

IDENTIFIKASI KARAKTERISTIK KASUS RAWAN PANGAN DI PROVINSI ACEH PADA PENERIMA PROGRAM BANTUAN SOSIAL

S Mustaqima¹, E Ramadhani^{2*}, A F Hadi³, B Sartono⁴

¹Department of Statistics, Syiah Kuala University, Banda Aceh, 23111, Indonesia

²Department Of Statistics, Bogor Agriculture University, Bogor, 16680, Indonesia

³Department Of Mathematics, Jember University, Jember, 68121, Indonesia

*E-mail: evi.ramadhani@unsyiah.ac.id

ABSTRACT

Food insecurity is a global issue related to community welfare. Food insecurity is a condition where an individual or household is unable to meet their adequate food needs. To overcome food insecurity, it is necessary to know the characteristics or factors that cause it. The social protection program is a policy that plays an important role in efforts to fulfill economic access for households in an effort to reduce the incidence of food insecurity. This study aims to determine the characteristics of food-insecure households based on the status of receiving social protection programs in Aceh Province in 2020. The SMOTE CART method can be applied to achieve this goal. The application of SMOTE proved to be able to balance the data class on the food insecurity status variable. The results obtained were that BPNT and PKH contributed the most in determining food insecurity status with an AUC value of 0.61.

KEYWORDS: *Food insecurity, Social protection, SMOTE, CART, AUC*

ABSTRAK

Rawan pangan merupakan isu global yang berkaitan dengan kesejahteraan masyarakat. Rawan pangan adalah kondisi dimana suatu individu atau rumah tangga tidak mampu memenuhi kebutuhan pangan yang cukup. Untuk mengatasi rawan pangan perlu diketahui karakteristik atau faktor-faktor penyebabnya. Program perlindungan sosial merupakan kebijakan yang memegang peran penting dalam upaya pemenuhan akses secara ekonomi bagi rumah tangga dalam upaya menurunkan kejadian rawan pangan. Penelitian ini bertujuan untuk mengetahui karakteristik rumah tangga rawan pangan berdasarkan status penerimaan program perlindungan sosial di Provinsi Aceh tahun 2020. Metode SMOTE CART dapat diterapkan untuk mencapai tujuan tersebut. Penerapan SMOTE terbukti mampu menyeimbangkan kelas data pada variabel status rawan pangan. Hasil yang diperoleh adalah BPNT dan PKH memberikan kontribusi paling besar dalam menentukan status kerawanan pangan dengan nilai AUC sebesar 0,61.

KATA KUNCI: *Rawan pangan, perlindungan Sosial, SMOTE, CART, AUC*

1. PENDAHULUAN

Sesuai dengan Tujuan Pembangunan Berkelanjutan, salah satu tujuan Pemerintah Indonesia dalam mencapai pembangunan sosial adalah menghilangkan kelaparan dan menurunkan angka rawan pangan. Rawan pangan merupakan kondisi ketidakmampuan dalam pemenuhan pangan yang cukup secara jumlah dan kualitas. Rawan pangan merupakan permasalahan yang dihadapi secara global, termasuk di Provinsi Aceh. Rawan pangan erat kaitannya dengan tingkat kesejahteraan masyarakat sehingga harus diatasi. Kejadian rawan pangan pada rumah tangga dapat diatasi dengan mengetahui karakteristik atau faktor-faktor penyebab rawan pangan. Keterbatasan ekonomi menjadi salah satu penyebab rumah tangga mengalami kejadian rawan pangan. Berdasarkan data Badan Pusat Statistik (2021), angka kemiskinan di Provinsi Aceh mencapai 15,43%. Peningkatan angka kemiskinan menjadikan Aceh sebagai provinsi termiskin di Pulau Sumatera. Isu kemiskinan di Provinsi Aceh berpotensi memunculkan kejadian rawan pangan.

Merespon permasalahan kemiskinan pemerintah menerapkan kebijakan program perlindungan sosial bagi masyarakat untuk membantu pemenuhan akses secara ekonomi bagi rumah tangga. Program perlindungan sosial merupakan upaya pemerintah untuk membantu pemenuhan akses secara pendapatan dan konsumsi bagi rumah tangga miskin yang meliputi jaminan sosial dan bantuan sosial. Adapun program yang termasuk dalam bantuan sosial diantaranya adalah Program Keluarga Harapan (PKH), Kartu Keluarga Sejahtera (KKS), Program Indonesia Pintar (PIP), Bantuan Pangan Non-Tunai (BPNT), dan Bantuan Pemerintah Daerah (Supriyanto et al., 2014). Tingkat kejadian kerawanan pangan rumah tangga berhubungan dengan program perlindungan sosial. Oleh karena itu, program perlindungan sosial memegang peran penting dalam upaya pemenuhan akses secara ekonomi bagi rumah tangga dalam upaya menurunkan kejadian rawan pangan.

Rawan atau tidaknya suatu rumah tangga dapat ditentukan oleh status penerimaan program perlindungan sosial. Kajian mengenai kebijakan program perlindungan sosial terhadap kejadian rawan pangan menarik untuk dilakukan. Salah satu metode pembelajaran mesin yang dapat diterapkan untuk mengkaji masalah tersebut adalah *Classification and Regression Tree* (CART). CART merupakan salah satu metode klasifikasi berupa pohon keputusan yang sering digunakan untuk prediksi. Metode CART mampu menunjukkan karakteristik yang ditimbulkan oleh masing-masing tipe pada variabel *output* sehingga memudahkan dalam hal pengklasifikasian. Kelebihan dari metode ini adalah hasilnya mudah diinterpretasikan dan tidak terikat oleh asumsi. Ketepatan hasil pengklasifikasian pada CART dipengaruhi oleh keseimbangan pada data yang digunakan. Kelas data yang tidak seimbang (*imbalanced data*) dapat menurunkan nilai akurasi prediksi. Oleh karena itu,

pada penelitian ini diterapkan metode *Synthetic Minority Oversampling Technique* SMOTE untuk menangani masalah imbalanced data. Penerapan SMOTE terbukti berhasil mengatasi kasus data tidak seimbang.

Penelitian ini dilakukan untuk mengklasifikasikan rumah tangga rawan pangan di Provinsi Aceh. Variabel *input* yang akan digunakan berjumlah tujuh variabel, BPJS, PKH, BPNT, KKS, Jamkesda, Bantuan Pemda, dan PIP. Tujuan dari penelitian ini adalah untuk mengetahui karakteristik rumah tangga rawan pangan berdasarkan status penerimaan program perlindungan social di Provinsi Aceh. Hasil klasifikasi terbaik ditentukan berdasarkan nilai AUC. Variabel yang paling berpengaruh terhadap kejadian rawan pangan dilihat berdasarkan skor *permutation importance*.

2. TINJAUAN TEORITIS

Ketidakseimbangan data merupakan masalah yang sering ditemukan pada data. Data dikatakan tidak seimbang ketika salah satu kelas memiliki jumlah sampel yang lebih banyak dibandingkan kelas lainnya (Arifiyanti et al., 2020). Kelas dengan jumlah sampel yang lebih banyak disebut kelas mayor, sedangkan kelas dengan jumlah sampel yang lebih sedikit disebut dengan kelas minor. Penggunaan data yang tidak seimbang untuk membangun model akan memberikan pengaruh yang besar terhadap model tersebut. Oleh karena itu perlu dilakukan penyeimbangan kelas data. Secara umum, ketidakseimbangan kelas pada level data dapat diperbaiki dengan menggunakan teknik undersampling dan oversampling. Undersampling bekerja menyeimbangkan data dengan menghapus atau mengeliminasi sampel pada kelas mayor sehingga jumlah sampel pada kelas tersebut berkurang. Oversampling merupakan kebalikan dari metode undersampling. Metode oversampling bekerja dengan menambahkan jumlah sampel pada kelas minor sehingga kelas data menjadi seimbang.

SMOTE merupakan lanjutan dari teknik *oversampling*. Berbeda dengan *oversampling*, SMOTE bekerja dengan melakukan penambahan data pada data kelas minor secara acak sehingga menghasilkan jumlah data yang sama dengan data pada kelas mayor dengan cara replikasi atau membangkitkan data buatan (sintetis). Metode ini mampu mengurangi masalah *overfitting* dan tidak menyebabkan hilangnya informasi penting pada data. Metode yang digunakan pada algoritma SMOTE adalah *k-nearest neighbors* (*k*-tetangga terdekat) yang dipilih secara acak. Jumlah *k*-tetangga terdekat ditentukan dengan mempertimbangkan kemudahan dalam melaksanakannya. Pembangkitan data buatan yang berskala numerik berbeda dengan kategorik. Data numerik diukur jarak kedekatannya dengan jarak *Euclidean*

sedangkan data kategorik diukur menggunakan nilai modus (Akbar dan Hayati, 2020). Penyeimbangan data menggunakan metode SMOTE pada data kategorik dilakukan langkah sebagai berikut.

a. Menghitung jarak antar amatan pada kelas minor dengan menggunakan rumus *Value Difference Metric* dengan persamaan berikut.

$$\Delta(X, Y) = w_x w_y \sum_{i=1}^N \delta(X_i Y_i)^R, \dots\dots\dots (1)$$

dengan

- $\Delta(X, Y)$: jarak antara amatan X dan Y
- $w_x w_y$: bobot amatan (dapat diabaikan)
- N : banyaknya peubah penjelas
- R : bernilai 1 (jarak Manhattan) atau 2 (jarak *Euclidean*)
- $\delta(x_i, y_i)^r$: jarak antar kategori

Untuk mendapatkan nilai jarak antar kategori digunakan persamaan berikut.

$$\delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{c_{1i}}{c_1} - \frac{c_{2i}}{c_2} \right|^k, \dots\dots\dots (2)$$

dengan

- $\delta(V_1, V_2)$: jarak antar nilai V_1 dan V_2
- C_{1i} : banyaknya V_1 yang masuk kelas i
- C_{2i} : banyaknya V_2 yang masuk kelas i
- I : banyaknya kelas; $i = 1, 2, \dots, m$
- C_1 : banyaknya nilai 1 yang terjadi
- C_2 : banyaknya nilai 2 yang terjadi
- n : banyaknya kategori
- k : konstanta (biasanya 1)

- b. Menentukan nilai k dan persentase *over sampling*.
- c. Memilih satu sampel secara acak dari kelas minor.
- d. Menentukan amatan k tetangga terdekat pada kelas minor.
- e. Melakukan sintetis data.
- f. Mengulangi proses hingga banyaknya data yang ingin disintetis tercapai.

CART merupakan salah satu metode pengklasifikasian *machine learning* yang dilakukan dengan teknik pohon keputusan yang dikembangkan pertama kali oleh Breiman

pada awal tahun 1980-an. CART akan menghasilkan pohon klasifikasi (*classification tree*) jika variabel *output* yang digunakan berskala kategorik dan akan menghasilkan pohon regresi (*regression tree*) jika variabel *output* yang digunakan berskala numerik (Breiman at al., 1983). CART memiliki beberapa keunggulan salah satunya adalah tidak terikat oleh asumsi dan algoritma pembentukan pohonnya mudah dipahami. proses pembentukan pohon pada CART terdiri atas empat langkah dasar. Langkah pertama adalah pembentukan *root node* (simpul utama) dengan melakukan pemilihan pemilah, penentuan simpul terminal, dan pemberian label kelas. Langkah kedua terdiri penghentian proses pembentukan pohon. Langkah ketiga merupakan proses pemangkasan pohon (*pruning*) dengan melakukan pemotongan *node*. Langkah terakhir adalah penentuan pohon optimal. Fungsi yang digunakan untuk menentukan simpul induk dihitung berdasarkan indeks gini dengan persamaan sebagai berikut.

$$I(t) = \sum_{i,j=1} p(j | t)p(i | t), i \neq j, \dots\dots\dots (3)$$

dengan

$I(t)$: fungsi heterogenitas atau impuritas simpul t

$p(i | t)$: proporsi kelas i pada simpul t

$p(j | t)$: proporsi kelas j pada simpul t

Selanjutnya pemilah terbaik dipilih berdasarkan kriteria *goodness of split*. *Goodness of split* dengan fungsi persamaan sebagai berikut.

$$\phi(s, t) = \Delta I - p_L I(t_L) - p_R I(t_R), \dots\dots\dots (4)$$

dengan

$\phi(s, t)$: nilai *goodness of split*

$I(t)$: fungsi heterogenitas pada simpul t

p_L : proporsi pengamatan simpul kiri

p_R : proporsi pengamatan simpul kanan

$I(t_L)$: fungsi heterogenitas pada simpul anak kiri

$I(t_R)$: fungsi heterogenitas pada simpul anak kanan

Karakteristik rumah tangga rawan pangan diukur menggunakan skor *permutation importance*. *Permutation Importance* (PIMP) bekerja dengan cara menghitung rata-rata pengurangan impuritas yang dihasilkan oleh suatu variabel *input* pada pohon klasifikasi. PIMP merupakan tahapan untuk mengoreksi bias *gini importance*. PIMP bekerja dengan melakukan permutasi pada variabel *input* yang digunakan untuk membangun model (Altmann et al., 2010). Variabel yang berkontribusi besar akan sangat berpengaruh terhadap

performa model jika dirubah susunan datanya. Variabel yang berkontribusi kecil akan berpengaruh kecil terhadap performa model jika dirubah susunan datanya. Besarnya nilai kontribusi diukur menggunakan performa matriks ROC-AUC.

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data SUSENAS Provinsi Aceh tahun 2020. Data diperoleh dari Badan Pusat Statistik Provinsi Aceh. Data penelitian terdiri atas 12.971 amatan. Variabel yang digunakan dalam penelitian ini terdiri atas 1 variabel *output*, yaitu status rawan pangan dan 7 variabel *input*, yaitu PKH, KKS, BPNT, Bantuan Pemda, BPJS, Jamkesda, dan PIP.

4. HASIL DAN PEMBAHASAN

Tahapan awal dari analisis data adalah mengecek ketidakseimbangan data. Diketahui terdapat 7.304 rumah tangga dengan status tidak rawan dan sisanya sebanyak 1.775 rumah tangga dengan status rawan. Hal tersebut menunjukkan adanya ketidakseimbangan data. Peningkatan jumlah amatan pada kelas minor dilakukan dengan cara replikasi berdasarkan aturan k-tetangga terdekat menggunakan metode SMOTE. Karena data yang digunakan berskala nominal, maka fungsi komputasi yang diterapkan untuk menangani kelas data tidak seimbang adalah SMOTEN. Tabel 4.2 menunjukkan hasil sebelum dan setelah dilakukan penyeimbangan data.

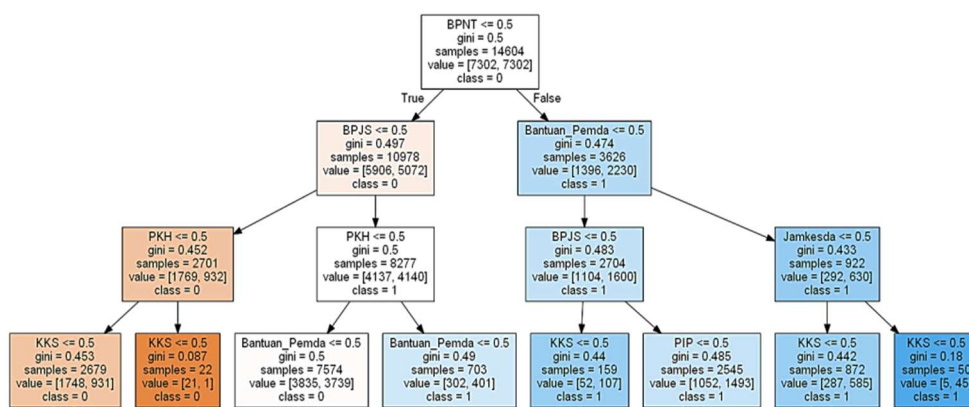
Tabel 1. Jumlah amatan sebelum dan setelah replikasi

Kelas Data	Jumlah Amatan	
	Sebelum	Setelah Replikasi
Mayor	1,10	1,15
Minor	1,05	1,10
Tinggi	1,00	1,05

Tabel 1 menunjukkan perbandingan jumlah amatan pada variabel output sebelum dan setelah dilakukan penyeimbangan data menggunakan fungsi SMOTEN. Jarak ditentukan berdasarkan nilai fitur yang overlap terhadap seluruh vektor fitur. Kemudian dipilih lima tetangga terdekat untuk menentukan sampel acak. Sampel acak yang terpilih pada kelas minor dilakukan replikasi sebanyak empat kali dari jumlah semula. Persentase oversampling yang dilakukan pada kelas minor adalah sebesar 400% sehingga jumlahnya

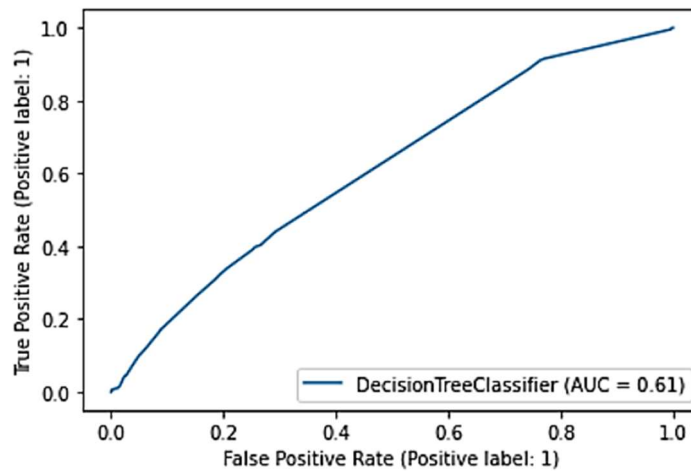
mendekati kelas mayor.

Setelah data seimbang, tahapan analisis selanjutnya adalah pembentukan pohon klasifikasi menggunakan metode CART. Pohon klasifikasi dibangun dengan menentukan *hyper-parameter*. Parameter-parameter tersebut berperan dalam mengatur kriteria penghentian pohon sehingga pohon klasifikasi yang dihasilkan optimal. Pemangkasan pohon bertujuan untuk mendapatkan pohon klasifikasi optimal tanpa menurunkan nilai akurasi prediksi. Pembangunan pohon klasifikasi dibatasi oleh kriteria nilai *min split* sebesar 3, nilai *min sample leaf* sebesar 2, dan nilai *max depth* sebesar 3. Berikut adalah visualisasi dari pohon klasifikasi yang terbentuk.



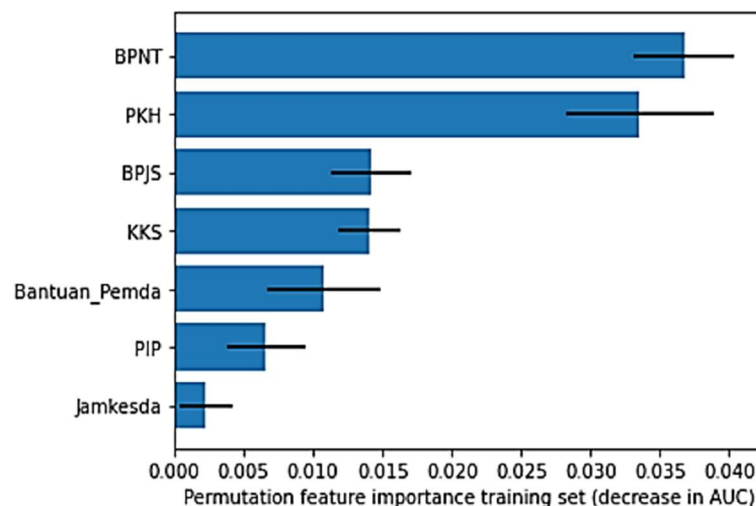
Gambar 1. Pohon klasifikasi

Selanjutnya dilakukan evaluasi performa pohon berdasarkan pohon klasifikasi yang terbentuk. Evaluasi pohon bertujuan untuk mengetahui seberapa akurat hasil klasifikasi. Suatu pohon klasifikasi dikatakan baik jika data testing menghasilkan nilai akurasi yang tinggi. Penelitian ini menggunakan nilai AUC sebagai ukuran evaluasi model. AUC merupakan nilai luas persegi di bawah kurva ROC. Kurva ROC pada Gambar 2 menunjukkan hasil bahwa nilai AUC dari pohon klasifikasi yang terbentuk adalah sebesar 0,61.



Gambar 2. Kurva ROC

Tahapan akhir dari analisis data adalah menentukan karakteristik rumah tangga rawan pangan berdasarkan skor PIMP. PIMP merupakan ukuran yang digunakan untuk mengetahui urutan kepentingan dari suatu variabel *input* yang digunakan pada penelitian untuk mengetahui kontribusi setiap variabel yang digunakan dalam pembentukan model. Karakteristik rawan pangan pada rumah tangga di Provinsi Aceh dapat dilihat berdasarkan nilai PIMP yang dihasilkan dari pohon klasifikasi. Urutan kepentingan variabel pada pohon klasifikasi ditunjukkan pada gambar berikut.



Gambar 3. Urutan kepentingan variabel

Gambar 3 menunjukkan urutan kepentingan variabel yang mendeskripsikan

karakteristik rawan pangan pada rumah tangga di Provinsi Aceh. Karakteristik rawan pangan pada rumah tangga di Provinsi Aceh secara berurutan dapat ditentukan berdasarkan status penerimaan BPNT, PKH, BPJS, KKS, Bantuan Pemda, PIP, dan Jamkesda. Semakin rendah urutan kepentingan variabel, maka semakin tidak ada pengaruh variabel tersebut terhadap pohon klasifikasi. BPNT dan PKH adalah dua variabel teratas dalam urutan kepentingan variabel. Hal ini menunjukkan variabel BPNT dan PKH memiliki pengaruh yang sangat besar terhadap pohon klasifikasi yang terbentuk.

5. KESIMPULAN DAN REKOMENDASI

Berdasarkan hasil analisis data diperoleh kesimpulan sebagai berikut:

1. Metode SMOTE terbukti mampu menyeimbangkan kelas data pada variabel status rawan pangan dengan persentase *oversampling* sebesar 400%.
2. Metode CART mampu mengklasifikasikan rumah tangga rawan pangan di Provinsi Aceh tahun 2020 dengan akurasi sebesar 0,61 berdasarkan nilai AUC.
3. Karakteristik rawan pangan pada rumah tangga di Provinsi Aceh secara berurutan dapat ditentukan berdasarkan status penerimaan BPNT, PKH, BPJS, KKS, Bantuan Pemda, PIP, dan Jamkesda. Variabel yang paling besar pengaruhnya dalam mengklasifikasikan rawan pangan pada rumah tangga di Provinsi Aceh tahun 2020 adalah BPNT dan PKH.

DAFTAR PUSTAKA

Badan Pusat Statistik. (2021). *Statistik Indonesia 2021*. Badan Pusat Statistik, Jakarta.

Supriyanto, R. W., Ramdhani, E. R., & Rahmadan, E. (2014). *Perlindungan Sosial di Indonesia : Tantangan dan Arah ke Depan*. Badan Perencanaan Pembangunan Nasional, Jakarta

Arifiyanti, Amalia Anjani & Wahyuni, Eka Dyar. (2020). SMOTE : Metode Penyeimbangan Kelas pada Klasifikasi Data Mining. *SCAN*, 15(1), 34-39.

Akbar, K. & Hayaty, M. (2020). Data Balancing untuk Mengatasi Imbalance Dataset pada Prediksi Produksi Padi. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 2(02), 1-14.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1983). *Classification and Regression Trees*. Taylor & Francis Group, United States of Americ

Altmann, A., Tolo, L., Sander, O., & Lengauer, T. (2010). Permutation importance : a corrected feature importance measure. *Bioinformatics Original Paper 2010*, 26(10), 1340-1347.