

Klasifikasi Penerima Kartu Keluarga Sejahtera (KKS) di Kota Bandung Tahun 2024 dengan Menggunakan Metode *Categorical Boosting (CatBoost)*

Sabrina Naura Saputra *, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

sabrinasaputra18@email.com, abdul.kudus@unisba.ac.id

Abstract. The welfare of the Indonesian people, including in Bandung City, still faces serious challenges, primarily due to poverty. According to data from the Central Statistics Agency (BPS) as of March 2024, the number of poor people in Bandung City reached 101,100 people, or approximately 3.87 percent. The government responded to this condition through various social assistance programs, one of which is the Prosperous Family Card (KKS), which aims to ease the burden on low-income families. This study aims to classify KKS recipients in Bandung City in 2024 using the Categorical Boosting (CatBoost) algorithm, taking into account socio-economic variables. The data used comes from the 2024 SUSENAS, obtained through the West Java Province BPS. The modeling process includes pre-processing, initial training, hyperparameter tuning with grid search, and determining the optimal threshold using the Youden's J approach on the ROC curve. The results show that CatBoost delivers high performance with an initial accuracy of 87% and an AUC of 0.9378, increasing to 88% and an AUC of 0.9400 after tuning and selecting the optimal threshold. The model also shows a balanced F1-score across both classes. A total of 183 households were identified as KKS recipients based on the optimal threshold of 0.6819, generally having female heads of household, aged 50–59, divorced, without mobile phone ownership, and a small number of family members. This model is proven to be effective for fair and targeted predictions.

Keywords: *Classification, Prosperous Family Card, Catboost.*

Abstrak. Kesejahteraan masyarakat Indonesia, termasuk di Kota Bandung, masih menghadapi tantangan serius, terutama akibat kemiskinan. Berdasarkan data Badan Pusat Statistik (BPS) per Maret 2024, jumlah penduduk miskin di Kota Bandung mencapai 101.10 ribu jiwa atau sekitar 3.87 persen. Pemerintah merespons kondisi ini melalui berbagai program bantuan sosial, salah satunya Kartu Keluarga Sejahtera (KKS) yang ditujukan untuk meringankan beban keluarga kurang mampu. Penelitian ini bertujuan untuk mengklasifikasikan penerima KKS di Kota Bandung tahun 2024 menggunakan algoritma Categorical Boosting (CatBoost), dengan mempertimbangkan variabel sosial-ekonomi. Data yang digunakan berasal dari SUSENAS 2024, yang diperoleh melalui BPS Provinsi Jawa Barat. Proses pemodelan mencakup tahapan pre-processing, pelatihan awal, tuning hyperparameter dengan grid search, serta penentuan threshold optimal menggunakan pendekatan Youden's J pada ROC curve. Hasil menunjukkan bahwa CatBoost memberikan performa tinggi dengan akurasi awal 87% dan AUC 0.9378, meningkat menjadi 88% dan AUC 0.9400 setelah tuning dan pemilihan threshold optimal. Model juga menunjukkan F1-score seimbang pada kedua kelas. Sebanyak 183 rumah tangga teridentifikasi sebagai penerima KKS berdasarkan threshold optimal 0.6819, umumnya memiliki karakteristik kepala keluarga perempuan, berusia 50–59 tahun, cerai mati, tanpa kepemilikan HP, dan jumlah anggota keluarga sedikit. Model ini terbukti efektif untuk prediksi yang adil dan tepat sasaran.

Kata Kunci: *Klasifikasi, Kartu Keluarga Sejahtera, CatBoost.*

A. Pendahuluan

Kesejahteraan masyarakat merupakan kondisi ketika kebutuhan dasar seperti pangan, sandang, papan, pendidikan, dan kesehatan terpenuhi, serta individu mampu berpartisipasi secara produktif dalam kehidupan sosial. Dalam Undang-Undang No. 11 Tahun 2009 disebutkan bahwa kesejahteraan sosial adalah kondisi terpenuhinya kebutuhan material, spiritual, dan sosial warga negara untuk dapat hidup layak, mengembangkan diri, serta melaksanakan fungsi sosialnya. Meskipun berbagai upaya telah dilakukan, kesejahteraan masyarakat Indonesia masih menghadapi tantangan besar, terutama terkait permasalahan sosial seperti kemiskinan, pendidikan rendah, pengangguran, dan gizi buruk (Sugandi, 2016).

Kemiskinan menjadi salah satu persoalan utama yang dihadapi oleh hampir seluruh negara berkembang, termasuk Indonesia. (Dulkiah & Nurjanah, 2018) menjelaskan bahwa kemiskinan bukan hanya persoalan keterbatasan pendapatan, tetapi juga mencerminkan ketidakmampuan individu atau rumah tangga dalam memenuhi kebutuhan dasarnya. Di Kota Bandung, masalah ini masih cukup signifikan. Berdasarkan data Badan Pusat Statistik (BPS, 2024a) jumlah penduduk miskin di Kota Bandung per Maret 2024 tercatat sebanyak 101,10 ribu orang atau sekitar 3,87% dari total penduduk. Selain itu, garis kemiskinan di kota ini meningkat dari Rp591.124,- pada Maret 2023 menjadi Rp614.707,- per kapita per bulan pada Maret 2024.

Untuk menanggulangi masalah ini, pemerintah telah meluncurkan berbagai program bantuan sosial. Salah satunya adalah Kartu Keluarga Sejahtera (KKS) yang ditujukan kepada masyarakat miskin dan rentan. Program ini memberikan akses kepada bantuan lain seperti Program Keluarga Harapan (PKH) dan Bantuan Pangan Non-Tunai (BPNT), dengan harapan dapat meningkatkan kesejahteraan dan kemampuan ekonomi masyarakat (Holifah, 2022). Namun, dalam implementasinya, ditemukan kendala utama yaitu ketidaktepatan sasaran penerima bantuan. Banyak keluarga yang layak menerima tidak tercakup dalam program, sementara keluarga yang kurang memenuhi kriteria justru menerima bantuan. Hal ini disebabkan oleh keterbatasan data, perubahan kondisi ekonomi penerima, dan kelemahan dalam sistem verifikasi.

Menurut (Holifah, 2022) kemiskinan dipengaruhi oleh beberapa faktor utama, antara lain: (1) rendahnya tingkat pendidikan yang menyebabkan keterbatasan keterampilan di pasar kerja, (2) kurangnya motivasi kerja yang berdampak pada rendahnya produktivitas dan kesejahteraan, (3) terbatasnya ketersediaan lapangan pekerjaan yang meningkatkan angka pengangguran, serta (4) tingginya beban tanggungan keluarga yang menyulitkan pemenuhan kebutuhan dasar. Seseorang dikategorikan miskin apabila pendapatan per kapita bulannya berada di bawah garis kemiskinan dan tidak mencukupi untuk kebutuhan hidup sehari-hari.

Dalam penentuan penerima bantuan sosial seperti Kartu Keluarga Sejahtera (KKS), penting untuk mempertimbangkan berbagai faktor sosial-ekonomi agar bantuan tepat sasaran. Karakteristik rumah tangga miskin seperti jumlah anggota keluarga, status perkawinan, jenis kelamin, usia, tingkat pendidikan kepala keluarga, serta kepemilikan telepon seluler menjadi indikator penting dalam proses klasifikasi. Jumlah anggota keluarga berpengaruh terhadap beban ekonomi rumah tangga, semakin banyak anggota, semakin besar risiko kemiskinan, terutama bila banyak anggota belum bekerja (Hutahaean & Sitorus, 2022). Status perkawinan juga berpengaruh, kepala keluarga yang menikah umumnya memiliki pengeluaran lebih besar dan lebih rentan terhadap kemiskinan.

Faktor usia turut menentukan kondisi ekonomi, usia tua mungkin membawa pengalaman, namun juga tantangan dalam memperoleh pekerjaan. Sebaliknya, kepala keluarga yang terlalu muda bisa kekurangan keterampilan dan pengalaman. Tingkat pendidikan menjadi penentu penting lainnya, di mana pendidikan rendah berkorelasi kuat dengan kemiskinan (Sa'diyah & Arianti, 2012). Selain itu, kepemilikan telepon seluler menjadi indikator kesejahteraan yang relevan dalam era digital (Rosiana, 2021). HP dapat membuka akses terhadap informasi dan peluang ekonomi, serta mencerminkan tingkat akses terhadap teknologi dan layanan digital (BPS, 2024b).

Dengan mempertimbangkan berbagai aspek, machine learning seperti CatBoost menjadi pilihan tepat untuk meningkatkan akurasi klasifikasi penerima KKS. CatBoost merupakan algoritma Gradient Boosting Decision Trees (GBDT) yang dikembangkan oleh Yandex dan dirancang secara khusus untuk menangani fitur kategorikal tanpa perlu one-hot encoding. Fitur ordered boosting yang dimilikinya efektif dalam mengurangi bias dan overfitting. Pemilihan CatBoost dalam penelitian ini didasarkan pada berbagai studi sebelumnya yang menunjukkan performa unggul dibandingkan metode lain. Penelitian (Jasman et al., 2022) membuktikan bahwa CatBoost memberikan akurasi lebih tinggi

dibandingkan Gradient Boosting dan AdaBoost dalam klasifikasi kualitas air. Temuan serupa juga diperoleh dari penelitian yang dilakukan oleh (Ilmi & Kusri, 2025) dalam Jurnal Buffer Informatika dengan judul analisis “Perbandingan Kinerja Algoritma Machine Learning dalam Deteksi Potensi Risiko HIV”, yang menyatakan bahwa CatBoost lebih akurat dibandingkan XGBoost dan LightGBM dalam mendeteksi risiko HIV. Dengan demikian, tujuan dari penelitian ini adalah untuk menerapkan algoritma CatBoost dalam mengklasifikasikan penerima KKS di Kota Bandung tahun 2024 serta mengevaluasi performa algoritma tersebut agar bantuan sosial dapat lebih tepat sasaran.

B. Metode

Penelitian ini menggunakan data SUSENAS 2024 dari BPS Provinsi Jawa Barat sebanyak 1.125 rumah tangga di Kota Bandung. Berikut merupakan cuplikan data penelitian:

Tabel 1. Data Penelitian

N o.	Status Perkawinan Kepala Keluarga	Usia Kepala Keluarga	Jenis Kelamin Kepala Keluarga	Jenjang Pendidikan Tertinggi Kepala Keluarga	Kepe milikan HP	Jumlah anggota keluarga	Kepe milikan KKS
1	Kawin	34	Laki-Laki	Tidak sekolah	YA	3	TIDAK
2	Kawin	59	Laki-Laki	Tidak sekolah	YA	4	TIDAK
3	Cerai Mati	72	Perempuan	Tidak sekolah	TIDAK	3	TIDAK
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	Kawin	45	Laki-Laki	Tidak sekolah	YA	4	TIDAK
12							
15							

Sumber: BPS Provinsi Jawa Barat, 2024.

Analisis dilakukan dengan algoritma CatBoost melalui tahapan prapemrosesan data, seperti penanganan missing value, outlier, kategori langka, dan penyeimbangan kelas menggunakan SMOTE-NC yang dapat dihitung berdasarkan rumus berikut:

$$x_{syn} = x_i + (x_{knn} - x_i) \times \gamma \quad \dots (1)$$

Dimana x_{syn} adalah data sintetis yang akan diciptakan, x_i adalah data yang akan direplikasi, x_{knn} adalah data yang memiliki jarak terdekat dari data yang akan direplikasi, dan γ adalah nilai random antara 0 sampai 1.

Untuk fitur kategorik dipilih berdasarkan modus, yaitu nilai kategorik yang paling sering muncul pada vektor utama serta pada vektor tetangga terdekatnya, dilanjutkan dengan pemilihan fitur sebagaimana yang dapat dihitung menggunakan rumus:

$$Feature\ Importance_F = \sum_{trees, leaf \in S_F} (v_1 - avr)^2 \cdot c_1 + (v_2 - avr)^2 \cdot c_2 \quad \dots (2)$$

$$avr = \frac{v_1 \cdot c_1 + v_2 \cdot c_2}{c_1 + c_2} \quad \dots (3)$$

Keterangan:

S_F : Kumpulan *split nodes* yang menggunakan fitur F.

v_1, v_2 : Prediksi nilai di daun kiri dan kanan setelah *split*.

c_1, c_2 : Bobot jumlah sampel (*object weight*) di daun kiri dan kanan. Bobot ini sama dengan jumlah objek di setiap daun jika bobot tidak ditentukan untuk *dataset*.

avr : Rata-rata prediksi terimbang dari daun kiri dan kanan.

dan encoding kategorikal yang dapat dihitung berdasarkan rumus berikut:

$$x_{\sigma_p, K} = \frac{\sum_{j=1}^{p-1} [x_{\sigma_j, k} = x_{\sigma_p, k}] y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1} [x_{\sigma_j, k} = x_{\sigma_p, k}] + a} \quad \dots (4)$$

Keterangan:

- σ : Permutasi acak dari data (urutan acak baris)
- $x_{\sigma_j, k}$: Nilai fitur kategorikal ke- k pada posisi ke- j setelah data diacak dengan permutasi σ
- $x_{\sigma_p, k}$: Nilai fitur kategorikal ke- k pada posisi ke- p setelah data diacak dengan permutasi σ
- y_{σ_j} : Label target untuk data ke- j dalam urutan permutasi
- a : Sebuah parameter yang lebih besar dari 0. Untuk memastikan tidak ada pembagian dengan 0 ketika nilai $x_{\sigma_j} \neq x_{\sigma_p}$. Tidak ada saran yang pasti dalam menentukan nilai a .
- P : Prior yang biasanya dibentuk dari nilai rata-rata label dalam dataset
- $x_{\sigma_p, K}$: Nilai numerik hasil encoding fitur kategori ke- K pada baris ke- p

Data dibagi menjadi 80% data latih dan 20% data uji. Pemodelan awal dilakukan dengan parameter iterasi 200, depth 6, dan learning rate 0.1, kemudian dilakukan tuning hyperparameter menggunakan grid search. Model dilatih secara iteratif berdasarkan gradien yang dapat dihitung menggunakan rumus:

$$\frac{\partial L}{\partial f(x)} = y - \hat{y}_t \quad \dots (5)$$

Di mana:

- y : Label aktual (0 atau 1)
 - $\hat{y}$: Prediksi Probabilitas dari model untuk kelas 1
- dan *cosine similarity* yang dapat dihitung menggunakan rumus:

$$Cosine = \frac{\sum_i^n w_i \cdot \Delta_i \cdot g_i}{\sqrt{\sum_{i=1}^n w_i \cdot \Delta_i^2} \cdot \sqrt{\sum_{i=1}^n w_i g_i^2}} \quad \dots (6)$$

Keterangan:

- w_i : Bobot objek ke- i
- g_i : Gradien pada objek ke- i yang sesuai dengan *log-loss*
- Δ_i : Nilai daun pada objek ke- i

Lalu dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, AUC, serta penentuan threshold optimal dengan pendekatan Youden's J menggunakan rumus berikut:

1. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \dots (7)$$

2. Precision

$$Precision = \frac{TP}{(TP+FP)} \quad \dots (8)$$

3. Recall

$$Recall = \frac{TP}{(TP+FN)} \quad \dots (9)$$

4. F1-Score

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \dots (10)$$

5. False Positive Rate (FPR)

$$FPR = \frac{FP}{FP + TN} \quad \dots (11)$$

6. Youden's J Statistic

$$J = \left(\frac{TP}{TP + FN} \right) + \left(\frac{TN}{TN + FP} \right) - 1 \quad \dots (12)$$

C. Hasil Penelitian dan Pembahasan

Feature Selection menggunakan Feature Importance

Perhitungan feature importance dilakukan untuk memilih fitur terbaik untuk model. Feature Importance dihitung berdasarkan Prediction Values Change, yaitu metode yang mengukur seberapa besar perubahan nilai prediksi yang dihasilkan setiap kali suatu fitur digunakan untuk melakukan pembelahan (split) pada node pohon keputusan selama proses boosting.

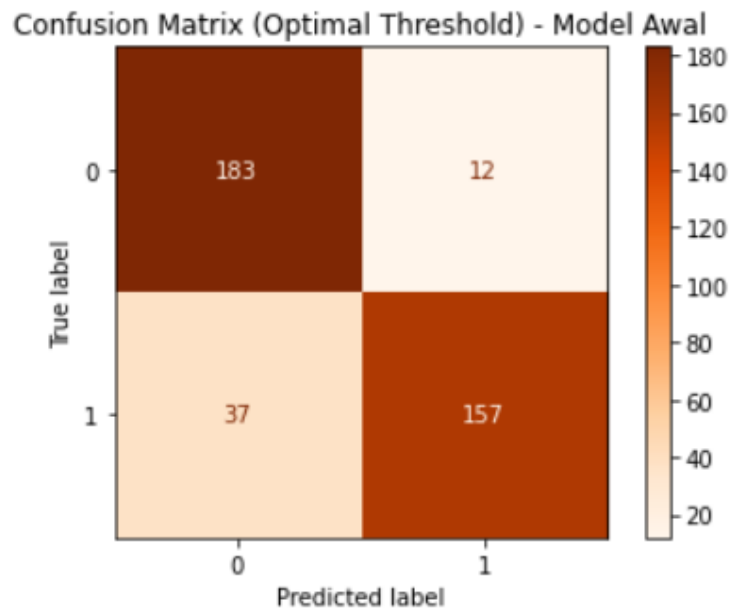
Tabel 2. Feature Importance

<i>Feature</i>	<i>Importance</i>
Jumlah anggota keluarga	66.130933
Usia Kepala Keluarga	18.819896
Kepemilikan HP	6.475987
Status Perkawinan Kepala Keluarga	6.472833
Jenis Kelamin Kepala Keluarga	2.057210
Jenjang Pendidikan Tertinggi Kepala Keluarga	0.043140

Berdasarkan Tabel 1. fitur jumlah anggota keluarga memiliki pengaruh paling besar terhadap prediksi model. Oleh karena itu, fitur dengan nilai feature importance di atas 1 dipilih untuk pelatihan model guna meningkatkan akurasi dan efisiensi. Fitur yang digunakan meliputi jumlah anggota keluarga, usia kepala keluarga, kepemilikan HP, status perkawinan, dan jenis kelamin, sementara jenjang pendidikan kepala keluarga dikeluarkan karena kontribusi yang kurang signifikan terhadap hasil prediksi.

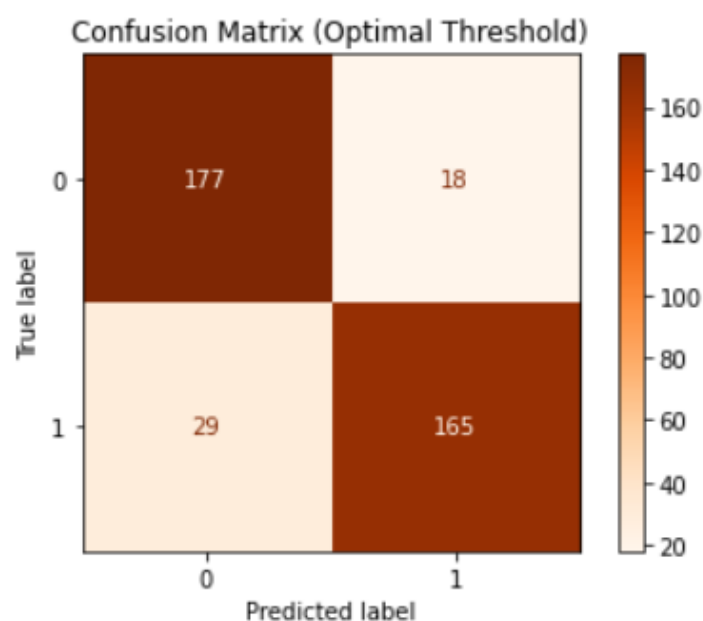
Implementasi Categorical Boosting (CatBoost)

Berikut merupakan penelitian mengenai klasifikasi penerima Kartu Keluarga Sejahtera (KKS) di kota Bandung tahun 2024 menggunakan metode Categorical Boosting (CatBoost) dengan menggunakan Python. Model CatBoost dilatih menggunakan 80% data latih, sedangkan 20% sisanya digunakan sebagai data uji. Pada penelitian ini model dilatih dengan menggunakan model awal yaitu iterasi sebesar 200 dan kedalaman pohon sebesar 6. Selanjutnya, dilakukan tuning hyperparameter menggunakan metode grid search, yaitu teknik pencarian sistematis terhadap kombinasi parameter terbaik berdasarkan metrik evaluasi seperti akurasi dan AUC. Setelah proses pelatihan selesai, dilakukan tahap prediksi dan evaluasi model menggunakan data uji. Evaluasi ini mencakup perhitungan confusion matrix, baik sebelum maupun sesudah tuning, untuk mengetahui secara rinci jumlah prediksi benar dan salah pada masing-masing kelas. Guna menghasilkan klasifikasi yang lebih seimbang antara True Positive Rate dan True Negative Rate, digunakan metode Youden's J untuk menentukan nilai threshold optimal, yaitu titik potong probabilitas yang memberikan performa terbaik dalam memisahkan dua kelas. Dengan threshold optimal ini, model tidak hanya mengandalkan akurasi keseluruhan, tetapi juga mempertimbangkan keseimbangan sensitivitas dan spesifisitas. Hasil evaluasi model dan perbandingan kinerjanya ditampilkan secara visual dalam Gambar 1 dan Gambar 2, yang memuat confusion matrix model sebelum dan sesudah dilakukan tuning dan pemilihan threshold optimal berdasarkan Youden's J. Hasil dapat dilihat pada gambar 1 dan 2.



Gambar 1. *Confusion Matrix* sebelum tuning menggunakan *threshold* optimal

Berdasarkan Gambar 1. menunjukkan model berhasil memprediksi 183 data kelas 0 dengan benar, sebanyak 12 data kelas 0 salah diklasifikasikan sebagai 1, Sebanyak 37 data kelas 1 salah diklasifikasikan sebagai 0, serta model benar mengklasifikasikan 157 data kelas 1. Threshold optimal ini dipilih karena memberikan keseimbangan yang baik untuk tujuan klasifikasi tanpa bias terhadap salah satu kelas, namun masih terdapat jumlah kesalahan prediksi yang cukup tinggi pada false positive dan false negative. Maka dari itu dilakukan tuning, berikut merupakan hasil confusion matrix setelah dilakukan tuning menggunakan threshold optimal.



Gambar 2. *Confusion Matrix* setelah tuning menggunakan *threshold* optimal

Berdasarkan Gambar 2. model berhasil mengklasifikasikan secara tepat 177 data dari kelas 0 dan 165 data dari kelas 1, dengan kesalahan prediksi terdiri dari 18 false positive dan 29 false negative. Hasil ini menunjukkan bahwa model memiliki tingkat kesalahan yang rendah dan cukup andal. Penerapan threshold optimal terbukti mampu meningkatkan kualitas klasifikasi secara keseluruhan,

terutama dalam menjaga keseimbangan antara dua jenis kesalahan. Pendekatan ini relevan diterapkan pada kebijakan distribusi bantuan seperti penetapan penerima Kartu Keluarga Sejahtera (KKS), karena menghasilkan klasifikasi yang lebih adil dan representatif dibandingkan penggunaan threshold default.

Tabel 3. Hasil evaluasi *threshold* optimal model awal sebelum tuning

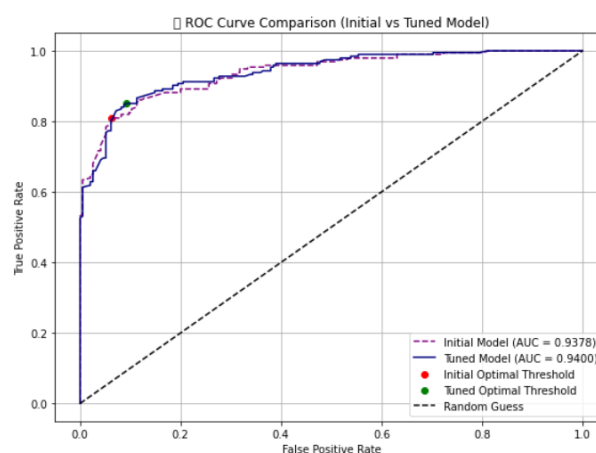
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
0	0.83	0.94	0.88	195
1	0.93	0.81	0.87	194
<i>Accuracy</i>			0.87	389
<i>Threshold Optimal (Youden's J)</i>	0.6028			

Berdasarkan Tabel 3. hasil evaluasi pada threshold optimal (Youden's J = 0,6028) menunjukkan bahwa model klasifikasi memiliki performa yang seimbang dalam membedakan kedua kelas. Model mencapai akurasi sebesar 87%, dengan precision yang tinggi pada kelas positif (0,93) dan recall yang tinggi pada kelas negatif (0,94). Capaian ini mengindikasikan bahwa model telah berada pada titik keseimbangan yang baik antara sensitivitas (True Positive Rate) dan spesifisitas (True Negative Rate), sebagaimana ditunjukkan oleh nilai F1-score yang relatif seimbang di kedua kelas.

Tabel 4. Hasil evaluasi *threshold* optimal model awal setelah tuning

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
0	0.86	0.91	0.88	195
1	0.90	0.85	0.88	194
<i>Accuracy</i>			0.88	389
<i>Threshold Optimal (Youden's J)</i>	0.6819			

Tabel 4. menunjukkan hasil evaluasi model setelah tuning menggunakan threshold optimal sebesar 0.6819 berdasarkan metode Youden's J. Model menunjukkan kinerja yang seimbang dan kuat di kedua kelas, dengan *precision* dan *recall* tinggi baik pada kelas 0 (0.86 dan 0.91) maupun kelas 1 (0.90 dan 0.85). Nilai F1-Score yang setara sebesar 0.88 di kedua kelas menunjukkan konsistensi model, sementara akurasi keseluruhan sebesar 88% menegaskan performa klasifikasi yang baik terhadap seluruh data.



Gambar 3. Kurva ROC AUC sebelum dan sesudah tuning pada threshold optimal

Berdasarkan Gambar 3. menampilkan kurva ROC sebagai evaluasi performa model sebelum dan sesudah tuning. Kurva ROC hasil tuning tampak sedikit lebih unggul dari model awal,

menunjukkan adanya peningkatan performa klasifikasi. Perbedaan ini juga tercermin dari nilai AUC yang meningkat setelah *tuning*. Titik threshold optimal yang ditandai pada kurva menunjukkan bahwa model hasil tuning memiliki keseimbangan yang lebih baik antara sensitivitas dan spesifisitas. Hal ini menegaskan bahwa tuning dan penyesuaian threshold memberikan dampak positif dalam meningkatkan akurasi dan efisiensi model, khususnya dalam konteks klasifikasi penerima bantuan seperti Kartu Keluarga Sejahtera (KKS). Hasil prediksi akhir menggunakan threshold optimal ditampilkan pada tabel 2.

Tabel 5. Hasil Prediksi Akhir

<i>y_actual</i>	<i>y_proba_tuned</i>	<i>y_pred_optimal_tuned</i>
1	0.329979747	0
1	0.94479775	1
0	0.188263063	0
0	0.090141586	0
0	0.000988059	0
⋮	⋮	⋮
389	0.024010245	0

Berdasarkan Tabel 5. menunjukkan hasil prediksi akhir sebanyak 389 data, yang terdiri dari label aktual, probabilitas hasil prediksi setelah tuning, dan prediksi akhir berdasarkan threshold optimal. Prediksi ditentukan dengan membandingkan probabilitas dengan nilai threshold 0.6819. Jika probabilitas di bawah threshold, maka diprediksi sebagai 0, sebaliknya diprediksi sebagai 1. Proses ini menghasilkan klasifikasi akhir, termasuk identifikasi kasus true positive, false negative, dan lainnya sesuai perbandingan antara label aktual dan prediksi.

Pemeriksaan kelas prediksi berdasarkan setiap fitur

Analisis ini bertujuan untuk mengidentifikasi fitur-fitur yang paling berkontribusi terhadap pengambilan keputusan model dalam memprediksi rumah tangga penerima Kartu Keluarga Sejahtera (KKS). Identifikasi ini dilakukan berdasarkan hasil pelatihan model menggunakan algoritma Categorical Boosting (CatBoost) yang mampu menangani data kategorikal secara langsung dan mempertimbangkan interaksi antar fitur secara efektif dalam proses pembentukan pohon keputusan. Dengan mengevaluasi nilai feature importance yang dihasilkan oleh model, dapat diketahui fitur mana saja yang paling signifikan dalam memengaruhi klasifikasi penerima bantuan sosial.

Berdasarkan hasil analisis tersebut, karakteristik penerima KKS yang paling dominan dapat diuraikan sebagai berikut: mayoritas berasal dari rumah tangga dengan jumlah anggota keluarga yang sangat sedikit, yaitu hanya 1 hingga 2 orang. Selain itu, kepala keluarga umumnya berusia antara 50 hingga 59 tahun, berada pada masa transisi menuju usia tidak produktif. Faktor lainnya yang memengaruhi keputusan model adalah ketiadaan kepemilikan telepon seluler (HP), yang dapat mengindikasikan keterbatasan akses terhadap teknologi dan informasi. Di samping itu, status perkawinan kepala keluarga yang dominan adalah cerai mati, yang mencerminkan kemungkinan hilangnya dukungan finansial dari pasangan. Jenis kelamin perempuan juga yang secara umum lebih rentan terhadap kondisi sosial ekonomi tertentu. Hasil ini menunjukkan bahwa model mampu mengidentifikasi pola kerentanan sosial ekonomi secara akurat, sehingga diharapkan dapat membantu pemerintah dalam penyaluran bantuan yang lebih tepat sasaran.

Pemeriksaan kelas prediksi berdasarkan seluruh fitur

Setelah menganalisis masing-masing fitur utama, seluruh fitur digabungkan dalam proses prediksi model *CatBoost*. Model ini memproses fitur secara simultan melalui pembentukan decision trees, di mana setiap observasi melewati pemisahan berdasarkan fitur-fitur yang paling berkontribusi. Hasil pemrosesan berupa logit, yang kemudian dikonversi menjadi probabilitas menggunakan fungsi sigmoid, yakni dapat dihitung menggunakan rumus:

$$\hat{y} = \frac{1}{1 + e^{-f(x)}} \quad \dots (12)$$

Probabilitas tersebut dibandingkan dengan threshold optimal (0.6819) untuk menentukan klasifikasi akhir. Proses ini memungkinkan model secara otomatis mengatur prioritas fitur berdasarkan kontribusinya dalam mengurangi loss function.

D. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa Penelitian ini menunjukkan bahwa algoritma CatBoost berhasil diterapkan untuk klasifikasi penerima Kartu Keluarga Sejahtera (KKS) di Kota Bandung tahun 2024. Pemodelan dilakukan melalui tahapan pre-processing, pelatihan awal, tuning hyperparameter dengan grid search, dan evaluasi model. Seleksi fitur berdasarkan feature importance menghasilkan lima fitur utama, yaitu Jumlah Anggota Keluarga, Usia Kepala Keluarga, Kepemilikan HP, Status Perkawinan, dan Jenis Kelamin Kepala Keluarga. Threshold optimal ditentukan menggunakan metode Youden's J.

Model awal memiliki akurasi 87% dan AUC 0,9378, namun kurang optimal pada kelas minoritas. Setelah tuning dan penyesuaian threshold, performa meningkat dengan AUC 0,9400 dan akurasi 88%, serta F1-score yang lebih seimbang. Sebanyak 183 rumah tangga diklasifikasikan sebagai penerima KKS, dengan karakteristik dominan: kepala keluarga berusia 50–59 tahun, hanya memiliki 1–2 anggota keluarga, tidak memiliki HP, berstatus cerai mati, dan berjenis kelamin perempuan.

Ucapan Terimakasih

Penulis mengucapkan syukur kepada Allah SWT atas rahmat dan hidayah-Nya, serta terima kasih kepada orang tua tercinta atas doa dan dukungannya. Ucapan terima kasih juga disampaikan kepada Dekan FMIPA, Ketua Program Studi Statistika, dan dosen pembimbing atas bimbingan dan arahnya selama penyusunan skripsi. Seluruh dosen Statistika UNISBA, keluarga, serta teman-teman kuliah dan SMA yang telah memberikan semangat dan motivasi juga mendapatkan apresiasi khusus atas dukungan mereka dalam menyelesaikan tugas akhir ini.

Daftar Pustaka

- BPS. (2024a, August). *Persentase Penduduk Miskin Kota Bandung Maret 2024 adalah sebesar 3,87 persen*. Bps.Go.Id. <https://bandungkota.bps.go.id/id/pressrelease/2024/08/01/1566/persentase-penduduk-miskin-kota-bandung-maret--2024-adalah-sebesar-3-87-persen.html>
- BPS. (2024b, October). *Persentase Rumah Tangga yang Memiliki/Menguasai Telepon Seluler Menurut Provinsi dan Klasifikasi Daerah, 2022-2023*. Bps.Go.Id.
- Diyandini, S. M., & Kurniati, E. (2024). Analisis Maksimalisasi Utilitas Konsumen Melalui Grafik dan Persamaan Rumus. *DataMath: Journal of Statistics and Mathematics*, 2(1), 1–10.
- Dulkiah, M., & Nurjanah. (2018). Pengaruh Kemiskinan terhadap Tingkat Tindak Kriminalitas di Kota Bandung. *Jurnal Ilmu Sosial Dan Ilmu Politik*, 8(1), 36–57.
- Holifah, N. (2022). Mekanisme Penyaluran Program Bantuan Kartu Keluarga Sejahtera Berbasis Cashless di Kelurahan Lewoleba. *Indonesian Journal of Political Studies*, 2(2), 74–97.

- Hutahaean, Y. M., & Sitorus, J. R. H. (2022). Faktor-Faktor yang Memengaruhi Kemiskinan Rumah Tangga Bekerja di Pulau Jawa. *Semina Nasional Official Statistics 2022*.
- Ilmi, M. B., & Kusriani. (2025). Perbandingan Kinerja Algoritma Machine Learning dalam Deteksi Potensi Risiko HIV. *Jurnal Buffer Informatika*, 11(1), 36–46.
- Jasman, T. Z., Fadhlullah, M. A., Pratama, A. L., & Rismayani. (2022). Analisis Algoritma Gradient Boosting, AdaBoost dan CatBoost dalam Klasifikasi Kualitas Air . *Jurnal Teknik Informatika Dan Sistem Informasi* , 8(2), 392–402.
- Noor, F. R., & Sirodj, D. A. N. (2024). Penerapan Support Vector Machine untuk Klasifikasi Opini Masyarakat Terhadap Isu Bullying. *Jurnal Riset Statistika*, 57–66. <https://doi.org/10.29313/jrs.v4i1.3877>
- Novian, W., & Herlina, M. (2022). Faktor-faktor yang Mempengaruhi Perilaku Masyarakat terhadap Penggunaan Internet pada Desa Digital Sukaraja. *Jurnal Riset Statistika*, 161–168. <https://doi.org/10.29313/jrs.v2i2.1466>
- Rosiana, N. (2021). Penggunaan Telepon Seluler untuk Pemasaran serta Dampaknya terhadap Kesejahteraan Petani. *AGRISSEP*, 20(1), 25–42. <https://doi.org/10.31186/jagrisep.20.1.25-40>
- Sa'diyah, Y. H., & Arianti, F. (2012). Analisis Kemiskinan Rumah Tangga melalui Faktor-Faktor yang Mempengaruhinya di Kecamatan Tugu Kota Semarang. *Diponogoro Journal of Economics*, 1(1), 1–11.
- Sugandi, Y. S. (2016). Program Penanggulangan Kemiskinan Perkotaan Pengalaman Kota Bandung. *Jurnal Analisis Dan Pelayanan Publik*, 2(1), 109–128.