

## Komparasi Algoritma *Decision Tree* dan *Gradient Boosting* untuk Prediksi Penerima Bantuan Sosial di Desa Kalijati Barat

<sup>1</sup>Fitri Akmaliah, <sup>2</sup>Yuyun Umaidah, <sup>3</sup>Iqbal Maulana

<sup>1,2,3</sup>Program Studi Informatika, Universitas Singaperbangsa Karawang, Karawang

E-mail: <sup>1</sup>2210631170020@student.unsika.ac.id, <sup>2</sup>yuyun.umaidah@staff.unsika.ac.id,  
<sup>3</sup>iqbal.maulana@staff.unsika.ac.id

### ABSTRAK

Kemiskinan masih menjadi persoalan sosial utama di Indonesia yang menghambat pemerataan kesejahteraan masyarakat. Program *Bantuan Langsung Tunai Dana Desa (BLT-DD)* menjadi instrumen pemerintah dalam mengatasi kemiskinan, namun penentuan penerimanya masih dilakukan secara manual dan subjektif sehingga rawan ketidaktepatan sasaran. Penelitian ini membandingkan performa algoritma *Decision Tree* dan *Gradient Boosting* dalam memprediksi penerima bantuan sosial di Desa Kalijati Barat, Kabupaten Subang, menggunakan pendekatan *Knowledge Discovery in Databases (KDD)*. Data yang digunakan berupa *Buku Induk Penduduk (BIP)* Tahun 2019 dan data penerima *BLT-DD* Tahun 2025. Ketidakseimbangan kelas ekstrem (rasio  $\pm 101:1$ ) ditangani menggunakan *SMOTE*. Model dioptimasi dengan *GridSearchCV* dan *Stratified K-Fold Cross Validation* ( $k=5$ ) pada tiga skenario pembagian data. Hasil menunjukkan kedua algoritma menghasilkan *accuracy* tinggi (0,91–0,98), namun berdasarkan *recall* dan *F1-score*, *Decision Tree* lebih unggul pada skenario 80:20 (*recall*=0,3333; *F1-score*=0,1579), sedangkan *Gradient Boosting* gagal mendeteksi kelas minoritas pada skenario 90:10. Analisis *feature importance* menunjukkan tingkat pendidikan kepala keluarga, jenis pekerjaan, dan rasio tanggungan sebagai faktor paling berpengaruh.

**Kata kunci :** *Bantuan Sosial, Decision Tree, Gradient Boosting, KDD, Machine Learning, SMOTE*

### ABSTRACT

Poverty remains a major social issue in Indonesia that hinders the equitable distribution of public welfare. The Village Fund Direct Cash Assistance (BLT-DD) program has become a government instrument to address poverty; however, the recipient selection process is still conducted manually and subjectively, leading to potential targeting inaccuracies. This study compares the performance of the Decision Tree and Gradient Boosting algorithms in predicting social assistance recipients in Kalijati Barat Village, Subang Regency, using the Knowledge Discovery in Databases (KDD) approach. The data used consist of the Population Master Book (BIP) of 2019 and BLT-DD recipient data of 2025. Extreme class imbalance ( $\pm 101:1$  ratio) was addressed using SMOTE. The models were optimized using GridSearchCV and Stratified K-Fold Cross Validation ( $k=5$ ) across three data-splitting scenarios. The results show that both algorithms achieved high accuracy (0.91–0.98). However, based on recall and F1-score, Decision Tree performed better in the 80:20 scenario (recall=0.3333; F1-score=0.1579), while Gradient Boosting failed to detect the minority class in the 90:10 scenario. Feature importance analysis indicates that the head of household's education level, occupation type, and dependency ratio are the most influential factors.

**Keyword :** *Social Assistance, Decision Tree, Gradient Boosting, KDD, Machine Learning, SMOTE*

## 1. PENDAHULUAN

Kemiskinan masih menjadi salah satu persoalan sosial utama di Indonesia dan menjadi faktor penghambat peningkatan kesejahteraan masyarakat. Berdasarkan data *Badan Pusat Statistik (BPS)*, persentase penduduk miskin nasional per Maret 2025 berada pada angka 8,47%. Pemerintah meluncurkan berbagai program bantuan sosial seperti *Program Keluarga Harapan (PKH)*, *Bantuan Pangan Non-Tunai (BPNT)*, serta *Bantuan Langsung Tunai Dana Desa (BLT-DD)* yang ditujukan kepada masyarakat miskin dan rentan (BPS, 2025; Kemenko Perekonomian, 2024).

Keberhasilan program bantuan sosial sangat bergantung pada ketepatan sasaran penerima bantuan. Namun, pada praktiknya proses penentuan penerima bantuan di banyak daerah masih dilakukan secara manual dan subjektif sehingga rawan ketidaktepatan sasaran (Qadrini et al., 2021; BPS Jawa Barat, 2025). Permasalahan serupa juga terjadi di Desa Kalijati Barat, Kabupaten Subang, di mana tingkat kemiskinan meningkat dari 8,12% pada tahun 2019 menjadi 9,52% pada tahun 2023 (BPS, 2025).

Perkembangan teknologi *machine learning* memberikan peluang untuk membantu proses klasifikasi penerima bantuan sosial berdasarkan karakteristik sosial dan kependudukan masyarakat. *Machine learning* mampu memproses data dalam jumlah besar dan menghasilkan model prediksi yang lebih objektif dan konsisten dibandingkan proses manual (Firmansyah & Abidin, 2022; Khan et al., 2024).

Pemilihan kedua algoritma tersebut didasarkan pada perbedaan pendekatan yang digunakan. *Decision Tree* bekerja sebagai *single learner* dengan membangun satu pohon keputusan utama, sedangkan *Gradient Boosting* bekerja sebagai *ensemble learner* yang mengombinasikan banyak pohon keputusan kecil secara bertahap

(Ramadhon et al., 2024; Khan et al., 2024). Perbandingan kedua algoritma menjadi penting karena karakteristik data bantuan sosial pada tingkat desa cenderung memiliki distribusi kelas yang sangat tidak seimbang antara penerima dan non-penerima bantuan.

Penelitian sebelumnya oleh Qadrini et al. menerapkan algoritma *Decision Tree* dan *AdaBoost* untuk klasifikasi penerima bantuan sosial, namun belum membandingkannya dengan metode *Gradient Boosting* modern (Qadrini et al., 2021; Nugraha & Pradana, 2022). Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan membandingkan langsung kedua algoritma, dilengkapi *hyperparameter tuning* menggunakan *GridSearchCV* dan penanganan ketidakseimbangan kelas menggunakan *SMOTE* (Almarogi et al., 2025).

## 2. LANDASAN TEORI

### 2.1 Data Mining

*Data mining* merupakan proses penggalian informasi, pola, dan pengetahuan baru dari kumpulan data berukuran besar menggunakan teknik statistik, *machine learning*, dan kecerdasan buatan (Senubekti, 2022). *Data mining* merupakan komponen utama dari proses *Knowledge Discovery in Databases (KDD)* yang mengubah data mentah menjadi pengetahuan bernilai (Nugraha & Pradana, 2022). Dalam bidang sosial, *data mining* banyak digunakan untuk menyelesaikan permasalahan klasifikasi dan prediksi, termasuk dalam penentuan penerima bantuan sosial berbasis data kependudukan (Fauziah & Hernadi, 2025).

### 2.2 Decision Tree (CART)

*Decision Tree* merupakan salah satu algoritma klasifikasi yang membangun model prediksi berbentuk struktur pohon keputusan (Maulana et al., 2023). Algoritma ini bekerja dengan membagi

data ke dalam beberapa subset berdasarkan atribut paling optimal sehingga menghasilkan node keputusan dan *leaf* yang merepresentasikan kelas output (Ramadhani & Kurniawan, 2022).

Pada penelitian ini digunakan algoritma *CART* (*Classification and Regression Tree*). *CART* membangun pohon keputusan biner menggunakan kriteria *Gini Impurity* untuk menentukan atribut pemisah terbaik pada setiap node (Hidayat et al., 2023). Rumus Gini Impurity dinyatakan sebagai berikut:

$$Gini(t) = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

dengan:

- $p_i$  = proporsi data pada kelas ke- $i$
- $k$  = jumlah kelas

Proses pemilihan *split* terbaik dilakukan dengan meminimalkan nilai *impurity* setelah pemisahan data:

$$Gini_{split} = \sum_{j=1}^n \frac{n_j}{n} \times Gini(t_j) \quad (2)$$

dengan:

- $n_j$  = jumlah data pada cabang ke- $j$
- $n$  = total data pada *node* induk
- $t_j$  = *node* hasil *split*

*Decision Tree* cocok digunakan pada kasus klasifikasi bantuan sosial karena mampu menghasilkan aturan keputusan yang transparan dan mudah diinterpretasikan (Pratama, 2024).

### 2.3 Gradient Boosting

*Gradient Boosting* merupakan metode *ensemble learning* yang membangun model prediksi secara bertahap dengan menggabungkan beberapa *weak learner*, biasanya berupa pohon keputusan kecil (Rahmayani et al., 2024). Setiap model baru dibangun untuk memperbaiki kesalahan residual dari model sebelumnya menggunakan fungsi *loss gradient*. Model *Gradient Boosting* dapat dinyatakan sebagai:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \quad (3)$$

dengan:

- $F_m(x)$  = model pada iterasi ke- $m$
- $F_{m-1}(x)$  = model sebelumnya

- $h_m(x)$  = *weak learner* pada iterasi ke- $m$
- $\eta$  = learning rate

*Gradient Boosting* memiliki kemampuan menghasilkan *accuracy* dan *ROC-AUC* yang tinggi, terutama pada data dengan pola kompleks (Boldini et al., 2022). Namun, metode ini memiliki tingkat interpretabilitas yang lebih rendah dan lebih sensitif terhadap ketidakseimbangan kelas dibandingkan *Decision Tree* (Kurniawan, 2020).

### 2.4 SMOTE

*SMOTE* (*Synthetic Minority Oversampling Technique*) merupakan metode *oversampling* yang digunakan untuk menangani permasalahan ketidakseimbangan kelas (*imbalanced dataset*) (Aisyah & Amin, 2021). *SMOTE* bekerja dengan menghasilkan sampel sintesis baru dari kelas minoritas berdasarkan interpolasi antar tetangga terdekat.

Pada penelitian ini, *SMOTE* diterapkan pada data *training* dalam *pipeline machine learning* untuk meningkatkan kemampuan model dalam mendeteksi kelas minoritas, yaitu penerima bantuan sosial. Integrasi *SMOTE* dalam *pipeline* memastikan tidak terjadi *data leakage* pada data *testing* (Hatu, 2024).

### 2.5 Cross Validation dan Stratified K-Fold

*Cross Validation* merupakan teknik evaluasi model yang dilakukan dengan membagi *dataset* ke dalam beberapa subset untuk memperoleh hasil evaluasi yang lebih stabil dan objektif (Universitas Padjadjaran, 2022). Pada penelitian ini digunakan *Stratified K-Fold Cross Validation*. Metode ini mempertahankan proporsi distribusi kelas pada setiap *fold* sehingga cocok digunakan pada *dataset* yang tidak seimbang (Aggarwal, 2023).

### 2.6 Knowledge Discovery in Database (KDD)

*Knowledge Discovery in Database* (KDD) merupakan proses sistematis

untuk menemukan pola dan pengetahuan baru dari data dalam jumlah besar. KDD terdiri dari beberapa tahapan utama, yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*.

Pada penelitian ini, metode KDD digunakan sebagai kerangka kerja penelitian untuk membangun model prediksi penerima bantuan sosial. Tahapan KDD dimulai dari pemilihan data kependudukan dan data penerima BLT-DD, dilanjutkan dengan *preprocessing* dan transformasi data, penerapan algoritma klasifikasi, hingga evaluasi performa model

### 2.7 Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa algoritma klasifikasi dalam memprediksi penerima bantuan sosial. Pada penelitian ini digunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*. Pada kondisi *dataset* tidak seimbang, *recall* dan *F1-score* lebih representatif karena mampu mengukur kemampuan model dalam mendeteksi kelas minoritas (Luque et al., 2022).

*Accuracy* merupakan metrik evaluasi yang digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data (Lobo et al., 2022).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

dengan:

- TP = *True Positive*
- TN = *True Negative*
- FP = *False Positive*
- FN = *False Negative*

*Precision* digunakan untuk mengukur tingkat ketepatan model dalam memprediksi kelas positif (Lobo et al., 2022).

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

*Recall* digunakan untuk mengukur kemampuan model dalam mendeteksi

seluruh data positif yang sebenarnya (Lumbaa, 2022).

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

*F1-Score* merupakan *harmonic mean* antara *precision* dan *recall* (Lumbaa, 2022).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

*ROC-AUC* digunakan untuk mengukur kemampuan model dalam membedakan kelas positif dan negatif, serta banyak digunakan pada kasus klasifikasi dengan distribusi kelas tidak seimbang (Sokolova & Lapalme, 2021; Powers, 2020).

## 3. METODE PENELITIAN

### 3.1 Sumber Data

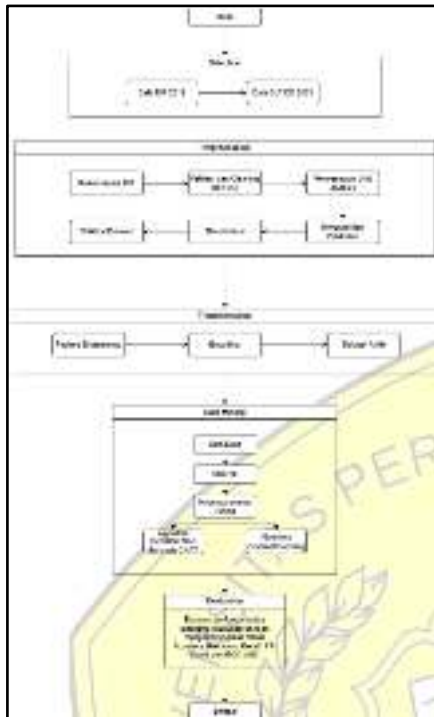
Penelitian ini menggunakan data sekunder yang diperoleh dari Kantor Desa Kalijati Barat, Kecamatan Kalijati, Kabupaten Subang. Data yang digunakan terdiri dari *Buku Induk Penduduk (BIP)* Tahun 2019 dengan total 4.542 kartu keluarga dan data penerima *BLT Dana Desa* Tahun 2025 yang berisi 44 penerima bantuan.

Variabel yang digunakan meliputi jenis kelamin kepala keluarga, tingkat pendidikan kepala keluarga, jenis pekerjaan kepala keluarga, jumlah anggota keluarga, jumlah tanggungan, dan rasio tanggungan yang merupakan hasil *feature engineering*.

### 3.2 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Databases (KDD)*. *KDD* merupakan metode pengolahan data yang digunakan untuk menemukan pola dan pengetahuan baru dari data dalam jumlah besar (Chicco & Jurman, 2020). Tahapan *KDD* yang diterapkan meliputi *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Alur dari

rancangan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Rancangan penelitian

## 4. HASIL DAN PEMBAHASAN

### 4.1 Data Selection

Tahapan *data selection* dilakukan dengan memilih dan mengintegrasikan data *Buku Induk Penduduk (BIP)* Tahun 2019 dan data penerima *BLT-DD* Tahun 2025 dari Kantor Desa Kalijati Barat. *Dataset BIP* memiliki 18.474 baris awal yang direduksi menjadi 13.932 individu setelah penghapusan baris header administrasi kependudukan.

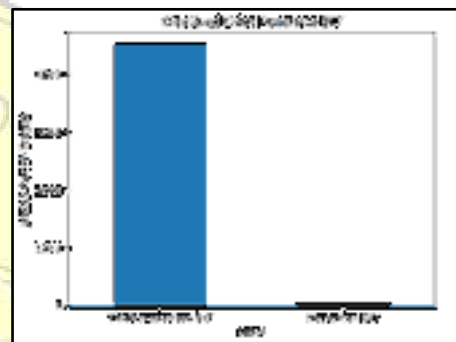
Variabel yang digunakan pada penelitian ini meliputi jenis kelamin kepala keluarga, tingkat pendidikan kepala keluarga, jenis pekerjaan kepala keluarga, jumlah anggota keluarga, jumlah tanggungan, dan label target berupa status penerima *BLT-DD*.

### 4.2 Data Preprocessing

Tahapan *preprocessing* dilakukan untuk membersihkan dan menyesuaikan data sebelum proses pemodelan. Proses

*preprocessing* meliputi penghapusan header administrasi, normalisasi format NIK, penyesuaian status *SHDK*, dan integrasi data *BLT-DD* dengan *BIP* melalui pencocokan NIK untuk menghasilkan label kelas.

Hasil *preprocessing* menunjukkan bahwa distribusi kelas *dataset* sangat tidak seimbang. Dari total 4.542 KK, hanya 44 KK (0,97%) yang merupakan penerima *BLT-DD*, sementara 4.498 KK (99,03%) bukan penerima.



Gambar 2. Distribusi kelas penerima BLT

Distribusi kelas yang sangat tidak seimbang menyebabkan model berpotensi bias terhadap kelas mayoritas. Oleh karena itu, penelitian ini menerapkan metode *SMOTE* untuk menyeimbangkan distribusi kelas pada data *training*.

### 4.3 Data Transformation

Pada tahap *transformation* dilakukan proses *feature engineering* dan *encoding* pada variabel kategorikal. *Feature engineering* dilakukan dengan membentuk fitur baru berupa rasio tanggungan, yaitu perbandingan antara jumlah tanggungan dengan total anggota keluarga.

Selain itu, dilakukan proses *encoding* menggunakan *ordinal encoding* pada variabel pendidikan, *binary encoding* pada jenis kelamin, dan *one-hot encoding* pada variabel pekerjaan. Setelah proses *encoding*, jumlah fitur meningkat dari 6 menjadi 25 fitur.

#### 4.4 Data Mining

Tahap *data mining* dilakukan menggunakan algoritma *Decision Tree* dan *Gradient Boosting* dengan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. *SMOTE* diterapkan dalam *pipeline machine learning* dan *hyperparameter tuning* dilakukan menggunakan *GridSearchCV* dengan *Stratified K-Fold Cross Validation* ( $k=5$ ).

Hasil optimasi parameter terbaik untuk algoritma *Decision Tree* ditunjukkan pada Tabel 1.

Tabel 1. Parameter terbaik algoritma *Decision Tree*

Skenario	Max_depth	Min_samples_split	Min_samples_leaf
70:30	5	2	1
80:20	10	2	1
90:10	10	4	1

Sementara itu, parameter terbaik algoritma *Gradient Boosting* ditunjukkan pada Tabel 2.

Tabel 2. Parameter terbaik algoritma *Gradient Boosting*

Skenario	Learning rate	Max_depth	N_estimators	Subsample
70:30	0,1	3	200	1,0
80:20	0,1	5	100	1,0
90:10	0,05	5	200	1,0

#### 4.5 Evaluation

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*. Hasil perbandingan performa kedua algoritma ditunjukkan pada Tabel 3.

Tabel 3. Hasil perbandingan performa algoritma *Decision Tree* dan *Gradient Boosting*

Algoritma	Skenario	Accuracy	Precision	Recall	F1-Score	ROC-AUC
DT	70:30	0.92	0.04	0.31	0.07	0.64
	80:20	0.96	0.10	0.33	0.16	0.65
	90:10	0.97	0.10	0.25	0.14	0.77
GB	70:30	0.98	0.12	0.23	0.15	0.70
	80:20	0.97	0.09	0.22	0.13	0.77
	90:10	0.99	0.00	0.00	0.00	0.70

Berdasarkan hasil evaluasi, kedua algoritma menghasilkan *accuracy* yang tinggi pada rentang 0,91–0,98. Namun, tingginya nilai *accuracy* tidak dapat dijadikan satu-satunya acuan evaluasi pada *dataset* yang sangat tidak seimbang

karena model dapat mencapai *accuracy* tinggi dengan hanya memprediksi seluruh data sebagai kelas mayoritas.

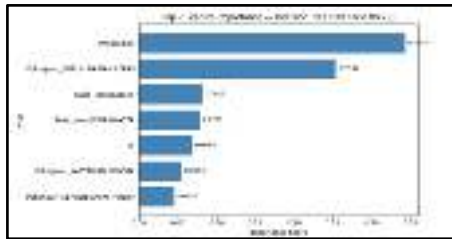
Algoritma *Decision Tree* menghasilkan performa terbaik pada skenario 80:20 dengan *recall* sebesar 0,3333 dan *F1-score* sebesar 0,1579. Sementara itu, *Gradient Boosting* mengalami penurunan performa yang signifikan pada skenario 90:10, di mana model tidak mampu mendeteksi kelas minoritas sama sekali (*recall* = 0,000).

#### 4.6 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* lebih efektif dalam mendeteksi kelas minoritas dibandingkan *Gradient Boosting* pada *dataset* dengan ketidakseimbangan kelas yang sangat ekstrem. Keunggulan *Decision Tree* pada kondisi ini dipengaruhi oleh karakteristik algoritmanya yang lebih sederhana dan tidak terlalu sensitif terhadap ketidakseimbangan kelas dibandingkan metode *ensemble* yang kompleks.

Meskipun *Gradient Boosting* menghasilkan *accuracy* dan *ROC-AUC* yang lebih tinggi, model cenderung bias terhadap kelas mayoritas sehingga kurang optimal untuk kasus deteksi penerima bantuan sosial yang membutuhkan tingkat *recall* yang tinggi.

Selain itu, analisis *feature importance* menunjukkan bahwa tingkat pendidikan kepala keluarga, jenis pekerjaan, dan rasio tanggungan menjadi faktor paling berpengaruh dalam penentuan kelayakan penerima bantuan sosial. Hasil ini konsisten dengan indikator kemiskinan yang umum digunakan dalam kebijakan sosial di Indonesia.



Gambar 3. *Feature Importance* algoritma *Decision Tree* (80:20)

Hasil penelitian ini menunjukkan bahwa pendekatan *machine learning* dapat dimanfaatkan sebagai sistem pendukung keputusan berbasis data dalam proses seleksi penerima bantuan sosial di tingkat desa, sehingga dapat mengurangi subjektivitas dan meningkatkan ketepatan sasaran program bantuan.

## 5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Decision Tree* dan *Gradient Boosting* berhasil diterapkan untuk prediksi penerima bantuan sosial menggunakan pendekatan Knowledge Discovery in Databases (KDD) pada data kependudukan Desa Kalijati Barat. Hasil evaluasi menunjukkan bahwa kedua algoritma menghasilkan nilai *accuracy* yang tinggi pada seluruh skenario pengujian. Namun, karena *dataset* memiliki distribusi kelas yang sangat tidak seimbang, *recall* dan *F1-score* digunakan sebagai metrik utama dalam mengevaluasi performa model. Algoritma *Decision Tree* menunjukkan performa terbaik pada skenario pembagian data 80:20 dengan nilai *recall* sebesar 0,3333 dan *F1-score* sebesar 0,1579. Sementara itu, *Gradient Boosting* menghasilkan nilai *accuracy* dan *ROC-AUC* yang lebih stabil, namun kurang optimal dalam mendeteksi kelas minoritas pada skenario tertentu. Hasil analisis *feature importance* menunjukkan bahwa tingkat pendidikan kepala keluarga, jenis pekerjaan, dan rasio tanggungan menjadi faktor paling berpengaruh dalam menentukan status penerima bantuan sosial. Berdasarkan

hasil tersebut, algoritma *Decision Tree* dinilai lebih sesuai digunakan pada kasus prediksi penerima bantuan sosial karena memiliki kemampuan deteksi kelas minoritas yang lebih baik serta menghasilkan aturan keputusan yang mudah diinterpretasikan.

Penelitian selanjutnya disarankan menggunakan data kependudukan yang lebih mutakhir dan menambahkan variabel sosial-ekonomi lain seperti pendapatan, kondisi rumah, dan kepemilikan aset agar hasil prediksi lebih akurat. Selain itu, penelitian berikutnya dapat membandingkan algoritma lain seperti *Random Forest*, *XGBoost*, atau *LightGBM* untuk memperoleh performa klasifikasi yang lebih optimal pada *dataset* tidak seimbang. Pengembangan metode penanganan *imbalanced dataset* seperti *ADASYN* atau *cost-sensitive learning* juga dapat dilakukan untuk meningkatkan kemampuan model dalam mendeteksi kelas minoritas. Model hasil penelitian ini juga berpotensi dikembangkan menjadi sistem pendukung keputusan berbasis web guna membantu pemerintah desa dalam proses seleksi penerima bantuan sosial secara lebih objektif dan berbasis data.

## 6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kantor Desa Kalijati Barat, Kecamatan Kalijati, Kabupaten Subang yang telah memberikan izin dan menyediakan data kependudukan untuk keperluan penelitian ini. Penulis juga menyampaikan apresiasi kepada seluruh pihak yang telah memberikan dukungan, bantuan, dan kontribusi selama proses penelitian hingga penyusunan penelitian ini dapat diselesaikan dengan baik.

## DAFTAR PUSTAKA

- BPS. (2025). Profil kemiskinan di Indonesia Maret 2025. Badan Pusat Statistik.
- Kementerian Koordinator Bidang Perekonomian RI. (2024). Program bantuan sosial untuk pengentasan kemiskinan. Kemenko Perekonomian RI.
- Qadrini, M., Mustafa, S., Amir, A., & Mattjik, A. (2021). Klasifikasi penerima bantuan sosial menggunakan *Decision Tree* dan *AdaBoost*. *Jurnal Teknologi Informasi*, 8(2), 120–128.
- BPS Jawa Barat. (2025). Persentase penduduk miskin Kabupaten Subang tahun 2019–2024. Badan Pusat Statistik Provinsi Jawa Barat.
- Ramadhon, R., et al. (2024). Implementasi *Decision Tree* pada klasifikasi data sosial. *Jurnal Informatika*, 11(1), 45–52.
- Firmansyah, A., & Abidin, M. (2022). Perbandingan *Decision Tree* dan *Gradient Boosting* pada data klasifikasi. *Jurnal Informatika*, 10(1), 33–41.
- Khan, S., Ahmad, S., & Ahmad, M. (2024). Evaluation metrics for imbalanced classification problems. *International Journal of Data Science*, 7(2), 55–64.
- Almarogi, A., Izzatillah, M., Lestari, M., & Timur, J. (2025). *Gradient Boosting* for social assistance prediction. *International Journal of Data Science*, 5(1), 15–24.
- Nugraha, A., & Pradana, R. (2022). Penerapan *data mining* pada analisis data sosial. *Jurnal Sistem Informasi*, 9(2), 87–95.
- Senubekti, M. A. (2022). Tahapan *data mining* dalam pengolahan data. *Jurnal Teknologi Informasi*, 7(1), 12–19.
- Fauziah, A., & Hernadi, J. (2025). Klasifikasi data tak seimbang menggunakan algoritma *Random Forest* dengan *SMOTE* dan *SMOTE-ENN*. *Teknomatika: Jurnal Informatika dan Komputer*, 17(2), 38–47.
- Maulana, R., Nugroho, A., & Prasetyo, D. (2023). Implementasi *machine learning* untuk klasifikasi penerima bantuan sosial. *Jurnal Teknologi Informasi dan Komputer*, 9(2), 88–97.
- Ramadhani, F., & Kurniawan, A. (2022). Penerapan algoritma *CART* pada klasifikasi data kependudukan. *Jurnal Informatika*, 10(1), 41–49.
- Hidayat, M., Sari, N., & Putra, R. (2023). Analisis performa *Decision Tree* pada klasifikasi data tidak seimbang. *Jurnal Sistem Komputer dan Informatika*, 11(3), 112–120.
- Pratama, W. (2024). Optimasi *hyperparameter* menggunakan *GridSearchCV*. *Jurnal Machine Learning Indonesia*, 3(1), 44–52.
- Rahmayani, R., et al. (2024). Pengaruh *hyperparameter tuning* terhadap performa model klasifikasi. *Jurnal Data Science*, 4(2), 71–79.
- Boldini, D., Friedrich, L., Kuhn, D., & Sieber, S. A. (2022). *Gradient Boosting* on imbalanced bioassay modelling with custom loss functions. *International Journal of Artificial Intelligence*, 12(3), 88–97.
- Kurniawan, A. (2020). Bantuan sosial sebagai instrumen pengentasan

- kemiskinan. *Jurnal Sosial dan Humaniora*, 5(2), 31–39.
- Aisyah, N., & Amin, M. (2021). Evaluasi penyaluran *BLT Dana Desa*. *Jurnal Administrasi Publik*, 6(1), 15–24.
- Hatu, R. (2024). Pemanfaatan *Buku Induk Penduduk* dalam administrasi desa. *Jurnal Pemerintahan Desa*, 4(2), 55–63.
- Universitas Padjadjaran. (2022). Digitalisasi data kependudukan desa. Universitas Padjadjaran.
- Aggarwal, C. C. (2023). *Data mining: The textbook* (2nd ed.). Springer International Publishing.
- Chicco, D., & Jurman, G. (2020). The advantages of the *Matthews correlation coefficient (MCC)* over *F1 score* and *accuracy* in binary classification evaluation. *BMC Genomics*, 21(6), 1–13.
- Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2022). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition Letters*, 116, 3–12.
- Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2022). *AUC*: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 21(4), 124–133.
- Lumbaa, S. T. (2022). Klasifikasi penerima bantuan sosial menggunakan *SVM* dan *Gradient Boost*. *Jurnal Komputasi*, 9(1), 41–50.
- Sokolova, M., & Lapalme, G. (2021). Systematic analysis of multi-label classification performance metrics. *Information Processing & Management*, 58(3), 1–15.
- Powers, D. M. W. (2020). Evaluation: From *precision, recall* and *F-factor* to *ROC, informedness, markedness* and *correlation*. *arXiv:2010.16061*.