

PREDIKSI STUNTING PADA BALITA DENGAN MENGGUNAKAN ALGORITMA KLASIFIKASI K-NEAREST NEIGHBORS

STUNTING PREDICTION IN CHILDREN USING K-NEAREST NEIGHBORS CLASSIFICATION ALGORITHM

Hindratmo Hady Sutarno¹, Roswan Latuconsina², Ashri Dinimaharawati³

^{1,2,3} Universitas Telkom, Bandung

¹hindratmohs@student.telkomuniversity.ac.id, ²Roswan@telkomuniversity.co.id,

³ashridini@telkomuniversity.ac.id

Abstrak

Stunting merupakan salah satu kegagalan pertumbuhan pada anak akibat ketidakcukupan nutrisi gizi yang berlangsung dalam waktu lama. Dari data yang didapat mulai dari bulan Januari hingga bulan Oktober pada tahun 2020 yang diambil dari UPT Puskesmas Pitu berada pada Kabupaten Ngawi. Dari data tersebut akan dilakukan klasifikasi dengan menggunakan machine learning untuk menentukan prediksi stunting dan diharapkan dapat menjadi solusi untuk pencegahan kasus terjadinya stunting. Penentuan algoritma pada penelitian Tugas Akhir ini menggunakan algoritma K- Nearest Neighbors yang digunakan untuk menentukan hasil prediksi stunting berdasarkan parameter yang digunakan pada dataset. Dari sistem yang dibangun menunjukkan hasil akurasi yang didapat sebesar 97% berdasarkan dari pengujian partisi data menggunakan confusion matrix dengan menggunakan partisi data 90% data training dan 10% data testing dengan tetangga terdekat k=5.

Kata kunci : K-Nearest Neighbors, Machine Learning, Stunting.

Abstract

Stunting is one of the growth failures in children due to inadequate nutritional nutrition that lasts for a long time. From the data obtained from January to October in 2020 which was taken from the Pitu Health Center UPT located in Ngawi Regency. From this data, a classification will be carried out using machine learning to determine stunting predictions and is expected to be a solution for preventing stunting cases. Determination of the algorithm in this final project using the algorithm K-Nearest Neighbors which is used to determine stunting prediction results based on the parameters used in the dataset. The system built shows the accuracy obtained by 97% based on data partition testing using a confusion matrix using 90% data partitioning data training and 10% data testing with the nearest neighbor k=5.

Keywords: K-Nearest Neighbors, Machine Learning, Stunting.

1. Pendahuluan

Stunting merupakan salah satu kegagalan pertumbuhan pada anak akibat akumulasi ketidakcukupan nutrisi gizi yang berlangsung dalam waktu lama. Banyak faktor yang menyebabkan tingginya kejadian stunting pada balita pada usia 0 sampai 60 bulan. Penyebab langsung adalah kurangnya asupan gizi makanan dan pengetahuan ibu yang kurang, pola asuh yang salah, sanitasi dan hygiene yang buruk dan rendahnya pelayanan Kesehatan [1]. Stunting masih menjadi masalah utama yang terjadi di dunia. Hampir setengah dari angka kematian pada anak di bawah 5 tahun di Asia dan Afrika karena malnutrisi. Ini menyebabkan kematian dari 3 juta anak per tahun. Pada tahun 2011, Indonesia berada di peringkat 5 dari 81 negara dengan terbesar Jumlah anak stunting di dunia yang mencapai 7.547.000 anak. Pada tahun 2013, Indonesia memiliki kecenderungan prevalensi balita berdasarkan berat badan kurang indikator berat badan per umur 19,6% yang terbagi dalam status gizi kurang 5,7% dan gizi kurang status 13%, dan stunting berdasarkan indikator tinggi badan terhadap usia 37,2% [2].

Sedangkan dari kabupaten ngawi sendiri, yang dikutip dari website suara.ngawikab.go.id pada tahun 2019 dilansir dari mca-indonesia.go.id menyebutkan ada 8,9 juta anak Indonesia yang kurang gizi, dengan prevalensi stunting sebesar 37,2 %, Kondisi inilah yang sekarang menjadi perhatian serius Pemerintah Kabupaten Ngawi, untuk melakukan percepatan penurunan stunting. Pasalnya, di Kabupaten Ngawi ditemukan 200 anak dengan tumbuh kembang terlambat.

Pada penelitian Tugas Akhir ini diharapkan mampu dalam memberikan solusi pencegahan kasus terjadinya stunting di atas dengan membuat sistem prediksi menggunakan metode *machine learning* yang diimplementasikan kedalam aplikasi berbasis web. Data stunting diambil dari puskesmas pitu dari bulan Januari sampai dengan bulan Oktober yang berada pada kabupaten ngawi yang akan dilakukan klasifikasi dengan menggunakan *machine learning* untuk menentukan prediksi stunting. *machine learning* sendiri merupakan kecerdasan buatan dengan kemampuan untuk mempelajari data dengan metode mempelajari dari algoritma yang sudah ada. Banyak algoritma yang digunakan dalam pengolahan data klasifikasi. Penentuan algoritma pada penelitian Tugas Akhir ini menggunakan algoritma *K-Nearest Neighbors* adalah salah satu algoritma klasifikasi yang diambil dari tingkat akurasi yang bagus akan digunakan untuk memprediksi keadaan stunting pada balita dari data stunting berdasarkan hasil standar antropometri penilaian status gizi anak dalam keputusan Menteri kesehatan nomor 1995/MENKES/SK/XII/2010 pada tanggal 30 Desember 2010 parameter terkait nilai Z-score dalam tabel kategori dan ambang batas status gizi anak berdasarkan Z-score dengan batas -3 SD sampai dengan kurang dari -2 SD[3].

2. Perancangan Sistem

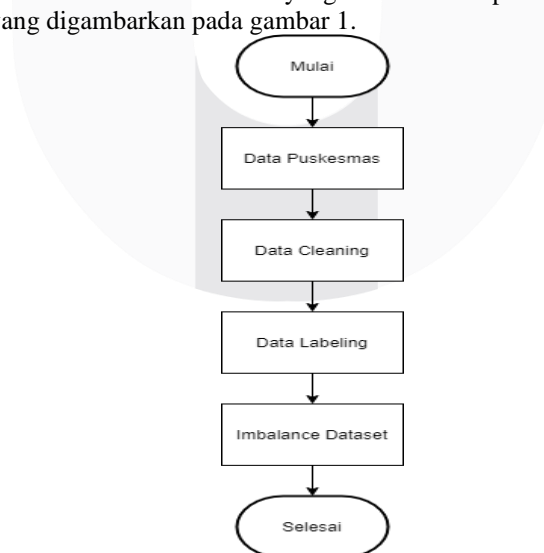
2.1 Pemodelan Machine Learning

Pengumpulan Data

Pada penelitian ini data yang digunakan diambil dari UPT Puskesmas Pitu, Jl. Raya Ngancar, Kecamatan Pitu, Kabupaten Ngawi, Jawa Timur. Data yang digunakan mulai dari bulan Januari sampai dengan bulan Oktober tahun 2020 pada 10 desa yang berbeda setelah dilakukan rekap data. Data yang diperoleh berjumlah 22855 data sebelum data digunakan pada pemodelan data akan dilakukan pembersihan data dengan menggunakan preprocessing.

Data Preprocessing

Preprocessing adalah teknik yang sering digunakan dalam machine learning dalam pengolahan data dengan mengubah data mentah yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan dalam pengolahan selanjutnya. Sebelum membangun model machine learning data yang digunakan perlu dibersihkan terlebih dahulu dengan menggunakan data preprocessing sehingga data lebih mudah untuk diproses oleh model machine learning dan meningkatkan performa akurasi dari model yang dibuat. Adapun beberapa tahapan data preprocessing seperti yang digambarkan pada gambar 1.



Gambar 1. Diagram Alir *Preprocessing*

Algoritma K-Nearest Neighbors

Setelah melakukan pembersihan dan pelabelan serta menyeimbangkan data maka pada tahap selanjutnya data yang sudah dibersihkan dapat digunakan untuk melakukan proses klasifikasi dengan menggunakan sklearn dengan algoritma klasifikasi k-nearest neighbors. K-Nearest neighbors adalah salah satu algoritma klasifikasi data menyimpan seluruh data kasus yang ada akan

diklasifikasikan menjadi data kasus-kasus baru berdasarkan kedekatan lokasi (jarak) [4]. Algoritma k-nearest neighbors merupakan suatu metode yang merupakan algoritma supervised. Pada supervised learning bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru [5]. Pada Penelitian ini dalam pemilihan algoritma k-nearest neighbors ini karena memiliki beberapa kelebihan, tangguh terhadap training data yang noisy dan efektif apabila training data-nya dalam jumlah besar [6]. Cara mencari kasus diklasifikasikan berdasarkan pencocokan bobot tetangga terdekatnya. Dalam menghitung jarak k-nearest neighbors menggunakan rumus jarak euclidean distance.

1. Tentukan parameter K (jumlah dari tetangga terdekat).
2. Hitung jarak antara data.
3. Urutkan jarak yang terbentuk (urut naik).
4. Tentukan jarak terdekat sampai urutan K.
5. Pasangkan kelas yang bersesuaian.
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

x_i : data latih ke-i
 y_i : data uji ke-i
 i : variabel data
 D : jarak
 n : dimensi data

Uji Performa

Pada tahap terakhir ini merupakan proses pengujian skenario performa pada proses ini terdapat beberapa proses pencarian nilai Precision, Recall, F1 score dan Accuracy. Dalam mengukur kinerja suatu algoritma klasifikasi dalam melakukan prediksi pada penelitian ini menggunakan *confusion matrix* dengan rumus sebagai berikut

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

$$Precision : \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$Recall : \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$F1 \text{ Score} : 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (5)$$

Keterangan:

TP : True Positive
 TN : True Negatif
 FP : False Positive
 FN : False Negatif

Selanjutnya mengukur kinerja dari algoritma dengan melakukan pengujian dengan menggunakan *k-fold cross validation*. *k-fold cross validation* merupakan salah satu dari jenis pengujian yang dilakukan untuk menilai kinerja dari proses sebuah metode algoritma dengan cara membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K pada k-fold. Kemudian salah satu dari kelompok k-fold tersebut akan dijadikan sebagai data testing sedangkan kelompok yang lain akan dijadikan sebagai data training [7].

3. Hasil dan Pengujian Sistem

3.1. Pengujian performa K-Nearest Neighbors

Berdasarkan rumus confusion matrix dalam mengukur kinerja pada algoritma klasifikasi dengan menggunakan nilai tetangga terdekat $k=5$ dilakukan dengan pengujian partisi data maka didapatkan nilai confusion matrix sebagai berikut.

Tabel 1. Confusion matrix partisi data

Pengujian	Data Training	Data Testing			Class Predicted	
					Negative	Positive
1	50%	50%	Class Actual	Negative	1855	92
				Positive	31	1925
2	70%	30%	Class Actual	Negative	1131	49
				Positive	19	1143
3	90%	10%	Class Actual	Negative	375	15
				Positive	6	385

Pada Tabel 1 diatas adalah hasil dari proses klasifikasi yang didapat dengan menggunakan metode k-nearest neighbors. Berdasarkan dari hasil tersebut setelah dilakukan perhitungan dengan menggunakan rumus *confusion matrix* maka didapat nilai *Precision*, *Recall*, *F1 Score* dan *Accuracy* seperti yang terlihat pada tabe 2 berikut.

Tabel 2. Hasil uji partisi data

Pengujian	Data Training	Data Testing	Accuracy	Precision	Recall	F1-Score
1	50%	50%	96,84%	95,43%	98,41%	96,89%
2	70%	30%	97,09%	95,88%	98,36%	97,10%
3	90%	10%	97,31%	96,25%	98,46%	97,34%

Berdasarkan hasil pada Tabel 2 diatas pengujian yang paling optimal adalah pengujian ke 3 dengan perbandingan 90% data training dan 10% data testing dengan nilai akurasi sebesar 97,31%, presisi sebesar 96,25%, recall sebesar 98,46% dan F1-Score 97,34%.

3.2. Pengujian Evaluasi Performa k-fold Cross Validation

Pengujain ini dilakukan untuk evaluasi kinerja dari metode *k-nearest neighbors* dengan menggunakan *K-fold Cross Validation* dengan cara membagi sampel data secara acak dan mengelompokkan data nilai K-fold sebanyak 3 kali pengujian dengan nilai *k-fold cross validation* yaitu 3,5,10 dengan menggunakan 90% data training dan 10% data testing pada data tersebut menggunakan nilai tetangga terdekat $k = 3$ berdasarkan hasil parameter terbaik dengan pencarian parameter menggunakan *grid search*. Hasil nilai rata-rata akurasi dapat dilihat pada Tabel 3.

Tabel 3. Hasil Rata-Rata Akurasi

Pengujian k-fold	Hasil rata-tara akurasi
3 k-fold Cross Validation	93,85%
5 k-fold Cross Validation	95,26%
10 k-fold Cross Validation	94,62%

Berdasarkan pada Tabel 3 diatas setelah melakukan 3 kali pengujian *k-fold Cross Validation* pada tabel di atas pada data *training* 90% dan data *testing* 10% didapatkan pada pengujian k-fold ke 5 memiliki hasil nilai optimal yaitu dengan rata-rata akurasi 95,26%. Dari hasil nilai rata-rata akurasi yang didapat sudah cukup baik sehingga dapat digunakan dalam melakukan prediksi stunting.

4. Kesimpulan

Berdasarkan hasil dari pengujian pada penelitian Tugas Akhir ini, penulis dapat menarik kesimpulan sebagai berikut:

1. Sistem program yang dibangun dapat melakukan prediksi dengan baik berdasarkan hasil pengujian validasi dengan perbandingan perhitungan manual dan keluaran aplikasi web mendapatkan hasil output yang sama dan diharapkan dapat sebagai langkah awal dalam pencegahan kasus stunting.
2. Performa dari algoritma klasifikasi k-nearest neighbors berdasarkan data stunting mendapatkan performa terbaik dengan menggunakan perbandingan partisi data training 90% dan data testing 10% menghasilkan nilai akurasi terbaik sebesar 97,31%, presisi sebesar 96,25%, recall sebesar 98,46% dan F1-Score 97,34%.

Referensi:

- [1] W. Mutika and D. Syamsul, "ANALISIS PERMASALAHAN STATUS GIZI KURANG PADA BALITA DI PUSKESMAS TEUPAH SELATAN KABUPATEN SIMEULEU Analysis Of Malnutritional Status Problems On Toddlers At South Teupah Health Center Simeulue," *J. Kesehat. Glob.*, vol. 1, no. 3, pp. 127–136, 2018.
- [2] M. Ohyver, J. V. Moniaga, K. R. Yunidwi, and M. I. Setiawan, "Logistic Regression and Growth Charts to Determine Children Nutritional and Stunting Status: A Review," *Procedia Comput. Sci.*, vol. 116, pp. 232–241, 2017, doi: 10.1016/j.procs.2017.10.045.
- [3] Kemenkes, "Keputusan Menteri Kesehatan RI tentang Standar Antropometri Penilaian Status Gizi Anak," *Standar Antropometri Penilaian Status Gizi Anak*. p. 40, 2011.
- [4] N. Sagala and H. Tampubolon, "Komparasi Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 2, p. 98, 2018, doi: 10.23917/khif.v4i2.7061.
- [5] R. T. Prasetyo, A. A. Rismayadi, and I. F. Anshori, "Implementasi Algoritma Genetika pada k-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang," *J. Inform.*, vol. 5, no. 2, pp. 186–194, 2018, doi: 10.31311/ji.v5i2.4123.
- [6] Sunaryono, "Penelitian Komparasi Algoritma Klasifikasi dalam Menentukan Website Palsu," *Teknikom*, vol. 1, no. 1, pp. 1–12, 2017.
- [7] T. & I. Hutapea, "Penerapan Algoritma Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3957–3961, 2018.

Lampiran