

Classification of Poor Households in Padang City Using the Naïve Bayes Algorithm with Synthetic Minority Oversampling Technique

Anice Kartika Putri, Dina Fitria*, Syafriandi, Tessa Octavia Mukhti

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: dinafitria@fmipa.unp.ac.id

Submitted : 12 Agustus 2024

Revised : 13 Agustus 2024

Accepted : 14 Agustus 2024

ABSTRACT

Poverty is a condition where a person is unable to meet minimum basic needs or a condition caused by the influence of development policies that have not been able to reach all levels of society. In Indonesia, the government has designed various programs to overcome poverty, but these programs are often not on target. One method to improve the effectiveness of the program is through proper classification of poor and non-poor households. This study uses the Naïve Bayes classification method which is popular in data mining to predict data categories based on the probability distribution of its features. However, challenges arise when the data is unbalanced between different classes. To overcome this, the Synthetic Minority Oversampling Technique (SMOTE) method is used to balance the data. Based on the analysis that has been carried out To determine the performance of Naïve Bayes using SMOTE and without SMOTE in classifying poor households in Padang City in 2023, classification using the Naïve Bayes method without SMOTE produced an accuracy value of 98%, precision of 0%, and recall of 0%. Meanwhile, the classification using the Naïve Bayes method with SMOTE produces an accuracy value of 90%, precision of 87%, and recall of 92% and the results of the criteria for poor households in Padang City in 2023 using Naïve Bayes can be seen from the results that the probability of poor households is much greater than that of non-poor households, therefore the data is classified as group of households that are classified as poor.

Keywords: Imbalance data, Naïve Bayes, Poor Households, Synthetic Minority Oversampling Technique, Naïve Bayes.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Kemiskinan termasuk dalam permasalahan yang cukup meresahkan bagi masyarakat di seluruh dunia. Permasalahan selain meliputi ketidakmampuan ekonomi, tetapi juga mencerminkan ketidakseimbangan dalam akses terhadap sumber daya dasar seperti pangan, pendidikan, dan kesehatan. Kemiskinan merupakan keadaan kehidupan yang dialami oleh seseorang atau rumah tangga yang tidak memiliki sumber daya yang memadai dan tidak mampu memenuhi kebutuhan dasar mereka. Miskin atau tidaknya seseorang dapat ditentukan oleh garis kemiskinan. Garis kemiskinan juga dapat diartikan sebagai pendapatan minimum yang harus dicapai oleh seseorang untuk memperoleh standar hidup yang layak di suatu daerah yang diukur dengan satuan rupiah/kapita/bulan.

Kota Padang adalah salah satu kota di wilayah Provinsi Sumatera Barat yang telah berhasil meningkatkan taraf kehidupan masyarakatnya melalui pembangunan di berbagai aspek (Mulia & Saputra, 2020). Meningkatnya perekonomian di Kota Padang maka meningkat pula garis kemiskinan. Garis kemiskinan di Kota Padang tinggi akibat inflasi yang meningkatkan biaya hidup dan pembangunan yang tidak merata, sehingga meskipun ekonomi tumbuh, banyak penduduk tetap terjebak dalam kemiskinan. Hal tersebut dilihat dari jumlah penduduk miskin tertinggi di Provinsi Sumatera Barat adalah Kota Padang. Untuk mengatasi kemiskinan Pemerintah telah mengembangkan berbagai program untuk mengatasi kemiskinan, salah satunya Kartu Indonesia Pintar (KIP) kuliah. Akan tetapi, program pemerintah yang disalah gunakan oleh masyarakat. Sehingga diperlukan pengawasan yang lebih ketat serta klasifikasi yang tepat agar pemerintah dapat membedakan rumah tangga miskin dan tidak miskin sehingga dana bantuan tersebut digunakan sesuai tepat sasaran.

Klasifikasi merupakan salah satu teknik yang terdapat dalam data mining. Tujuan klasifikasi adalah mengelompokkan data dalam kelas-kelas dengan mempelajari pola berdasarkan kesamaan karakteristik data. Salah satu

metode klasifikasi yang paling populer digunakan adalah metode *Naïve Bayes*. Menurut Singh & Singh (2019), *Naïve Bayes* merupakan metode klasifikasi probabilistik didasarkan pada teorema bayes, digunakan untuk menentukan kelas atau kategori dari data berdasarkan distribusi probabilitas dari fitur-fiturnya. *Naïve Bayes* memprediksi probabilitas keanggotaan kelas dengan mengasumsikan independensi antar atribut.

Namun, salah satu kelemahan dari metode klasifikasi adalah jika data yang digunakan untuk pelatihan tidak seimbang di antara kelas-kelas yang berbeda. Ketidakseimbangan ini dapat menyebabkan model lebih cenderung memprediksi kelas yang dominan, sehingga mengurangi akurasi dan keandalan dalam mengklasifikasikan data dari kelas minoritas. Hal ini bisa berakibat buruk pada kinerja model, terutama dalam situasi di mana pengenalan kelas minoritas sangat penting (Sun et al., 2009). alam konteks data tidak seimbang, yaitu jumlah sampel dalam satu kelas jauh lebih banyak dibandingkan kelas lainnya. Kelas mayoritas memiliki banyak data dalam suatu kelas, sedangkan kelas minor memiliki sedikit data pada suatu kelas.

Metode *Synthetic Minority Oversampling Technique* (SMOTE) digunakan untuk mengatasi ketidakseimbang kelas minoritas pada sebaran data, dianggap lebih baik karena tidak menghilangkan informasi penting. Pada analisis ini memilih metode SMOTE yaitu data yang digunakan tidak seimbang atau *imbalance* maka perlunya penyeimbangan terlebih dahulu dan metode ini mampu meningkatkan kinerja model tanpa menyebabkan overfitting pada data.

Tujuan dari analisis ini adalah untuk mengetahui hasil perbandingan klasifikasi dari rumah tangga miskin di Kota Padang menggunakan algoritma *Naïve Bayes* dengan *Synthetic Minority Oversampling Technique* dan tanpa *Synthetic Minority Oversampling Technique*, serta mengetahui kriteria rumah tangga miskin di Kota Padang tahun 2023 yang dihasilkan oleh *Naïve Bayes*. Sehingga analisis ini dapat digunakan oleh pihak pemerintah dalam mengatasi masalah kemiskinan dengan mempertimbangkan hasil klasifikasi rumah tangga miskin yang dilakukan.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data sekunder yang digunakan pada analisis ini bersumber dari hasil Survey Sosial Ekonomi Nasional (Susenas) pada Modul Konsumsi dan Pengeluaran di Kota Padang tahun 2023 dari Badan Pusat Statistik (BPS).

No.	Atribut	Kategori
1	Luas lantai bangunan (X_1)	1= $<8m^2$; 2= $\geq 8m^2$
2	Jenis lantai tempat tinggal (X_2)	1=Marmer/granit; 2=Keramik; 3=Parket/vinil/karpet; 4=Ubin/tegel/teras; 5=Kayu/papan; 6=Semen/bata merah; 7=Bambu; 8=Tanah
3	Jenis dinding tempat tinggal (X_3)	1=Tembok; 2=Plesteran anyaman ambu/kawat; 3=Kayu/papan ; 4=Anyaman bambu; 5=Batang kayu; 6=Bambu; 7=Lainnya.
4	Fasilitas buang air besar (X_4)	1=Ada, digunakan hanya ART sendiri ; 2=Ada, digunakan bersama rumah tangga tertentu; 3=Ada, MCK komunal; 4=Ada, MCK umum; 5=Ada, ART tidak menggunakan; 6=Tidak ada fasilitas
5	Sumber penerangan (X_5)	1=PLN dengan meteran ; 2=PLN tanpa meteran ; 3=Non PLN ; 4=Bukan listrik
6	Sumber air minum (X_6)	1=Air kemasan bermerk ; 2=Air isi ulang ; 3=Leding meteran; 4=Leding eceran ; 5=Sumur bor/pompa; 6=Sumur terlindung; 7=Sumur tak terlindung ; 8=Mata air terlindung ; 9=Mata air tak terlindung ; 10=Air permukaan (sungai/ danau/waduk/kolam/irigasi); 11=Air Hujan
7	Bahan bakar untuk memasak (X_7)	0=Tidak memasak di rumah; 1=Listrik; 2=Elpiji 12 kg; 3=Elpiji 5.5 kg; 4=Elpiji 3 kg; 5=Gas kota; 6=Biogas; 7=Minyak tanah; 8=Briket; 9=Arang; 10=Kayu bakar; 11=Lainnya
8	Pendidikan tertinggi kepala keluarga (X_8)	0=Tidak memiliki ijazah; 1=Tamat SD/ sederajat; 2=Tamat SLTP/ sederajat; 3=Tamat SLTA/ sederajat; 4=D1/D2/D3/D4/S1; 5=Profesi/S2/S3
9	Kepemilikan Tabungan (X_9)	1=Ada; 5=Tidak
10	Pernah kehabisan makanan (X_{10})	1=Ya; 5=Tidak

11	Rumahan tangga berdasarkan garis kemiskinan (Y)	1=Rumah tangga tidak miskin (RT); 2=Rumah tangga miskin (RTM)
----	---	---

B. Teknik Analisis Data

Dengan menggunakan bantuan *software google colab*, berikut ini langkah-langkah analisis menggunakan metode *Naive Bayes* dengan SMOTE:

1. Pengumpulan data Susenas dari Badan Pusat Statistik Provinsi Sumatera Barat tahun 2023.
2. *Preprocessing* data merupakan proses awal yang dilakukan pada data mentah untuk mempersiapkan pengolahan data utama atau analisis selanjutnya. Dengan melabelkan data pada peubah sesuai dengan label yang akan digunakan untuk klasifikasi.
3. Menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan pada data

Pembentukan data sintesis berbeda antara data berskala numerik dan kategorik. Pada data numerik menggunakan metrik Euclidean, sedangkan pada data kategorik jarak diukur dengan menggunakan rumus *Value Difference Metric* (VDM). Persamaan *Value Difference Metric* (VDM) digunakan untuk menghitung jarak antar contoh kelas minor yang variabel prediktornya kategorik (Azmatul et al., 2013).

$$\Delta(X, Y) = \sum_{i=1}^N \delta(V_1, V_2)$$

dengan:

$$\delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|$$

- $\Delta(X, Y)$: Jarak antara amatan X dan Y
- N : Banyaknya variabel predictor
- $\delta(V_1, V_2)$: Jarak antara amatan X dan Y untuk setiap variabel yang dihitung
- C_{1i} : Banyaknya kategori ke-1 yang termasuk pada variabel ke- i
- C_{2i} : Banyaknya kategori ke-2 yang termasuk pada variabel ke- i
- C_1 : Banyaknya kategori ke-1 terjadi
- C_2 : Banyaknya kategori ke-2 terjadi
- n : Banyaknya kategori pada variabel ke- i

Teknik SMOTE dianggap sebagai solusi dalam mengatasi ketidakseimbangan data yang hampir samaa dengan teknik *oversampling*. SMOTE dilakukan untuk meningkatkan jumlah data kelas yang lebih sedikit atau minor untuk menyeimbangkan sebaran data dengan menghasilkan data sintesis baru berdasarkan k tetangga terdekat yang dipilih secara acak.

$$X_{new} = X_i + (\hat{X}_k - X_i). \delta \tag{1}$$

Nilai k tetangga terdekat bisa dipilih secara acak, namun pada umumnya k yang digunakan adalah $k=5$, dan bilangan acak untuk nominal adalah 0 atau 1 berbeda dengan data numerik nilai bilangan acak bernilai 0 sampai 1.

4. Melakukan klasifikasi menggunakan metode *Naive Bayes*.

Langkah klasifikasi menggunakan algoritma *Naive Bayes* sebagai berikut.

- a. Memecah data menjadi data latih dan data uji dengan perbandingan 80%:20%.

Penggunaan data uji untuk menguji performa dan evaluasi akurasi model setelah pelatihan. Sedangkan melatih model yang akan dikembangkan digunakan data latih.

- b. Melakukan perhitungan probabilitas *prior*.

Menurut Jiang dkk., (2012), metode *Naive Bayes* menggunakan statistik dan probabilitas berdasarkan pengalaman di masa sebelumnya untuk memprediksi peluang dimasa depan. Jika kelas *prior* probabilitas sebelumnya tidak diketahui, maka umumnya diasumsikan bahwa *prior* probabilitas kelas-kelas tersebut kemungkinan memiliki nilai yang sama. Kelas *prior* probabilitas dapat diduga dengan persamaan berikut :

$$P(C_i) = \frac{S_i}{S} \tag{2}$$

Dimana $P(C_i)$: *Prior* (probabilitas C)

S_i : banyak data latih pada kelas C_i
 C : label kelas tertentu
 S : total banyaknya data latih.

- c. Masing-masing kelas dihitung probabilitas atribut berdasarkan jenis peubah kategorik. Selanjutnya menghitung probabilitas atribut, dengan persamaan berikut (Han & Kamber, 2006: 312) :

$$P(X|C_i) = \prod_{k=1}^n P(x_k | C_i) \tag{3}$$

Peluang $P(x_1|C_i).P(x_2|C_i)...P(x_n|C_i)$ dapat diduga dari data *training* menggunakan persamaan berikut. Jika A_k atribut kategorik, maka (Han & Kamber, 2006: 312) :

$$P(x_k|C_i) = \frac{S_{ik}}{S_i} \tag{4}$$

Dimana S_{ik} adalah banyaknya data latih pada kelas C_i yang memiliki nilai x_k untuk A_k dan S_i banyaknya data latih pada kelas C_i .

- d. Perkalian probabilitas *prior* dihitung dengan masing-masing kelas pada probabilitas atribut.
 e. Perkalian probabilitas *prior* dengan probabilitas atribut untuk mencari nilai maksimal.
 5. Membandingkan akurasi model dengan *confusion matrix* sebelum SMOTE dan sesudah SMOTE.

Ukuran evaluasi model yang sering digunakan untuk masalah klasifikasi adalah *confusion matrix*. *Confusion matrix* merupakan suatu metode yang memiliki fungsi untuk melakukan analisis sejauh mana *machine learning* dapat mengklasifikasikan dengan baik untuk mengidentifikasi *tuple* di berbagai kelas yang berbeda. Empat kemungkinan hasil prediksi klasifikasi mungkin untuk variable respon dua kelas yang disajikan pada Tabel 1.

Tabel 1. *Confusion Matrix*

Classification		Predicted Class	
		Class = YES	Class = NO
Actual Class	Class = YES	True Positif (TP)	False Positif (FP)
	Class = NO	False Negatif (FN)	True Negatif (TN)

Keterangan :

- a. TP (*True Positive*) merupakan kumpulan data yang kelas positif data aktual dan prediksi.
 b. FN (*False Negative*) merupakan kumpulan data yang kelas aktual adalah kelas positif dan kelas prediksi kelas negatif.
 c. FP (*False Positive*) merupakan kumpulan data yang kelas aktual adalah kelas negatif dan kelas prediksi kelas positif.
 d. TN (*True Negative*) merupakan kumpulan data yang kelas negatif data actual dan prediskis
 Nilai-nilai dapat dihitung berdasarkan nilai-nilai *confusion matrix*
 a. *Accuracy* digunakan untuk mengevaluasi sejauh mana model klasifikasi berhasil dalam memprediksi dengan benar seluruh kelas yang dimaksud.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{6}$$

- b. *Precision* mengukur seberapa akurat prediksi positif model. Dihitung sebagai rasio antara jumlah prediksi positif benar (*True Positives*) dengan total jumlah prediksi positif (*True Positives + False Positives*).

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

- c. *Recall* atau *True Positive Rate (TPR)*. *Recall* mengukur kemampuan model untuk menemukan kasus positif yang ada. Ini adalah rasio antara jumlah prediksi positif yang benar (*True Positives*) dengan total jumlah instance positif yang sebenarnya (*True Positives + False Negatives*).

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

6. Menarik kesimpulan.

III. HASIL DAN PEMBAHASAN

Ketidakseimbangan data menjadi sebuah masalah yang dapat ditangani dengan metode *Synthetic Minority Oversampling Technique*. Menghasilkan data sintesis baru dengan k tetangga terdekat untuk mengimbangi jumlah kelas minor dan kelas mayor merupakan metode *Synthetic Minority Oversampling Technique*. Data kelas minor dalam penelitian ini dibangkitkan sehingga memiliki jumlah yang seimbang dengan kelas mayor, seperti pada Tabel 2.

Tabel 2. Perbandingan Data Sebelum dan Sesudah SMOTE

Kategori	Data Penelitian	
	Sebelum	Sesudah
1 (Tidak Miskin)	789 (98,5%)	789 (50%)
2 (Miskin)	12 (1,5%)	789 (50%)
Jumlah	801 (100%)	1578 (100%)

Berdasarkan Tabel 2 data yang digunakan pada proses klasifikasi dikumpulkan dari data kategori rumah tangga tidak miskin dan miskin di Kota Padang pada tahun 2023 menggunakan metode naïve bayes dengan SMOTE. Data dibagi menjadi 80% untuk data latih dan 20% untuk data uji.

Langkah pertama, dalam klasifikasi menggunakan *Naïve Bayes* yaitu menghitung probabilitas $P(C_i)$ Probabilitas pada masing-masing kategori rumah tangga disajikan pada Tabel 3.

Tabel 3. Probabilitas Kategori Rumah Tangga

Tidak Miskin	Miskin
0.5	0.5

Tabel 3. di atas menunjukkan bahwa probabilitas kategori rumah tangga tidak miskin telah seimbang (sama). Langkah selanjutnya menghitung probabilitas Probabilitas X bersyarat C_i dihitung dengan menggunakan data *training* sebanyak 1262 data dengan menggunakan variabel predictor X_1 sampai X_{10} .

Tabel 4. Probabilitas Kategori Rumah Tangga berdasarkan Luas Lantai Bangunan Tempat Tinggal.

Luas Lantai Bangunan Tempat Tinggal	Jenis Jumlah		Probabilitas ($P(X C_i)$)	
	Tidak Miskin	Miskin	Tidak Miskin	Miskin
$<8m^2$	3	0	0.0051	0
$\geq 8m^2$	629	632	0.9949	1
Jumlah	632	632	1	1

Dapat dilihat dari Tabel 4 diketahui bahwa Probabilitas Kategori Rumah Tangga berdasarkan Luas Lantai Bangunan Tempat Tinggal probabilitas kategori rumah tangga tidak miskin dengan luas lantai bangunan $\geq 8m^2$ lebih besar dibandingkan kategori lain. Maka, luas lantai bangunan dengan $\geq 8m^2$ rumah tangga memiliki probabilitas besar sebagai kategori rumah tangga tidak miskin.

Tabel 5. Probabilitas Kategori Rumah Tangga berdasarkan Variabel X_2 sampai dengan X_{10}

Kriteria miskin	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Tidak Miskin	0.3523	0.8783	0.7909	0.9785	0.6477	0.8327	0.4639	0.3549	0.9594
Miskin	0.6464	0.7402	0.8251	0.864	0.3447	0.7313	0.2459	0.6717	0.6641

Diperoleh informasi dari Tabel 5 untuk nilai untuk nilai probabilitas Rumah Tangga Miskin berdasarkan kriteria X_2 (jenis lantai tempat tinggal Semen/bata merah), X_3 (jenis dinding tempat tinggal tembok/ bata merah), X_4 (fasilitas buang air besar digunakan hanya ART sendiri), X_5 (sumber penerangan PLN dengan meteran), X_6 (sumber air minum isi ulang), X_7 (bahan bakar untuk memasak elpiji), X_8 (pendidikan tertinggi kepala rumah tangga Tamat SLTA/ sederajat), X_9 (tidak ada tabungan), X_{10} (tidak pernah kehabisan makanan)

Selanjutnya Uji Validasi Model dilakukan dengan memilih salah satu data uji dengan kategori rumah tangga miskin dengan kriteria luas lantai $\geq 8m^2$, jenis lantai semen/bata merah, jenis dinding tembok, fasilitas buang air ada digunakan hanya ART sendiri, sumber penerangan PLN tanpa meteran, sumber air minum sumur bor/pompa, bahan bakar untuk memasak elpiji 3 kg, pendidikan tertinggi kepala rumah tangga adalah tamat SLTA/ sederajat, memiliki tabungan, dan tidak pernah kehabisan makanan. Berdasarkan perhitungan probabilitas pada kategori rumah tangga miskin didapatkan nilai peluang sebesar 0.000132 sedangkan rumah tangga tidak miskin didapatkan nilai peluang sebesar 0.000668. Hasil probabilitas peluang rumah tangga miskin lebih besar dari pada tidak miskin, oleh karena itu data tersebut diklasifikasikan ke dalam kelas rumah tangga miskin.

Selanjutnya melakukan Evaluasi model dilakukan perbandingan *Confusion Matrix Data Sebelum dan Sesudah SMOTE* membagi data latih sebesar 80% dan data uji sebesar 20%.

Tabel 6. *Confusion Matrix Data Sebelum dan Sesudah SMOTE*

Kelas Aktual	Kelas Prediksi			
	Sebelum SMOTE		Setelah SMOTE	
	Miskin	Tidak Miskin	Miskin	Tidak Miskin
Miskin	0	1	133	12
Tidak Miskin	2	158	20	151

Berdasarkan pada Tabel 6 dapat dilihat bahwa sebelum dilakukan SMOTE kategori rumah tangga miskin yang diprediksi dengan benar sebagai miskin sebanyak 0, kategori miskin yang diprediksi tidak miskin sebanyak 1, kategori tidak miskin yang diprediksi miskin sebanyak 2, dan kategori tidak miskin yang diprediksi dengan benar sebagai tidak miskin sebanyak 158. Sedangkan pada data setelah SMOTE didapatkan bahwa kategori rumah tangga miskin yang diprediksi dengan benar sebagai miskin sebanyak 133, kategori miskin yang diprediksi tidak miskin sebanyak 12, kategori tidak miskin yang diprediksi miskin sebanyak 20, dan kategori tidak miskin yang diprediksi dengan benar sebagai tidak miskin sebanyak 151.

Evaluasi model untuk data sebelum dilakukan SMOTE diperoleh hasil nilai *accuracy* sebesar 98%, *precision* sebesar 0%, dan *recall* sebesar 0%. Sedangkan evaluasi model untuk data setelah dilakukan SMOTE diperoleh hasil nilai *accuracy* sebesar 90%, *precision* sebesar 87%, dan *recall* sebesar 92%.

IV. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, klasifikasi metode *Naïve Bayes* tanpa SMOTE menghasilkan performa *accuracy* sebesar 98%, namun *precision* dan *recall* menghasilkan sebesar 0%. Sedangkan klasifikasi menggunakan metode *Naïve Bayes* dengan SMOTE menghasilkan performa nilai *accuracy* sebesar 90%, *precision* sebesar 87%, dan *recall* sebesar 92%. Dari hasil perbandingan klasifikasi rumah tangga miskin di Kota Padang tahun 2023, dapat disimpulkan bahwa metode SMOTE berhasil mengklasifikasikan kedua kelas yaitu kelas positif (miskin) dan negative (tidak miskin) dengan baik. Hasil kriteria rumah tangga miskin di Kota Padang tahun 2023 oleh *Naïve Bayes* dapat dilihat dari hasil probabilitas rumah tangga miskin jauh lebih besar dari pada tidak miskin, oleh karena itu data tersebut diklasifikasikan ke dalam kelas rumah tangga yang dianggap miskin. Dari hasil analisis yang telah dilakukan diharapkan kepada pemerintah untuk menangani masalah kemiskinan dengan mengetahui peluang-peluang hasil probabilitas klasifikasi rumah tangga miskin yang dilakukan.

DAFTAR PUSTAKA

- Azmatul, B. R., Dina, S. I., & M. Efendi, F. (2013). Penerapan Synthetic Minority Oversampling Technique (SMOTE) terhadap data tidak seimbang pada Pembuatan Model Komposisi Jamu. *Xplore*, 1(1), 1–6.
- BAPPENAS. (2018). *Analisis Wilayah dengan Kemiskinan Tinggi*. Kementrian PPN/Bappenas.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Han J, & Kamber M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

- Mulia, R. A., & Saputra, N. (2020). Analisis Faktor-Faktor yang Mempengaruhi Kesejahteraan Masyarakat Kota Padang. *Jurnal El-Riyasah*, 11(1), 67–83.
- Singh, N., & Singh, P. (2019). A novel Bagged Naïve Bayes-Decision Tree approach for multi-class classification problems. *Journal of Intelligent & Fuzzy Systems*, 36(3), 2261–2271. <https://doi.org/10.3233/JIFS-169937>
- Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of Imbalance Data : A Review . *International Journal of Pattern Recognition and Artificial Intelligence*, 23(04), 687–719. <https://doi.org/10.1142/S0218001409007326>