



Tersedia Online : <http://e-journals.unmul.ac.id/>

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : <http://e-journals2.unmul.ac.id/index.php/atasi/index>



Perbandingan Metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda

Moch. Rizky Yuliansyah ^{1)*}, Muslimin B ²⁾, Annafi Franz ³⁾

Prodi Teknologi Rekayasa Perangkat Lunak, Teknik dan Informatika, Politeknik Pertanian Negeri Samarinda
E-Mail : mochrizkyuliansyah44@gmail.com ¹⁾; muslimin@politansamarinda.ac.id ²⁾;
annafifranz@politansamarinda.ac.id ³⁾;

ARTICLE INFO

Article history:

Received : 23 April 2022
Revised : 14 Juni 2022
Accepted : 27 Juni 2022
Available online : 30 Juni 2022

Keywords:

Classification,
Nutritional Status,
K-Nearest Neighbors (KNN), Naïve
Bayes Classifier (NBC)

Kata Kunci :

Klasifikasi,
Status Gizi,
K-Nearest Neighbors (KNN), Naïve
Bayes Classifier (NBC)

APA style in citing this article:

Moch. Rizky Yuliansyah, B. M., &
Franz, A. (2022). Perbandingan Metode
K-Nearest Neighbors dan Naïve Bayes
Classifier Pada Klasifikasi Status Gizi
Balita di Puskesmas Muara Jawa Kota
Samarinda. *Adopsi Teknologi Dan Sistem
Informasi (ATASI)*, 1(1), 08 - 20.
<https://doi.org/10.30872/atasi.v1i1.25>

ABSTRACT

Nutritional status is a measure of success in fulfilling nutrition for children as indicated by the child's weight and height and defined as health status resulting from a balance between nutritional needs and inputs. The K-Nearest Neighbors (KNN) method is a method for classifying based on the proximity of a data location (distance) to other data. The Naïve Bayes Classifier (NBC) method is a classification method that uses probability theory to predict future probabilities based on past experiences. The purpose of this study was to compare the results of the Classification of Nutritional Status of Children Under Five using the methods K-Nearest Neighbors and Naïve Bayes Classifier. Based on research results from the performance comparison between the method K-Nearest Neighbors and the Naïve Bayes Classifier using the *f1* score as the main benchmark for classification performance. It was found that the method was K-Nearest Neighbors superior to the *f1* score with a large enough difference, namely 13.42%. Concluded that in the problem of classifying the nutritional status of children under five, the method K-Nearest Neighbors outperforms the Naïve Bayes Classifier.

ABSTRAK

Status Gizi merupakan ukuran keberhasilan dalam pemenuhan nutrisi untuk anak yang diindikasikan oleh berat badan dan tinggi badan anak dan status kesehatan yang dihasilkan oleh keseimbangan antara kebutuhan dan masukan nutrisi. Metode *K-Nearest Neighbors* (KNN) merupakan sebuah metode untuk melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. Metode *Naïve Bayes Classifier* (NBC) merupakan sebuah metode klasifikasi yang memanfaatkan teori probabilitas untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Adapun tujuan dari penelitian ini adalah untuk melakukan perbandingan hasil Klasifikasi Status Gizi Balita dengan menggunakan metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*. Berdasarkan hasil penelitian perbandingan performa antara metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* menggunakan *f1 score* sebagai patokan utama performa klasifikasi. Didapatkan hasil bahwa metode *K-Nearest Neighbors* unggul pada *f1 score* dengan selisih cukup besar yakni 13,42 %. Kesimpulannya bahwa pada masalah klasifikasi status gizi balita metode *K-Nearest Neighbors* mengungguli *Naïve Bayes Classifier*.

2022 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

1. PENDAHULUAN

Puskesmas Muara Jawa adalah puskesmas yang berada di kecamatan muara jawa kabupaten kutai kartanegara, yang beralamat di Jl. Ir. Soekarno, Kelurahan Muara Jawa Pesisir, Kecamatan Muara Jawa, Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Wilayah kerja Puskesmas Muara Jawa terdiri dari Kelurahan Teluk Dalam, Kelurahan Muara Jawa Ilir, Kelurahan Muara Jawa Tengah, Kelurahan Muara Jawa Ulu, Kelurahan Muara Jawa Pesisir, Kelurahan Dondang, Kelurahan Tamapole dan Kelurahan Muara Kembang.

*) Corresponding Author

<https://doi.org/10.30872/atasi.v1i1.25>

Klasifikasi adalah salah satu pembelajaran yang paling umum di dalam data mining. Klasifikasi dapat didefinisikan sebagai bentuk dari analisis data yang digunakan untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas yang terdapat dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas yang terdapat dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap Pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan dalam penggunaan model klasifikasi yang digunakan untuk memprediksi label kelas pada data (Sartika & Sensuse, 2017).

Status Gizi adalah ukuran keberhasilan dalam pemenuhan nutrisi untuk anak yang diindikasikan oleh berat badan dan tinggi badan anak. Status gizi juga dapat didefinisikan sebagai status kesehatan yang dihasilkan oleh keseimbangan antara kebutuhan dan masukan nutrisi. Seorang anak bisa dikatakan balita pada saat dia berusia antara 0 sampai 5 tahun. Karena istilah BALITA merupakan suatu kata kepanjangan dari “Bayi Dibawah Lima Tahun”. Pada usia balita merupakan usia pertumbuhan, yaitu dimana seorang balita pasti aktif dan energik dalam melakukan suatu tindakan. Balita aktif dan energik dalam melakukan tindakan dikarenakan dalam pikirannya muncul rasa penasaran terhadap sesuatu yang dia temui. Status gizi balita dapat diklasifikasikan sebagai berikut: Gizi Buruk, Gizi Kurang, Gizi Baik dan Gizi Lebih (Hariri & Pamungkas, 2016).

Perbandingan Dalam Kamus Besar Bahasa Indonesia (KBBI) Perbandingan adalah perbedaan (selisih) kesamaan dan persamaan, selanjutnya perbandingan adalah suatu upaya untuk mengamati perbedaan atau persamaan yang dimiliki oleh dua buah objek atau lebih yang memiliki suatu kesamaan tertentu (Setiawan, 2012). Maka dibutuhkan sebuah pendekatan metode dalam proses klasifikasi status gizi balita. Penulis menggunakan perbandingan 2 metode, yaitu metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*. Metode *K-Nearest Neighbors* (KNN) merupakan sebuah metode untuk melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. *K-Nearest Neighbors* termasuk algoritma *supervised learning*, Prinsip kerja *K-Nearest Neighbors* (KNN) adalah mencari jarak terdekat antara data yang dievaluasi dengan (K) tetangga terdekatnya dalam data pelatihan (Nugraha dkk., 2017).

Metode *Naïve Bayes Classifier* (NBC) merupakan sebuah metode klasifikasi yang memanfaatkan teori probabilitas untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Klasifikasi-klasifikasi Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. (Hariri & Pamungkas, 2016). Berdasarkan uraian latar belakang diatas, maka penulis akan melakukan Klasifikasi Status Gizi Balita Di Puskesmas Muara Jawa. Pada penelitian ini penulis akan melakukan perbandingan 2 metode pengklasifikasian yaitu metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*. Oleh karena itu penulis akan melakukan sebuah penelitian dengan judul “Perbandingan Metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* Pada Klasifikasi Status Gizi Balita Di Puskesmas Muara Jawa”.

2. TINJAUAN PUSTAKA

A. Studi Literatur

Penelitian yang dilakukan oleh Saleh dkk., (2019) yang berjudul Klasifikasi Status Gizi Balita Menggunakan Metode *K-Nearest Neighbor*. Sistem yang dirancang dalam penelitian ini adalah sistem penentuan status gizi balita dengan menggunakan metode K-NN (*K-Nearest Neighbor*), dimana metode K-NN (*K-Nearest Neighbor*) merupakan metode pengklasifikasian atau mengelompokkan data uji yang belum diketahui kelasnya ke beberapa tetangga terdekat dengan melakukan perhitungan jarak dengan menggunakan rumus. Variabel yang digunakan dalam sistem ini berdasarkan data pada Anthropometri atau pengukuran tubuh manusia yaitu U (Umur), BB (Berat Badan), TB (Tinggi Badan), LK (Lingkar Kepala). Hasil yang didapatkan dari sistem ini adalah status gizi berdasarkan BB/U (Berat Badan menurut Umur) yaitu gizi buruk, gizi kurang, gizi baik, dan gizi lebih.

Penelitian yang dilakukan oleh Zubair & Muksin (2018) yang berjudul Penerapan Metode *Naïve Bayes* Untuk Klasifikasi Status Gizi (Studi Kasus Di Klinik Bromo Malang). Pada penelitian ini Status gizi adalah hal yang penting untuk kesehatan tiap-tiap orang. Dengan cara mengetahui status gizi seseorang dapat mengatur pola makan sehingga dapat mencapai status gizi yang ideal. Oleh karena itu, penelitian ini menggunakan data mining untuk membuat klasifikasi berdasarkan lima jenis status gizi. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naïve Bayes*. Hasil uji akurasi dari penelitian ini berhasil sangat baik. Hal ini ditunjukkan oleh akurasi klasifikasi yang mencapai 98 persen. Selain itu akurasi yang mencapai 98 persen adalah klasifikasi dengan menghilangkan atribut umur, berat badan, lingkar perut, dan visceral fat.

Penelitian yang dilakukan oleh Pramitarini dkk., (2013) yang berjudul Analisa Rekam Medis Untuk Menentukan Status Gizi Anak Balita Menggunakan *Naïve Bayes Classifier*. Rekam medis merupakan informasi identitas dan riwayat berobat oleh pasien pada suatu balai pengobatan (klinik, rumah sakit, ataupun puskesmas). Rekam medis mempunyai informasi yang lengkap dari pasien dan intepretasinya sudah sesuai dengan standar WHO dimana ada status gizi baik, status gizi sedang, status gizi kurang, status gizi lebih dan status gizi buruk. Adapun atribut yang digunakan adalah jenis kelamin, umur, tinggi badan, berat badan, dan tiga buah komorbir yaitu diare, edema, dan pneumonia. Metode klasifikasi yang di terapkan dalam pencarian pola adalah algoritma *Naïve Bayes Classifier* dengan jumlah data bersih sebanyak 198 data. Hasil dari penelitian ini adalah mengklasifikasikan data yang ada sesuai dengan inteprestasinya. Pengujian dilakukan dengan membagi data sampel menjadi 70%:30%, 50%:50% dan 30%:70%. Untuk uji validitasnya diperoleh bila prontase 70:30 maka

besar nilai akurasi 88%, jika prosentasenya 50:50 maka besar nilai akurasi 55% dan jika presentasi validasinya 30:70 maka besar nilai akurasi 28%.

Penelitian yang dilakukan oleh Haryati dkk., (2016) yang berjudul Klasifikasi Jenis Batubara Menggunakan Jaringan Saraf Tiruan Dengan Algoritma *Backpropagation*. Pada penelitian ini batubara adalah batuan sedimen yang dapat terbakar yang terbentuk dari endapan organik, utamanya adalah sisa-sisa tumbuhan dan terbentuk melalui proses pembatubaraan. Dalam penelitian ini membuat sistem klasifikasi jenis batubara menggunakan Jaringan Saraf Tiruan (JST) dengan Algoritma *Backpropagation*. Klasifikasi jenis batubara ini menggunakan tiga lapisan. Lapisan pertama yaitu input layer sebanyak tujuh neuron, lapisan kedua hidden layer sebanyak tiga neuron, dan lapisan ketiga output layer sebanyak lima neuron. Sistem klasifikasi ini menghasilkan lima kelas keluaran diantara yaitu Antrashit, Sub-Bituminous, Bituminous, Lignit dan Gambut. Berdasarkan hasil pengujian pada data uji sebanyak 200 data diperoleh akurasi sebesar 98% dengan learning rate 0,2, dan toleransi error 0,001. Sedangkan pengujian terhadap 500 data latih menghasilkan akurasi sebesar 99%.

Penelitian yang dilakukan oleh Putri dkk., (2014) yang berjudul Perbandingan Metode Klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor* Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012. Pada Penelitian ini akan melakukan klasifikasi status pekerjaan pada jumlah penduduk di angkatan kerja di Kabupaten Demak untuk tahun 2012 dengan menggunakan dua metode yaitu *Naïve Bayes* dan *K-Nearest Neighbor*. *Naïve Bayes* adalah metode klasifikasi berdasarkan pada perhitungan probabilitas sederhana, sedangkan *K-Nearest Neighbor* adalah metode klasifikasi berdasarkan pada perhitungan kedekatan. Variabel yang digunakan dalam menentukan apakah status pekerjaan seseorang menganggur atau tidak adalah jenis kelamin, status dalam rumah tangga, status perkawinan, pendidikan, dan usia. Hasil dari status pekerjaan dari metode pemrosesan data *Naïve Bayes* dengan akurasi yang diperoleh adalah sebesar 94,09% dan metode *K-Nearest Neighbor* yang diperoleh adalah dengan akurasi 96,06%. Untuk mengevaluasi hasil klasifikasi digunakan perhitungan Tekan Q dan APER. Berdasarkan analisis, nilai Q Press yang diperoleh menunjukkan bahwa kedua metode sudah baik dalam melakukan klasifikasi data. Berdasarkan perhitungan APER, klasifikasi data menggunakan metode *K-Nearest Neighbor* memiliki tingkat kesalahan lebih kecil dari metode *Naïve Bayes*. Dari analisis ini dapat disimpulkan bahwa metode *K-Nearest Neighbor* bekerja lebih baik dibandingkan dengan metode *Naïve Bayes* untuk data status pekerjaan di Kabupaten Demak untuk tahun 2012.

B. Landasan Teori

1. Data Mining

Menurut Kusriani & Luthfi (2009). *Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi suatu informasi yang bermanfaat dan pengetahuan yang terkait dalam database besar. Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan cara memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Pernyataan tersebut menegaskan bahwa dalam *data mining* otomatisasi tidak menggantikan campur tangan manusia. Manusia harus ikut aktif dalam setiap fase dalam proses *data mining*. Kehebatan kemampuan algoritma *data mining* yang terdapat di dalam perangkat lunak analisis yang terdapat saat ini memungkinkan terjadinya kesalahan dalam penggunaan yang berakibat fatal. Pengguna mungkin menerapkan analisis yang tidak tepat terhadap kumpulan data dengan menggunakan pendekatan yang berbeda. Oleh karena itu, dibutuhkan pemahaman tentang statistik dan struktur model matematika yang mendasari kerja perangkat lunak. *Data mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dulu. *data mining* memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik, database, dan juga *information retrieval*.

2. Klasifikasi

Klasifikasi adalah salah satu pembelajaran yang paling umum di dalam data mining. Klasifikasi dapat didefinisikan sebagai bentuk dari analisis data yang digunakan untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas yang terdapat dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas yang terdapat dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap Pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan dalam penggunaan model klasifikasi yang digunakan untuk memprediksi label kelas pada data (Sartika & Sensuse, 2017).

3. Perbandingan

Dalam Kamus Besar Bahasa Indonesia (KBBI) Perbandingan adalah perbedaan (selisih) kesamaan dan persamaan, selanjutnya perbandingan adalah suatu upaya untuk mengamati perbedaan atau persamaan yang dimiliki oleh dua buah objek atau lebih yang memiliki suatu kesamaan tertentu (Setiawan, 2012).

4. Status Gizi Balita

Status Gizi adalah ukuran keberhasilan dalam pemenuhan nutrisi untuk anak yang diindikasikan oleh berat badan dan tinggi badan anak. Status gizi juga dapat didefinisikan sebagai status kesehatan yang dihasilkan oleh keseimbangan antara kebutuhan dan masukan nutrisi. Seorang anak bisa dikatakan balita pada saat dia berusia antara 0 sampai 5 tahun. Karena istilah BALITA merupakan suatu kata kepanjangan dari "Bayi Dibawah Lima

Tahun". Pada usia balita merupakan usia pertumbuhan, yaitu dimana seorang balita pasti aktif dan energik dalam melakukan suatu tindakan. Balita aktif dan energik dalam melakukan tindakan dikarenakan dalam pikirannya muncul rasa penasaran terhadap sesuatu yang dia temui. Status gizi balita dapat diklasifikasikan sebagai berikut: Gizi Buruk, Gizi Kurang, Gizi Baik dan Gizi Lebih (Hariri & Pamungkas, 2016).

5. Metode *K-Nearest Neighbors*

Metode *K-Nearest Neighbors* (KNN) merupakan sebuah metode untuk melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. *K-Nearest Neighbors* termasuk algoritma *supervised learning*, Prinsip kerja *K-Nearest Neighbors* (KNN) sendiri adalah mencari jarak terdekat antara data yang dievaluasi dengan (K) tetangga terdekatnya dalam data pelatihan. Sebelum mencari jarak terdekat antara data yang dievaluasi, pada algoritma *K-Nearest Neighbors* harus dilakukan preprocessing atau normalisasi terlebih dahulu. Preprocessing sendiri bertujuan untuk mendapatkan standar nilai pada semua atribut atau indikator dalam perhitungan (Nugraha dkk., 2017).

Adapun rumus dari metode *K-Nearest Neighbors* adalah sebagai berikut:

$$d(X_i, X_j) = \sqrt{\sum_{l=1}^N (\text{diff}(X_{il}, X_{jl}))^2} \dots\dots\dots(1)$$

Keterangan:

- X_{il} : data testing ke- i pada variabel ke- l
- X_{jl} : data training ke- i pada variabel ke- l
- d(X_i, X_j) : *jarak euclidean*
- N : dimensi data variabel bebas
- diff(X_{il}, X_{jl}) : *difference* atau ketidaksamaan

6. Metode *Naïve Bayes Classifier*

Metode *Naïve Bayes Classifier* (NBC) merupakan sebuah metode klasifikasi yang memanfaatkan teori probabilitas untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Klasifikasi-klasifikasi Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. (Hariri & Pamungkas, 2016). *Naïve Bayes Classifier* memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada database yang besar. *Naïve Bayes Classifier* termasuk ke dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar yang akan dijadikan label kelas data masukan tersebut (Sartika & Sensuse, 2017).

Adapun rumus dari metode *Naïve Bayes Classifier* untuk variabel diskrit adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots(2)$$

Keterangan:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data *class* spesifik
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
- P(H) : Probabilitas hipotesis H (*prior probability*)
- P(X|H) : Probabilitas X berdasarkan kondisi H (*likelihood probability*)
- P(X) : Probabilitas X (*evidence probability*)

Adapun rumus dari metode *Naïve Bayes Classifier* untuk variabel kontinyu adalah sebagai berikut:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \dots\dots\dots(3)$$

Keterangan:

- f(x) : fungsi x
- π : pi
- σ : standar deviasi
- μ : rata - rata
- e : eksponen

7. Diagram Konteks

Diagram konteks adalah diagram yang terdiri dari suatu proses input dan output dalam menggambarkan ruang lingkup suatu sistem. Diagram konteks akan memberikan gambaran tentang keseluruhan sistem, diagram konteks merupakan diagram yang menggambarkan interaksi sistem dengan penggunanya. Pengguna memasukkan data kedalam sistem dan dapat menerima informasi dari sistem. Data yang dimasukkan oleh pengguna selanjutnya

akan diolah oleh sistem kemudian akan di tampilkan menjadi informasi yang berguna bagi pemakainya (Sihombing dkk., 2012).

8. Python

Python merupakan bahasa pemrograman serbaguna yang dikembangkan oleh Guido van Rossum pada tahun 1990. Python merupakan kelanjutan dari bahasa pemrograman ABC. Nama python dipilih oleh Guido sebagai nama bahasa ciptaannya karena kecintaan Guido pada acara televisi *Monty Python's Flying Circus*. Python secara umum berbentuk pemrograman berorientasi objek, pemrograman imperative, dan pemrograman fungsional. Saat ini, pengembangan Python terus dilakukan oleh sekelompok programmer yang dikoordinasi oleh Guido dan Python Software Foundation. Distribusi Python sudah mencapai versi 2.6.1 dan versi 3.0 sampai saat ini. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan analisis data termasuk *Data Mining* dan juga dapat berjalan di berbagai platform sistem operasi (Enterprise, 2017).

9. Scikit-Learn

Scikit-Learn adalah proyek open source, artinya bebas untuk menggunakan dan mendistribusikan, dan siapa pun dapat dengan mudah mendapatkan kode sumber. Scikit-learn berisi sejumlah algoritma machine learning yang canggih, serta dokumentasi komprehensif tentang masing-masing algoritma. scikit-learn adalah alat yang sangat populer, dan merupakan library python yang paling menonjol untuk machine learning. scikit-learn banyak digunakan di industri dan akademisi dan banyak tutorial serta cuplikan kode tersedia secara online (Müller & Guido, 2017).

10. Jupyter Notebook

Jupyter Notebook adalah lingkungan interaktif untuk menjalankan kode di browser dan merupakan tools yang digunakan untuk mendokumentasikan sebuah pekerjaan. Jupyter notebook adalah software yang baik untuk melakukan analisis data eksplorasi dan banyak digunakan oleh para ilmuwan data. jupyter notebook memudahkan untuk memasukkan kode, teks, dan gambar. Sementara itu, jupyter notebook mendukung banyak bahasa pemrograman tidak hanya bahasa pemrograman python (Müller & Guido, 2017).

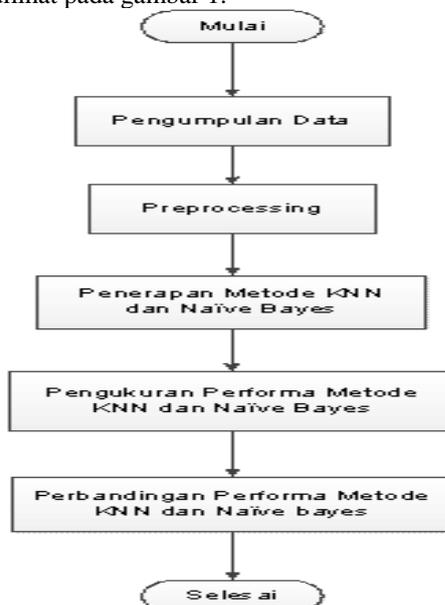
11. Flask

Flask adalah sebuah web framework yang di tulis menggunakan bahasa Python dan tergolong sebagai jenis microframework. Flask merupakan kerangka kerja kecil menurut sebagian besar standar. Flask berfungsi sebagai kerangka kerja dari aplikasi dan tampilan dari suatu web. Dengan menggunakan Flask dan bahasa Python, pengembang dapat membuat sebuah web yang terstruktur dan dapat mengatur behavior suatu web dengan lebih mudah. Flask memiliki dua dependensi utama yaitu, routing, debugging, dan Web Server Gateway Interface (WSGI). Selain itu, meskipun Flask disebut sebagai microframework, bukan berarti Flask mempunyai kekurangan dalam hal fungsionalitas. Microframework disini bahwa Flask bermaksud untuk membuat core dari aplikasi ini sesederhana mungkin, tetapi tetap dapat dengan mudah ditambahkan dan digunakan. Dengan begitu, fleksibilitas serta skalabilitas dari Flask dapat dikatakan cukup tinggi dibandingkan dengan framework lainnya (Grinberg, 2014).

3. METODE PENELITIAN

A. Prosedur Penelitian

Adapun prosedur penelitian dapat dilihat pada gambar 1.



Gambar 1. Flowchart Prosedur Penelitian

Dari flowchart prosedur penelitian pada gambar 1, dapat dijelaskan tahapan prosedur penelitian sebagai berikut:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data status gizi balita. Data status gizi balita di peroleh dari Puskesmas Muara Jawa, Kecamatan Muara Jawa, Kabupaten Kutai Kartanegara, data yang di ambil adalah data status gizi balita tahun 2019.

2. Preprocessing

Preprocessing atau Pembersihan Data merupakan tahapan awal dalam mengolah data input sebelum memasuki proses yaitu proses dan langkah dalam melakukan pengolahan data untuk membuat data mentah menjadi data yang berkualitas dan mendapatkan input yang baik. Ada beberapa cara melakukan preprocessing seperti : menghapus data yang memiliki nilai kosong atau mengisi nilai yang kosong tersebut dengan rata-rata.

3. Penerapan Metode K-Nearest Neighbors dan Naïve Bayes Classifier

Adapun tahapan metode K-Nearest Neighbors adalah sebagai berikut:

- a. Tentukan jumlah tetangga terdekat K
- b. Hitung jarak test data ke training data
- c. Urut data berdasarkan data yang mempunyai jarak Euclid terkecil
- d. Tentukan kelompok test data berdasarkan label mayoritas pada K

Adapun tahapan metode Naïve Bayes Classifier adalah sebagai berikut:

Jika Data Numerik:

- a. Menu tiap parameter
- b. Standar deviasi tiap parameter
- c. Tabel mean dan standar deviasi

Jika Data Kategorik:

- a. Jumlah dan probabilitas setiap kejadian
- b. Tabel probabilitas

4. Pengukuran Performa Metode K-Nearest Neighbors dan Naïve Bayes Classifier

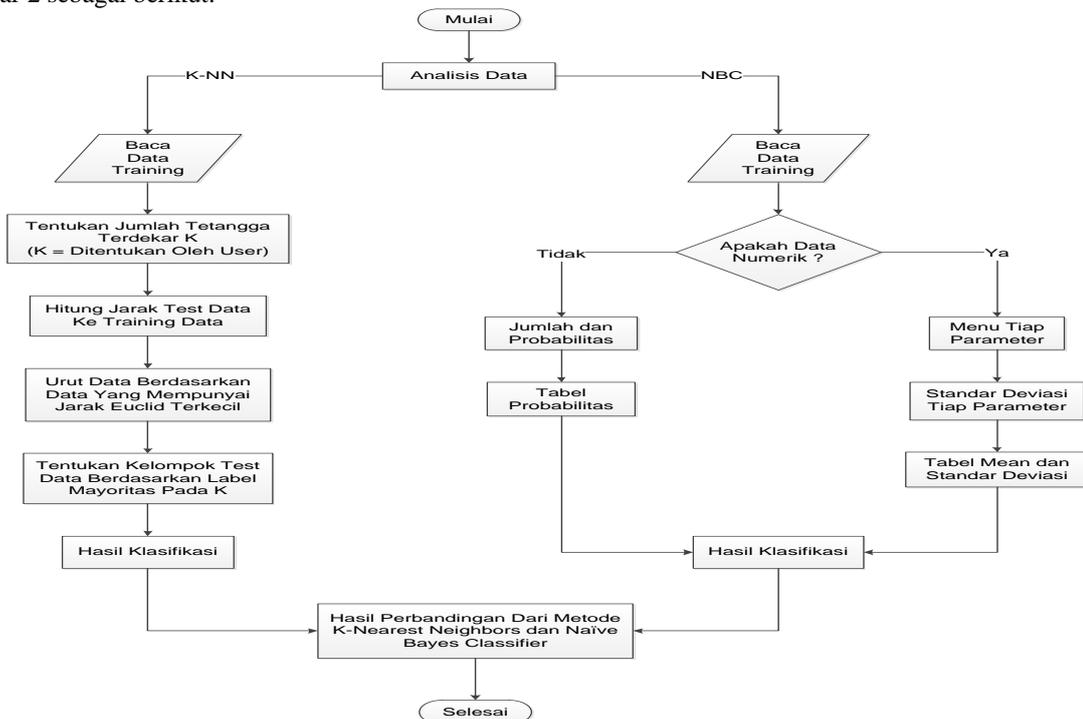
Pengukuran performa metode dapat dilakukan menggunakan rumus Accuracy, Recall dan F Score. Pengukuran performa metode dilakukan untuk menghitung tingkat performa pada metode K-Nearest Neighbors dan Naïve Bayes Classifier sehingga di dapatkan kesamaan atau kedekatan suatu hasil perhitungan dengan angka atau data yang sebenarnya.

5. Perbandingan Performa Metode K-Nearest Neighbors dan Naïve Bayes Classifier

Perbandingan performa metode dilakukan setelah tahap pengukuran performa, perbandingan performa metode dilakukan untuk mengetahui hasil perbandingan dari metode K-Nearest Neighbors dan Naïve Bayes Classifier, sehingga dapat diketahui metode manakah yang memiliki performa yang lebih baik dalam melakukan klasifikasi status gizi balita.

B. Analisis Perancangan Model

Adapun tahapan model dari metode K-Nearest Neighbors dan Naïve Bayes Classifier dapat di lihat pada gambar 2 sebagai berikut:



Gambar 2. Analisis Perancangan Model

Dari analisis perancangan model pada gambar 2, dapat dijelaskan tahapan analisis perancangan model *K-Nearest Neighbors* dan *Naive Bayes Classifier* sebagai berikut:

1. Analisis Perancangan Model *K-Nearest Neighbors*

Adapun analisis perancangan model *K-Nearest Neighbors* adalah sebagai berikut:

- a. Tentukan Jumlah Tetangga Terdekat K
Menentukan jumlah tetangga terdekat K merupakan tahapan awal dalam proses metode *K-Nearest Neighbors*.
- b. Hitung Jarak Tes Data Ke Training Data
Hitung jarak tes data ke training data dapat dilakukan menggunakan rumus 1, dengan cara menghitung jumlah keseluruhan jarak, sehingga dapat diketahui jarak manakah yang mempunyai euclid terkecil.
- c. Urut Data Berdasarkan Data Yang Mempunyai Jarak Euclid Terkecil
Setelah melakukan perhitungan jarak euclid terhadap tes data ke training data langkah selanjutnya adalah mengurutkan data dari hasil perhitungan jarak euclid dari yang terkecil hingga terbesar, sehingga dapat diketahui jarak euclid manakah yang paling kecil.
- d. Tentukan Kelompok Test Data Berdasarkan Label Mayoritas Pada K
Menentukan kelompok tes data berdasarkan label mayoritas pada K merupakan tahapan akhir dalam proses metode *K-Nearest Neighbors*, tahapan ini dilakukan untuk mengelompokkan semua label berdasarkan voting mayoritas sebagai prediksi data test.

2. Analisis Perancangan Model *Naive Bayes Classifier*

Adapun analisis perancangan model *Naive Bayes Classifier* adalah sebagai berikut:

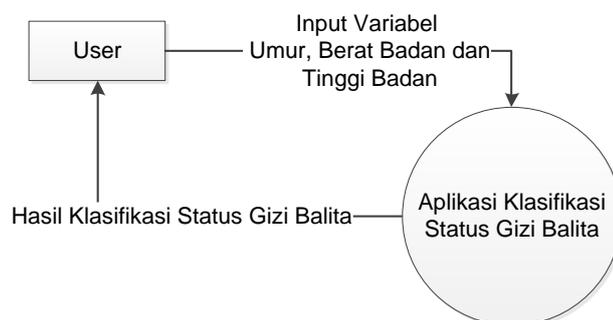
Jika Data Numerik:

- a. Menu Tiap Parameter
Menu tiap parameter menampilkan parameter atau variabel apa saja yang akan diproses menggunakan *Naive Bayes Classifier*.
- b. Standar Deviasi Tiap Parameter
Standar deviasi tiap parameter dapat dilakukan menggunakan rumus 3, merupakan tahapan untuk mencari nilai standar deviasi yang terdapat pada setiap variabel.
- c. Tabel Mean dan Standar Deviasi
Tabel mean dan standar deviasi dapat dilakukan menggunakan rumus 3, merupakan sebuah tahapan untuk mencari nilai rata-rata dan standar deviasi pada setiap variabel berdasarkan kelas.

Jika Data Kategorik:

- a. Jumlah dan Probabilitas Setiap Kejadian
Jumlah dan probabilitas setiap kejadian dapat dilakukan menggunakan rumus 2, merupakan tahapan yang dilakukan untuk menghitung jumlah berapa banyak probabilitas setiap kejadian.
- b. Tabel Probabilitas
Tabel probabilitas merupakan sebuah tabel hasil dari perhitungan probabilitas setiap kelas.

C. Analisis Perancangan Sistem



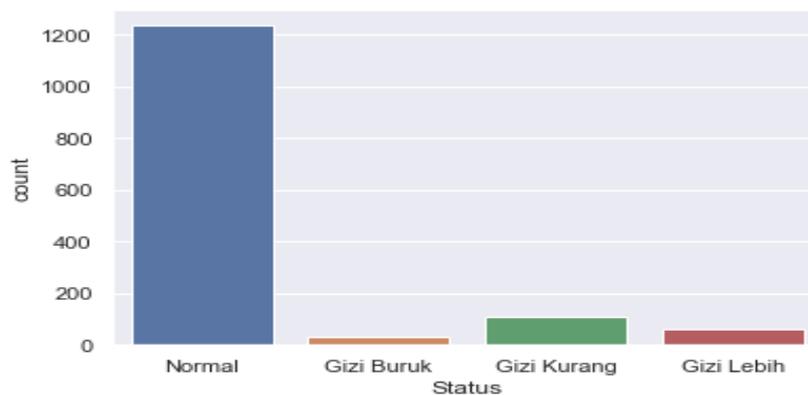
Gambar 3. Diagram Konteks

Gambar 3 dapat dijelaskan alur data dari proses aplikasi klasifikasi status gizi balita menggunakan metode *K-Nearest Neighbors*. User memasukkan variabel-variabel kedalam aplikasi klasifikasi status gizi balita seperti variabel umur, berat badan dan tinggi badan. Kemudian aplikasi akan melakukan prediksi dari variabel-variabel yang telah dimasukkan dan akan menampilkan hasil klasifikasi status gizi balita.

4. HASIL DAN PEMBAHASAN

A. Penerapan Metode *K-Nearest Neighbors* dan *Naive Bayes Classifier*

Pada penelitian ini penulis berfokus pada variabel/fitur yang sifatnya numerik seperti Umur, Berat dan Tinggi jadi variabel yang tidak diperlukan pada proses penerapan metode akan dihapus seperti : Nama dan Jenis Kelamin. Selanjutnya melihat distribusi dari status gizi balita. Adapun distribusi status gizi balita dapat dilihat pada gambar 4. Distribusi status gizi balita.



Gambar 4. Distribusi Status Gizi Balita

Dapat dilihat pada gambar 4, distribusi status gizi normal/baik sangat mendominasi dibandingkan dengan status gizi yang lainnya. Sehingga data status gizi balita mengalami masalah *imbalanced class*/kelas tidak seimbang. Maka, pada penelitian ini penulis berfokus untuk menyeimbangkan antara *precision* dan *recall* dengan melihat nilai dari *f1 score* yang merupakan rata-rata harmonik dari *precision* dan *recall* dan bukan pada akurasi karena kelas tidak seimbang. Jumlah data yang digunakan dalam penelitian ini adalah 1443 balita dengan proporsi pembagian data yaitu 80:20, dimana 80% untuk *training* dan 20% untuk *testing*. Yang berarti ada 1151 *data training* dan 288 *data testing*. Adapun data *training* dan data *testing* dapat dilihat pada tabel 1 dan 2.

Tabel 1. Data Training

Index	Umur	Berat	Tinggi	Status
628	18	9,2	74	Normal
409	26	10,7	79	Normal
672	8	4,5	54	Normal
9	20	9	78	Normal
1190	20	7,1	76	Gizi Buruk
.....
480	59	24	111	Gizi Lebih
850	23	8	80	Gizi Buruk
1315	34	12,9	88,5	Normal
605	20	9,1	87	Gizi Buruk
587	27	8,2	76	Normal

Tabel 2. Data Testing

Index	Umur	Berat	Tinggi	Status
462	26	9,5	95	Gizi Buruk
1115	58	11	110	Gizi Buruk
258	39	12	90	Normal
1023	6	5	57	Normal
65	25	9,4	79	Normal
.....
144	35	10,2	78	Normal
1359	58	15	100	Normal
1312	22	7,7	73	Normal
2	53	12	87	Normal
1168	9	5	60	Normal

Pada tabel 1 dan 2, dapat dilihat bahwa setelah data dibagi menjadi *training* dan *testing* maka data akan diacak urutannya. Dimana index merupakan representasi dari nomor balita.

1. Penerapan Metode *K-Nearest Neighbors*

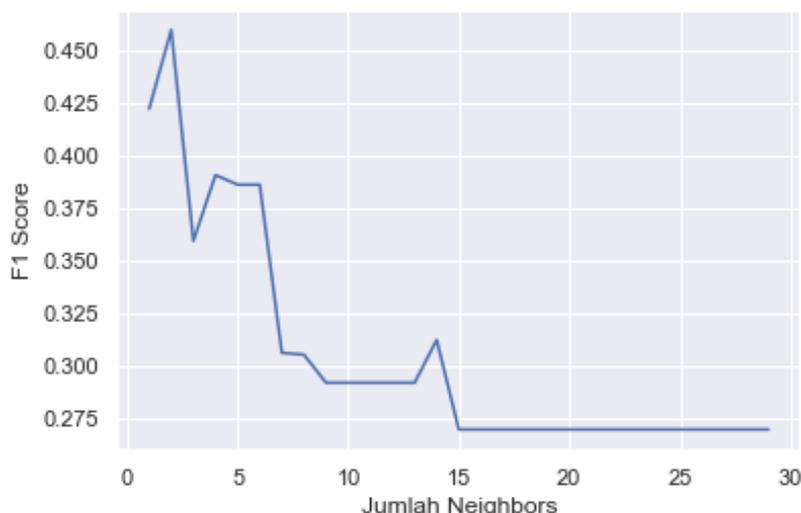
Jika kita lihat pada tabel 1 dan 2, skala pengukuran pada setiap variabel berbeda-beda seperti pada variabel Berat pada index 1115 memiliki nilai sebesar 11 sementara variabel Tinggi memiliki nilai sebesar 110. Jelas

bahwa perbedaan skala diantara keduanya besar, ini akan menjadi masalah pada metode *K-Nearest Neighbors*. Karena metode *K-Nearest Neighbors* merupakan metode berbasis jarak sehingga kita tidak menginginkan perbedaan skala antar variabel yang cukup jauh. Cara untuk menyelesaikan permasalahan ini membutuhkan proses *min-max scaling* untuk menyamakan skala antar variabel dengan rentang 0 – 1. Adapun hasil *min-max scaling* pada data training dapat dilihat pada tabel 3.

Tabel 3. Hasil Min-Max Scaling Data Training

Index	Umur	Berat	Tinggi	Status
628	0,280702	0,067497	0,383562	Normal
409	0,421053	0,083074	0,452055	Normal
672	0,105263	0,018692	0,109589	Normal
9	0,315789	0,065421	0,438356	Normal
1190	0,315789	0,045691	0,410959	Gizi Buruk
.....
480	1	0,221184	0,890411	Gizi Lebih
850	0,368421	0,055036	0,465753	Gizi Buruk
1315	0,561404	0,105919	0,582192	Normal
605	0,315789	0,066459	0,561644	Gizi Buruk
587	0,438596	0,057113	0,410959	Normal

Adapun untuk data testing skalanya juga sama yakni dari 0 – 1. Setelah melakukan *scaling* maka, langkah selanjutnya adalah menerapkan metode *K-Nearest Neighbors*. Pada metode *K-Nearest Neighbors* kita dapat mencari jumlah K atau *neighbors* yang dapat memberikan performa yang paling optimal terhadap data kita. Dalam penelitian ini *f1 score* akan dijadikan patokan utama performa klasifikasi. Adapun hasil mencari jumlah k optimal dapat dilihat pada gambar 5.



Gambar 5. Mencari Jumlah K Optimal

Dapat dilihat pada gambar 5, jumlah K atau *neighbors* optimal adalah 2 yang memberikan nilai performa *f1 score* sebesar 0,45. Pada penelitian ini akan digunakan *confusion matrix* untuk melihat hasil klasifikasi. Adapun *confusion matrix* untuk *K-Nearest Neighbors* yang dibuat menjadi tabel agar lebih mudah dipahami dapat dilihat pada tabel 4.

Tabel 4. Confusion Matrix K-Nearest Neighbors

		Aktual			
		Gizi Buruk	Gizi Kurang	Gizi Lebih	Normal
Prediksi	Gizi Buruk	2	1	0	1
	Gizi Kurang	3	5	0	28
	Gizi Lebih	0	0	4	3
	Normal	2	16	8	215

Dapat dilihat pada tabel 4, baris adalah prediksi sementara kolom adalah aktual. Model klasifikasi yang dibangun menggunakan metode *K-Nearest Neighbors* masih banyak melakukan kesalahan prediksi, itu dapat

dilihat dari angka-angka yang terisi pada selain diagonal tabel. Namun, dapat dilihat bahwa model *K-Nearest Neighbors* dapat memprediksi gizi buruk yang ternyata memang benar gizi buruk sebanyak 2 dengan 5 kesalahan. Prediksi gizi kurang yang ternyata memang benar gizi kurang sebanyak 5 dengan 17 kesalahan. Prediksi gizi lebih yang ternyata memang benar gizi lebih sebanyak 4 dengan 8 kesalahan. Prediksi gizi normal yang ternyata memang benar gizi normal sebanyak 215 dengan 32 kesalahan. Model klasifikasi menggunakan *K-Nearest Neighbors* banyak mengklasifikasikan kelas minoritas sebagai kelas mayoritas ini dapat dilihat pada banyaknya kesalahan prediksi gizi normal ternyata bukan gizi normal, inilah ciri khas masalah klasifikasi dengan kelas yang tidak seimbang.

2. Penerapan Metode *Naïve Bayes Classifier*

Naïve Bayes Classifier tidak memerlukan standarisasi seperti metode *K-Nearest Neighbors*, karena metode *Naïve Bayes Classifier* sendiri berbasis probabilitas bersyarat. Namun, metode *Naïve Bayes Classifier* tidak dapat dilakukan pencarian parameter apapun yang dapat memberikan performa yang paling optimal seperti yang telah dilakukan pada metode *K-Nearest Neighbors*. Jadi, kita dapat melakukan proses training tanpa mengatur parameter apapun. Kembali lagi, confusion matrix akan digunakan untuk melihat hasil klasifikasi. Adapun confusion matrix untuk *Naïve Bayes Classifier* dapat dilihat pada tabel 5.

Tabel 5. Confusion Matrix *Naïve Bayes Classifier*

		Aktual			
		Gizi Buruk	Gizi Kurang	Gizi Lebih	Normal
Prediksi	Gizi Buruk	0	0	0	0
	Gizi Kurang	0	0	0	0
	Gizi Lebih	0	0	3	1
	Normal	7	22	9	246

Dapat dilihat pada tabel 5, model klasifikasi yang dibangun menggunakan metode *Naïve Bayes Classifier* memiliki performa yang buruk. Dengan prediksi gizi buruk yang ternyata memang benar gizi buruk sebanyak 0 (tidak ada yang benar) dengan kesalahan 7. Prediksi gizi kurang yang ternyata memang benar gizi kurang juga sebanyak 0 (tidak ada yang benar) dengan kesalahan 22. Prediksi gizi lebih ternyata memang benar gizi lebih sebanyak 3 dengan kesalahan 9. Prediksi gizi normal ternyata memang benar gizi normal sebanyak 246 dengan kesalahan 1. Model klasifikasi menggunakan metode *Naïve Bayes Classifier* banyak mengklasifikasikan kelas minoritas sebagai kelas mayoritas dan kesalahannya lebih besar dari *K-Nearest Neighbors*, ditambah lagi dengan tidak ada prediksi yang benar sama sekali untuk gizi buruk dan kurang.

B. Pengukuran Performa Metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*

Setelah melakukan penerapan metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* untuk klasifikasi status gizi balita maka, selanjutnya akan dilakukan pengukuran performa klasifikasi baik untuk *K-Nearest Neighbors* maupun *Naïve Bayes Classifier*. Pengukuran performa yang akan digunakan pada penelitian ini adalah *precision*, *recall*, *f1 score* dan *accuracy* walaupun *accuracy* hanya sebagai tambahan karena pengukuran performa akan lebih fokus ke *f1 score*.

1. Performa Metode *K-Nearest Neighbors*

Adapun performa model klasifikasi yang dibangun menggunakan metode *K-Nearest Neighbors* dapat dilihat pada tabel 6.

Tabel 6. Performa *K-Nearest Neighbors*

Metode	Accuracy	Precision	Recall	F1 Score
<i>K-Nearest Neighbors</i>	78,47	52,56	42,91	45,95

Dapat dilihat pada tabel 6, model klasifikasi yang dibangun menggunakan metode *K-Nearest Neighbors* mendapatkan *accuracy* sebesar 78,47, *precision* sebesar 52,56, *recall* sebesar 42,91 dan yang paling penting *f1 score* sebesar 45,95.

2. Performa Metode *Naïve Bayes Classifier*

Adapun performa model klasifikasi yang dibangun menggunakan metode *Naïve Bayes Classifier* dapat dilihat pada tabel 7.

Tabel 7. Performa *Naïve Bayes Classifier*

Metode	Accuracy	Precision	Recall	F1 Score
<i>Naïve Bayes Classifier</i>	86,45	40,40	31,14	32,53

Dapat dilihat pada tabel 7, model klasifikasi yang dibangun menggunakan metode *Naïve Bayes Classifier* mendapatkan *accuracy* sebesar 86,45, *precision* sebesar 40,40, *recall* sebesar 31,14 dan *f1 score* sebesar 32,53.

C. Perbandingan Performa Metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*

Setelah pengukuran performa selesai tahapan selanjutnya yaitu melakukan perbandingan performa model klasifikasi yang dibangun menggunakan metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*. Tujuan dari perbandingan performa ini adalah untuk mengetahui metode manakah yang memiliki performa lebih baik dalam masalah klasifikasi status gizi balita. Adapun hasil perbandingan metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* dapat dilihat pada tabel 8.

Tabel 8. Perbandingan Performa *K-Nearest Neighbors* dan *Naive Bayes Classifier*

Metode	Accuracy	Precision	Recall	F1 Score
<i>K-Nearest Neighbors</i>	78,47	52,56	42,91	45,95
<i>Naïve Bayes Classifier</i>	86,45	40,40	31,14	32,53

Dapat dilihat pada tabel 8, metode *Naïve Bayes Classifier* mengungguli *K-Nearest Neighbors* dengan *accuracy* sebesar 86,45 % namun, kembali penulis ingatkan bahwa pada penelitian ini tidak berfokus pada *accuracy* namun pada *f1 score*. Jika dilihat dari *precision*, *recall* dan terutama *f1 score* *K-Nearest Neighbors* mengungguli *Naïve Bayes Classifier* dengan selisih *f1 score* sebesar 13,42 %. Jadi, dapat dikatakan bahwa pada masalah klasifikasi status gizi balita metode *K-Nearest Neighbors* mengungguli *Naïve Bayes Classifier*.

D. Pembuatan Aplikasi

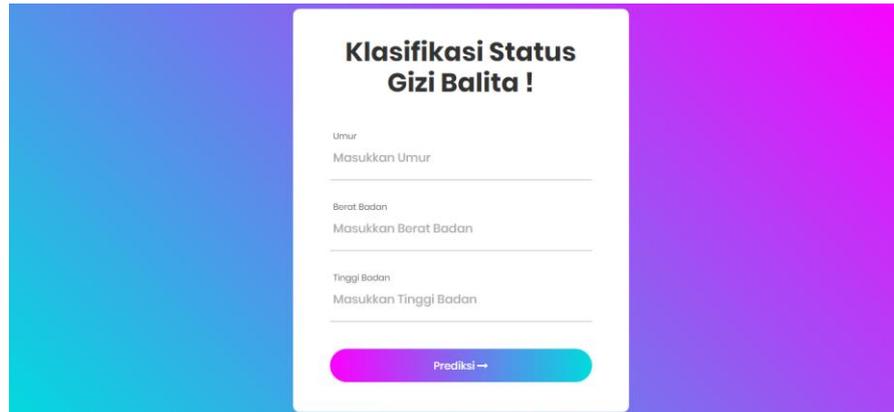
Setelah proses pemodelan selesai, proses selanjutnya adalah mendeploy sebuah aplikasi, sebelum aplikasi di deploy model yang sudah dibuat dimodel selection terlebih dahulu untuk memilih model manakah yang memiliki performa terbaik. Dapat dilihat pada performa model *K-Nearest Neighbors* mengungguli *Naïve Bayes Classifier*. Jadi, untuk model yang akan di deploy menjadi sebuah aplikasi adalah model *K-Nearest Neighbors*. Berikut ini adalah tampilan aplikasi yang telah di buat atau di deploy menggunakan metode *K-Nearest Neighbors*.

1. Tampilan Aplikasi Klasifikasi Status Gizi Balita

Berikut adalah tampilan dari aplikasi klasifikasi status gizi balita yang terdiri dari tampilan prediksi dan tampilan hasil prediksi.

a. Tampilan Prediksi

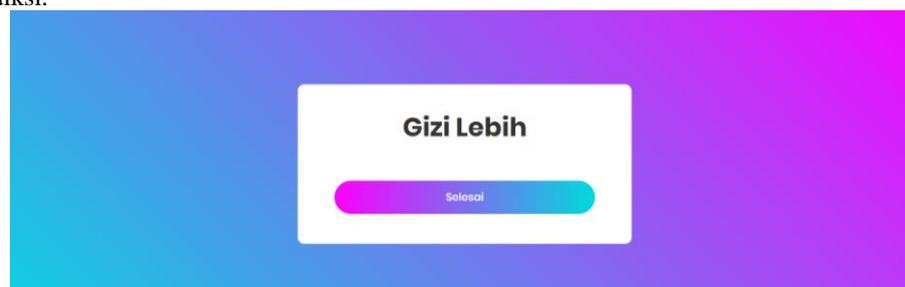
Berdasarkan pada gambar 6, tampilan prediksi digunakan untuk melakukan penginputan pada variabel umur, berat badan dan tinggi badan, sebelum melakukan prediksi atau klasifikasi.



Gambar 6. Tampilan Prediksi

b. Tampilan Hasil Prediksi

Berdasarkan pada gambar 7, tampilan hasil prediksi di gunakan untuk menampilkan hasil prediksi atau klasifikasi, setelah melakukan penginputan pada variabel umur, berat badan dan tinggi badan yang tersedia pada tampilan prediksi.



Gambar 7. Tampilan Hasil Prediksi

E. Pengujian Sistem

Pengujian sistem menggunakan *black box* dilakukan untuk menemukan kesalahan pada sistem. Adapun hasil pengujian sistem dapat dilihat pada tabel 9.

Tabel 9. Pengujian Sistem

Pengujian	Keluaran Yang Dihasilkan	Keterangan
Tampilan Home	Menampilkan halaman home yang berisi informasi masukkan umur, berat badan dan tinggi badan.	Berhasil
Klik Masukkan Umur	Menampilkan halaman untuk memasukkan umur.	Berhasil
Klik Masukkan Berat Badan	Menampilkan halaman untuk memasukkan berat badan.	Berhasil
Klik MasukkanTinggi Badan	Menampilkan halaman untuk memasukkan tinggi badan.	Berhasil
Klik Prediksi	Menampilkan hasil prediksi.	Berhasil
Klik Selesai	Kembali ke tampilan home.	Berhasil

F. Pengujian Kuisisioner

Pengujian kuisisioner merupakan pengujian yang dilakukan untuk mengambil responden atau pendapat. Pengujian dalam bentuk kuisisioner ini terdiri dari 6 pertanyaan terkait dengan aplikasi yang dibuat yang disebar pada 30 responden yang berbeda profesi dan pendidikan. Perhitungan dibuat menggunakan skala likert. Skala likert adalah skala yang digunakan untuk mengukur data yang didapatkan secara kualitatif, sehingga menghasilkan data yang akurat dan teruji kebenarannya. Adapun rumus skala likert adalah sebagai berikut:

$$\Sigma \text{ Nilai persentasi} = \frac{x}{\text{responden}} \times 100\% \dots\dots\dots(4)$$

Berikut ini adalah jumlah responden setiap pertanyaan, dapat dilihat pada tabel 10.

Tabel 10. Jumlah Responden

No.	Pertanyaan	Jumlah Responden		
		SS (%)	S (%)	CS (%)
1.	Bagaimana pendapat anda mengenai tampilan aplikasi tersebut?	43,3	50	0
2.	Aplikasi ini mudah digunakan?	36,7	63,3	0
3.	Aplikasi ini berjalan dengan baik dan sesuai fungsinya?	53,3	36,7	10
4.	Aplikasi ini dapat dipahami dengan jelas?	40	46,7	13,3
5.	Aplikasi ini dapat berjalan dengan lancar?	43,3	56,7	0
6.	Aplikasi klasifikasi status gizi balita ini dapat membantu mempermudah proses klasifikasi status gizi balita?	33,3	60	0
Rata-Rata (%)		41,65	52,23	3,883

Berdasarkan hasil analisa pada tabel 10 dengan total responden sebanyak 30 orang responden, dari hasil uji coba kuisisioner yang telah dilakukan pada sistem aplikasi klasifikasi status gizi balita, maka didapatkan nilai yaitu 41,65 % dengan jawaban Sangat Setuju (SS), 52,23 % dengan jawaban Setuju (S) dan 3,883 % dengan jawaban Cukup Setuju (CS). Maka dapat dijelaskan bahwa sistem aplikasi ini dapat berjalan dengan baik.

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat disimpulkan bahwa Metode *K-Nearest Neighbors* memiliki cara kerja klasifikasi dengan menghitung jarak antara setiap variabel pada data training dengan data testing. Kemudian akan dicari jumlah K atau *neighbors* yang memiliki jarak terdekat. Sementara metode *Naïve Bayes Classifier* memiliki cara kerja klasifikasi dengan membuat fungsi pendekatan berdasarkan probabilitas bersyarat yang didapatkan dari data training. Metode *K-Nearest Neighbors* performa model klasifikasi akan sangat dipengaruhi oleh jumlah K. Sehingga dalam penerapan metode *K-Nearest Neighbors* kita dapat mencari jumlah K yang dapat memberikan performa paling optimal. Berdasarkan hasil perbandingan performa antara metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* menggunakan *f1 score* sebagai patokan utama performa klasifikasi. Didapatkan hasil bahwa metode *K-Nearest Neighbors* unggul pada *f1 score* dengan selisih cukup besar yakni 13,42 %. Sehingga pada masalah klasifikasi status gizi balita metode *K-Nearest Neighbors* mengungguli *Naïve Bayes Classifier*. Berdasarkan hasil pengujian pada aplikasi yang telah dibuat menggunakan model dari metode *K-Nearest Neighbors*, didapatkan hasil 30 responden 41,65 % dengan jawaban Sangat Setuju (SS), 52,23 % dengan jawaban Setuju (S) dan 3,883 % dengan jawaban Cukup Setuju (CS). Saran pengembangan selanjutnya dapat dikembangkan dengan menerapkan metode yang dapat digunakan untuk mengatasi masalah klasifikasi dengan kelas tidak seimbang. Penambahan variabel yang bersifat kategori dapat diikutsertakan ke dalam proses pemodelan.

6. DAFTAR PUSTAKA

- Enterprise, J. (2017). *Otodidak Pemrograman Python*. Jakarta: Elex Media Komputindo.
- Grinberg, M. (2014). *Flask Web Development*. (Meghan Blanchette and Rachel Roumeliotis, Ed.). Sebastopol: O'Reilly Media.
- Hariri, F. R., & Pamungkas, D. P. (2016). Implementasi Naïve Bayes Classifier Untuk Diagnosa Status Gizi Balita. *Seminar Nasional Teknologi Informasi Dan Multimedia*, Hal. 19-24.
- Haryati, D. F., Abdullah, G., & Hadiana, A. I. (2016). Klasifikasi Jenis Batubara Menggunakan Jaringan Syaraf Tiruan Dengan Algoritma Backpropagation. *Seminar Nasional Teknologi Informasi Dan Komunikasi*, Hal. 557-562.
- Kusrini, & Emha, L. T. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python*. (Dawn Schanafelt, Ed.). Sebastopol: O'Reilly Media.
- Nugraha, S. D., Putri, R. R. M., & Wihandika, R. C. (2017). Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol.1(No. 9), Hal. 925-932.
- Pramitarini, Y., Purnama, I. K. E., & Purnomo, M. H. (2013). Analisa Rekam Medis Untuk Menentukan Status Gizi Anak Balita Menggunakan Naive Bayes Classifier. *Prosiding Seminar Nasional Manajemen Teknologi XVII*, Hal. 1-8.
- Putri, R. E., Suparti, & Rahmawati, R. (2014). Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012. *Jurnal Gaussian*, Vol. 3(No. 4), Hal. 831-838.
- Saleh, H., Faisal, M., & Musa, R. I. (2019). Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor. *Jurnal Sistem Informasi Dan Teknik Komputer*, Vol. 4(No. 2), Hal. 120-126.
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes , Nearest Neighbour , dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jatisi*, Vol. 1(No. 2), Hal. 151-161.
- Setiawan, E. (2012). Kamus Besar Bahasa Indonesia (KBBI).
- Sihombing, E. S., Honggowibowo, A. S., & Nugraheny, D. (2012). Implementasi Data Mining Menggunakan Metode Apriori Pada Transaksi Penjualan Barang (Studi Kasus Di Chorus Minimarket). *Compiler*, Volume 1(Nomor 1).
- Zubair, A., & Muksin, M. (2018). Penerapan Metode Naive Bayes Untuk Klasifikasi Status Gizi (Studi Kasus Di Klinik Bromo Malang). *Seminar Nasional Sistem Informasi*, Hal. 1204-1208.