

BAB II TINJAUAN PUSTAKA

2.1 Keluarga

Ulfa, M., & Na'imah, N. (2020) menjelaskan bahwa keluarga adalah sebuah lingkungan sosial terkecil dan secara tidak langsung lingkungan tersebut dapat mempengaruhi suatu proses perkembangan dan proses pertumbuhan seorang anak.

Menurut buku Panduan Pelaksanaan Pendampingan Keluarga dalam Upaya Percepatan Penurunan *Stunting* di Tingkat Desa/Kelurahan, keluarga merupakan unit paling kecil dalam suatu masyarakat yang terdiri dari suami dan istri, atau terdiri dari suami, istri, dan anak, atau hanya terdiri dari ayah dan anak atau ibu dan anak.

Dapat diambil kesimpulan bahwa keluarga merupakan salah satu organisasi atau lingkup lingkungan sosial yang paling kecil yang ada pada suatu tatanan Masyarakat yang biasanya terdiri dari minimal 2 orang dan terbentuk atas ikatan darah dan keturunan.

2.2 *Stunting*

Stunting merupakan salah satu permasalahan gizi terhadap balita yang saat ini sedang diperhatikan oleh seluruh dunia. Kasus *stunting* ini disebabkan karena kurangnya asupan nutrisi dalam makanan, dan minuman yang tidak sesuai dengan kebutuhan gizi seharusnya (Wiraguna, 2022). Berdasarkan Peraturan Presiden Nomor 72 Tahun 2021 tentang Pendekatan Pendampingan dan Strategi Penanganan *Stunting*, menyatakan bahwa *stunting* merupakan sebuah kondisi gagal pertumbuhan yang disebabkan oleh gizi buruk kronis dan infeksi berulang pada anak-anak berusia antara 0-59 bulan dengan ciri-ciri tinggi badan berada dibawah rata-rata standar yang telah ditetapkan pemerintah (sumber: LN.2021/No.172, jdih.setneg.go.id : 23 hlm.).

Kondisi ini membuat pertumbuhan anak menjadi terhambat atau terhenti secara permanen, yang berdampak pada perkembangan fisik, kognitif, dan psikomotorik anak. *Stunting* juga dapat menyebabkan sistem kekebalan tubuh menjadi lebih lemah dan meningkatkan risiko penyakit saat anak tumbuh dewasa. Selain itu, menurut Badan Kesehatan Dunia (World Health Organization) dan UNICEF (United Nations International Children's Emergency Fund), *stunting* adalah sebuah kondisi ketika tinggi badan seorang anak yang memiliki rentang usia dibawah 5 tahun mempunyai tinggi badan yang kurang dari dua standar deviasi dari median atau nilai tengah tinggi badan anak sehat yang seumurannya. Salah satu upaya dalam menangani keluarga

beresiko *stunting* adalah dengan memberikan pendampingan dan bantuan oleh pemerintah. Namun, masih banyak program pemerintah yang belum sesuai kriteria yang diharapkan dan belum tepat sasaran (Susanti, R. 2023).

Berdasarkan pendapat para ahli diatas dapat disimpulkan *stunting* adalah kondisi pertumbuhan anak yang berbeda dengan anak seumuran lainnya. Anak cenderung memiliki perbedaan pertumbuhan dengan anak lainnya namun dalam hal kekurangan seperti kurangnya tinggi badan dari mayoritas anak seumurannya dan kekurangan gizi untuk mendukung pertumbuhannya. Sedangkan keluarga beresiko *stunting*, adalah kondisi sebuah keluarga yang memungkinkan keturunannya akan mengalami *stunting* dengan berbagai faktor yang dimiliki.

2.3 Prediksi

Prediksi adalah dugaan atau prediksi tentang sebuah peristiwa atau kejadian yang akan datang. Prediksi dapat bersifat kualitatif (tanpa angka) atau kuantitatif (dengan angka). Prediksi kualitatif sulit diperoleh hasil yang baik karena variabel kualitatif bersifat relative (Novianty, D et al. 2021). Tujuan prediksi adalah untuk mendapatkan informasi terkait apa akan terjadi di masa mendatang dengan tingkat kemungkinan terbesar bahwa hal itu akan terjadi.

Sedangkan menurut (T. Syahputra 2018 et.al.), prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Sumber: Jurnal Sistem Informasi Triguna Dharma (JURSI TGD), 1(4), 275-281.)

Berdasarkan keterangan para ahli diatas, prediksi adalah sebuah proses peramalan atau perkiraan terhadap suatu hal yang akan terjadi pada masa mendatang dengan menggunakan metode ilmiah. Karena sifatnya berupa perkiraan, maka bisa saja perkiraan tersebut tidak sesuai. Namun dengan prediksi ini, seseorang dapat memiliki informasi berupa kemungkinan terbesar suatu hal akan terjadi pada masa mendatang.

2.4 Algoritma Naïve Bayes

Menurut (Akbar et al., 2022) Naïve Bayes Classifier (NBC) adalah sebuah algoritma dalam proses pengklasifikasian probabilitas sederhana dengan menerapkan Teorema Bayes yang mana teori ini berasumsi tidak ada ketergantungan (independent) yang tinggi. Bentuk persamaan umum Naïve Bayes menurut (Akbar et al., 2022) adalah sebagai berikut:

$$P(C|F_1, \dots, F_n) = \frac{p(C) \cdot p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (2.1)$$

C adalah representasi dari kelas, sementara $F_1 \dots F_n$ adalah representasi berbagai karakteristik petunjuk yang digunakan untuk klasifikasi. Persamaan tersebut berarti bahwa peluang masuk sebuah data dengan karakteristik ke dalam kelas C adalah peluang munculnya kelas C sebelum masuknya data dikalikan dengan peluang muncul karakteristik data secara global.

Arisandi, R.R., Warsito, B., & Hakim, A.R. (2022) menjelaskan Naïve Bayes Classifier adalah salah satu jenis algoritma yang berbasis klasifikasi. Naïve Bayes Classifier berasumsi bahwa ada atau tidaknya ciri khusus dari sebuah kelas, tidak ada kaitannya dengan ciri dari kelas lainnya.

Naïve Bayes menurut pendapat para ahli diatas, adalah salah satu algoritma klasifikasi data yang memiliki asumsi bahwa suatu kelas tidak ada kaitannya dengan kelas lain. Algoritma Naïve Bayes ini cocok diterapkan pada dataset dengan dimensi tinggi karena sifat algoritmanya sederhana dan cepat. Naïve Bayes sering diaplikasikan pada topik klasifikasi teks, rekomendasi produk, pengenalan pola, diagnosis medis, analisis sentiment, dan topik lainnya. Namun, Naïve Bayes tidak selamanya menjadi pilihan yang tepat. Dalam pemilihan model tentunya harus diselaraskan dengan jenis, tujuan, dan karakteristik dataset.

2.5 Python

Python adalah salah satu dari sekian banyak bahasa pemrograman tingkat tinggi. Menurut Enterprise, J. (2019) dalam bukunya, python adalah bahasa pemrograman interpretatif yang bersifat mudah dipelajari dan berfokus pada keterbacaan kode. Menurut pengertian lain, *python* adalah bahasa pemrograman tingkat tinggi yang dirilis tahun 1991 oleh Guido Van Rossum dan telah menjadi salah satu bahasa paling populer (Alfarizi et al., 2023).

Menurut keterangan diatas, python merupakan sebuah bahasa pemrograman yang berfokus kepada pembacaan kode dan termasuk bahasa pemrograman yang mudah dipelajari. Python merupakan bahasa pemrograman Tingkat tinggi, lengkap, dan mudah untuk dipahami sehingga saat ini termasuk salah satu bahasa pemrograman yang populer karena fleksibel dan memiliki komunitas yang besar.

Bahasa pemrograman python tidak hanya mendukung pemrograman berorientasi objek, dan skrip saja, tetapi juga memiliki banyak library dan framework yang memungkinkan suatu pengembangan, seperti pengembangan web, analisis data, kecerdasan buatan, pengembangan perangkat lunak, dan fungsi lainnya.

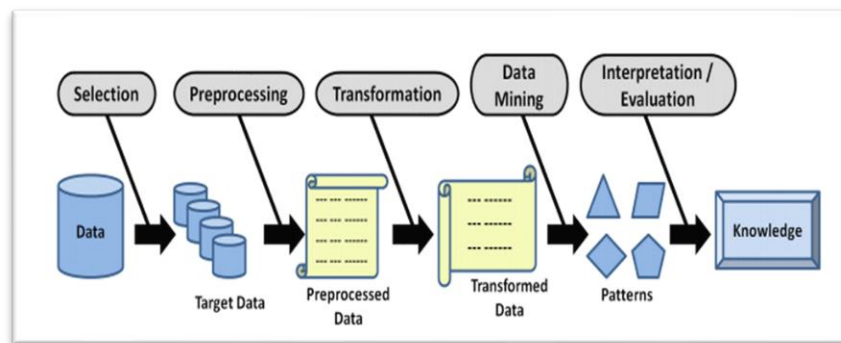
2.6 SMOTE

SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan sebuah metode yang sering dipakai dalam *machine learning* dalam mengatasi ketidakseimbangan kelas variabel data. Dalam hal ini biasanya terdapat satu kelas mayoritas dan satu kelas minoritas. Kondisi seperti ini disebut *imbalance*. Dalam kondisi *imbalance* ini, SMOTE bekerja dengan menciptakan sampel baru pada kelas minoritas hingga data menjadi seimbang.

Dalam penjelasan lain yang dijelaskan oleh Kasanah, A. N., Muladi, M., & Pujiyanto, U. (2019), SMOTE adalah sebuah metode machine learning yang dapat membangkitkan record data minoritas menjadi sebanyak data mayoritas. Menurut penjelasan diatas, SMOTE merupakan metode untuk menangani ketidakseimbangan data, lalu data tersebut diolah menjadi seimbang dengan membuat sampel kelas pada data yang lebih sedikit.

2.7 Knowledge Discovery In Database (KDD)

KDD merupakan sebuah metode yang digunakan untuk memperoleh sebuah pengetahuan yang bersumber dari suatu database. Hasil dari pengetahuan yang diperoleh dapat dijadikan sebuah basis pengetahuan (*knowledge base*) yang selanjutnya dapat digunakan sebagai landasan dalam proses pengambilan keputusan (Gustientiedina, G., 2019).



Gambar 2. 1 Tahap Knowledge Discovery In Database (KDD)

Menurut Sembiring, M. A dkk, KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data besar. Sedangkan menurut Fakhri, D. A., & Defit, S. (2021), *Knowledge Discovery in Database* (KDD) merupakan sebuah proses analisa terstruktur yang bertujuan untuk mendapatkan informasi yang baru dan benar, menemukan pola dari data yang kompleks, dan bermanfaat.

Menurut penjelasan diatas, *Knowledge Discovery In Database* (KDD) adalah sebuah tahapan dalam memproses data dengan tujuan untuk memahami sebuah database agar dapat menghasilkan sebuah pengetahuan yang dapat dijadikan landasan dalam proses pengambilan keputusan.

2.8 Evaluasi Model

Evaluasi merupakan sebuah tahapan pengecekan dalam setiap hal dengan tujuan untuk melihat progress, dan menilai progress. Dalam analisis data, evaluasi merupakan hal yang sangat penting karena dengan evaluasi sebuah model dapat diukur kualitasnya dan progresnya. Untuk mengukur kualitas model ada beberapa metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Dengan hal tersebut sebuah dapat diketahui kualitas dan kekurangannya.

Tabel 2. 1 *Confusion Matrix*

Kategori		True Value	
		True	False
Hasil Klasifikasi	True	TP	FN
	False	FP	TN

Keterangan:

1. TP (True Positive) adalah jumlah sampel yang sebenarnya positif dan diprediksi positif.
2. FN (False Negative) adalah jumlah sampel yang sebenarnya positif tetapi diprediksi negative.
3. FP (False Positive) adalah jumlah sampel yang sebenarnya negatif tetapi diprediksi positif.
4. TN (True Negative) adalah jumlah sampel yang sebenarnya negatif dan diprediksi negatif.

2.8.1 Accuracy

Akurasi dihitung sebagai rasio antara jumlah prediksi yang benar dan total jumlah prediksi. Ini menunjukkan seberapa baik model klasifikasi mampu mengklasifikasikan data dengan benar.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.2)$$

2.8.2 Precision

Presisi dapat mengukur jumlah prediksi positif yang benar atau *True Positive* (TP) dengan total prediksi positif adalah rasio presisinya.

$$Precision = \frac{TP}{TP+FP} \quad (2.3)$$

2.8.3 Recall

Recall, juga disebut Sensitivitas atau *True Positive Rate* (TPR). *Recall* dapat mengukur sejauh mana model mengidentifikasi semua kasus positif. Ini adalah rasio antara jumlah prediksi positif yang benar (TP) dengan total jumlah kasus positif yang sebenarnya.

$$Recall = \frac{TP}{TP+FN} \quad (2.4)$$

2.8.4 F1-Score

F1-Score dapat berfungsi ketika ingin mencari keseimbangan antara presisi dan *recall* karena *f1-score* merupakan metrik yang menggabungkan presisi dan *recall*.

$$F1-Score = \frac{2 \times precision \times recall}{precision+recall} \quad (2.5)$$

2.9 Penelitian Terdahulu

Dalam proses penyelesaian penelitian ini, peneliti merujuk pada beberapa referensi penelitian dengan topik yang relevan berupa 5 jurnal nasional dan 2 jurnal internasional.

Tabel 2. 2 Penelitian Terdahulu

No	Penulis	Judul	Metode	Hasil
1	Arisandi, R.R., Warsito, B., & Hakim, A.R. (2022).	APLIKASI NAÏVE BAYES CLASSIFIER (NBC) PADA KLASIFIKASI STATUS GIZI BALITA STUNTING DENGAN PENGUJIAN K-	Naïve Bayes	Penelitian ini melakukan proses klasifikasi status gizi balita <i>stunting</i> dengan menggunakan <i>10-Fold Cross Validation</i> pada algoritma Naïve Bayes Classifier dan memperoleh akurasi tertinggi sebesar 95,14%. Hasil klasifikasi digunakan menjadi model untuk proses prediksi.

		FOLD CROSS VALIDATION		
2	Argina, A. M. (2020).	Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes.	K-Nearest Neighbor	Hasil akurasi tertinggi yaitu 39% pada K=3, presisi tertinggi yaitu 65% pada K=3 dan K=5, recall tertinggi yaitu 36% pada K=3, dan F-Measure tertinggi yaitu 46% pada K=3. Nilai yang diperoleh tidak cukup baik dikarenakan jumlah data yang digunakan cukup kecil.
3	Titimeidara, M.Y., & Hadikurniawati, W. (2021)	IMPLEMENTASI METODE NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI STATUS GIZI STUNTING PADA BALITA	Naïve Bayes	Penerapan metode Naïve Bayes pada kalsifikasi status gizi <i>stunting</i> balita ini mendapat nilai akurasi sebesar 88% dari total data 300 yang dibagi menjadi 275 data sebagai data latih dan 25 data sebagai data uji.
4	Kurniadi, D., Nuraeni, F., & Lestari, S. (2022)	Implementasi Algoritma Naïve Bayes Menggunakan Feature Forward Selection dan SMOTE Untuk Memprediksi Ketepatan Masa Studi Mahasiswa Sarjana	Naïve Bayes	Pada penelitian ini diterapkan proses <i>Feature Forward Selection</i> dan SMOTE yang dapat mempengaruhi nilai akurasi menjadi lebih tinggi jika dibandingkan dengan yang tidak memakai 2 fitur tersebut. Performa terbaik pada model ini yaitu nilai <i>Accuracy</i> sebesar 87,13%, nilai <i>Recall</i> sebesar 83,82% dan nilai <i>Precision</i> sebesar 89,76%.

5	Larassati, D., Zaidiah, A., & Afrizal, S. (2022)	Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naive Bayes	Naïve Bayes	Pada objek prediksi penyakit jantung coroner ini, Naïve Bayes mendapatkan akurasi tertinggi sebesar 83,1% dengan perbandingan data <i>training</i> dan <i>testing</i> 60:40
6	Ressan, M. B., & Hassan, R. F. (2022)	Naive-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets	Naïve Bayes	Model Naïve Bayes multinomial berhasil mencapai akurasi klasifikasi lebih tinggi dibandingkan model lain (mencapai akurasi sebesar 91,6% pada dataset pertama dan akurasi 87,6% pada dataset kedua).
7	Putri, V. M., Masjkur, M., & Suhaeni, C. (2021)	Performance of SMOTE in a random forest and naive Bayes classifier for imbalanced Hepatitis-B vaccination status	Random Forest & Naïve Bayes	Penerapan SMOTE pada random forest & naïve bayes menjadikan identifikasi status non vaksinasi Hepatitis B meningkat akurasinya sebesar 30,08% dan 26,09% dibanding tidak menerapkan SMOTE.

Berdasarkan beberapa penelitian terdahulu di atas, topik *stunting* sudah menjadi pembahasan yang cukup familiar, sehingga peneliti memilih topik berbeda dengan konteks yang masih berhubungan, yaitu keluarga beresiko *stunting* sebagai tahap pencegahan dalam kasus *stunting*. Peneliti juga memilih algoritma Naïve Bayes, karena dari rata-rata penelitian terdahulu hasil akurasi Naïve Bayes memiliki akurasi yang tinggi. Selain itu, algoritma Naïve Bayes termasuk mudah dan cocok dengan karakteristik dataset keluarga beresiko *stunting*.

Proses *oversampling* SMOTE digunakan untuk mengatasi proses *imbalance data*. Terbukti dengan penelitian terdahulu dalam proses pengolahan data dapat diolah dengan baik dan meningkatkan performa akurasi model. SMOTE dapat membuat data sintesis untuk menyamakan jumlah antara atribut yang terdapat pada kolom penentu sehingga jumlahnya menjadi seimbang.