkan hasil dari pemodelan algoritma, didapatkan hasil akurasi rata-rata terbaik yang didapat adalah K=5 yang memperoleh akurasi sebesar 66%, kemudian diikuti oleh K=7 dengan akurasi rata-rata 65%, K=9 dengan akurasi rata-rata 64%, K=3 dengan akurasi rata-rata 63% dan K=11 dengan akurasi rata-rata paling rendah yaitu 62%.
2. Fitur yang dianggap paling berpengaruh menurut perhitungan *Correlation Pearson* pada data nilai akhir Bahasa Arab mahasiswa adalah *Time spent on course*, *Progress* dan *Course completed*.
3. Setelah melalui tahap seleksi fitur, disimpulkan bahwa *Correlation Pearson* tidak begitu efisien dalam meningkatkan akurasi dari KNN pada data nilai mahasiswa Bahasa Arab di UMKT. Hanya ada satu akurasi rata-rata yang mengalami peningkatan, yaitu K=3 yang mengalami peningkatan sebesar 1%, kemudian K=7 justru malah mengalami penurunan yang cukup drastis yakni 2%, dan sisanya tidak mengalami perubahan atau tetap.

### Saran

Saran untuk peneliti selanjutnya adalah:

1. Pada penelitian selanjutnya, diharapkan untuk melakukan seleksi fitur dengan metode lain pada algoritma KNN menggunakan data nilai Bahasa Arab mahasiswa UMKT seperti ANOVA, *Adaboost* dan lain sebagainya untuk melihat perbandingan hasil seleksi fiturnya.
2. Penelitian selanjutnya diharapkan untuk dapat melakukan pengembangan sebuah sistem dengan menerapkan prinsip kerja KNN yang mampu mengklasifikasi data nilai mahasiswa di UMKT.

## DAFTAR PUSTAKA

- [1] Litbang Diklat Kemenag, "Laporan Survei Pelaksanaan Belajar Dari Rumah Pada Masa Covid 19 Di Madrasah Dan Sekolah," pp. 1–14, 2020, [Online]. Available: [https://simlitbangdiklat.kemendiknas.go.id/simlitbang/spdata/upload/dokumen-penelitian/1592449941Laporan\\_Hasil\\_Survei\\_Penda.pdf](https://simlitbangdiklat.kemendiknas.go.id/simlitbang/spdata/upload/dokumen-penelitian/1592449941Laporan_Hasil_Survei_Penda.pdf)
- [2] "Penilaian MKDU (17 - 22)."
- [3] I. Magdalena, A. Ridwanita, and B. Aulia, "Evaluasi Belajar Peserta Didik," *J. Pendidik. dan Dakwah*, vol. 2, no. 1, pp. 117–127, 2020, [Online]. Available: <https://ejournal.stitpn.ac.id/index.php/pandawa>
- [4] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Appl. Comput. Informatics*, vol. 12, no. 1, pp. 90–108, 2016, doi: 10.1016/j.aci.2014.10.001.
- [5] D. Prasetyawan and R. Gatra, "Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 56–67, 2022, doi: 10.14421/jiska.2022.7.1.56-67.
- [6] Z. Pan, Y. Wang, and Y. Pan, "A new locally adaptive k-nearest neighbor algorithm based on discrimination class," *Knowledge-Based Syst.*, vol. 204, p. 106185, 2020, doi: 10.1016/j.knosys.2020.106185.
- [7] A. Jhamtani, R. Mehta, and S. Singh, "Size of wallet estimation: Application of K-nearest neighbour and quantile regression," *IIMB Manag. Rev.*, vol. 33, no. 3, pp. 184–190, 2021, doi: 10.1016/j.iimb.2021.09.001.
- [8] Y. Mu, X. Liu, and L. Wang, "A Pearson's correlation coefficient based decision tree and its parallel implementation," *Inf. Sci. (Ny.)*, vol. 435, pp. 40–58, 2018, doi: 10.1016/j.ins.2017.12.059.
- [9] A. Sharma and S. Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis," *Int. J. Comput. Appl.*, no. June, pp. 15–20, 2012, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Performance+Investigation+of+Feature+Selection+Methods+and+Sentiment+Lexicons+for+Sentiment+Analysis#0>
- [10] R. A. Manullang and F. A. Sianturi, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa: Bahasa Indonesia," *J. Ilmu Komput. Dan ...*, vol. 4, no. 2, pp. 15–23, 2021, [Online]. Available: <http://ejournal.sisfokomtek.org/index.php/jikom/article/view/155>
- [11] M. T. F. Imaning Dyah Larasati, Ahmad Afif Supianto, "Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Metode Modified K-Nearest Neighbor (MK-NN)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4558–4563, 2019, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5284>

- [12] Z. Rahman, "Optimalisasi dalam Mengidentifikasi Selekasi Mahasiswa Jalur Cepat (Fast-track) Menggunakan Metode K-Nearest Neighbor," *J. Sistim Inf. dan Teknol.*, vol. 5, pp. 49–54, 2023, doi: 10.37034/jsisfotek.v5i2.166.
- [13] A. E. Gumanti, Taslim, S. Handayani, and T. Dafwen, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Topik Skripsi Mahasiswa di Fakultas Ilmu Komputer," *J. Inf. Technol. Comput. Sci.*, vol. 1, pp. 44–50, 2022, [Online]. Available: <http://journal.unilak.ac.id/index.php/JITACS/article/view/10037>
- [14] S. Wiyono and T. Abidin, "Implementation of K-Nearest Neighbour (Knn) Algorithm To Predict Student'S Performance," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 873–878, 2018, doi: 10.24176/simet.v9i2.2424.
- [15] A. Nugraheni, R. D. Ramadhani, A. B. Arifa, and A. Prasetiadi, "Perbandingan Performa Antara Algoritma Naive Bayes Dan K-Nearest Neighbour Pada Klasifikasi Kanker Payudara," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 2, no. 1, pp. 11–20, 2022, doi: 10.20895/dinda.v2i1.391.
- [16] S. Rahmatullah, "Prediksi Tingkat Kelulusan Tepat Waktu Dengan Metode Naïve Bayes Dan K-Nearest Neighbor," *J. Inf. dan Komput.*, vol. 7, no. 1, pp. 7–16, 2019, doi: 10.35959/jik.v7i1.118.
- [17] Taslim, Yuhelmi, and D. Toresa, "Optimasi Nilai k Pada Algoritma k Nearest Neighbor Untuk Prediksi Akademik Mahasiswa yang Bekerja," *Indones. J. Comput. Sci.*, vol. 10, no. 2, pp. 379–388, 2021, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3005>
- [18] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019, doi: <https://doi.org/10.24176/simet.v10i2.2882>.
- [19] T. A. Yoga Siswa, *Data Mining - Mengupas Tuntas Analisis Data dengan Metode Klasifikasi Hingga Deployment Aplikasi Menggunakan Python*. Samarinda: UMKT PRESS, 2023.
- [20] L. Muflikhah, D. Eka Ratnawati, and R. Regasari Mardi Putri, *Data Mining*. UB PRes, 2018.
- [21] M. kom. Buulolo, Efori, S.kom., *Data Mining untuk Perguruan Tinggi*. Yogyakarta: Deepublish Publisher, 2020. [Online]. Available: [https://books.google.co.id/books?hl=id&lr=&id=K\\_SDwAAQBAJ&oi=fnd&pg=PP1&dq=buku+data+mining&ots=KdGw3Sq0Tq&sig=OZhh5Sn4DHuYWfGY\\_xhNqZIoBDI&redir\\_esc=y#v=onepage&q=buku data mining&f=false](https://books.google.co.id/books?hl=id&lr=&id=K_SDwAAQBAJ&oi=fnd&pg=PP1&dq=buku+data+mining&ots=KdGw3Sq0Tq&sig=OZhh5Sn4DHuYWfGY_xhNqZIoBDI&redir_esc=y#v=onepage&q=buku data mining&f=false)
- [22] R. Primartha, *Algoritma Machine Learning*. Bandung: Penerbit Informatika, 2021.
- [23] A. M. Habibi and R. R. Santika, "Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Jurusan Menggunakan Metode Euclidean Distance Berbasis Web Pada SMP Setia Gama," *Skanika*, vol. 3, no. 4, pp. 7–14, 2020, [Online]. Available: <https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/2735>
- [24] J. Indriyanto, *Algoritma K-Nearest Neighbor untuk Prediksi Nasabah Asuransi*. Penerbit NEM, 2021.
- [25] B. Santosa and A. Umam, *Big Data dan Analytics*. Yogyakarta: Penebar Media Pustaka, 2018.
- [26] D. Nofriansyah and G. Widi Nurcahyo, *Algoritma Data Mining dan Pengujian*. Medan: Deepublish Publisher, 2015. [Online]. Available: [https://books.google.co.id/books?hl=id&lr=&id=Fn-QDwAAQBAJ&oi=fnd&pg=PR4&dq=algoritma+data+mining+dan+pengujian&ots=oYxCMiHLSF&sig=UUMGKBvPzSBB4w5LDwDCJznekEQ&redir\\_esc=y#v=onepage&q=algoritma data mining dan pengujian&f=false](https://books.google.co.id/books?hl=id&lr=&id=Fn-QDwAAQBAJ&oi=fnd&pg=PR4&dq=algoritma+data+mining+dan+pengujian&ots=oYxCMiHLSF&sig=UUMGKBvPzSBB4w5LDwDCJznekEQ&redir_esc=y#v=onepage&q=algoritma data mining dan pengujian&f=false)
- [27] I. Daqiqil, *MACHINE LEARNING: Teori, Studi Kasus dan Implementasi Menggunakan Python*. Riau: UR PRESS, 2021. [Online]. Available: [https://www.researchgate.net/publication/353338909\\_Machine\\_Learning\\_Teori\\_Studi\\_Kasus\\_dan\\_Implementasi\\_Menggunakan\\_Python](https://www.researchgate.net/publication/353338909_Machine_Learning_Teori_Studi_Kasus_dan_Implementasi_Menggunakan_Python)
- [28] Mustika *et al.*, *Data Mining dan Aplikasinya*. Bandung: Widina Bhakti Persada Bandung, 2021. [Online]. Available: [https://books.google.co.id/books?hl=id&lr=&id=53FXEAAAQBAJ&oi=fnd&pg=PR4&dq=data+mining+dan+aplikasinya&ots=srGF\\_JBBDK&sig=5jYZ8kNGavn\\_0qIh9HAT\\_jATu0&redir\\_esc=y#v=onepage&q=data mining dan aplikasinya&f=false](https://books.google.co.id/books?hl=id&lr=&id=53FXEAAAQBAJ&oi=fnd&pg=PR4&dq=data+mining+dan+aplikasinya&ots=srGF_JBBDK&sig=5jYZ8kNGavn_0qIh9HAT_jATu0&redir_esc=y#v=onepage&q=data mining dan aplikasinya&f=false)
- [29] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [30] N. T. Romadloni and Hilman F Pardede, "Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 505–510, 2019, doi: 10.29207/resti.v3i3.1189.
- [31] S. Glen, "Koefisien Korelasi: Definisi Sederhana, Rumus, Langkah Mudah," *StatisticsHowTo.com : Statistik Dasar untuk Kita Semua!*, 2023. <https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/> (accessed Mar. 30, 2023).
- [32] R. Safitri, W., "Analisis Korelasi Dalam Menentukan Hubungan Antara Kejadian Demam Berdarah Dengue Dengan Kepadatan Penduduk Di Kota Surabaya Pada Tahun 2012 - 2014," *J. Kesehat. Masy.*, vol. 1, no. 3, pp. 1–9, 2014.



**SURAT KETERANGAN ARTIKEL PUBLIKASI**

*Assalamu'alaikum Warahmatullahi wabarakatuh*

Saya yang bertanda tangan dibawah ini:

Nama	:	Wawan Joko Pranoto, S.Kom., M.T.I
NIDN	:	1102057701
Nama	:	Putra Saymen
NIM	:	1911102441011
Fakultas	:	Sains dan Teknologi
Progam Studi	:	S1 Teknik Informatika

Manyatakan bahwa artikel ilmiah yang berjudul "Analisis Seleksi Fitur Berbasis *Correlation Pearson* pada Algoritma *K-Nearest Neighbor* Dalam Klasifikasi Nilai Mahasiswa" telah di submit pada Jurnal Informatika Teknologi dan Sains (JINTEKS) pada tahun 2023. <http://www.jurnal.uts.ac.id/index.php/JINTEKS/>.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

*Wassalamu'alaikum Warahmatullahi wabarakatuh*

Mahasiswa



Putra Saymen

Samarinda, Kamis, 20 Juli 2023

Dosen Pembimbing



Wawan Joko Pranoto, S.Kom., M.T.I

NIDN. 1102057701