



Prediction of Stunting Risk In Families Using Naïve Bayes Classifier and Chi-Square

Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square

**Umairah Rizkya Gurning^{1*}, Sania Fitri Octavia², Dwi Ratna Andriyani³,
Nurainun⁴, Inggih Permana⁵**

^{1,2,3,5}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

⁴Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

^{1,2,3,4,5}Puzzle Research Data Technology (Predatech), Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹11950320687@students.uin-suska.ac.id, ²11950321581@students.uin-suska.ac.id,

³111950324645@students.uin-suska.ac.id, ⁴11950125156@students.uin-suska.ac.id,

⁵inggihpermana@uin-suska.ac.id

Received Jul 16th 2023; Revised Nov 18th 2023; Accepted Dec 24th 2023

Corresponding Author: Umairah Rizkya Gurning

Abstract

Stunting is something that is dangerous to humans because it can cause growth and development inhibition of other organs including the brain, heart and kidneys. The increasing cases of stunting in toddlers requires an effort in early handling and prevention. There are 17 attributes in the stunting data that must be considered, with so many attributes it is difficult to find the most influential attribute in predicting stunting. In this study, feature selection was applied using Chi Square and the Naïve Bayes Algorithm was applied to find the attributes that should be prioritized in predicting stunting. Prediction results using only Naive Bayes in this study obtained an accuracy value of 94.3%, a recall value of 93.9% and a precision value of 93.93% with a time of 0.07 seconds. Meanwhile, by applying Chi square feature selection in this study, the 5 attributes that most influenced the prediction of stunting were obtained which could increase the speed of model formation of the Naive Bayes Algorithm with a time of 0.01 seconds, but could not increase accuracy, recall and precision. It is hoped that related agencies can pay more attention to and prioritize these 5 attributes as monitoring of stunting predictions in the City of Dumai.

Keyword: Chi Square , Feature Selection, Naive Bayes Classifier, Prediction, Stunting

Abstrak

Stunting merupakan sesuatu yang berbahaya pada manusia karena dapat menyebabkan terjadinya hambatan pertumbuhan serta perkembangan organ lainnya termasuk otak, jantung dan ginjal. Meningkatnya kasus stunting pada balita memerlukan suatu upaya dalam penanganan dan pencegahan secara dini. Terdapat 17 atribut pada data stunting yang harus diperhatikan, dengan banyaknya atribut tersebut menyebabkan sulitnya menemukan atribut yang paling berpengaruh dalam memprediksi stunting. Pada penelitian ini diterapkan seleksi fitur menggunakan Chi Square dan menerapkan Algoritma Naïve Bayes untuk menemukan atribut yang harus diprioritaskan dalam memprediksi stunting. Hasil prediksi dengan menggunakan Naive bayes saja pada penelitian ini didapatkan nilai akurasi sebesar 94,3 %, nilai recall sebesar 93,9 % dan nilai precision sebesar 93,93% dengan waktu 0,07 detik. Sedangkan dengan menerapkan seleksi fitur Chi square pada penelitian ini diperoleh 5 atribut yang paling berpengaruh terhadap prediksi stunting yang dapat meningkatkan kecepatan pembentukan model Algoritma Naive Bayes dengan waktu 0,01 detik, namun tidak dapat meningkatkan akurasi, recall dan presisi. Harapannya instansi terkait dapat lebih memperhatikan dan memprioritaskan ke-5 atribut tersebut sebagai pemantauan prediksi stunting di Kota Dumai.

Kata Kunci: Chi Square, Naive Bayes Classifier, Prediksi, Seleksi Fitur, Stunting

1. PENDAHULUAN

Stunting adalah gangguan pertumbuhan dan perkembangan pada anak akibat kekurangan gizi kronis dan infeksi berulang yang ditandai dengan tinggi badan yang dibawah standar [1]. Stunting merupakan sesuatu yang berbahaya pada manusia. Hal ini dikarenakan stunting dapat menyebabkan fisik pendek serta terjadinya hambatan pertumbuhan dan perkembangan organ lainnya termasuk otak, jantung dan ginjal [2]. Oleh karena itu anak yang mengalami stunting akan mengalami penurunan fungsi organ, gangguan psikomotor, penurunan Intelligence Quotient (IQ), kemampuan motorik dan neurosensory . Selain itu, stunting dapat memberikan dampak buruk untuk jangka panjang dan jangka pendek. Dampak jangka panjang dari stunting adalah mengurangi kemampuan kognitif dan prestasi belajar, menurunkan kekebalan tubuh sehingga lebih rentan sakit, gangguan kesehatan reproduksi, resiko tinggi terhadap munculnya penyakit diabetes, kanker, stroke, jantung dan pembuluh darah serta mengalami disabilitas pada usia tua, sedangkan dampak jangka pendeknya adalah terganggunya kecerdasan anak, pertumbuhan fisik, metabolisme dan pencernaan yang tidak normal [3]. Pada Desember 2021 tercatat angka prevalensi stunting di Indonesia sebesar 24,4%, angka tersebut masih belum memenuhi standar yang diberikan WHO terhadap target stunting yaitu maksimal 20% [4].

Stunting dapat dideteksi dengan segera, misalnya dengan pemantauan status gizi, sehingga dapat dilakukan peringatan dini [5]. Early warning sangat penting untuk stunting. Melalui early warning, permasalahan dapat diidentifikasi lebih awal sehingga pengelolaan risiko dapat juga dilakukan lebih awal [6]. Maka, manfaat early warning stunting adalah untuk mengetahui munculnya permasalahan gizi yang dapat dijadikan early warning, agar dapat diatasi dengan segera [7]. Salah satu contoh program stunting yaitu pemantauan tumbuh kembang anak, berperan penting dalam memastikan anak bebas dari masalah kesehatan dan status gizi. Program tersebut dapat menjadi early warning jika ada hambatan dalam tumbuh kembang anak. Misalnya, mengukur tinggi badan dapat digunakan untuk melihat apakah seorang anak mengalami stunting. Demikian pula, timbang berat badan anak untuk melihat apakah berat badan anak dalam kisaran normal. Dalam hal ini, jika hasil pengukuran atau penimbangan menunjukkan gejala gagal tumbuh, tentu dapat segera dilakukan pencegahan [8].

Early Warning pada stunting sangat sulit dilakukan. Terbatasnya akan pengetahuan masyarakat tentang kesehatan serta informasi stunting dapat menyebabkan stunting menjadi sulit untuk terdeteksi dikarenakan istilah stunting masih terdengar asing dikalangan masyarakat [9]. Selain itu rendahnya kepedulian masyarakat dan pemahaman masyarakat akan pengetahuan tentang stunting seperti kurangnya kesadaran masyarakat dalam melakukan pemeriksaan dapat menyebabkan sulitnya pengontrolan dan pencegahan masalah stunting [10]. Keterbatasan akan sarana dan prasarana termasuk kualitas layanan kesehatan yang belum optimal juga berpengaruh dalam mendeteksi tumbuh kembangnya balita seperti terbatasnya alat pengukur tinggi badan dan berat badan yang dapat mengakibatkan sulitnya untuk melakukan pemeriksaan [11]. Faktor lainnya yaitu keterbatasannya tenaga kesehatan yang menyebabkan daya cakup pelayanan kesehatan menjadi belum optimal serta keterbatasan seorang pakar juga membuat masyarakat kesulitan dalam mengambil tindakan saat terkena stunting karena kurangnya pengetahuan yang dimiliki [11].

Penelitian ini menggunakan Algoritma Naive Bayes Classifier (NBC) untuk mengatasi sulitnya *early warning* stunting. NBC adalah salah satu algoritma klasifikasi pada bidang *machine learning* (ML). Hasil klasifikasi dari NBC bisa digunakan sebagai *early warning stunting*. ML sendiri adalah salah satu cabang dalam kecerdasan buatan yang dapat bekerja dengan data dalam jumlah yang besar dan memungkinkan sistem untuk belajar sendiri tanpa harus diprogram berulang kali [12]. Algoritma NBC yaitu sebuah algoritma yang mengimplementasikan metode probabilitas dan statistik untuk memprediksi peluang pada masa yang akan datang berdasarkan pengalaman pada masa lalu [13]. Kelebihan dari Algoritma NBC adalah proses klasifikasi masih dapat dilakukan meskipun terdapat nilai yang hilang pada suatu atribut [14]. Hal inilah yang menjadi alasan Algoritma NBC digunakan pada penelitian ini. Hal ini disebabkan pada data yang digunakan pada penelitian ini banyak terdapat missing value pada atributnya. Pada data tersebut terdapat sekitar lebih dari 50% record memiliki missing value, dengan atribut yang memiliki missing value paling banyak adalah atribut ‘ada anak 7-15 tahun tidak sekolah’. Algoritma NBC juga telah digunakan pada berbagai penelitian tentang penyakit, seperti: penyakit ginjal kronik [15], penyakit kulit [16], penyakit stroke [17] dan penyakit asma [18].

Permasalahan “Curse Of Dimensionality” atau banyaknya dimensi data dapat mengurangi akurasi terhadap deteksi awal penyakit stunting. Untuk melakukan deteksi awal stunting BKKBN memiliki 17 indikator sebagai dasar dari prediksi deteksi awal stunting. Akurasi algoritma dapat mengalami penurunan yang diakibatkan oleh beberapa faktor seperti banyaknya atribut, atribut yang redundant dan kurang relevannya atribut yang ada [19] [20]. Selain itu banyaknya atribut atau dimensi pada algoritma juga dapat menurunkan kinerja suatu algoritma karena dari banyaknya fitur tersebut mempengaruhi algoritma baik dari segi waktu maupun proses pengenalan pola datanya, namun penghapusan fitur secara acak dapat menyebabkan kehilangan informasi pada data [21]. Sehingga Curse of dimensionality ini dapat menyebabkan kompleksitas pemrosesan data dari segi ruang dan waktu [22]. Banyaknya dimensi data membutuhkan

peningkatan kinerja algoritma pada pengolahan data dan sebuah metode yang dapat menghitung tingkat dependensi fitur terhadap suatu kelas sehingga algoritma yang diterapkan dapat berjalan secara optimal [22]

Penelitian ini menggunakan Chi-Square sebagai reduksi dimensi untuk mengatasi permasalahan curse of dimensionality. Menurut [23], [24], dan [25] reduksi dimensi data dapat meningkatkan akurasi algoritma NBC. Salah satu metode reduksi dimensi adalah metode seleksi fitur. Seleksi fitur adalah proses pemilihan subset variabel input yang relevan dari dataset besar yang akan digunakan untuk membangun model [24]. Pemilihan fitur bertujuan untuk mengurangi fitur yang tidak relevan dan berlebihan, membutuhkan lebih sedikit waktu untuk melatih model, dan dapat membantu meningkatkan kinerja pengklasifikasi yang dihasilkan [26]. Metode seleksi fitur yang digunakan pada penelitian ini adalah Chi-Square. Penerapan model chi-square adalah untuk mengurangi tingkat kesalahan dari terlalu banyak atribut dan untuk mengurangi penyertaan atribut-atribut tersebut dalam variabel klasifikasi penyakit. Selanjutnya, langkah pengurangan merupakan langkah penting dalam pengenalan pola untuk mengklasifikasikan fitur diagnostik pada atribut [27]. Pada penelitian [28] dan [29] NBC yang menggunakan Chi-Square memiliki akurasi yang lebih tinggi dari NBC yang tanpa menggunakan Chi-Square.

Penelitian ini terdiri dari empat sub bab, yaitu pendahuluan, data dan metode, hasil dan pembahasan, serta kesimpulan. Pada sub bab pendahuluan menjelaskan mengenai latar belakang dari penelitian dengan judul: “Prediksi Risiko Stunting Pada Keluarga Menggunakan Naïve Bayes Classifier Dan Chi-Square”. Selain itu juga dijelaskan secara singkat mengenai metode yang akan digunakan. Sub bab data dan metode menjelaskan mengenai data yang akan digunakan beserta metodenya, yaitu menggunakan algoritma NBC dengan seleksi fitur chi-square hingga pengukuran performa yang akan dilakukan. Sub bab hasil dan pembahasan menjelaskan mengenai hasil dari penelitian yang sudah dilakukan dengan data dan metode sebelumnya. Dan sub bab kesimpulan menjelaskan mengenai kesimpulan akhir yang didapatkan dari penelitian yang dilakukan.

2. DATA DAN METODE

2.1 Data

Data stunting besumber dari Dinas Pengendalian Penduduk dan Keluarga Brenecana Kota Dumai Provinsi Riau. Pada penelitian ini menggunakan data stunting pada 10 kelurahan Kota Dumai, yaitu Kelurahan Bagan Besar, Kelurahan Bagan Keladi, Kelurahan Basilan Baru, Kelurahan Guntung, Kelurahan Jaya Mukti, Kelurahan Kayu Kapur, Kelurahan Purnama, Kelurahan STDI, dan Kelurahan Teluk Binjai.

Pada penelitian ini data yang digunakan adalah data stunting pada Dinas Pengendalian Penduduk dan Keluarga Berencana Kota Dumai Provinsi Riau dengan jumlah 18 atribut dan 5000 record data. Dimana, atribut yang digunakan yaitu (A1) Baduta (0-23 bulan), (A2) Balita (24-59 bulan), (A3) PUS, (A4) PUS Hamil, (A5) Anak 7-15 tahun tidak sekolah, (A6) Tidak ada anggota keluarga multi sumber penghasilan, (A7) Jenis lantai tanah, (A8) Tidak setiap anggota keluarga makan makanan beragam, (A9) Keluarga Pra sejahtera keluarga, (A10) Keluarga tidak mempunyai air minum utama, (A11) Keluarga tidak mempunyai jamban yang layak, (A12) Keluarga tidak mempunyai rumah layak huni, (A13) Pendidikan terakhir ibu dibawah SLTP, (A14) Terlalu muda (Umur istri < 20 tahun), (A15) Terlalu tua (Umur istri > 75 tahun), (A16) Terlalu dekat (<2 tahun), (A17) Terlalu banyak (3 anak) dan atribut terakhir yaitu sebagai class dengan atribut (A18) Kategori keluarga berpotensi resiko stunting.

Tabel 1. Atribut Data Stunting

No	Item	Kode	Penjelasan	Jumlah Data	Value (X)	Value (V)	Value (-)	Missing Value
1.	Baduta	A1	Bawah Dua Tahun (0-23 bulan)	5021	4270	750	0	1
2.	Balita	A2	Bawah Lima Tahun (24-59 bulan)	5021	3471	1549	0	1
3.	PUS	A3	Pasangan Usia Subur	5021	100	4921	0	0
4.	PUS Hamil	A4	Pasangan Usia Subur Hamil	5021	4785	236	0	0
5.	Anak 7-15 Tahun	A5	Ada Anak 7-15 Tahun Tidak Sekolah	5021	142	221	4657	1
6.	Sumber Penghasilan	A6	Tidak Ada Anggota Keluarga Memiliki Kebutuhan Pokok Untuk Memenuhi Kebutuhan Pokok Per Bulan	5021	4954	67	0	0
7.	Jenis Lantai Tanah	A7	Jenis Lantai Rumah Tanah, Semen, atau Keramik	5021	4977	44	0	0

No	Item	Kode	Penjelasan	Jumlah Data	Value (X)	Value (V)	Value (-)	Missing Value
8.	Jenis Makanan	A8	Tidak Setiap Anggota Keluarga Makan “Makanan Beragam” Paling Sedikit 2 Kali Sehari	5021	4993	28	0	0
9.	Keluarga Pra Sejahtera	A9	Keluarga Yang Tidak Mampu Memenuhi Kebutuhan Dasarnya, Seperti Kebutuhan Pendidikan, Agama, Sandang, Pangan, dan Kesehatan.	5021	4714	307	0	0
10.	Sumber Air Minum	A10	Keluarga Yang Tidak Mempunyai Sumber Air Minum Utama Yang Layak	5021	4915	106	0	0
11.	Jenis Jamban	A11	Keluarga Yang Tidak Mempunyai Kamar Mandi Yang Layak	5021	4662	359	0	0
12.	Jenis Rumah	A12	Keluarga Tidak Mempunyai Rumah Layak Huni	5021	4067	954	0	0
13.	Pendidikan Ibu	A13	Pendidikan Terakhir Ibu Dibawah SLTP	5021	3766	1246	0	0
14.	Terlalu Muda (<20 Tahun)	A14	Umur Istri Kurang Dari 20 Tahun	5021	4816	130	74	1
15.	Terlalu Tua (>35 Tahun)	A15	Umur Istri Lebih Dari 35 Tahun	5021	2047	2490	119	2
16.	Terlalu Dekat (< 2 Tahun)	A16	Umur Anak Jaraknya Hanya 2 Tahun	5021	3205	252	1558	6
17.	Terlalu Banyak (3 Anak)	A17	Memiliki Anak Yang Terlalu Banyak	5021	2956	1964	86	2

2.2 Metode

Langkah-langkah penelitian ini dapat dilihat pada Gambar 1.

2.2.1 Naïve Bayes Classifier

Naive Bayes Classifier merupakan metode pengklasifikasian probabilistik sederhana untuk menghitung beberapa kemungkinan dari suatu dataset dengan dijumlahkannya frekuensi dan kombinasi nilai [30]. Arti lain dari *Naive Bayes* yang dikemukakan oleh *Thomas Bayes* yaitu metode klasifikasi berdasarkan probabilitas statistik yang memprediksi peluang dimasa depan berdasarkan pengalaman dari masa yang lalu [31]. *Teorema Bayes* memiliki bentuk umum pada persamaan 1 [32] :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dimana:

- X = Data class yang belum diketahui
- H = Hipotesis Data
- P(H|X) = Probabilitas hipotesis H berdasar kondisi X
- P(H) = Probabilitas hipotesis H
- P(X|H) = Probabilitas X berdasar kondisi pada hipotesis H
- P(X) = Probabilitas X

2.2.2 Chi - Square

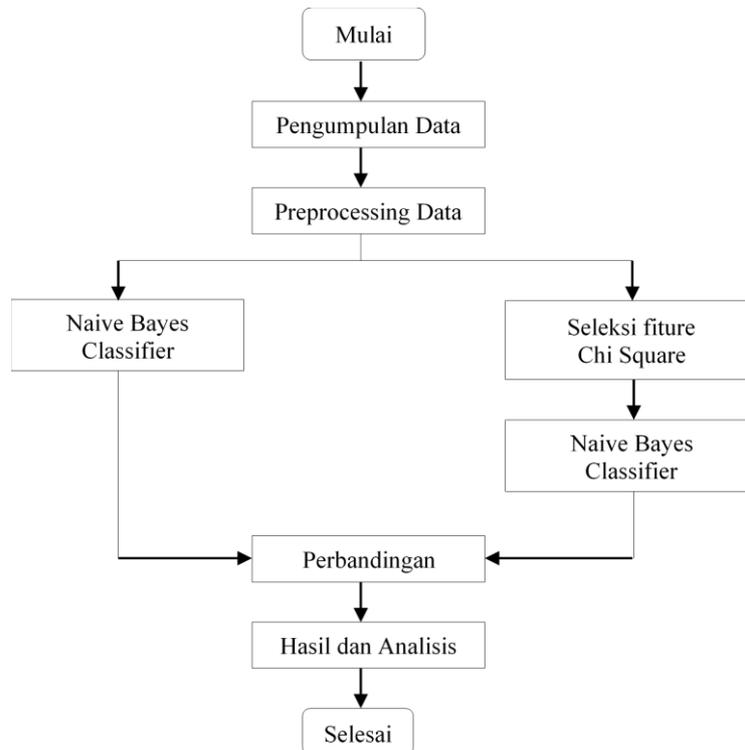
Chi-Square merupakan salah satu metode yang digunakan dalam bidang statistika [33]. Uji *Chi-square* merupakan jenis uji perbandingan non parametris yang paling umum dilakukan dengan menggunakan dua variabel, dengan variabel kedua memiliki skala data nominal. Jika hanya ada satu variabel dengan skala nominal dari dua variabel, maka analisis *Chi-square* dilakukan, menekankan bahwa metode ini perlu

digunakan pada derajat yang tidak seimbang [34]. Berikut rumus dari *Chi-Square* pada persamaan 2 adalah[35]:

$$\chi^2 = \sum_{i=0}^n \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

Dimana:

- χ^2 = distribusi Chi Square
- O_i = frekuensi observasi pengamatan ke-i
- E_i = frekuensi harapan pengamatan ke-i



Gambar 1. Metodologi Penelitian

2.2.3 *K-fold Cross Validation*

K-Fold Cross Validation adalah jenis validasi paling umum yang digunakan untuk mengukur efisiensi metode algoritma dengan menyimpan data sampel dan menganalisis data dalam peningkatan *K-fold Cross Validation*. Pada penelitian ini dilakukan dengan percobaan $K=10$. Oleh karena itu, satu-satunya kelompok *K-fold Cross Validation* yang bersangkutan akan ditetapkan sebagai data uji, sedangkan sisa kelompok lainnya akan ditetapkan sebagai data latih [36].

2.3 Pengukuran Performa

2.3.1 Akurasi

Menurut Kabir dan Hasan pada penelitian [37], Akurasi adalah suatu metode yang dapat dilakukan untuk menguji atau mengukur performa dengan melihat tingkat kedekatan antara nilai prediksi dengan nilai yang sebenarnya. Akurasi hasil prediksi didapatkan dari hasil analisis yang berupa jumlah data yang dapat diklasifikasikan secara benar. Adapun persamaan akurasi sebagai berikut.

$$Accuracy = \frac{N_{benar}}{N} \times 100\% \quad (3)$$

2.3.2 Recall

Menurut Kabir dan Hasan pada penelitian [37], Recall adalah suatu metode yang dapat dilakukan untuk menguji atau mengukur performa dengan melihat perbandingan antara jumlah informasi relevan yang sistem dapatkan dengan jumlah seluruh informasi relevan yang ada pada koleksi informasi (baik yang terambil maupun tidak terambil oleh sistem). Adapun persamaan recall sebagai berikut.

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

2.3.3 Precision

Menurut Kabir dan Hasan pada penelitian [37], Precision atau presisi adalah suatu metode yang dapat dilakukan untuk menguji atau mengukur performa dengan melihat perbandingan antara jumlah informasi relevan yang sistem dapatkan dengan jumlah seluruh informasi yang terambil pada sistem baik yang relevan maupun tidak. Adapun persamaan presisi sebagai berikut.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

2.4 Tools

Pada penelitian ini menggunakan tools yang digunakan yaitu Weka. Weka adalah aplikasi data mining open source berbasis Java. Weka terdiri dari kumpulan algoritme pembelajaran mesin yang dapat digunakan untuk menggeneralisasi atau membuat rumus dari sampel data [38].

2.5 Hasil Seleksi Fitur

Hasil seleksi fitur menggunakan chi-square didapatkan yaitu (A9) Keluarga Pra sejahtera keluarga, (A12) Keluarga tidak mempunyai rumah layak huni, (A13) Pendidikan terakhir ibu dibawah SLTP, (A15) Terlalu tua (Umur istri > 75 tahun), dan (A17) Terlalu banyak (3 anak).

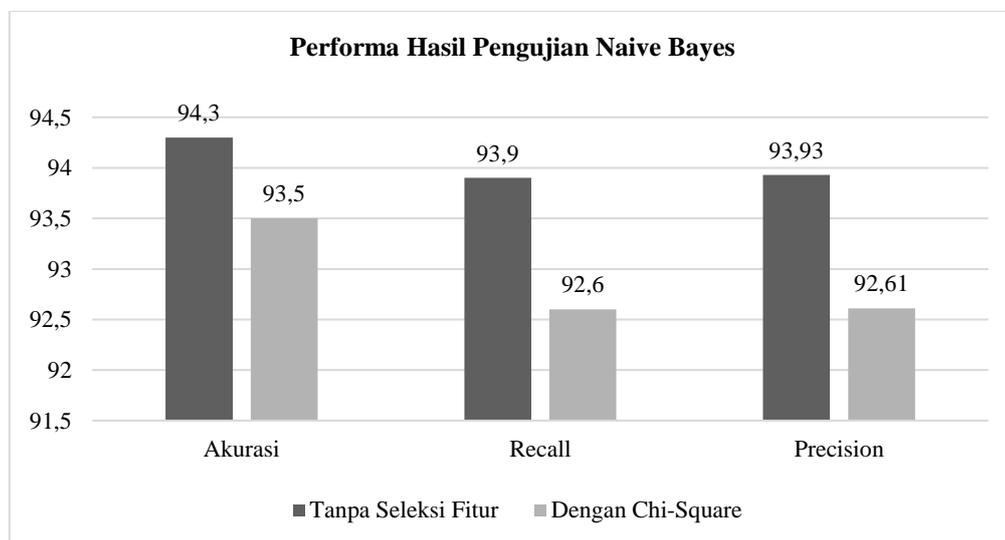
3. HASIL DAN PEMBAHASAN

3.1 Hasil Seleksi Fitur

Hasil seleksi fitur menggunakan chi-square didapatkan yaitu (A9) Keluarga Pra sejahtera keluarga, (A12) Keluarga tidak mempunyai rumah layak huni, (A13) Pendidikan terakhir ibu dibawah SLTP, (A15) Terlalu tua (Umur istri > 75 tahun), dan (A17) Terlalu banyak (3 anak).

3.2 Klasifikasi Menggunakan NBC

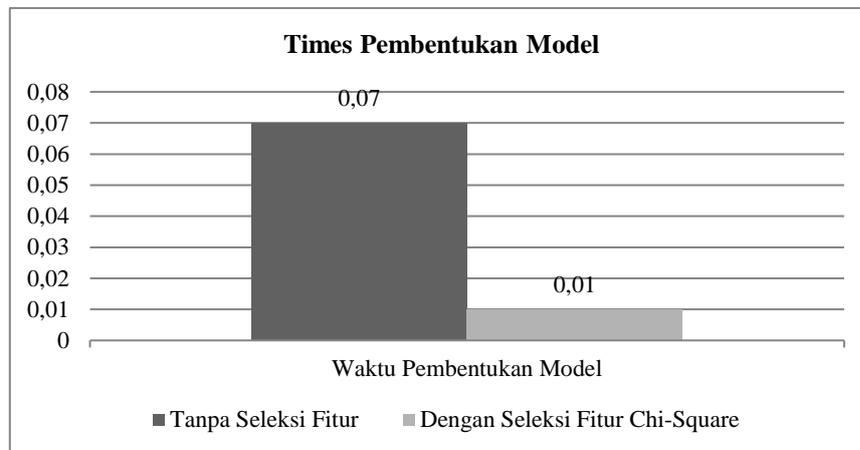
Berdasarkan hasil yang didapatkan dari seleksi fitur menggunakan Chi-square, data dibagi menggunakan *K-fold Cross Validation* untuk membagi data latih dan data uji. Sehingga didapatkan hasil precision, recall dan akurasi yang dapat dilihat pada Gambar 2 serta waktu pembentukan model algoritma NBC dapat dilihat pada gambar 3.



Gambar 2. Performa Hasil Pengujian Naive Bayes

Berdasarkan gambar 2 dan gambar 3 diketahui bahwa nilai rata-rata akurasi yang diperoleh pada kedua percobaan yang dilakukan adalah 93,9 dengan nilai akurasi tanpa seleksi fitur lebih tinggi 0,8% daripada percobaan dengan seleksi fitur chi-square. Tidak hanya nilai akurasi, nilai recall dan precision pada percobaan tanpa seleksi fitur juga lebih baik daripada dengan seleksi fitur dengan selisih 1,3 untuk recall dan 1,32 untuk nilai precision. Sehingga dapat diketahui bahwa penerapan seleksi fitur chi-square pada algoritma NBC untuk prediksi stunting di Kota Dumai tidak meningkatkan nilai akurasi, namun penerapan seleksi fitur dapat mempersingkat waktu pengolahan data pembentukan model NBC pada aplikasi Weka dengan 14

atribut dan 5 atribut mengalami penurunan waktu 0,06 detik. Pembentukan model dengan seleksi fitur chi-square lebih cepat daripada tanpa seleksi fitur. Dan dari penelitian ini dapat diketahui bahwa algoritma NBC dapat mengklasifikasikan data dengan baik meskipun data tersebut bersifat inbalance, dapat dilihat pada tabel 1 nilai pada false V lebih banyak daripada value lainnya.



Gambar 3. Times Pembentukan Model Algoritma

4. KESIMPULAN

Berdasarkan hasil dan analisis yang dilakukan dapat disimpulkan bahwa penerapan seleksi fitur Chi-Square pada prediksi stunting dapat meningkatkan kecepatan pembentukan model algoritma NBC namun tidak dapat meningkatkan nilai akurasi, recall dan presisi. Sehingga diperoleh nilai akurasi terbaik 94,3, recall 93,9 dan nilai precision 93,93 pada pengolahan data tanpa seleksi fitur. Seleksi fitur yang dilakukan pada penelitian ini menyeleksi 14 atribut menjadi 5 atribut yaitu (A9) Keluarga Pra sejahtera keluarga, (A12) Keluarga tidak mempunyai rumah layak huni, (A13) Pendidikan terakhir ibu dibawah SLTP, (A15) Terlalu tua (Umur istri > 75 tahun), dan (A17) Terlalu banyak (3 anak). Dengan demikian diperoleh 5 atribut yang paling berpengaruh terhadap prediksi stunting, harapannya instansi terkait dapat lebih memperhatikan dan memprioritaskan ke-5 atribut tersebut sebagai pemantauan prediksi stunting di Kota Dumai.

5. UCAPAN TERIMAKASIH

Ucapan terimakasih kepada Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) serta Dinas Pengendalian Penduduk dan Keluarga Berencana (DPPKB) yang telah memberikan data untuk penelitian ini. Terimakasih juga kepada Organisasi Puzzle Research Data Technology (Predatech) yang selalu memberikan dukungan, menjadi wadah pembelajaran dan memberikan motivasi dalam penyelesaian penelitian ini.

REFERENCES

- [1] Bimrew Sendekie Belay, "No Title", *הארץ*, vol. 2, no. 8.5.2017, pp. 2003–2005, 2022.
- [2] D. G. Shauma, Udzu Nabila & Purbaningrum, "Implementasi Kebijakan Percepatan Pencegahan Stunting," *J. Kebijak. Publik*, vol. 13, no. 2, pp. 97–104, 2022.
- [3] R. A. Saputri and J. Tumangger, "Hulu-Hilir Penanggulangan Stunting Di Indonesia," *J. Polit. Issues*, vol. 1, no. 1, pp. 1–9, 2019, doi: 10.33019/jpi.v1i1.2.
- [4] A. D. N. Yadika, K. N. Berawi, and S. H. Nasution, "Pengaruh stunting terhadap perkembangan kognitif dan prestasi belajar," *J. Major.*, vol. 8, no. 2, pp. 273–282, 2019.
- [5] E. M. Sari, M. Juffrie, N. Nurani, and M. N. Sitaresmi, "296266261," vol. 12, no. 4, 2016.
- [6] W. I. Susanti, A. P. Widodo, and S. A. Nugraheni, "Pengembangan Sistem Informasi Pencatatan dan Pelaporan Status Gizi Balita Stunting di Kelurahan Gajah Mungkur," *J. Manaj. Kesehat. Indones.*, vol. 7, no. 1, pp. 67–74, 2019, doi: 10.14710/jmki.7.1.2019.67-74.
- [7] W. Ghylfh *et al.*, "\$ FFHOHUDWLQJ LQIRUPDWLRQ RI WVXQDPL GLVDVWHU XVLQJ HDUO \ ZDUQLQJ V \ VWHP (: 6 ' HYL FHV V PDUW VROXWLRQ IRU FRPPXQLFDWLRQ QDYLDJWLRQ DQG VXUYHLOODQFH GXH WR WLGDO VHD," vol. 07008, 2021.
- [8] P. Priyono, "Strategi Percepatan Penurunan Stunting Perdesaan (Studi Kasus Pendampingan Aksi Cegah Stunting di Desa Banyumundu, Kabupaten Pandeglang)," *J. Good Gov.*, vol. 16, no. 2, pp. 149–174, 2020, doi: 10.32834/gg.v16i2.198.
- [9] A. Z. Rahmadhani, "Literature Review Pengalaman Ibu dengan Anak Stunting," *Repos. UNISA*

- Yogyakarta*, 2020.
- [10] N. Mellinia, H. Kurniawan, and A. Bhagaswanda, "Upaya peningkatan kunjungan posyandu di dusun serai serumpun desa sumber agung," vol. 1, 2022.
 - [11] K. Isni and S. M. Dinni, "Pelatihan Pengukuran Status Gizi Balita Sebagai Upaya Pencegahan Stunting Sejak Dini Pada Ibu Di Dusun Randugunting, Sleman, Diy," *Panrita Abdi - J. Pengabd. pada Masy.*, vol. 4, no. 1, p. 60, 2020, doi: 10.20956/pa.v4i1.7299.
 - [12] A. Rozaq, Y. Yunitasari, K. Sussolaikah, E. R. N. Sari, and R. I. Syahputra, "Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors Dan Decision Tree," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 746, 2022, doi: 10.30865/mib.v6i2.3554.
 - [13] D. Iskandar and Y. K. Suprpto, "Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4 . 5 dan Naïve Bayes Clasifier," *JAVA J. Electr. Electron. Eng.*, vol. 11, no. 1, pp. 14–17, 2013.
 - [14] J. Sihombing, "Klasifikasi Data Antropometri Individu Menggunakan Algoritma Naïve Bayes Classifier," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 1–10, 2021, doi: 10.37148/bios.v2i1.15.
 - [15] Q. A'yuniyah *et al.*, "Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 72, 2022, doi: 10.30865/json.v4i1.4781.
 - [16] I. S. Hafisah and P. N. Andono, "Deteksi Otomatis Penyakit Kulit Menggunakan Algoritma Naive Bayes," *J. Kesehat.*, no. 5, pp. 1–6, 2015.
 - [17] I. saifudin Subaeri, Triawan A. C, "Klasifikasi Penyakit Diabetes Pada Wanita Menggunakan Metode Naive Bayes," *Rec. Manag. J.*, vol. 1, no. 2, pp. 1–15, 2017, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.5042&rep=rep1&type=pdf%0Ahttps://www.ideals.illinois.edu/handle/2142/73673%0Ahttp://www.scopus.com/inward/record.url?eid=2-s2.0-33646678859&partnerID=40&md5=3ee39b50a5df02627b70c1bdac4a60ba%0Ahtt>
 - [18] S. Wiwit, "Metadata, citation and similar papers at core.ac.uk 4," *Донну*, vol. 5, no. December, pp. 118–138, 2015.
 - [19] Yunitasari, H. S. Hopipah, and R. Mayasari, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes," *Technomedia J.*, vol. 6, no. 1, pp. 99–110, 2021, doi: 10.33050/tmj.v6i1.1531.
 - [20] R. D. Widyantara, S. Zulaikah, F. B. Juangsa, B. A. Budiman, and M. Aziz, "Review on Battery Packing Design Strategies for Superior Thermal Management in Electric Vehicles," 2022.
 - [21] N. Indah, S. Mania, and N. Nursalam, "Peningkatan Kemampuan Literasi Matematika Siswa Melalui Penerapan Model Pembelajaran Problem Based Learning Di Kelas Vii Smp Negeri 5 Pallangga Kabupaten Gowa," *MaPan*, vol. 4, no. 2, pp. 200–210, 2016, doi: 10.24252/mapan.2016v4n2a4.
 - [22] K. Abend, "How Convolutional Neural Networks Defy the Curse of Dimensionality: Deep Learning Explained," 2022, doi: 10.36227/techrxiv.18316439.v1.
 - [23] O. Saini and S. Sharma, "A Review on Dimension Reduction Techniques in Data Mining," *Comput. Eng. Intell. Syst.*, vol. 9, no. 1, pp. 7–14, 2018.
 - [24] R. Aziz, C. K. Verma, and N. Srivastava, "Dimension reduction methods for microarray data: a review," *AIMS Bioeng.*, vol. 4, no. 1, pp. 179–197, 2017, doi: 10.3934/bioeng.2017.1.179.
 - [25] G. Chao, Y. Luo, and W. Ding, "Recent Advances in Supervised Dimension Reduction: A Survey," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 341–358, 2019, doi: 10.3390/make1010020.
 - [26] G. Kicska and A. Kiss, "Comparing swarm intelligence algorithms for dimension reduction in machine learning," *Big Data Cogn. Comput.*, vol. 5, no. 3, 2021, doi: 10.3390/bdcc5030036.
 - [27] V. Novalia *et al.*, "JITE (Journal of Informatics and Telecommunication Engineering) Application of Artificial Intelligence Chi-Square Model and Classification Of KNN in Heart Disease Detection," vol. 6, no. July, pp. 180–188, 2022.
 - [28] E. Karyadiputra, E. Noersasongko, and A. Marjuni, "Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial Menggunakan Algoritma Naive Bayes Berbasis Seleksi Atribut Chi Squared," *Cyberku J.*, vol. 12, no. 2, pp. 116–122, 2016, [Online]. Available: <http://research.pps.dinus.ac.id/index.php/Cyberku/article/view/13>
 - [29] W. Febriasto, Rizki, Satyaning, "Prediksi Kuat Tekan Semen Untuk Produk Feature Selection," *Sains Dan Seni*, vol. 8, no. 2, 2019.
 - [30] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
 - [31] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," *EVOLUSI J. Sains dan Manaj.*,

-
- vol. 8, no. 1, pp. 93–105, 2020, doi: 10.31294/evolusi.v8i1.7535.
- [32] H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [33] J. Reynaldo, P. P. Adikara, and R. C. Wihandika, “Analisis Sentimen Mengenai Produk Toyota Avanza Menggunakan Metode Learning Vector Quantization Versi 3 (LVQ 3) dengan Seleksi Fitur Chi Square, Lexicon ...,” ... *Tekno. Inf. dan ...*, vol. 4, no. 3, pp. 830–839, 2020.
- [34] H. Mardiansyah, R. Widia Sembiring, and S. Efendi, “Handling Problems of Credit Data for Imbalanced Classes using SMOTEXGBoost,” *J. Phys. Conf. Ser.*, vol. 1830, no. 1, 2021, doi: 10.1088/1742-6596/1830/1/012011.
- [35] R. Palupi and S. M. S. Winarsih, “Pengaruh Disiplin Ilmu Terhadap Kecenderungan Mahasiswa Dalam Mengakses Informasi Melalui Media Sosial Menggunakan Metode Chi Square,” vol. 9, no. 1, pp. 1–6, 2021.
- [36] F. Tangguh Admojo, “Indonesian Journal of Data and Science Klasifikasi Aroma Alkohol Menggunakan Metode KNN,” vol. 1, no. 2, pp. 34–38, 2020.
- [37] G. Balita, S. Dengan, and P. K. Cross, “Aplikasi naïve bayes classifier (nbc) pada klasifikasi status gizi balita stunting dengan pengujian k-fold cross validation,” *J. Gaussian*, vol. 11, pp. 130–139, 2022.
- [38] M. Bakri, “Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means,” *J. Teknoinfo*, vol. 11, no. 1, p. 6, 2017, doi: 10.33365/jti.v11i1.3.