



## IDENTIFIKASI STATUS PERLINDUNGAN ANAK MENGGUNAKAN ALGORITMA KLASIFIKASI K-NEAREST NEIGHBORS

### *Identification Of Child Protection Status Using K-Nearest Neighbors Classification Algorithm*

Muh. Jabal Nur<sup>1</sup>, Tri Wahyudi Budi Sastra<sup>2</sup>, Muhammad Faisal<sup>3\*</sup>

<sup>1,2,3</sup>STMIK Profesional Makassar

E-mail : muhfaisal@stmikprofesional.ac.id

#### **Abstract**

The United Nations Children's Fund (UNICEF) has defined the child protection system as "certain formal and informal structures, functions and capacities that have been pooled to prevent and respond to violence, abuse, neglect and exploitation of children". Problems identified by the Sulawesi Community Foundation group (CSF) is the absence of digital records that accommodate the results of direct interviews with children in terms of measuring experiences or actions that have been experienced by a child as a source of information that can be used for decision-making needs. The KNN algorithm is one of the most popular data mining algorithms. It has been widely and successfully applied to data analysis applications across various research topics in computer science. Based on the value of Classification Y, obtained as much as 1 data Category = Unprotected, and 3 data Category = Protected, then the resulting decision is that the identification of child protection status in an area is based on attribute values  $X1=5$ ,  $X2=3$ ,  $X3=50$  and  $X4 =$  Does not result in a decision that the majority of children have guaranteed protection status. The decisions made are in accordance with the facts, therefore the Sulawesi Community Foundation NGO can apply the results of analysis and testing into an integrated system so that it can become a valid source of information in digital format.

**Keywords:** *knn, child protection, classification, algorithm, NGOs, identification.*

#### **Abstrak**

United Nations Children's Fund (UNICEF) telah mendefinisikan sistem perlindungan anak sebagai "struktur formal dan informal tertentu, fungsi dan kapasitas yang telah dikumpulkan untuk mencegah dan menanggapi kekerasan, pelecehan, penelantaran, dan eksploitasi anak". Permasalahan yang ditemukan oleh kelompok Sulawesi Community Foundation (CSF) adalah belum adanya catatan digital yang menampung hasil wawancara langsung terhadap anak dalam hal pengukuran pengalaman atau tindakan yang pernah dialami oleh seorang anak sebagai sumber informasi yang dapat digunakan untuk kebutuhan pengambilan keputusan. Algoritma KNN adalah salah satu algoritma data mining yang paling populer. Ini telah secara luas dan berhasil diterapkan pada aplikasi analisis data di berbagai topik penelitian dalam ilmu komputer. Berdasarkan nilai Klasifikasi Y, maka diperoleh sebanyak 1 data Kategori = Tidak Terlindungi, dan 3 data Kategori = Terlindungi, maka keputusan yang dihasilkan bahwa identifikasi status perlindungan anak terhadap suatu wilayah berdasarkan nilai atribut  $X1=5$ ,  $X2=3$ ,  $X3=50$  dan  $X4=Tidak$  menghasilkan keputusan Bahwa mayoritas anak mendapat status perlindungan Terjamin. keputusan yang dihasilkan sesuai dengan fakta, oleh karena itu LSM Sulawesi Community Foundation dapat menerapkan hasil analisa dan pengujian

kedalam sistem terintergrasi sehingga dapat menjadi sumber informasi yang valid dalam format digital.

**Kata Kunci:** knn, perlindungan anak, klasifikasi, algoritma, lsm, identifikasi

## **PENDAHULUAN**

United Nations Children's Fund (UNICEF) telah mendefinisikan sistem perlindungan anak sebagai "struktur formal dan informal tertentu, fungsi dan kapasitas yang telah dikumpulkan untuk mencegah dan menanggapi kekerasan, pelecehan, penelantaran, dan eksploitasi anak". Komponen utama sistem perlindungan anak meliputi hukum dan kebijakan, sumber daya manusia dan keuangan, tata kelola, sarana pengumpulan data dan pemantauan sistem, layanan perlindungan dan tanggapan anak, dan dukungan nonformal dari keluarga dan masyarakat, oleh karena itu lembaga swadaya masyarakat Sulawesi Community Foundation (SCF) yang aktif dalam kegiatan perlindungan anak terus menggalakan edukasi terhadap masyarakat tentang pentingnya perlindungan terhadap anak.

Penelitian yang dilakukan oleh Floor Middel menyatakan bahwa Anak berhak untuk berpartisipasi dalam investigasi perlindungan anak, dimana kunci pendekatan berbasis hak dalam upaya melindungi anak adalah fokus pada hak anak untuk terlibat dalam pengambilan keputusan dalam penelitian yang mempengaruhi kehidupan mereka. Permasalahan yang ditemukan oleh kelompok Sulawesi Community Foundation (CSF) adalah belum adanya catatan digital yang menampung hasil wawancara langsung terhadap anak dalam hal pengukuran pengalaman atau tindakan yang pernah dialami oleh seorang anak sebagai sumber informasi yang dapat digunakan untuk kebutuhan pengambilan keputusan. Selayaknya sebagai lembaga swadaya masyarakat formal maka LSM-CSF harus memiliki banyak sumber informasi yang dapat digunakan baik untuk kebutuhan internal maupun external.

Berdasarkan uraian dan permasalahan yang ditemukan maka penulis akan melakukan uraian topik yang membahas tentang metode identifikasi status perlindungan anak menggunakan algoritma klasifikasi K-NN sebagai alat ukur melalui bidang teknologi informasi.

## **METODE**

Pada bagian metodologi dijelaskan secara teknis tentang cara, metode atau trik agar dapat diketahui komponen yang dilibatkan serta dapat menjadi pedoman dalam proses penelitian yang.

Kegiatan penelitian menggunakan metode dalam bentuk ceramah, mengisi kuesioner, tanya jawab. Adapun tahapan-tahapan dalam pelaksanaan kegiatannya antara lain :

- a) Ceramah digunakan untuk menyampaikan pengetahuan secara umum tentang hak-hak anak, kebijakan perlindungan anak, perlindungan anak dan tindak kekerasan terhadap anak.

- b) Kuesioner diberikan untuk menggali informasi tentang kejadian yang menjadi paramter klasifikasi antara lain pelecehan seksual, bullying ,exploitas anak dan kekerasan fisik.
- c) Tanya jawab digunakan untuk melengkapi hal-hal yang belum terakomodasi dari kegiatan sebelumnya.
- d) Diskusi Terbatas melibatkan stakeholders yang berkepentingan langsung dengan perlindungan anak guna memastikan diadakannya kegiatan lanjutan berupa pelatihan aktifis /kader perlindungan anak terpadu.

Untuk memudahkan penelitian ini dalam penerapan metode atau teknik yang digunakan maka akan dijelaskan tentang teori-teori yang dihimpun dari beberapa sumber.

**Algoritma K-NEAREST NEIGHBORS(K-NN)**

Algoritma KNN adalah salah satu algoritma data mining yang paling populer. Ini telah secara luas dan berhasil diterapkan pada aplikasi analisis data di berbagai topik penelitian dalam ilmu komputer[4]. Algoritma metode k-nearest neighbor (KNN) bekerja berdasarkan jarak terdekat dari query instance ke training sample untuk menentukan keputusan. Training sample diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi training sample. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance, dituliskan dalam formula :

$$D(a,b) = \sqrt{\sum_{k=1}^d (X_k - Y_k)^2} \dots\dots\dots (1)$$

$X^1 = [X_1, X_2, \dots X_p]$  dan  $Y^1 = [Y_1, Y_2, \dots Y_p]$

Adalah

$$D2(x,y) = (X1 - Y1)2 + (X2 - Y2)2 + \dots\dots\dots + (Xp - Yp)2$$

$$= (X - Y) (X - Y) = (Y - X) (Y - X)$$

Dimana matriks  $D(a,b)$  adalah jarak skalar dari kedua vektor  $a$  dan  $b$  dari matriks dengan ukuran  $d$  dimensi. Semakin besar nilai  $D$  akan semakin jauh tingkat keserupaan antara kedua individu dan sebaliknya jika nilai  $D$  semakin kecil maka akan semakin dekat tingkat keserupaan antar individu tersebut.

K-Nearest Neighbor (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN. Menentukan parameter k (jumlah tetangga paling dekat).

K-NN merupakan salah satu algoritma yang digunakan dalam klasifikasi atau prediksi data baru. Prinsip kerja tetangga nilai K-NN adalah untuk menemukan jarak terdekat antara data yang akan dievaluasi oleh tetangga terdekat dalam data pelatihan, dituliskan dalam formula sebagai berikut :

$$\cos \alpha = \frac{AxB}{|A||B|} = \dots\dots\dots (2)$$

$$\frac{\sum_{i=1}^n AixBi}{\sqrt{\sum_{i=1}^n (Ai)^2} \times \sqrt{\sum_{i=1}^n (Bi)^2}}$$

cos  $\alpha$  : kesamaan *cosinus*

A : vektor A, yang akan dibandingkan kemiripannya

B : vektor B, yang akan dibandingkan kemiripannya

A x B : dot product antara vektor A dan vektor B

| A | : panjang vektor A

| B | : panjang vektor B

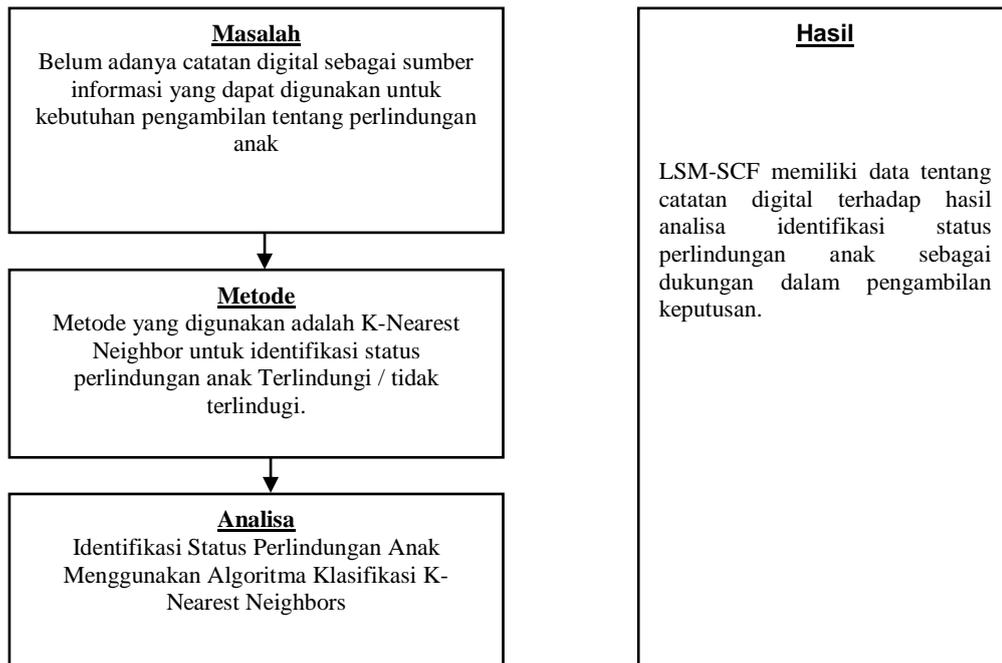
| A || B | : cross product antara | A | dan | B |

Tahapan algoritma K-NN adalah sebagai berikut :

- a) Menentukan parameter K (jumlah tetangga paling dekat).
- b) Menghitung kuadrat jarak euclid (query instance) masing-masing objek terhadap data sampel yang diberikan.
- c) Kemudian mengurutkan objek–objek tersebut kedalam kelompok yang mempunyai jarak euclid terkecil.
- d) Mengumpulkan kategori Y (Klasifikasi nearest neighbor)
- e) Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat dipredisikan nilai query instance yang telah dihitung.
- f) Menghitung kuadrat jarak eucliden objek terhadap data training yang diberikan.
- g) Mengurutkan hasil no 2 secara ascending (berurutan dari nilai tinggi ke rendah).

**Kerangka Konseptual**

Kerangka konseptual merupakan sebuah alur pemikiran terhadap suatu hubungan antar konsep satu dengan konsep yang lainnya untuk dapat memberikan gambaran dan mengarahkan asumsi terkait dengan variable-variable yang akan diteliti. Kerangka konseptual dalam penelitian ini, akan digambarkan dalam bentuk diagram yang diperlihatkan pada gambar 1.



Gambar 1. Kerangka Konseptual

## HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan proses analisa secara ilmiah melalui data sample yang diuji untuk mendapatkan informasi yang akan digunakan untuk menunjang diperolehnya hasil penelitian. Pada bagian ini akan dilakukan analisa indentifikasi status perlindungan anak melalui pengujian data yang dihimpun secara acak. Selanjutnya akan dilakukan identifikasi data baru menggunakan K-NN yang menjadi pedoman terhadap inputan data selanjutnya yang bekerja secara otomatis.

Ketepatan algoritma KNN sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur agar performa klasifikasi menjadi lebih baik. Terdapat beberapa data yang berasal dari survey questioner tentang status perlindungan anak apakah Terlindungi atau Tidak-Terlindungi melalui data pengujian menggunakan 4 atribut yaitu pelecehan seksual, bullying, exploitas anak dan kekerasan fisik. Klasifikasi data training dapat dilihat pada tabel 1.

Tabel 1. Data Training Kasus Perlindungan Anak

(X1)	(X2)	(X3)	(X4)	Y =
<b>Pelecehan Seksual</b>	<b>Bullying</b>	<b>Exploitas Anak</b>	<b>Kekerasan Fisik</b>	<b>Klasifikasi</b>

4	2	60	Ya	Terjamin
2	4	40	Tidak	Tidak Terjamin
2	2	30	Ya	Tidak Terjamin
6	5	85	Tidak	Terjamin
8	6	100	ya	Terjamin
3	3	65	Ya	Terjamin

Berdasarkan klasifikasi data training yang telah ditampung kemudian akan diidentifikasi status perlindungan anak berdasarkan inputan atribut  $X_1=5$ ,  $X_2=3$ ,  $X_3=50$  dan  $X_4=Tidak$ , maka dapat diklasifikasikan status perlindungan anak tersebut termasuk dalam kategori Terjamin atau Tidak Terjamin. Adapun prosedur algoritma K-NN adalah sebagai berikut :

- a) Menentukan parameter K (Jumlah tetangga paling dekat), misalkan kita menggunakan  $K = 4$ .
- b) Menghitung kuadrat jarak euclid (query-instance) masing-masing objek terhadap sampel data yang diberikan.

Koordinat query-instance adalah (5,3,50) dimana nilai tersebut berasal dari sub atribut yang akan diidentifikasi. Perhitungan Jarak Euclid dapat dilihat pada Tabel 2.

Tabel 2. Perhitungan Jarak Euclid

(X1)	(X2)	(X3)	(X4)	Square distance to query distance (50,3,40)
4	2	60	Ya	$(4-5)^2 + (2-3)^2 + (60-50)^2 = 102$
2	4	40	Tidak	$(2-5)^2 + (4-3)^2 + (40-50)^2 = 110$
2	2	30	Ya	$(2-5)^2 + (2-3)^2 + (30-50)^2 = 410$
6	5	85	Tidak	$(6-5)^2 + (5-3)^2 + (85-50)^2 = 905$
8	6	100	ya	$(8-5)^2 + (6-3)^2 + (100-50)^2 = 2518$
3	3	65	Ya	$(3-5)^2 + (3-3)^2 + (65-50)^2 = 229$

Langkah selanjutnya adalah mengurutkan objek-objek termasuk ke dalam kelompok yang mempunyai jarak euclid terkecil. Pengurutan jarak euclid terkecil dapat dilihat pada Tabel 3.

Tabel 3. Pengurutan jarak euclid terkecil

(X1)	(X2)	(X3)	(X4)	Jarak (50,3,40)	tan Jarak Terkecil	Keputusan h termasuk nearest neighbor (K)
4	2	60	Ya	$(4-5)^2 + (5-3)^2 + (60-50)^2 = 102$	1	Ya
2	4	40	Tidak	$(2-5)^2 + (4-3)^2 + (40-50)^2 = 110$	2	Ya
2	2	30	Ya	$(2-5)^2 + (2-3)^2 + (30-50)^2 = 410$	4	Ya
6	5	85	Tidak	$(6-5)^2 + (5-3)^2 + (85-50)^2 = 905$	5	Tidak
8	6	100	ya	$(8-5)^2 + (6-3)^2 + (100-50)^2 = 2518$	6	Tidak
3	3	65	Ya	$(3-5)^2 + (3-3)^2 + (65-50)^2 = 229$	3	Ya

Langkah yang dilakukan selanjutnya adalah Mengumpulkan kategori Y (klasifikasi nearest neighbor), seperti yang ditampilkan pada tabel 4.

Tabel 4. Klasifikasi Nearest Neighbor

(X1)	(X2)	(X3)	(X4)	Jarak (50,3,40)	tan Jarak Terkecil	Keputusan (K)	Klasifikasi Y
4	2	60	Ya	$(4-5)^2 + (5-3)^2 + (60-50)^2 = 102$	1	Ya	Terlindungi
2	4	40	Tidak	$(2-5)^2 + (4-3)^2 + (40-50)^2 = 110$	2	Ya	Terlindungi
2	2	30	Ya	$(2-5)^2 + (2-3)^2 + (30-50)^2 = 410$	4	Ya	Tidak Terlindungi
6	5	85	Tidak	$(6-5)^2 + (5-3)^2 + (85-50)^2 = 905$	5	Tidak	-
8	6	100	ya	$(8-5)^2 + (6-3)^2 + (100-50)^2 = 2518$	6	Tidak	-
3	3	65	Ya	$(3-5)^2 + (3-3)^2 + (65-50)^2 = 229$	3	Ya	Terlindungi

Dengan menggunakan kategori nearest neighbor yang paling mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung. Berdasarkan nilai Klasifikasi Y, maka diperoleh sebanyak 1 data Kategori = Tidak Terlindungi, dan 3 data Kategori = Terlindungi. karena nilai Y terbanyak adalah Terlindungi, maka keputusan yang dihasilkan bahwa identifikasi status perlindungan anak terhadap suatu wilayah berdasarkan nilai atribut X1=5, X2=3, X3=50 dan X4=Tidak menghasilkan keputusan Bahwa mayoritas anak mendapat status perlindungan Terjamin.

## KESIMPULAN

Berdasarkan pengujian berdasarkan data training yang diolah menggunakan seluruh tahapan pada algoritma K-NN, maka disimpulkan bahwa keputusan yang dihasilkan sesuai dengan fakta, oleh karena itu LSM Sulawesi



Community Foundation dapat menerapkan hasil analisa dan pengujian kedalam sistem terintergrasi sehingga dapat menjadi sumber informasi yang valid berbasis data digital.

#### **DAFTAR PUSTAKA**

- M. G. Wessells, “Bottom-up approaches to strengthening child protection systems: Placing children, families, and communities at the center,” *Child Abuse Negl.*, vol. 43, pp. 8–21, May 2015, doi: 10.1016/j.chiabu.2015.04.006.
- F. Middel, W. Post, M. López López, and H. Grietens, “Participation of Children Involved in the Child Protection System – Validation of the Meaningful Participation Assessment Tool (MPAT),” *Child Indic. Res.*, vol. 14, no. 2, pp. 713–735, Apr. 2021, doi: 10.1007/s12187-020-09772-2.
- J. Tobin and J. Cashmore, “Thirty years of the CRC: Child protection progress, challenges and opportunities,” *Child Abuse Negl.*, vol. 110, p. 104436, Dec. 2020, doi: 10.1016/j.chiabu.2020.104436.
- S. Zhang, “Challenges in KNN Classification,” *IEEE Trans. Knowl. Data Eng.*, 2021, doi: 10.1109/TKDE.2021.3049250.

