



## Klasifikasi Kesehatan Calon Pengantin Menggunakan Algoritma Modified K-Nearest Neighbor (MKNN)

Dewangga Aldiar Rozzaq\*, Lutfi Ali Muharom, Daryanto

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Email: barqydewa@gmail.com, lutfi.muhamrom@unmuhjember.ac.id, daryanto@unmuhjember.ac.id

### ABSTRAK

Kasus pernikahan dini masih sering terjadi di berbagai daerah dengan berbagai latar belakang. Oleh karena itu, kasus ini telah menjadi perhatian komunitas internasional. Pilihan gaya hidup seseorang dapat mempengaruhi perilaku seseorang, seperti gaya hidup baik dan sehat, maka orang yang memiliki gaya hidup seperti itu perilaku kesehariannya cenderung melakukan sesuatu hal positif dan bermanfaat dan hal itu juga dapat berpengaruh terhadap perilaku kesehariannya seperti dalam perilaku seksual seseorang yang belum menikah maupun setelah menikah. Maka dari itu kita harus lebih pintar dan bijak dalam memilih pasangan dengan memperhatikan, menyelidiki dan mengenal pasangan kita terlebih dahulu. Faktor yang dapat dijadikan pertimbangan ialah mengetahui riwayat kesehatan, kepribadian dan perilaku gaya hidupnya. Hal ini tentu berkaitan dengan pemeriksaan kesehatan pra nikah seperti penyelidikan, pengamatan dan pemeriksaan kondisi kesehatan jasmani dan rohani yang berguna untuk kelangsungan pernikahan. Tujuan dari penelitian ini adalah untuk mengklasifikasikan kesehatan calon pengantin yang “resiko rendah” dan “resiko tinggi”. Dalam penelitian ini menggunakan Algoritma Modified K-Nearest Neighbor (MKNN) dalam melakukan klasifikasi pada Kesehatan Calon Pengantin. Dengan performa akurasi sebesar 88%, nilai presisi sebesar 90,91%, serta nilai *recall* sebesar 100% yang didapat pada  $k=3$ .  
**Kata Kunci:** Kesehatan, Calon Pengantin, Klasifikasi, Modified K-Nearest Neighbor (MKNN).

### ABSTRACT

*Cases of early marriage still occur frequently in various regions with various backgrounds. Therefore, this case has attracted the attention of the International Community. A person's lifestyle choices can influence a person's behavior, such as a good and healthy lifestyle, so people who have a lifestyle like that, their daily behavior tends to do something positive and useful and this can also influence their daily behavior, such as the sexual behavior of someone who is not married. or after marriage. Therefore, we must be smarter and wiser in choosing a partner by paying attention, investigating and getting to know our partner first. Factors that can be taken into consideration are knowing their health history, personality and lifestyle behavior. This is of course related to pre-marital health checks such as investigations, observations and examinations of physical and spiritual health conditions which are useful for the continuity of the marriage. The aim of this research is to classify the health of prospective brides and grooms who are "low risk" and "high risk". In this research, an algorithm is used Modified K-Nearest Neighbor (MKNN) in classifying the health of prospective brides and grooms. With accuracy performance of 88%, precision value of 90.91%, as well as value recall equal to 100% which is obtained at  $k=3$ .*

**Keywords:** Health, Bride and Groom Candidates, Classification, Modified K-Nearest Neighbor (MKNN).

## 1. PENDAHULUAN

Kasus pernikahan dini masih sering terjadi di berbagai daerah dengan berbagai latar belakang. Oleh karena itu, kasus ini telah menjadi perhatian komunitas internasional. Oleh karena itu, diharapkan bahwa peran orang tua, termasuk dokter anak dan keluarga, akan meningkatkan kesadaran untuk mengurangi dan menghentikan pernikahan usia dini (Fadlyana & Larasaty, 2016).

Pilihan gaya hidup seseorang dapat mempengaruhi perilaku seseorang, seperti gaya hidup baik dan sehat, maka orang yang memiliki gaya hidup seperti itu perilaku kesehariannya cenderung

melakukan sesuatu hal positif dan bermanfaat dan hal itu juga dapat berpengaruh terhadap perilaku kesehariannya seperti dalam perilaku seksual seseorang yang belum menikah maupun setelah menikah. Maka dari itu, kita harus lebih pintar dan bijak dalam memilih pasangan dengan memperhatikan, menyelidiki dan mengenal pasangan kita terlebih dahulu sebelum memutuskan untuk menjalani kehidupan dalam berumah tangga.

Faktor yang dapat dijadikan pertimbangan untuk memilih pasangan ialah mengetahui riwayat kesehatan, kepribadian dan perilaku gaya hidupnya. Hal ini tentunya berkaitan dengan pemeriksaan kesehatan pra nikah seperti penyelidikan, pengamatan dan pemeriksaan kondisi kesehatan jasmani dan rohani yang berguna untuk kelangsungan pernikahan.

Terdapat penelitian yang menjadikan acuan dalam menggunakan Algoritma *Modified K-Nearest Neighbor* yaitu penelitian yang berjudul “Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode *Modified K-Nearest Neighbor*” dengan hasil pengujian tingkat akurasi tertinggi sebesar 93,33% dengan nilai  $K=3$  (Fernanda dkk., 2017).

Berdasarkan latar belakang tersebut, peneliti akan melakukan penelitian yang berjudul “Klasifikasi Kesehatan Calon Pengantin Menggunakan Algoritma *Modified K-Nearest Neighbor* (MKNN)” guna membantu SEKDA-DPPKB (Dinas Pemberdayaan Perempuan dan Keluarga Berkualitas) dalam mengevaluasi data calon pengantin.

## 2. KAJIAN PUSTAKA

### A. Kesehatan

Kesehatan sangat penting bagi manusia karena ketika mereka sehat, mereka dapat beraktivitas dengan nyaman dan berbagi kebaikan dengan sesama. Sehat fisik didefinisikan sebagai keadaan di mana seseorang dapat menjalani kegiatan sehari-hari secara normal tanpa mengalami gangguan fisik dan memungkinkan perkembangan mental atau psikologis dan sosial. Sehat sosial sendiri merupakan kehidupan dalam masyarakat, dimana seseorang memiliki cukup kemampuan untuk memelihara serta memajukan kehidupannya sendiri dan kehidupan keluarga atau lingkungan sekitarnya (Fuadi Husin, 2014).

### B. Data Mining

Data mining adalah bidang keilmuan yang menangani masalah pengambilan informasi dari database yang sangat besar. Ini melakukannya dengan menganalisis dan mengevaluasi kumpulan data untuk menemukan hubungan yang tidak terduga. Kemudian, informasi disimpulkan dengan cara yang berbeda dari sebelumnya sehingga mudah dipahami dan bermanfaat bagi pemilik data (Utomo & Mesran, 2020).

### C. Klasifikasi

Dengan menggunakan hasil pengamatan terhadap data dan atributnya, klasifikasi digunakan untuk merencanakan fungsi. Ini memungkinkan pemetaan data yang tidak memiliki kelas ke dalam data yang memiliki kelas, dan kemudian mereka dikelompokkan sesuai dengan aturan yang ditetapkan (Wafiyah dkk., 2017).

### D. Algoritma *Modified K-Nearest Neighbor* (MKNN)

Algoritma *Modified K-Nearest Neighbor* (MKNN) dimodifikasi dari metode *K-Nearest Neighbor* (KNN) dengan menambahkan proses seperti perhitungan bobot dan perhitungan nilai validitas. Dalam metode KNN, label kelas data dikaitkan dengan nilai  $k$ , dan validitas semua data pada data latih akan dihitung. Selanjutnya, validitas data digunakan untuk melakukan perhitungan *weight voting* pada semua data pengujian. Validitas data latih dapat menghasilkan tingkat akurasi yang tinggi. (Parvin dkk., 2008).

### E. *Euclidean Distance*

*Euclidean Distance* digunakan untuk menentukan jarak antara dua titik, yaitu titik pada data *training* ( $x$ ) dan titik pada data uji ( $y$ ), maka rumus *Euclidean* digunakan :

$$d(x,y) = \sqrt{\sum_i^n = 1(Xi - Yi)^2} \quad (1)$$

dengan  $d$  adalah jarak antara titik pada data *training*  $x$  dan titik data *testing* yang akan diklasifikasi, dimana  $x=x_1,x_2,\dots,x_i$  dan  $y=y_1,y_2,\dots,y_i$  dan  $i$  merepresentasikan nilai atribut serta  $n$  merupakan dimensi atribut (Wafiyah dkk., 2017).

#### F. Nilai Validitas

Untuk algoritma *Modified K-Nearest Neighbor*, setiap data dalam data pelatihan harus divalidasi terlebih dahulu, validitas setiap data bergantung pada setiap tetangganya. Setelah menentukan validitas data, nilai validitas digunakan untuk memberikan informasi tambahan tentang data (Wafiyah dkk., 2017).

Persamaan untuk menghitung nilai validitas pada setiap data *training* yaitu :

$$\text{Validity (i)} = \frac{1}{H} \sum_{i=1}^H S(\text{lbl}(x), \text{lbl}(N = N_i(x)))$$

Dimana :  $H$  : Jumlah titik terdekat  
 $\text{lbl}(x)$  : label kelas  $x$   
 $\text{lbl}(N_i(x))$  : label kelas titik terdekat  $x$

Fungsi  $S$  digunakan untuk menghitung kesamaan antara titik  $x$  dan data ke- $I$  dari tetangga terdekat. Yang dituliskan dalam persamaan

$$S(a,b) = \begin{cases} 1 & \text{jika } a = b \\ 0 & \text{jika } a \neq b \end{cases}$$

Keterangan :  $a$  = kelas  $a$  pada data training  
 $b$  = kelas lain selain  $a$  pada data training

#### G. Weight Voting

*Weight Voting* KNN adalah teknik KNN yang memanfaatkan tetangga terdekat, terlepas dari kelas data, dengan mendukung setiap data dalam data pelatihan. Berat *voting* diberikan kepada setiap data setara dengan beberapa fungsi jarak yang diketahui. Rumus yang digunakan untuk menghitung *weight* voting masing-masing tetangga dalam algoritma MKNN adalah sebagai berikut (Parvin dkk., 2010) :

$$W(i) = \text{Validity}(i) \times \frac{1}{de+a}$$

Keterangan :  $W(i)$  : Perhitungan *Weight Voting* (Bobot)  
 $\text{Validity}(i)$  : Nilai Validitas  
 $de$  : Jarak *Euclidean*  
 $a$  : Konstanta

#### H. K-Fold Cross Validation

*K-Fold Cross Validation* digunakan untuk menggeneralisasi kumpulan data independen, hasil statistik analisis dinilai dengan metode validasi model yang dikenal sebagai *cross validation*, atau estimasi rotasi. Ini juga digunakan untuk memprediksi model dan mengevaluasi seberapa akurat model prediktif saat diterapkan dalam praktiknya. *K-Fold Cross Validation* membagi data menjadi  $k$  bagian dari kumpulan data yang sama. Metode ini digunakan untuk menurunkan bias data. Instruksi dan penilaian diberikan sebanyak  $k$  kali. Dalam percobaan pertama, subset  $S_1$  digunakan sebagai data uji dan subset yang berbeda digunakan sebagai data latih, percobaan berikutnya subset  $S_1, S_3, \dots, S_k$  diubah menjadi data latih dan  $S_2$  diubah menjadi data uji dan seterusnya (Tempola dkk., 2018)

#### I. Google Colab

Google Colab merupakan *platform* pengembangan berbasis *cloud* yang memungkinkan penggunaannya menggunakan Python untuk melakukan pemrograman dan analisis data. (Guntara, 2023). Karena tidak perlu menginstal perangkat lunak pada komputer lokal, Google Colab membuat visualisasi data lebih mudah.

J. *Confusion Matrix*

Tabel menunjukkan jumlah data uji yang benar diklasifikasikan dan yang salah diklasifikasikan. *Confusion Matrix* adalah metode yang umum digunakan untuk menghitung akurasi konsep data *mining* (Rahman dkk., 2017).

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		1	2
Kelas Aktual	1	TP	FP
	2	FN	TN

Keterangan :

- *True Positive* (TP) merupakan jumlah data *positive* yang terklasifikasi benar oleh *system*.
- *True Negative* (TN) merupakan jumlah data *negative* yang terklasifikasi benar oleh *system*.
- *False Positive* (FP) merupakan jumlah data *positive* namun terklasifikasi salah oleh *system*.
- *False Negative* (FN) merupakan jumlah data *negative* namun terklasifikasi salah oleh *system*.

Untuk perhitungan akurasi, presisi dan *recall* digunakan dengan persamaan :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{2}$$

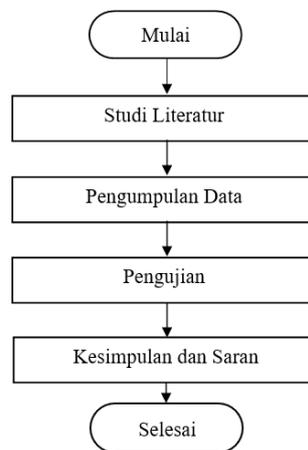
$$Presisi = \frac{TP}{TP+FP} \times 100\% \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{4}$$

3. METODE PENELITIAN

A. Tahapan Penelitian

Berikut tahapan penelitian yang akan dilakukan sebagai berikut:



Gambar 1. Tahapan Penelitian

B. Studi Literatur

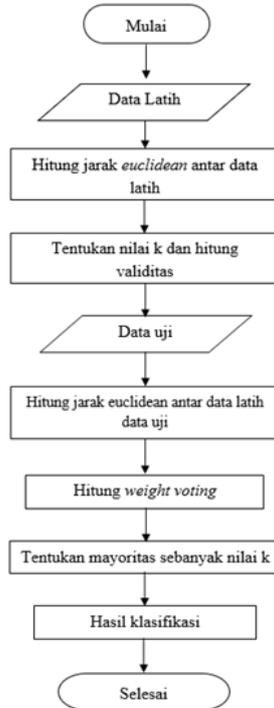
Studi literatur adalah tahap di mana informasi dan data dikumpulkan dengan membaca artikel dari buku atau jurnal yang berkaitan dengan penelitian pengelompokan dataset. Proses ini dilakukan dengan menggunakan algoritma *Modified K-Nearest Neighbor*.

C. Pengumpulan Data

Dalam penelitian ini menggunakan data sekunder yang diperoleh dari KUA Daerah Kabupaten Bondowoso pada Bulan Oktober hingga Desember tahun 2021 yang berjumlah 200 data. Dataset ini akan diolah menggunakan algoritma *Modified K-Nearest Neighbor*.

D. Implementasi Algoritma *Modified K-Nearest Neighbor*

Pada Implementasi algoritma ini ada beberapa langkah langkah yang harus dilakukan. Berikut adalah *flowchart* dari alur algoritma *Modified K-Nearest Neighbor* yang akan digunakan untuk klasifikasi pengaruh kesehatan pada calon pengantin:



Gambar2. Hasil Klasifikasi Diagram Alur Algoritma MKNN

4. HASIL DAN PEMBAHASAN

A. *Pre-processing* Data

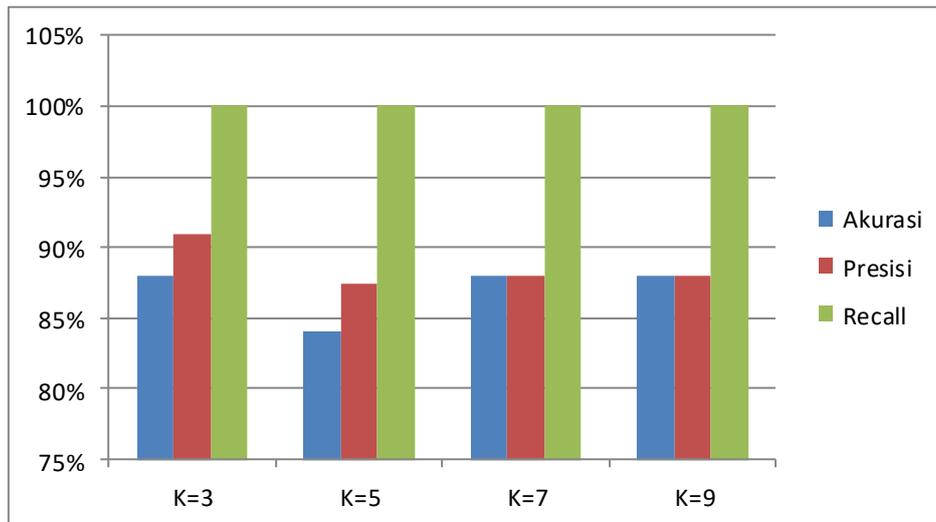
Sebelum data digunakan dalam proses pengujian, tahap *pre-processing* dilakukan setelah pengumpulan data. Proses ini akan membersihkan data dari elemen yang dianggap tidak penting atau suara. Hanya atribut yang dianggap penting yang akan digunakan dalam penelitian yang digunakan dalam tabel data. Pada data Calon Pengantin penentuan label diberikan pada kategori Resiko Rendah (0), Resiko Tinggi (1). Atribut yang akan digunakan pada proses pengujian yaitu diantaranya Merokok Pria, Merokok Wanita, Miras Pria, Miras Wanita, Pergaulan Pria, Pergaulan Wanita, Penyakit Pria dan Penyakit Wanita.

B. Hasil Klasifikasi

Pada proses klasifikasi ini dilakukan menggunakan Algoritma *Modified K-Nearest Neighbor* (MKNN). Selanjutnya adalah proses menghitung hasil klasifikasi *Modified K-Nearest Neighbor*, klasifikasi ini untuk mengetahui nilai akurasi, presisi dan *recall*. Dengan *K-Fold Cross Validation Fold 2, Fold 4, Fold 5, dan Fold 8*. Menggunakan nilai K=3, K=5, K=7 dan K=9.

Tabel 2. Hasil Pengujian pada k-fold 8

Nilai K	Akurasi	Presisi	Recall
3	88%	90.91%	100%
5	84%	87.50%	100%
7	88%	88%	100%
9	88%	88%	100%



Gambar 3. Hasil Klasifikasi

Berdasarkan tabel hasil akurasi, presisi, *recall* untuk *dataset* Kesehatan Calon Pengantin dengan algoritma *Modified K-Nearest Neighbor* menggunakan *K-fold Cross Validation*. Dengan menggunakan nilai K yang beragam, yaitu nilai K=3, K=5, K=7, dan K=9, pencarian mampu menghasilkan nilai persentase akurasi sebesar 88%, presisi sebesar 90.91%, dan *recall* sebesar 100%. Pada percobaan dengan nilai K=3, sistem menggunakan hasil klasifikasi yang seharusnya pada *matrix confusion* untuk mendapatkan nilai tertinggi.

## 5. KESIMPULAN DAN SARAN

### A. KESIMPULAN

Hasil penelitian yang dilakukan tentang Klasifikasi Kesehatan Calon Pengantin Menggunakan Algoritma *Modified K-Nearest Neighbor* (MKNN) dapat disimpulkan bahwa :

- 1) Hasil pengujian nilai akurasi tertinggi yang didapat dari algoritma *Modified K-Nearest Neighbor* pada klasifikasi kesehatan calon pengantin memperoleh presentasi tertinggi yaitu 88% pada percobaan dengan nilai K=3 pada k-fold 8 skenario 3.
- 2) Hasil pengujian nilai presisi tertinggi yang didapat dari algoritma *Modified K-Nearest Neighbor* pada klasifikasi kesehatan calon pengantin memperoleh presentasi tertinggi 90.91% pada percobaan dengan nilai K=3 pada k-fold skenario 5.
- 3) Hasil pengujian nilai *recall* tertinggi yang didapat dari algoritma *Modified K-Nearest Neighbor* pada klasifikasi kesehatan calon pengantin memperoleh presentasi tertinggi 100% pada percobaan dengan nilai K=3 pada k-fold skenario 3

### B. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, tentu masih terdapat beberapa kekurangan sehingga perlu adanya penyempurnaan oleh penelitian selanjutnya. Berikut beberapa rekomendasi untuk penelitian lanjutan yaitu :

- 1) Penelitian ini dapat digunakan sebagai referensi untuk penelitian berikutnya.
- 2) Meningkatkan jumlah data yang akan digunakan untuk penelitian selanjutnya.
- 3) Nilai k dan k-fold yang lebih beragam dapat digunakan untuk penelitian selanjutnya.

## 6. DAFTAR PUSTAKA

Anggian, F. C., Hidayat, N., & Furqon, M. T. (2019). Implementasi Metode *Modified K-Nearest Neighbor* untuk Klasifikasi Status Gunung Berapi. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(12), 11027-11033.

- Dasuki, M. (2021). Optimasi Nilai Bobot Algoritma Backpropagation Neural Network Dengan Algoritma Genetika. *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, 6(1), 38-44.
- Diniyati, L. S., & Jayatmi, I. (2017). Pengaruh Empat Variabel Terhadap Perilaku Pernikahan Dini Perempuan Pesisir. *Jurnal Ilmiah Kesehatan*, 16(02), 14-29.
- Fauzi, L., Dasuki, M., & Muharom, L. A. (2024). Penerapan Algoritma Support Vector Machine untuk Klasifikasi Sentimen Vaksin Booster pada Twitter. *JUSTIFY: Jurnal Sistem Informasi Ibrahimi*, 2(2), 135-143.
- Fauziatin, N., Kartini, A., & Nugraheni, S. A. (2019). Pengaruh Pendidikan Kesehatan Dengan Media Lembar Balik Tentang Pencegahan Stunting Pada Calon Pengantin. *VISI KES: Jurnal Kesehatan Masyarakat*, 18(2), 224-233.
- Fernanda, S. I., Ratnawati, D. E., & Adikara, P. P. (2017). Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K-Nearest Neighbor (MKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(6), 507-513.
- Guntara, R. G. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, 2(6), 2091-2100.
- Mandalika, E. A., Dasuki, M., & Umilasari, R. (2024). Implementasi Algoritma Naïve Bayes Menggunakan Pemilihan Atribut Information Gain Pada Penyakit Diabetes. *Jurnal Smart Teknologi*, 5(4), 430-438.
- Mughniy, M., Wihandika, R. C., & Prasetyo, B. H. (2018). Sistem Rekomendasi Psikotes Untuk Penjurusan Siswa SMA Menggunakan Metode Modified K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 282-287.
- Parvin, H., Alizadeh, H., & Minaei-Bidgoli, B. (2008, October). MKNN: Modified k-Nearest Neighbor. *In Proceedings of the world congress on engineering and computer science* (Vol. 1). Newswood Limited.
- Parvin, H., Alizadeh, H., & Minati, B. (2010). A Modification On K-Nearest Neighbor Classifier. *Global Journal of Computer Science and Technology*, 10(14), 37-41.
- Putra, M. I. P., Murdiansyah, D. T., & Aditsania, A. (2019). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) Untuk Klasifikasi Penyakit Kanker Payudara. *eProceedings of Engineering*, 6(1).