

# Implementasi Algoritma *K – Nearest Neighbor* untuk Mengklasifikasikan Faktor Pernikahan Dini

Siti Muntari<sup>1</sup>, Febriansyah<sup>2</sup>

<sup>1</sup>Intitut Teknologi Pagar Alam

Jl Masik Siagim Simpang Mbacang Kota Pagar Alam

<sup>2</sup>Institution/affiliation

, e-mail: [Muntariaza@gmail.com](mailto:Muntariaza@gmail.com)

## ARTICLE INFO

Article history:

Received March 2026

Revised April 2026

Accepted June 2026

Available online July 2026

## ABSTRACT

*Early Marriage in Pagar Alam City is currently still quite high, the Pagar Alam Religious Court only relies on a recap of the number of cases per year to draw conclusions about the early marriage data. This method has limitations in classifying early marriage factors, so the Religious Court has difficulty in monitoring and controlling the occurrence of Early Marriage in Pagar Alam City. The purpose of this research thesis is to produce a classification system for Early Marriage factors using the K-Nearest Neighbor Algorithm to find out what factors influence the occurrence of Early Marriage, the method used in this study is the Cross Industry Standard Process for Data Mining (CRISP-DM) which has 6 stages, namely: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The testing stage in this study uses Confusion Matrix and BlackBox Testing. The final results of this study indicate that the system can classify Early Marriage factors. The classification model built achieved an Accuracy of 94.12% and a Precise value of 85.71% and a Recall value of 100.00%, while testing using Black Box testing in the form of alpha obtained a feasibility value of 83.2%, so this system is very suitable for use.*

**Keywords:** *Early Marriage, K-Nearest Neighbor, RapidMiner, Confusion Matrix*

## 1. Introduction

Pernikahan Dini adalah perkawinan yang dilangsungkan oleh sepasang pria dan gadis remaja [1]. Nikah Muda masih menjadi tren yang sangat diminati oleh masyarakat sampai sekarang. Pernikahan itu terjadi akibat minimnya pemahaman mengenai Pernikahan Dini, yang berujung pada berbagai dampak negatif seperti trauma, kekerasan domestik, reproduksi, dan terganggunya kondisi Kesehatan [2]. Pernikahan Dini adalah pernikahan yang dilangsungkan oleh orang-orang yang masih di bawah umur dan belum siap secara fisik, mental, maupun finansial. [3].

Dampak pernikahan pada usia muda secara umum terbagi menjadi dua kelompok, yaitu dampak yang menguntungkan dan dampak yang merugikan [4]. Dari perspektif keagamaan, manfaatnya

*Received 10 Februari, 2026; Revised 7 April, 2026; Accepted 23 Juni, 2026*

yang baik adalah terjaga dari pergaulan bebas dan menjauhi tindakan zina. Menikah di usia muda dianggap dapat menghindari seks di luar nikah karena kebutuhan seksual yang terpuaskan. Pernikahan dini dapat mencegah perilaku zina dan menjauhkan dari pergaulan yang bebas [5]. Klasifikasi merupakan teknik pembelajaran data untuk menghasilkan prediksi nilai dari serangkaian atribut, Klasifikasi sering diterapkan untuk meramalkan kelas pada label tertentu [6], yaitu dengan mengelompokkan data (membuat model) berdasarkan set pelatihan dan nilai-nilai (*label kelas*) dalam mengelompokkan atribut tertentu [7]. pernikahan dini di Kota Pagar Alam masih menjadi perhatian mengingat Pernikahan Dini dapat memengaruhi kualitas hidup generasi mendatang. Salah satu dampak signifikan dari Pernikahan Dini cenderung berdampak negatif pada kehidupan individu, terutama perempuan. Dari sisi kesehatan, perempuan yang hamil pada usia muda memiliki risiko tinggi terhadap komplikasi kehamilan dan persalinan, termasuk kematian ibu dan bayi [8]. Selain itu, Pernikahan Dini sering menyebabkan putus sekolah karena terbatasnya kesempatan untuk melanjutkan pendidikan setelah menikah

Metode ini memiliki keterbatasan dalam mengklasifikasikan faktor Pernikahan Dini, sehingga Pengadilan Agama mengalami kesulitan dalam pemantauan dan pengendalian terjadinya Pernikahan Dini di Kota Pagar Alam. Oleh karena itu dibutuhkan sistem yang mampu menganalisis dan mengklasifikasikan data dengan lebih cepat dan akurat. Berdasarkan penelitian

K-Nearest Neighbor (KNN) merupakan salah satu algoritma dalam data mining yang bekerja dengan memilih sejumlah data terdekat ( $k$ ) dari suatu objek sebagai dasar penentuan klasifikasi [9]. Hasil klasifikasi pada algoritma ini sangat dipengaruhi oleh nilai  $k$  yang digunakan [10]. Untuk menentukan nilai  $k$  yang optimal, cara sederhana yang dapat dilakukan adalah dengan menjalankan algoritma secara berulang menggunakan berbagai variasi nilai  $k$ , kemudian memilih nilai  $k$  yang memberikan kinerja terbaik .

Berdasarkan hasil penelitian [11], dapat disimpulkan bahwa Pernikahan Usia Dini adalah pernikahan yang dilakukan oleh salah satu pasangan yang dianggap sebagai anak atau belum mencapai batas usia di bawah 19 tahun. Konsekuensi dari Pernikahan Usia Dini yang akan dirasakan anak meliputi efek buruk pada kesehatan, psikologi, sosial, dan perkembangan.

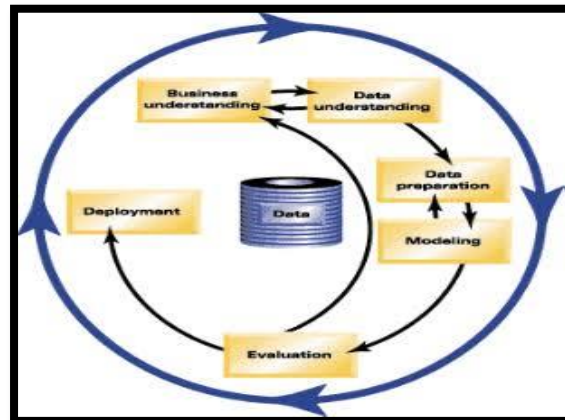
Berdasarkan hasil penelitian, disimpulkan bahwa permasalahan yang umum terjadi adalah akses data mengenai kasus Pernikahan Dini yang tidak tercatat atau terekam dalam dokumen resmi [12]. Penelitian ini mengandalkan pengakuan dari perempuan berusia 20-24 tahun saat survei yang menyatakan bahwa mereka telah menikah sebelum mencapai usia 18 tahun. Variabel yang diteliti dalam studi ini mencakup kondisi kemiskinan, klasifikasi tempat tinggal, status Pernikahan Dini kepala keluarga (KK), pendidikan KK, jenis pekerjaan KK, jumlah anggota keluarga, serta status pemanfaatan internet oleh perempuan usia 20-24 tahun.

Tujuan dari penelitian ini adalah menghasilkan sistem klasifikasi faktor Pernikahan Dini menggunakan Algoritma *K-Nearest Neighbor* guna mengetahui faktor apa saja yang mempengaruhi hingga terjadinya Pernikahan Dini.

## 2. Research Method

Metode *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) adalah sebuah kerangka kerja atau metodologi yang digunakan dalam data mining dan analisis data. *CRISP-DM* adalah pendekatan yang fleksibel dan dapat disesuaikan dengan berbagai proyek *data mining* [13] Tujuan dari metode *CRISP-DM* yaitu untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan. *CRISP-DM* memiliki tahapan dan kerangka kerja yang terstruktur sehingga pengguna metode ini akan lebih terarah dan mengetahui langkah yang harus dikerjakan

dalam penelitian [14].



### 1. Business Understanding

Proses pemahaman bisnis, memahami situasi dan kondisi pada saat penelitian dan menetapkan sebuah tujuan dari penelitian yang dilakukan ke dalam permasalahan yang diselesaikan dengan *data mining* melalui proses pengumpulan data, meliputi: observasi, wawancara dan studi pustaka.

### 2. Data Understanding

Peneliti melakukan pemahaman terhadap kebutuhan data terkait dengan kasus faktor Pernikahan Dini di kota Pagar Alam. Data diperoleh dari penyebaran kuesioner di wilayah Kota Pagar Alam.

### 3. Data Preparation

Proses pengolahan data terkait nilai-nilai yang hilang dan normalisasi data, data akan melalui proses identifikasi, pemilihan data, pembersihan data dan transformasi data untuk mempersiapkan pengembangan model.

### 4. Modelling

Pemodelan yang dilakukan pada penelitian ini adalah metode klasifikasi menggunakan algoritma *k-nearest neighbor* sehingga menghasilkan pola yang akan digunakan pada data penelitian.

### 5. Evaluation

Proses untuk melakukan pengukuran hasil evaluasi dari model yang telah diimplementasikan sebelumnya. Hasil evaluasi tersebut menggambarkan proses dari *data mining* yang telah dilakukan dan mengukur model yang paling baik untuk digunakan.

### 6. Deployment

Proses pemodelan dengan menerapkan teknik *data mining* klasifikasi. Model tersebut menghasilkan informasi yang diharapkan dapat membantu Pengadilan Agama Kota Pagar Alam mengetahui faktor Pernikahan Dini.

## 3.7 Tahapan K-Nearest Neighbor (KNN)

Pada penelitian ini model *data mining* yang akan digunakan yaitu metode klasifikasi dengan algoritma *K-Nearest Neighbor* (KNN) [15]. Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain:

### 1. Siapkan Dataset

*Dataset* diperoleh hasil penyebaran kuesioner wilayah kecamatan Pagar Alam sebanyak 85 Data. Atribut data yang digunakan pada sistem klasifikasi *K-Nearest Neighbor* yaitu nama, jenis kelamin, kecamatan, umur, status hubungan, tingkat pendidikan, berhenti sekolah,

pengasan ortu, pengaruh teman, lama pacaran, motivasi pribadi, komunikasi ortu, media sosial, konsumsi media, norma budaya, faktor ekonomi.

2. Menentukan Parameter K. yang sesuai dengan data yang digunakan, Nilai K minimal bernilai 1 dan maksimalnya jumlah dari data latih.
3. Menghitung Jarak antara data training dan data testing. Perhitungan jarak yang paling umum dipakai pada perhitungan algoritma K-NN adalah menggunakan perhitungan jarak *Euclidean*. Rumusannya adalah sebagai berikut:

$$euc = \sqrt{\sum (p_i - q_i)^2}$$

Dimana :

$p_i$  = sample data / data training

$q_i$  = data uji / data testing

$i$  = data variable

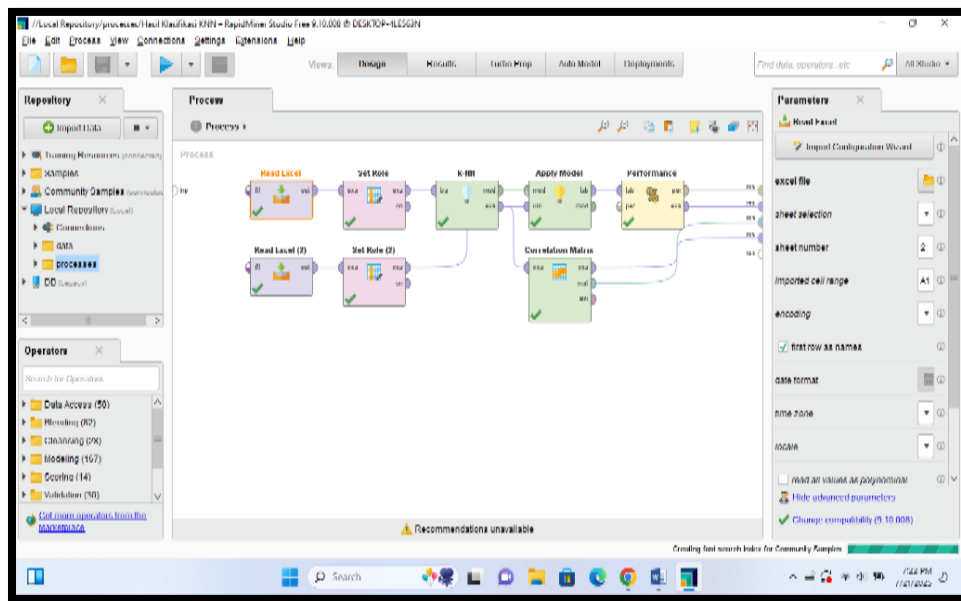
$n$  = dimensi data

4. Mengurutkan jarak yang terbentuk.
5. Menentukan jarak terdekat sampai urutan K.
6. Memasangkan kelas yang bersesuaian.
7. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

### 3. Results and Analysis

Hasil penelitian ini menghasilkan model yang dihasilkan Implementasi Algoritma *K-Nearest Neighbor (K-NN)* yang memiliki tingkat akurasi cukup tinggi sehingga dapat digunakan untuk mengklasifikasikan faktor Pernikahan Dini, dengan jumlah 85 *record* data untuk data *training* 68 data dan data *testing* 17 data dimana untuk hasil Pernikahan Dini yang akan diklasifikasikan. untuk klasifikasi faktor Pernikahan Dini dibagi menjadi 2 klasifikasi yaitu faktor pergaulan bebas dan faktor kemauan sendiri, sebelum melakukan klasifikasi data peneliti menggunakan bantuan aplikasi *RapidMiner*, untuk mendapatkan model. dan pada tahap pengujian dengan *confusion matrix* Hasil Evaluasi *Performance* model klasifikasi *K-Nearest Neighbor* di *Software RapidMiner*, menghasilkan nilai akurasi 94,12% dan nilai *precision* 85,71% dan nilai *recall* 100,00%.

3.1 Hasil dari prediksi tersebut dianalisis menggunakan operator "*Performance*" untuk mengevaluasi kinerja model. Di sisi lain, *dataset* kedua dianalisis menggunakan *Correlation Matrix* untuk mengetahui hubungan antar atribut dalam data. Alur ini menunjukkan proses lengkap dalam membangun, menerapkan, dan mengevaluasi model klasifikasi di *RapidMiner*.



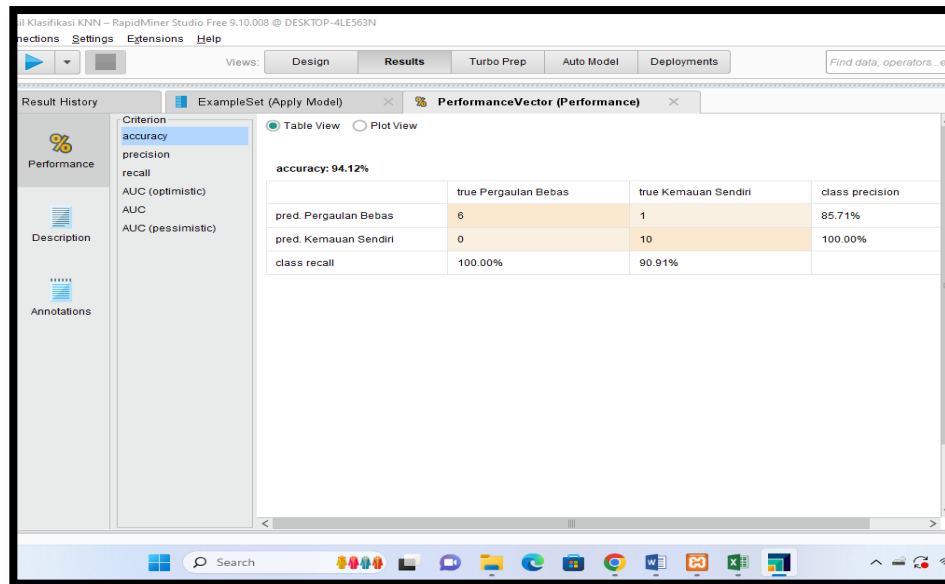
Gambar 2 Pemodelan (Modelling)

3.2 Setelah dilakukan proses dengan algoritma *k-nn* maka diperoleh hasil penerapan model klasifikasi pada data faktor Pernikahan Dini menggunakan *RapidMiner*. Tabel yang ditampilkan berada pada tab "Results" serta hasil prediksi model beserta tingkat kepercayaannya (*confidence*). Kolom "*prediction* (keterangan)" menunjukkan hasil klasifikasi, sedangkan "*confidence*" menunjukkan tingkat keyakinan model terhadap prediksi tersebut. Dari tampilan ini, pengguna dapat mengevaluasi seberapa akurat model dalam mengklasifikasikan faktor-faktor yang memengaruhi Pernikahan Dini berdasarkan data yang diberikan.

| Row No. | Name           | Faktor Pernikahan | prediction(Fak... | confidence... | confidence... | Kecamatan | Jenis Kelamin | Umr |
|---------|----------------|-------------------|-------------------|---------------|---------------|-----------|---------------|-----|
| 1       | Ast            | Pergaulan Bebas   | Pergaulan Bebas   | 1             | 0             | 2         | 5             | 16  |
| 2       | Yudaman        | Pergaulan Bebas   | Pergaulan Bebas   | 0.674         | 0.326         | 1         | 5             | 17  |
| 3       | Ard            | Pergaulan Bebas   | Pergaulan Bebas   | 0.665         | 0.315         | 1         | 1             | 16  |
| 4       | Thomas         | Kemauan Sendiri   | Kemauan Sendiri   | 0             | 1             | 1         | 5             | 17  |
| 5       | Peru           | Kemauan Sendiri   | Kemauan Sendiri   | 0.311         | 0.689         | 1         | 4             | 18  |
| 6       | Kusuma Wati    | Kemauan Sendiri   | Kemauan Sendiri   | 0.349         | 0.651         | 2         | 3             | 18  |
| 7       | Jamal Adhita   | Kemauan Sendiri   | Kemauan Sendiri   | 0.326         | 0.674         | 1         | 5             | 17  |
| 8       | Juwanti        | Pergaulan Bebas   | Pergaulan Bebas   | 1             | 0             | 2         | 4             | 15  |
| 9       | Eko Yulianto   | Kemauan Sendiri   | Kemauan Sendiri   | 0             | 1             | 1         | 3             | 19  |
| 10      | Bunga Lestari  | Kemauan Sendiri   | Kemauan Sendiri   | 0             | 1             | 2         | 2             | 19  |
| 11      | Jinan          | Pergaulan Bebas   | Pergaulan Bebas   | 1             | 0             | 2         | 4             | 15  |
| 12      | Eki Lorenzkyah | Kemauan Sendiri   | Kemauan Sendiri   | 0             | 1             | 2         | 3             | 18  |
| 13      | Ayo Fransiska  | Kemauan Sendiri   | Kemauan Sendiri   | 0             | 1             | 2         | 3             | 19  |
| 14      | Ace Azhan      | Pergaulan Bebas   | Pergaulan Bebas   | 0.520         | 0.520         | 1         | 3             | 17  |

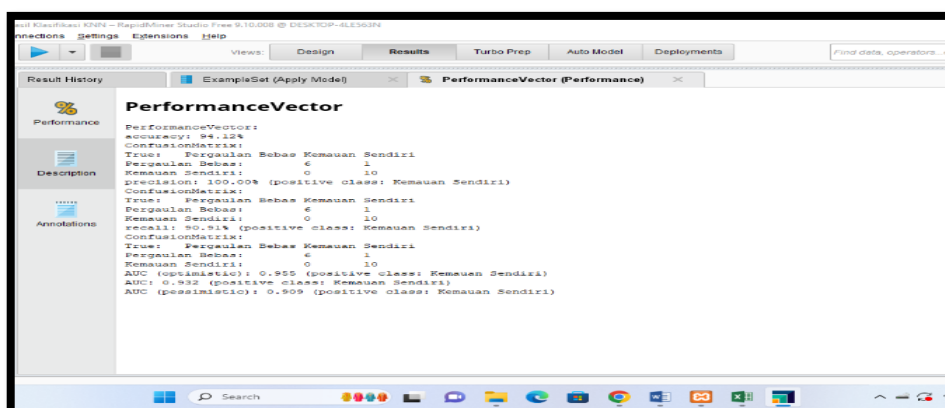
Gambar 3 Hasil Proses Algoritma K-NN

3.3. Hasil Evaluasi *Performance* model klasifikasi *K-Nearest Neighbor* di *Software RapidMiner*, menghasilkan nilai akurasi 94,12% *class precision* Kemauan Sendiri 85,71% *class precision* Pergaulan Bebas 100,00%. Dan *Class Recall* Kemauan Sendiri 90,91% *Class Recall* Pergaulan Bebas 100,00%. Seperti terdapat pada gambar dibawah ini.



Gambar 4 Hasil Akurasi

3.4 Tampilan hasil evaluasi *performance* model KNN di *RapidMiner* dengan nama *PerformanceVector* model menunjukkan tingkat akurasi klasifikasi 94,12%, *Confusion Matrix* dengan 6 data Pergaulan Bebas (*True Positive*), 1 data Kemauan Sendiri (*False Positive*), 10 data Pergaulan Bebas (*True Negative*), dan 0 data Pergaulan Bebas sebagai Kemauan Sendiri (*False Negatif*-tidak ada). *Precision* 85,71% benar-benar layak, *Recall* 100,00% tidak layak. Nilai AUC (*Area Under Curve*) sebesar 0.955. Seperti gambar di bawah ini.



Gambar 5 Evaluation Performance

#### 4. Conclusion

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini berhasil menghasilkan sebuah sistem klasifikasi yang digunakan untuk mengetahui faktor penyebab pernikahan dini di Kota Pagar Alam. Proses klasifikasi yang dilakukan menggunakan model K-Nearest Neighbor (KNN) melalui RapidMiner menunjukkan hasil yang konsisten dengan sistem yang telah dibangun, dengan tingkat akurasi yang tinggi sebesar 94,12%, sehingga model dinilai mampu mengklasifikasikan data dengan sangat baik. Selain itu, hasil pengujian menggunakan metode Black Box Testing pada tahap Alpha memperoleh nilai kelayakan sebesar 83,2%, yang menunjukkan bahwa sistem yang dikembangkan tergolong sangat layak dan sesuai untuk diterapkan.

#### References

- [1] L. N. Jamil, M. D. Lestari, and R. S. Utari, "Pengaruh Pernikahan Dini Terhadap Pendapatan Keluarga di Pulau Jawa," vol. 900, 2022.
- [2] U. M. Kudus, U. M. Kudus, and U. M. Kudus, "F y b p," vol. 11, no. 2, pp. 280–290, 2020.
- [3] S. Loviana and A. Wafiani, "DEDIKASI : Jurnal Pengabdian Masyarakat," vol. 4, no. 2, pp. 134–143, 2022.
- [4] D. E. Pangestika, D. A. Mustika, and A. A. Rahman, "Perbandingan Performa Metode Klasifikasi Regresi Logistik , Classification Tree , dan Random Forest Studi Kasus Perkawinan Anak pada Perempuan Usia Muda di Provinsi Nusa Tenggara," pp. 499–510, 2024.
- [5] E. R. Syalis and N. Nurwati, "REMAJA," vol. 3, pp. 29–38, 2020.
- [6] Y. Pinanda, W. F. Mahmudy, and E. Santoso, "Klasifikasi Risiko Penyakit pada Ibu Hamil menggunakan Metode Modified K-Nearest Neighbor ( MKNN )," vol. 6, no. 5, pp. 2116–2121, 2022.
- [7] I. Artikel, "Klasifikasi Dataset Diabetes menggunakan Algoritma K-Nearest," vol. 2, no. 1, pp. 36–42, 2024.
- [8] A. L. Muntamah *et al.*, "PERNIKAHAN DINI DI INDONESIA : FAKTOR DAN PERAN PEMERINTAH ( PERSPEKTIF PENEGAKAN DAN PERLINDUNGAN HUKUM BAGI ANAK )," vol. 21, no. 1, pp. 1–12, 2016.
- [9] V. Sianipar, D. Irmayani, and B. Bangun, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kelulusan Siswa Menggunakan Algoritma KNN," vol. 7, no. 1, 2025, doi: 10.47065/bits.v7i1.7386.
- [10] M. S. Mustafa and I. W. Simpen, "Implementasi Algoritma K-Nearest Neighbor ( KNN ) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," vol. VIII, no. 1, pp. 1–10.
- [11] D. Rahmania and R. Amelia, "MATH unesa," vol. 13, no. 0, pp. 165–171, 2025.
- [12] A. Hermambang, C. Ummah, E. S. Gratia, F. Sanusi, W. M. Ulfa, and R. Nooraeni, "Faktor-faktor yang memengaruhi pernikahan usia dini di Indonesia Factors affecting early marriage in Indonesia," vol. 16, no. 1, pp. 1–12, 2021, doi: 10.14203/jki.v16i1.428.
- [13] Feri sulianta, "Buku Dasar Data Mining from A to Z," no. January, pp. 15–15, 2023.
- [14] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [15] P. Pahrudin and K. Harianto, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial," vol. 4, no. 3, pp. 1241–1245, 2022, doi: 10.47065/bits.v4i3.2276.