

PENERAPAN MODEL ALGORITMA *XGBoost* DALAM PENENTUAN STATUS PENERIMA SUBSIDI LISTRIK DI PT PLN (PERSERO)

Muh Firman Syawal¹, Faisal², A. Muhammad Syafar³

¹²³Program Studi Teknik Informatika, Universitas Islam Negeri Alauddin Makassar

¹email : frmns77@gmail.com , ²email : faisal.rahman@uin-alauddin.ac.id ³email : andi.syafar@uin-alauddin.ac.id

Abstrak

Penyaluran subsidi listrik sering mengalami ketidaktepatan sasaran, di mana penerima berasal dari golongan mampu. Penelitian ini bertujuan untuk mengimplementasikan algoritma *XGBoost* (*Extreme Gradient Boosting*) guna menentukan status kelayakan penerima subsidi listrik studi kasus di PT. PLN (Persero) ULP Masamba. Metode penelitian yang digunakan berupa pendekatan eksperimen komputasional dengan pemodelan berbasis *machine learning*. Data mencakup variabel seperti daya listrik, tarif, keterangan keperluan, dan Padan Dukcapil. Hasil evaluasi model menunjukkan akurasi sebesar 84,30%, *recall* 85,52%, dan *AUC-ROC* 91,82%. Temuan ini menunjukkan bahwa *XGBoost* efektif dalam klasifikasi kelayakan subsidi listrik secara objektif dan adil.

Kata Kunci: *XGBoost*, Subsidi Listrik, PLN, *Machine Learning*

Abstract

The distribution of electricity subsidies often experiences mis-targeting, where recipients come from capable groups. This study aims to implement the XGBoost (Extreme Gradient Boosting) algorithm to determine the eligibility status of electricity subsidy recipients study case at PT. PLN (Persero) ULP Masamba. The research method used is a computational experimental approach with machine learning-based modeling. The data includes variables such as Electricity Capacity, Rates, Description of Needs, and Matching Dukcapil.. The evaluation results of the model show an accuracy of 84.30%, recall of 85.52%, and AUC-ROC of 91.82%. These findings indicate that XGBoost is effective in classifying electricity subsidy eligibility objectively and fairly.

Keywords: *XGBoost*, Electric Subsidiary, PLN, *Machine Learning*

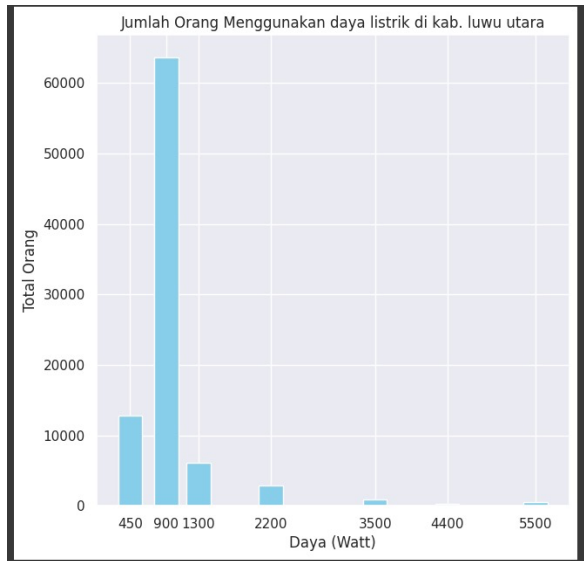
I. PENDAHULUAN

Subsidi listrik merupakan salah satu program pengendalian kemiskinan yang memberikan dana bantuan subsidi listrik kepada rumah tangga miskin dan tidak mampu yang dibayarkan oleh Pemerintah Indonesia kepada PT. PLN. Tujuan pemberian subsidi adalah untuk membantu pelanggan miskin dan mereka yang belum dihubungi oleh PT. PLN sehingga dapat ikut menikmati energi listrik (Yanto & Yunus, 2021).

Listrik menjadi salah satu industri terbesar di dunia. Dalam kehidupan modern, listrik merupakan salah satu kebutuhan vital yang sangat penting untuk keberlangsungan aktivitas manusia maupun kegiatan industri. Kebutuhan listrik diatur oleh Negara melalui PT. PLN. Kemajuan teknologi dan informasi yang dinamis mendorong manusia menggunakan peralatan dengan tenaga listrik yang menyebabkan konsumsi energi meningkat. Sehingga, setiap rumah tangga mengharapkan sambungan ke jaringan listrik tanpa gangguan agar kehidupan dapat berjalan dengan baik. Dalam hal ini, semakin banyak konsumsi tenaga listrik suatu rumah tangga, maka akan semakin banyak pula tarif listrik yang harus

dibayarkan. Upaya yang dilakukan oleh pemerintah dalam mengontrol harga tenaga listrik adalah penyesuaian tarif listrik dan pemberlakuan subsidi tepat sasaran. Kebijakan Pemerintah melalui Peraturan Menteri ESDM Nomor 29 Tahun 2016 mengenai Mekanisme Pemberian Subsidi Tarif Tenaga Listrik untuk Rumah Tangga mengatur penerapan subsidi tenaga listrik agar tepat sasaran, sehingga rumah tangga mampu seharusnya tidak lagi mendapatkan subsidi. Bantuan subsidi listrik yang diterima juga berbeda tiap rumah tangga tergantung pemakaian energi listrik rumah tangga tersebut. (Hutahaean & Wijayanto, 2022).

PT. PLN (Persero) ULP Masamba, Kab. Luwu Utara, Sulawesi Selatan, juga memiliki peran penting dalam memastikan bahwa subsidi listrik yang disediakan oleh pemerintah tersalurkan secara tepat sasaran kepada masyarakat yang berhak di wilayah Masamba dan sekitarnya. Sebagai Unit Layanan Pelanggan, PT. PLN (Persero) ULP Masamba bekerja sama dengan instansi pemerintah, seperti Kementerian Sosial dalam mengakses data ekonomi masyarakat untuk mendukung ketepatan sasaran.



Gambar 1. Grafik pelanggan listrik PT. PLN (Persero) ULP Masamba

Berdasarkan data pelanggan listrik PT. PLN (Persero) ULP Masamba, mayoritas masyarakat di Kabupaten Luwu Utara masih menggunakan daya 900 *Watt*, hal ini menandakan bahwa sebagian besar masyarakat di wilayah tersebut berpeluang untuk mengajukan subsidi listrik. Namun, beberapa masyarakat di wilayah tersebut mengeluhkan terkait proses penyaluran subsidi listrik, untuk itu dibutuhkan suatu sistem berbasis teknologi yang kiranya dapat membantu PT. PLN (Persero) ULP Masamba dalam menentukan status penerima subsidi listrik di wilayah tersebut secara akurat dan tepat sasaran berdasarkan data yang telah diterima dari Dinas Sosial. Salah satu sistem berbasis teknologi yang dimaksud adalah penggunaan algoritma *XGBoost*.

Tujuan utama dari penelitian ini adalah untuk menerapkan model algoritma *XGBoost* dalam penentuan status penerima subsidi listrik di PT. PLN (Persero) guna meningkatkan akurasi dan efisiensi dalam proses seleksi penerima subsidi. Dengan menggunakan algoritma berbasis *machine learning* seperti *XGBoost*, penelitian ini bertujuan untuk menghasilkan model yang dapat secara objektif mengidentifikasi masyarakat yang memenuhi syarat untuk menerima subsidi listrik, berdasarkan data pelanggan yang relevan. Selain itu, penelitian ini juga berupaya untuk memastikan bahwa penyaluran subsidi dilakukan secara adil, tepat sasaran, dan transparan, sesuai dengan prinsip keadilan yang diajarkan dalam Islam serta etika distribusi yang bertanggung jawab. Melalui pendekatan ini, peneliti berharap dapat membantu PLN dan lembaga terkait untuk mengoptimalkan

penggunaan sumber daya subsidi dan mencegah ketidaktepatan sasaran dalam distribusi bantuan.

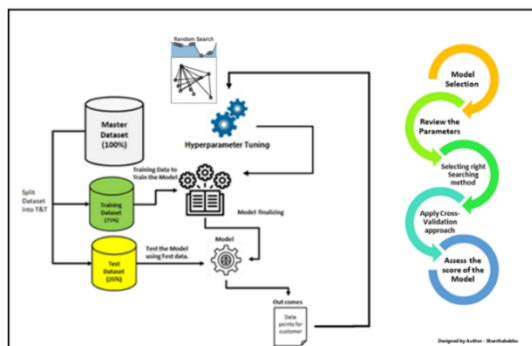
II. METODE PENELITIAN

Jenis penelitian yang digunakan adalah penelitian kuantitatif. Penelitian ini termasuk kuantitatif karena melibatkan pengumpulan dan analisis data numerik yang besar, seperti data Daya Listrik, Tarif dan Keterangan Keperluan pelanggan PLN, untuk menghasilkan prediksi yang akurat. Dalam penelitian ini, algoritma *XGBoost* digunakan untuk mengolah data dan melakukan prediksi status penerima subsidi, yang kemudian dievaluasi menggunakan metrik statistik seperti akurasi, *precision*, *recall*, dan *F1-Score*. Lalu dilakukan pendekatan eksperimen komputasional diterapkan dengan melakukan serangkaian uji coba pada model prediksi untuk menghasilkan model yang paling optimal.

Pada penelitian ini menggunakan metode *Hyperparameter tuning*, karena bertujuan untuk menemukan kombinasi parameter yang paling optimal bagi model. Dengan mengatur *Hyperparameter* seperti *learning rate*, *max depth*, dan *number of estimators*, model *XGBoost* dapat menghasilkan prediksi yang lebih akurat. Ini sangat penting dalam konteks penentuan penerima subsidi, di mana akurasi yang tinggi dapat berdampak signifikan pada keadilan distribusi subsidi.

Pengaturan *Hyperparameter* yang tepat dapat membantu model untuk mencapai keseimbangan yang baik antara kompleksitas dan generalisasi. *Overfitting* terjadi ketika model belajar terlalu banyak dari data pelatihan, sehingga tidak mampu menggeneralisasi pada data baru, sedangkan *underfitting* terjadi ketika model terlalu sederhana untuk menangkap pola dalam data. Dengan *Hyperparameter tuning*, risiko kedua masalah ini dapat diminimalkan.

Menggunakan *Hyperparameter tuning* dalam penelitian ini bukan hanya meningkatkan kualitas model, tetapi juga memastikan bahwa keputusan yang diambil berdasarkan model tersebut adalah tepat dan adil. Dalam konteks penentuan penerima subsidi listrik, hal ini sangat krusial untuk menghindari kesalahan yang dapat mempengaruhi banyak orang, serta untuk meningkatkan kepercayaan masyarakat terhadap sistem distribusi subsidi yang ada.



Gambar 2. Metode *Hyperparameter tuning*

Berikut adalah tahapan-tahapan utama dalam penerapan *Hyperparameter tuning*:

a. Pemahaman *Hyperparameter tuning*

Memahami bahwa *Hyperparameter* adalah parameter yang ditentukan sebelum proses pelatihan model dan tidak dioptimalkan oleh model itu sendiri. Dalam algoritma *XGBoost*, beberapa *Hyperparameter* penting termasuk:

learning_rate (eta): Mengontrol seberapa besar kontribusi setiap pohon ke model akhir.

max_depth: Menentukan kedalaman maksimum setiap pohon.

n_estimators: Jumlah total pohon yang akan dibangun.

subsample: Proporsi dari sampel yang digunakan untuk membangun setiap pohon.

colsample_bytree: Proporsi fitur yang digunakan untuk membangun setiap pohon.

gamma: Mengontrol kompleksitas model dengan mengatur minimum gain untuk membagi node.

b. Pemisahan data

Membagi dataset yang ada menjadi data pelatihan dan data validasi. Data pelatihan digunakan untuk melatih model, sedangkan data validasi digunakan untuk menguji performa model dengan kombinasi *Hyperparameter* yang berbeda. Serta Rasio umum pemisahan dapat menggunakan 80% untuk pelatihan dan 20% untuk validasi.

c. Menentukan rentang nilai untuk *Hyperparameter*

Pada tahap ini, kita akan menentukan rentang nilai untuk setiap *Hyperparameter* yang akan diuji. Antara lain: *learning_rate*, *max_depth*, *n_estimator*, *subsample*, dll.

d. Pemilihan metode tuning

Setelah menentukan rentang nilai untuk *Hyperparameter*, maka kita akan memilih metode yang akan digunakan pada penelitian ini.

e. Melaksanakan *Hyperparameter tuning*

Melaksanakan tuning dengan menggunakan metode yang dipilih. Setiap kombinasi *Hyperparameter* diuji dan model dilatih pada data pelatihan. Kemudian, model diuji pada data validasi untuk mengevaluasi kinerjanya menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-Score*.

f. Evaluasi Model

Mengevaluasi hasil dari setiap kombinasi *Hyperparameter*. Menyimpan metrik kinerja dari setiap iterasi dan mengidentifikasi kombinasi *Hyperparameter* yang memberikan hasil terbaik berdasarkan metrik evaluasi yang dipilih.

g. Menganalisis Hasil

Melakukan analisis lebih lanjut terhadap hasil prediksi dari model akhir dan mengevaluasi apakah ada peluang perbaikan lebih lanjut, baik dalam tuning *Hyperparameter* atau dalam pengolahan data.

III. HASIL DAN PEMBAHASAN

Penelitian ini dilatarbelakangi oleh perlunya peningkatan akurasi dalam proses identifikasi calon penerima subsidi listrik PT. PLN (Persero) ULP Masamba, agar penyaluran subsidi dari data Dinas Sosial menjadi lebih adil dan tepat sasaran.

1. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang mencerminkan kondisi riil pelanggan PLN di wilayah Masamba. Data tersebut mencakup berbagai informasi penting terkait karakteristik pelanggan, yang berpotensi mempengaruhi kelayakan mereka dalam menerima subsidi listrik. Variabel-variabel dalam dataset antara lain:

Tabel 1. Deskripsi Variabel Dataset Penelitian

No	Nama Variabel	Tipe Data	Deskripsi
1.	Daya Listrik	<i>Integer</i>	Kapasitas daya listrik pelanggan (contoh: 450VA,

			900VA, 1200VA,dst)
2.	Tarif	Kategori	Jenis pelanggan listrik (R1, R1M, R1T, R1MT, R2, R2T,dst.)
3.	Keterangan Keperluan	Kategori	Keterangan Keperluan penggunaan listrik (Rumah Tangga, Kost, Kontrakan, dst.)
4.	Padan Dukcapil	Kategori	Kecocokan data pelanggan listrik dan data Kependudukan dan Pencatatan Sipil (Padan dan Belum Padan)
5.	Status Penerima Subsidi	Biner	Target variabel klasifikasi: 1 = Layak Subsidi, 0 = Tidak Layak Subsidi.

A. Preprocessing Data

Sebelum data digunakan dalam proses pelatihan model *XGBoost*, dilakukan tahap *preprocessing* atau prapemrosesan data untuk memastikan bahwa dataset bersih, relevan, dan siap diproses oleh algoritma. *Preprocessing* merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah yang diproses oleh sistem (Nursyi'ah dkk., 2021).

Langkah pertama dalam *preprocessing* adalah mengidentifikasi dan menangani data yang kosong atau hilang (*missing value*). *Missing data* atau *missing value* adalah satu kondisi dimana terdapat nilai yang tidak lengkap atau kosong pada satu atau beberapa kriteria (Sudrajat & Cholid, 2023).. Proses ini dilakukan menggunakan aplikasi Weka, dengan fitur *built-in* yang secara otomatis mendeteksi nilai kosong pada atribut

tertentu. Setelah dilakukan pengecekan, atribut yang memiliki nilai kosong diperlakukan dengan dua cara, yaitu:

- 1) Menghapus entri (*instance*) yang memiliki nilai kosong pada atribut yang dianggap penting dan tidak bisa diimputasi.
- 2) Mengisi nilai kosong dengan nilai default atau *mean*/mode untuk atribut numerik atau kategorik, jika data tersebut masih relevan dan layak diselamatkan.

Penanganan *missing value* ini bertujuan untuk menghindari distorsi hasil pelatihan model, serta menjaga integritas data agar prediksi lebih akurat.

Beberapa atribut pada dataset awal tidak memiliki kontribusi signifikan terhadap proses klasifikasi status subsidi, atau bersifat non-numerik yang tidak dapat diproses langsung oleh algoritma *XGBoost*. Oleh karena itu, atribut-atribut tersebut dihapus secara manual melalui aplikasi Weka, dengan mempertimbangkan konteks bisnis dan relevansi terhadap label target. Atribut yang dihapus antara lain:

- 1) Thblmut : Merupakan informasi waktu pembaruan, tidak relevan langsung untuk klasifikasi.
- 2) Nama Pelanggan : Bersifat identitas, tidak memiliki nilai prediktif.
- 3) Alamat Pelanggan : Informasi teks bebas, tidak terstruktur, sulit dikonversi ke fitur numerik.

Penghapusan dilakukan untuk mengurangi kompleksitas data dan meminimalkan noise, sehingga hanya fitur yang relevan saja yang digunakan dalam proses pelatihan model.

B. Penerapan Model Algoritma *XGBoost*

Pada bagian ini, dijelaskan proses penerapan model algoritma *XGBoost* yang digunakan untuk memprediksi kelayakan pelanggan dalam menerima subsidi listrik berdasarkan data pelanggan PLN ULP Masamba. Penerapan model ini dilakukan secara bertahap mulai dari pemahaman algoritma, pengaturan *Hyperparameter*, hingga proses pelatihan model menggunakan data yang telah dipersiapkan sebelumnya.

XGBoost (*Extreme Gradient Boosting*) merupakan salah satu algoritma *machine learning* yang berbasis pada *ensemble learning*, khususnya *gradient boosting decision tree* (GBDT). Algoritma ini bekerja dengan membangun

sejumlah pohon keputusan (*decision tree*) secara bertahap, di mana setiap pohon baru memperbaiki kesalahan dari pohon sebelumnya.

XGBoost dikenal sangat efisien dan mampu menghasilkan prediksi yang akurat bahkan pada data yang tidak seimbang. Dalam konteks penelitian ini, *XGBoost* digunakan untuk melakukan klasifikasi terhadap status pelanggan apakah layak atau tidak layak menerima subsidi listrik berdasarkan berbagai parameter seperti daya, tarif, keterangan keperluan, dan lain-lain.

Pengaturan *Hyperparameter*

Agar model *XGBoost* bekerja secara optimal, diperlukan penyesuaian terhadap sejumlah *Hyperparameter*, yaitu parameter yang ditentukan sebelum proses pelatihan dimulai. Penyesuaian ini dilakukan melalui proses *Hyperparameter tuning*, baik secara manual maupun otomatis (misalnya dengan *GridSearch* atau *RandomSearch*).

C. Evaluasi Kinerja Model

Evaluasi model merupakan tahapan penting dalam proses pengembangan sistem prediktif berbasis machine learning. Pada penelitian ini, algoritma *XGBoost* telah diimplementasikan untuk memprediksi status kelayakan penerima subsidi listrik. Untuk mengetahui seberapa baik performa model yang telah dibangun, digunakan beberapa metrik evaluasi yang umum dalam klasifikasi biner, yaitu: *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC Curve*.

1. *Confusion Matrix*

Confusion Matrix merupakan representasi visual dari hasil prediksi model terhadap data uji, yang menggambarkan jumlah prediksi yang benar dan salah dalam setiap kategori.

Tabel 2. *Confusion Matrix*

	Prediksi: Tidak Layak (0)	Prediksi: Layak (1)
Sebenarnya: Tidak Layak (0)	TN = 2051	FP = 410
Sebenarnya: Layak (1)	FN = 280	TP = 1665

Interpretasi Hasil *Confusion Matrix*

Berdasarkan *Confusion Matrix* yang dihasilkan oleh model *XGBoost*, diketahui bahwa:

True Positive (TP = 2051) menunjukkan jumlah pelanggan yang diprediksi layak dan memang benar-benar layak menerima subsidi.

False Negative (FN = 280) adalah pelanggan yang sebenarnya layak, namun diprediksi tidak layak, sehingga berisiko tidak menerima subsidi yang menjadi haknya.

False Positive (FP = 410) merupakan pelanggan yang diprediksi layak, padahal sebenarnya tidak layak, sehingga subsidi diberikan kepada pihak yang tidak berhak.

True Negative (TN = 1665) menunjukkan prediksi yang tepat terhadap pelanggan tidak layak yang memang tidak seharusnya menerima subsidi.

2. Metrik Evaluasi

Untuk menilai sejauh mana model *XGBoost* dapat melakukan klasifikasi secara akurat terhadap status kelayakan penerima subsidi listrik, diperlukan proses evaluasi menggunakan berbagai metrik kinerja. Evaluasi ini bertujuan untuk mengetahui kemampuan model dalam membedakan pelanggan yang layak dan tidak layak subsidi berdasarkan data yang digunakan. Metrik evaluasi yang diterapkan meliputi *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan juga *AUC-ROC*. Setiap metrik memberikan perspektif yang berbeda mengenai kualitas prediksi, baik dari sisi ketepatan keseluruhan, kemampuan model meminimalkan kesalahan klasifikasi, hingga efektivitas dalam mengidentifikasi kelas positif secara akurat. Dengan demikian, evaluasi ini memberikan gambaran yang komprehensif mengenai performa model yang dihasilkan. Adapun hasil perhitungan metrik evaluasi dapat dilihat pada tabel berikut:

Tabel 3. Metrik

Metrik	Rumus	Nilai (%)
<i>Accuracy</i>	$(TP + TN) / \text{Total}$	$(1665 + 2051) / 4406 = 84.34$
<i>Precision</i>	$TP / (TP + FP)$	$1665 / (1665 + 410) = 80.24$

<i>Recall</i>	$TP / (TP + FN)$	$1665 / (1665 + 280) = 85.60$
<i>F1-Score</i>	$2 \times (Prec \times Rec) / (Prec + Rec)$	≈ 82.83

Setelah kita menjumlahkan perhitungan terhadap setiap metrik evaluasi berdasarkan hasil prediksi pada data uji. Hasil perhitungan ini mencerminkan seberapa baik model dalam meminimalkan kesalahan klasifikasi serta mengidentifikasi pelanggan yang benar-benar layak menerima subsidi, bisa kita lihat pada tabel berikut:

Tabel 4. Metrik Evaluasi

Metrik Evaluasi	Nilai Desimal	Persentase (%)
<i>Accuracy</i>	0.8430	84.30%
<i>Precision</i>	0.8014	80.14%
<i>Recall</i>	0.8552	85.52%
<i>F1 Score</i>	0.8275	82.75%
<i>AUC-ROC</i>	0.9182	91.82%

Hasil evaluasi menunjukkan bahwa model *XGBoost* memiliki performa yang sangat baik. Dengan akurasi sebesar 84,30%, model mampu memprediksi secara benar dalam sebagian besar kasus. Nilai *recall* yang tinggi (85,52%) sangat penting karena menunjukkan kemampuan model dalam mendeteksi pelanggan yang benar-benar layak menerima subsidi listrik suatu hal yang krusial untuk keadilan sosial.

Selain itu, *Precision* sebesar 80,14% menunjukkan bahwa prediksi layak subsidi yang dibuat oleh model cukup akurat dan tidak terlalu banyak memberikan subsidi kepada pelanggan yang tidak seharusnya menerima. Nilai *F1-Score* (82,75%) menunjukkan bahwa model memiliki keseimbangan yang baik antara kemampuan mendeteksi dan ketepatan prediksi. Terakhir, *AUC-ROC* sebesar 91,82% menandakan bahwa model sangat andal dalam membedakan dua kelas dengan jelas.

Secara keseluruhan, metrik evaluasi ini menunjukkan bahwa model *XGBoost* layak digunakan sebagai sistem bantu pengambilan keputusan dalam penentuan status penerima subsidi listrik.

a. Visualisasi *AUC ROC*

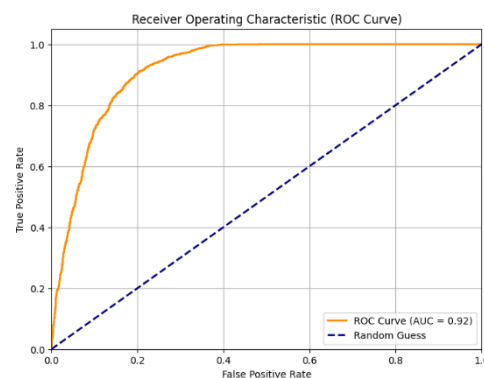
ROC Curve merupakan grafik yang menunjukkan kemampuan model dalam membedakan antara dua kelas (dalam hal ini: layak dan tidak layak menerima subsidi listrik). Grafik ini memplot:

- 1) *True Positive Rate* (TPR) atau *Recall* di sumbu vertikal (Y-axis)
- 2) *False Positive Rate* (FPR) di sumbu horizontal (X-axis)

Kurva *ROC* membandingkan performa model terhadap prediksi acak (*random guess*), yang digambarkan oleh garis diagonal putus-putus (dari titik (0,0) ke (1,1)).

b. Interpretasi Grafik

Untuk memperoleh pemahaman yang lebih komprehensif mengenai performa model *XGBoost* dalam membedakan pelanggan yang layak dan tidak layak menerima subsidi listrik, dilakukan analisis visual melalui grafik *Receiver Operating Characteristic* (ROC). Grafik ROC digunakan sebagai alat evaluasi untuk menggambarkan hubungan diantara *True Positive Rate* (TPR) dan juga *False Positive Rate* (FPR) pada berbagai nilai threshold. Melalui visualisasi ini, dapat dilihat secara jelas seberapa baik model dalam melakukan klasifikasi tanpa bergantung pada satu nilai ambang tertentu. Adapun grafik ROC hasil pemodelan dapat dilihat pada gambar berikut :



Gambar 3. Kurva ROC

Berdasarkan grafik *ROC* di atas, terlihat bahwa:

1) Kurva *ROC* model naik dengan tajam ke arah kiri atas (0,1), yang menandakan kemampuan klasifikasi yang sangat baik.

2) Model memiliki nilai *AUC* (*Area Under the Curve*) sebesar 0.92, yang menunjukkan bahwa model memiliki 92% kemungkinan untuk membedakan secara benar antara pelanggan yang layak dan tidak layak subsidi. Secara umum:

Tabel 5. Hasil Nilai *AUC*

Nilai <i>AUC</i>	Tingkat Kinerja Model
0.90 – 1.00	Sangat baik
0.80 – 0.90	Baik
0.70 – 0.80	Cukup
0.60 – 0.70	Kurang
0.50 – 0.60	Gagal

Dengan nilai *AUC* = 0.92, model *XGBoost* dalam penelitian ini dapat dikategorikan memiliki kemampuan klasifikasi yang sangat baik, bahkan mendekati ideal.

IV. KESIMPULAN

Hasil penelitian ini diharapkan dapat menjadi acuan bagi PT. PLN (Persero) ULP Masamba dalam penerapan model algoritma *XGBoost* untuk menentukan kelayakan penerima subsidi listrik secara objektif dan efisien, setelah mendapatkan data subsidi dari Dinas Sosial. Berdasarkan serangkaian tahapan, mulai dari pengumpulan dan preprocessing data hingga evaluasi model, diperoleh hasil bahwa model *XGBoost* memberikan performa yang sangat baik, pada hasil evaluasi menunjukkan bahwa model *XGBoost* memiliki performa yang sangat baik. Dengan nilai akurasi sebesar 84,30%, model mampu memprediksi secara benar dalam sebagian besar kasus. Nilai *recall* yang tinggi (85,52%) menunjukkan kemampuan model dalam mendeteksi pelanggan yang benar-benar layak menerima subsidi listrik. Nilai *Precision* sebesar 80,14% menunjukkan bahwa prediksi layak subsidi yang dibuat oleh model cukup akurat dan tidak terlalu banyak memberikan subsidi kepada pelanggan yang tidak seharusnya menerima. Nilai

F1-Score (82,75%) menunjukkan bahwa model memiliki keseimbangan yang baik antara kemampuan mendeteksi dan ketepatan prediksi. Terakhir, *AUC-ROC* sebesar 91,82% menandakan bahwa model sangat andal dalam membedakan dua kelas dengan jelas.

Hasil tersebut membuktikan bahwa model dapat mengklasifikasikan dengan cukup presisi dan konsisten siapa saja yang layak menerima subsidi, mengurangi risiko salah sasaran, serta mendukung keputusan yang adil dan efisien.

V. DAFTAR PUSTAKA

- Hutahaean, Y. M., & Wijayanto, A. W. (2022). Klasifikasi Rumah Tangga Penerima Subsidi Listrik di Provinsi Gorontalo Tahun 2019 dengan Metode K-Nearest Neighbor dan Support Vector Machine. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 10(1), 63. <https://doi.org/10.26418/justin.v10i1.51210>
- Nursyi'ah, S. Y., Erfina, A., & Warman, C. (2021). Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes. *Jurnal Media Informatika Budidarma*, 117–123.
- Sudrajat, W., & Cholid, I. (2023). K-Nearest Neighbor (K-Nn) Untuk Penanganan Missing value Pada Data Umkm. *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, 1(2), 54–63. <https://doi.org/10.59407/jrsit.v1i2.77>
- Yanto, B. H., & Yunus, Y. (2021). Evaluasi Penentuan Kelayakan Pemberian Subsidi Listrik dengan Metode MFEP. *Jurnal Informatika Ekonomi Bisnis*, 3, 109–114. <https://doi.org/10.37034/infv3i3.91>

