

PREDICTION OF ELECTRICITY BILL PAYMENT DELAYS FOR CUSTOMERS USING A MACHINE LEARNING APPROACH

PREDIKSI KETERLAMBATAN PEMBAYARAN REKENING LISTRIK PELANGGAN DENGAN PENDEKATAN MACHINE LEARNING

Dyah Puspita Sari Nilam Utami^{*1)}, Mochamad Iqbal Arifyanto²⁾

^{1),2)}Program Studi Magister Sains Komputasi, Fakultas Matematika dan Ilmu Pengetahuan,
Institut Teknologi Bandung

*Email Penulis: 20922327@mahasiswa.itb.ac.id¹, ikbal@as.itb.ac.id²

Abstract - Electricity is a vital necessity in modern life, and the management of electricity bill payments is crucial for the continuity of services and the financial stability of electricity providers like PLN. Identifying potential delays in payments by customers is a strategic step to enable effective preventive actions. This study aims to develop a prediction model for payment delays using two machine learning methods, namely Random Forest Regressor and Bidirectional Long Short-Term Memory, based on historical customer data from the period of 2018–2023. The research process includes data preprocessing to ensure consistency and accuracy, dividing the data into training and testing sets, and training the models using both algorithms. The results show that the Random Forest model performed the best in recognizing long-term statistical patterns with the lowest Mean Absolute Error value of 0.00387 on the 12-month Moving Average feature, as well as optimal efficiency with a number of trees between 100–200. On the other hand, the Bidirectional LSTM model demonstrated competitive ability in capturing temporal patterns of sequential data, with the best configuration yielding a validation error value of 0.243 and the highest validation accuracy of 56.2%. Both models are effective in predicting customers who are likely to delay their electricity bill payments. This research provides significant contributions to PLN in supporting data-driven decision-making and facilitating mitigation strategies such as early notifications or rescheduling payment plans to reduce the risk of overdue payments.

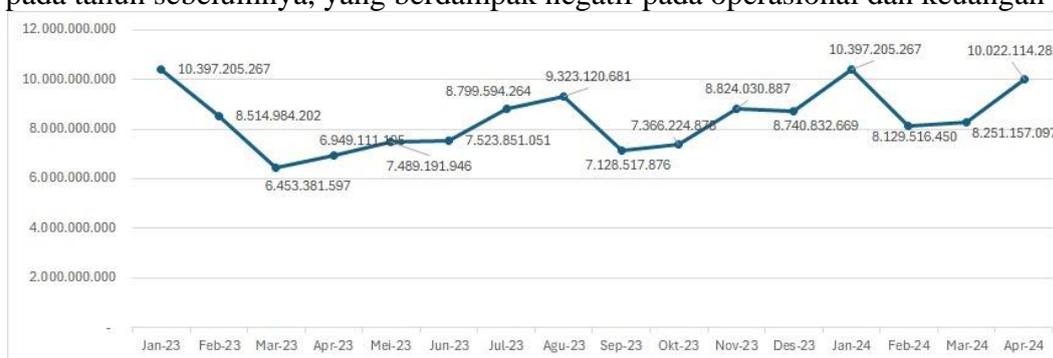
Keywords - Late Payments, Prediction, Random Forest, PLN, Moving Average

abstrak- Listrik merupakan kebutuhan vital dalam kehidupan modern, dan pengelolaan pembayaran rekening listrik menjadi hal yang sangat penting untuk kelangsungan layanan serta stabilitas finansial perusahaan penyedia listrik, seperti PLN. Identifikasi potensi keterlambatan pembayaran oleh pelanggan merupakan langkah strategis untuk memungkinkan tindakan preventif yang efektif. Penelitian ini bertujuan untuk mengembangkan model prediksi keterlambatan pembayaran menggunakan dua metode machine learning, yaitu Random Forest Regressor dan Bidirectional Long Short-Term Memory, berdasarkan data historis pelanggan dari periode 2018–2023. Proses penelitian mencakup preprocessing data untuk memastikan konsistensi dan akurasi, pembagian data menjadi set pelatihan dan set pengujian, serta pelatihan model menggunakan kedua algoritma. Hasil penelitian menunjukkan bahwa model Random Forest memberikan performa terbaik dalam mengenali pola statistik jangka panjang dengan nilai Mean Absolute Error terendah sebesar 0.00387 pada fitur Moving Average 12 bulan, serta efisiensi optimal dengan jumlah pohon antara 100–200. Di sisi lain, model Bidirectional LSTM menunjukkan kemampuan kompetitif dalam menangkap pola temporal data sekuensial dengan konfigurasi terbaik menghasilkan nilai validation error sebesar 0.243 dan validation accuracy tertinggi sebesar 56.2%. Kedua model ini efektif dalam memprediksi pelanggan yang berpotensi terlambat membayar tagihan listrik. Penelitian ini memberikan kontribusi signifikan bagi PLN dalam pengambilan keputusan berbasis data, serta mendukung strategi mitigasi seperti pemberitahuan dini atau pengaturan ulang jadwal pembayaran untuk mengurangi risiko tunggakan pembayaran.

Kata Kunci - Keterlambatan Pembayaran, Prediksi, Random Forest, PLN, Moving Average

I. PENDAHULUAN

Listrik adalah kebutuhan dasar yang sangat penting bagi kehidupan modern[1]. Selain mendukung aktivitas harian, seperti penerangan dan penggunaan alat elektronik, listrik juga berperan dalam perkembangan ekonomi dan industri[2]. Oleh karena itu, pengelolaan sistem pembayaran rekening listrik menjadi krusial untuk memastikan kontinuitas layanan dan stabilitas finansial perusahaan penyedia listrik[3]. PLN (Perusahaan Listrik Negara) sebagai BUMN memiliki tanggung jawab besar untuk menyediakan layanan kelistrikan yang terjangkau bagi seluruh masyarakat[4]. Namun, PLN UP3 Makassar Utara menghadapi permasalahan serius berupa keterlambatan pembayaran rekening listrik oleh pelanggan. Berdasarkan data internal PLN, sekitar 25% pelanggan melewati batas waktu pembayaran yang telah ditentukan, yaitu tanggal 20 setiap bulan, sebagaimana diatur dalam Surat Perjanjian Jual Beli Tenaga Listrik (SPJBTL). Tunggakan pembayaran menunjukkan tren peningkatan selama beberapa tahun terakhir, dengan total tunggakan mencapai Rp11 miliar pada April 2024. Angka ini mencerminkan peningkatan sebesar 18% dibandingkan dengan periode yang sama pada tahun sebelumnya, yang berdampak negatif pada operasional dan keuangan perusahaan.

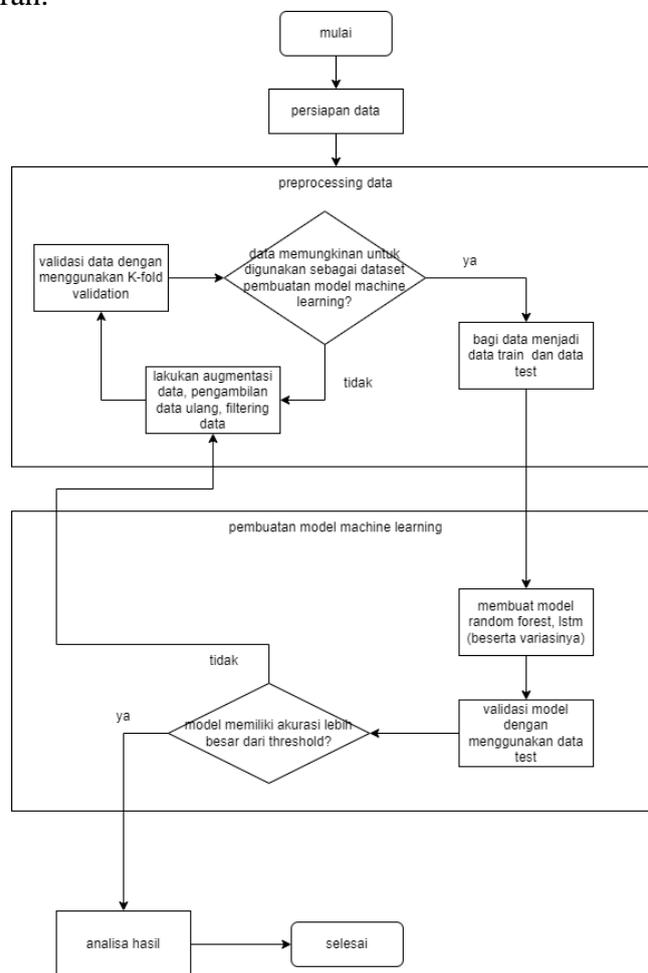


Gambar 1. Tren Pembayaran Rekening Listrik

Keterlambatan pembayaran ini disebabkan oleh beberapa faktor, termasuk perilaku pelanggan yang sering melewati batas waktu pembayaran serta sistem pengiriman invoice yang kurang efektif. Saat ini, sistem mengirimkan invoice secara massal tanpa mempertimbangkan pola pembayaran individu pelanggan, sehingga tidak selalu sampai tepat waktu, yang akhirnya meningkatkan saldo tunggakan[5]. Penelitian sebelumnya oleh [3] menunjukkan bahwa atribut seperti tanggal penerimaan gaji merupakan faktor utama yang menyebabkan konsumen terlambat membayar tagihan listrik. Penelitian menggunakan algoritma C4.5 dan SVM ini berhasil menganalisis faktor keterlambatan dengan tingkat akurasi yang baik. Selain itu, penelitian oleh [6] membandingkan metode seleksi fitur dalam memprediksi keterlambatan pembayaran biaya kuliah, di mana algoritma Random Forest dan C4.5 menunjukkan akurasi tertinggi, masing-masing dengan precision sebesar 65% dan recall sebesar 63%. Penelitian ini bertujuan untuk mengembangkan sistem prediksi keterlambatan pembayaran yang dapat mendukung PLN UP3 Makassar Utara dalam menentukan prioritas pengiriman invoice kepada pelanggan yang berisiko terlambat membayar. Penelitian ini memanfaatkan data historis untuk menganalisis pola keterlambatan pembayaran guna memastikan keakuratan data yang digunakan dalam analisis. Selanjutnya, model prediksi berbasis algoritma Random Forest dirancang untuk mendeteksi pelanggan dengan risiko keterlambatan pembayaran. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi langkah-langkah mitigasi yang dapat diterapkan, seperti pemberitahuan dini atau penjadwalan ulang pembayaran, guna membantu mengurangi tunggakan pembayaran. Hipotesis yang diajukan adalah bahwa algoritma Random Forest dapat memberikan prediksi yang akurat dalam mengidentifikasi pelanggan yang berisiko terlambat membayar, dengan nilai Mean Absolute Error (MAE) yang rendah ketika menggunakan fitur Moving Average.

II. SIGNIFIKASI STUDI

Penelitian ini memiliki signifikansi penting dalam membantu PLN UP3 Makassar Utara mengatasi permasalahan keterlambatan pembayaran listrik yang berdampak pada operasional dan stabilitas keuangan perusahaan. Dengan menggunakan algoritma Random Forest, penelitian ini berkontribusi dalam menyediakan model prediktif yang dapat mengidentifikasi pelanggan berisiko terlambat membayar secara akurat. Pemilihan Random Forest sebagai metode utama didasarkan pada beberapa kelebihan algoritma ini dalam menangani karakteristik data yang kompleks. Random Forest mampu mengolah data dengan banyak variabel kategori maupun numerik, serta data yang tidak terdistribusi secara normal. Algoritma ini juga efektif dalam mengidentifikasi pola dari dataset besar dengan fitur-fitur yang memiliki korelasi tinggi, seperti data pembayaran listrik pelanggan yang melibatkan elemen historis dan variabel demografis. Kemampuan algoritma untuk memitigasi overfitting melalui penggunaan pohon keputusan yang dibangun secara independen menjadikannya pilihan ideal untuk data deret waktu yang dianalisis dalam penelitian ini. Alur penelitian secara keseluruhan mengikuti tahapan yang ditunjukkan pada Gambar 2, dimulai dari pengumpulan data historis pelanggan hingga implementasi model prediktif berbasis Random Forest untuk meminimalkan keterlambatan pembayaran.



Gambar 2. Alur Penelitian

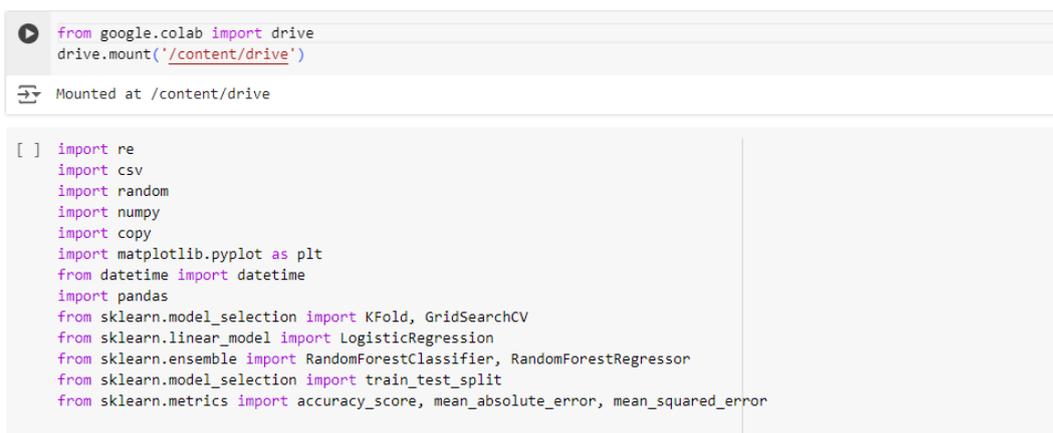
Hasil penelitian ini memiliki implikasi praktis yang signifikan bagi PLN maupun perusahaan penyedia layanan listrik lainnya. Model prediktif yang dikembangkan dapat diintegrasikan ke dalam sistem operasional PLN, misalnya dengan mengotomatisasi prioritas pengiriman invoice atau pemberian notifikasi dini kepada pelanggan yang teridentifikasi berisiko tinggi. Integrasi ini dapat meningkatkan efisiensi proses pengelolaan pembayaran dan mengurangi jumlah tunggakan secara signifikan.

Selain itu, penerapan model ini berpotensi memberikan dampak ekonomi yang positif, seperti menekan biaya operasional yang diakibatkan oleh penanganan keterlambatan pembayaran, sekaligus meningkatkan arus kas perusahaan. Dengan pengelolaan risiko yang lebih baik, PLN dapat mengalokasikan sumber daya secara lebih efisien dan mempertahankan stabilitas keuangan jangka panjang. Model ini juga dapat diadaptasi oleh perusahaan lain yang menghadapi tantangan serupa dalam mengelola pembayaran pelanggan, menjadikan hasil penelitian ini relevan secara luas dalam sektor utilitas dan layanan publik.

III. HASIL DAN PEMBAHASAN

Persiapan Data dan Normalisasi

Data yang digunakan dalam penelitian ini diimpor dari sumber yang relevan, mencakup riwayat pembayaran rekening listrik pelanggan tahun 2018-2023. Proses normalisasi dilakukan untuk memastikan konsistensi dan keakuratan data. Normalisasi ini membantu model memahami perbedaan antara pembayaran tepat waktu dan keterlambatan pembayaran. Nilai biner digunakan, di mana 0 menunjukkan pembayaran tepat waktu, dan 1 menunjukkan keterlambatan. Normalisasi ini menghasilkan dataset terstruktur yang memungkinkan algoritma untuk mengenali pola keterlambatan berdasarkan data historis.



```

from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[ ] import re
import csv
import random
import numpy
import copy
import matplotlib.pyplot as plt
from datetime import datetime
import pandas
from sklearn.model_selection import KFold, GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, mean_absolute_error, mean_squared_error

```

Gambar 3. Import Data Pelunasan Tahun 2018-2023.csv

Penggunaan *Moving Average* dan Pemilihan Fitur

Pada bagian ini, dilakukan proses pemodelan untuk memprediksi keterlambatan pembayaran listrik oleh pelanggan. Algoritma *Random Forest Regressor* digunakan dalam eksperimen ini, dengan beberapa fitur tambahan yang dibangun berdasarkan data historis keterlambatan pembayaran setiap pelanggan. Fitur *Moving Average* dipilih karena kemampuannya untuk menangkap pola historis dalam jangka waktu tertentu. Dalam penelitian ini, tiga periode *Moving Average* diuji (3 bulan, 6 bulan, dan 12 bulan). Fitur ini memberikan konteks tren keterlambatan dalam jangka pendek hingga jangka panjang.

Modelling Data dengan *Moving Average* 3 Bulan Mempersiapkan Data untuk Pemodelan

Langkah pertama dalam pemodelan ini adalah membentuk dataset yang siap untuk digunakan dalam model prediksi. Dataset terdiri dari beberapa fitur penting, yaitu:

1. **CustomerID**: ID pelanggan untuk setiap baris data.
2. **Date**: Tanggal pembayaran yang dihasilkan menggunakan fungsi `pandas.date_range()`, dengan frekuensi bulanan. Di sini, periode tanggal ditentukan dari Januari 2018 dengan jumlah bulan sesuai dengan data yang ada.

3. **Delay:** Nilai biner yang menunjukkan apakah pembayaran dilakukan terlambat (1) atau tepat waktu (0).

Setelah menentukan fitur dasar ini, data tambahan dibuat sebagai variabel fitur untuk model prediksi. Fitur yang ditambahkan meliputi:

1. **Lag Feature (Fitur Lag)** Setiap pelanggan memiliki data historis keterlambatan pembayaran, sehingga fitur lag diperkenalkan untuk mencerminkan keterlambatan pembayaran pada bulan-bulan sebelumnya. Fitur ini berguna untuk memberikan konteks historis dalam proses prediksi. Fitur lag ini dihitung menggunakan metode `shift()` berdasarkan `CustomerID`, sehingga setiap pelanggan memiliki catatan keterlambatan beberapa bulan sebelumnya (`lag_1`, `lag_2`, dst.).
2. **Moving Average (Rata-rata Bergerak)** Untuk mengukur tren keterlambatan pembayaran, fitur **moving average** 3 bulan dibuat. Fitur ini menghitung rata-rata keterlambatan selama tiga bulan sebelumnya dan digeser satu bulan ke depan untuk digunakan dalam prediksi bulan berikutnya. Fitur ini membantu model memahami pola keterlambatan dalam jangka waktu tertentu dan memperhitungkan perilaku yang mungkin terjadi berulang kali.
3. **Month (Bulan)** Fitur tambahan berupa bulan (1–12) dari setiap tanggal pembayaran digunakan, karena keterlambatan pembayaran mungkin bervariasi berdasarkan musim atau bulan tertentu. Fitur ini diambil langsung dari kolom `Date`.

Pembagian Data menjadi Train dan Test Set

Setelah fitur-fitur disiapkan, dataset dibagi menjadi **train set** dan **test set**. Dalam eksperimen ini, pembagian data dilakukan secara temporal, tanpa pengacakan, untuk mensimulasikan prediksi di masa depan berdasarkan data historis. Pembagian train set dan test set untuk model Random Forest ini adalah 80% dan 20%.

Pelatihan Model

Model yang digunakan dalam pemodelan ini adalah **Random Forest Regressor** dari `scikit-learn`, yang memiliki kemampuan untuk menangani masalah regresi dengan menggunakan beberapa pohon keputusan (`decision trees`) sebagai basisnya.

Model ini dilatih dengan berbagai jumlah estimator (pohon) untuk melihat bagaimana jumlah pohon mempengaruhi performa model.

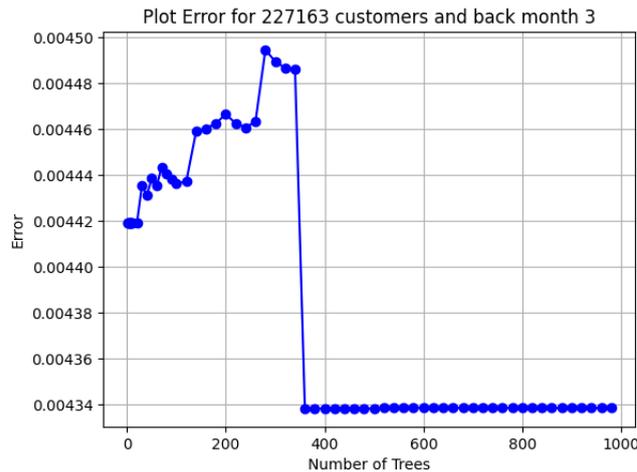
1. **RandomForestRegressor(`n_estimators=tree`,`random_state=42`)** digunakan untuk melatih model pada berbagai jumlah pohon, dari 1 hingga 280 pohon.
2. Parameter `random_state=42` digunakan untuk memastikan replikasi hasil pada setiap percobaan.

Prediksi dan Evaluasi Model

Setelah model dilatih, langkah berikutnya adalah melakukan prediksi pada **X_{test}**, yaitu data yang belum pernah dilihat oleh model. Prediksi ini kemudian dievaluasi menggunakan metrik **Mean Absolute Error (MAE)** untuk mengukur rata-rata kesalahan antara prediksi model dan nilai aktual keterlambatan pembayaran. **Moving Average 3 Bulan** memberikan hasil yang kurang optimal karena hanya menangkap pola jangka pendek, dengan MAE sebesar 0.0044..

Plotting Error terhadap Jumlah Pohon

Setelah model dilatih dan dievaluasi, langkah selanjutnya adalah memvisualisasikan hasil untuk memahami performa model dalam berbagai skenario jumlah pohon (estimators) pada **Random Forest Regressor**. Visualisasi ini dilakukan dengan menggunakan grafik garis untuk menampilkan hubungan antara jumlah pohon yang digunakan dalam model dan nilai **Mean Absolute Error (MAE)** sebagai metrik evaluasi. Berikut adalah beberapa grafik yang menampilkan hubungan antara **Mean Absolute Error (MAE)** dan jumlah pohon yang digunakan dalam model untuk dataset dengan keseluruhan data.



Gambar 4. Error untuk 227.163 Pelanggan dengan Back Month 3

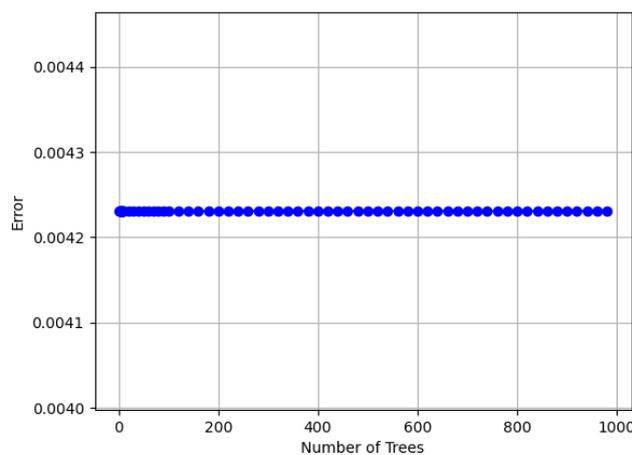
Grafik ini menunjukkan performa model untuk dataset besar dengan 227,163 pelanggan. Dari grafik tersebut, terlihat bahwa pada jumlah pohon awal (di bawah 50), nilai error relatif fluktuatif karena model belum cukup kompleks untuk menangkap pola dalam data yang besar.

Modelling Data dengan Moving Average 6 Bulan

Pemodelan menggunakan fitur **Moving Average 6 Bulan** dirancang untuk memahami dan memanfaatkan pola historis keterlambatan pembayaran dalam periode waktu yang lebih panjang. Pendekatan ini bertujuan untuk memberikan konteks perilaku pelanggan berdasarkan tren jangka menengah (enam bulan) dibandingkan pola bulanan yang lebih pendek.

Moving Average 6 Bulan lebih baik, dengan MAE sebesar 0.0042, tetapi masih terdapat fluktuasi pada jumlah pohon besar.

Plotting Error terhadap Jumlah Pohon Berikut adalah beberapa grafik yang menampilkan hubungan antara **Mean Absolute Error (MAE)** dan jumlah pohon yang digunakan dalam model untuk dataset dengan keseluruhan data.



Gambar 5. Error untuk 227.163 Pelanggan dengan Back Month 6

Grafik ini menunjukkan performa model dengan dataset yang sangat besar, yaitu 227.163 pelanggan. Error stabil sejak awal, tanpa fluktuasi berarti bahkan pada jumlah pohon yang sangat kecil. Setelah

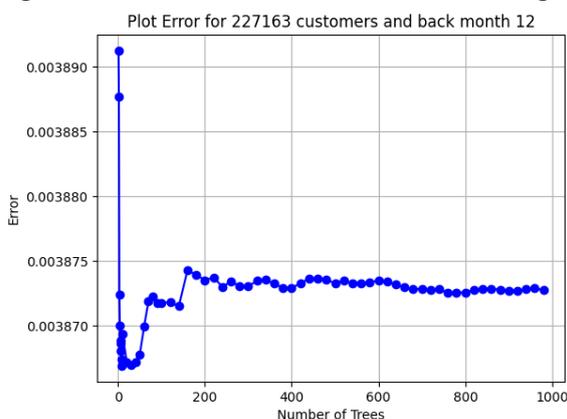
mencapai sekitar 100–200 pohon, error tetap konstan hingga 1000 pohon, menunjukkan bahwa menambah jumlah pohon lebih lanjut tidak memberikan perbaikan yang signifikan. Dataset besar memberikan cukup informasi bagi model untuk menghasilkan prediksi yang sangat konsisten dan stabil.

Modelling Data dengan Moving Average 12 Bulan

Pemodelan menggunakan fitur **Moving Average 12 Bulan** dirancang untuk memahami pola keterlambatan pembayaran pelanggan dalam periode waktu yang lebih panjang dibandingkan Moving Average 6 Bulan. Pendekatan ini memberikan gambaran lebih luas tentang perilaku pelanggan, sehingga membantu model memprediksi keterlambatan pembayaran dengan lebih akurat. Model menghasilkan prediksi yang konsisten dengan nilai **Mean Absolute Error (MAE)** yang tetap stabil. **Moving Average 12 Bulan** memberikan performa terbaik dengan MAE sebesar 0.00387. Hal ini menunjukkan bahwa periode historis yang lebih panjang membantu model memahami pola keterlambatan jangka panjang dengan lebih baik.

Plotting Error terhadap Jumlah Pohon

Berikut adalah beberapa grafik yang menampilkan hubungan antara **Mean Absolute Error (MAE)** dan jumlah pohon yang digunakan dalam model untuk dataset dengan keseluruhan data.



Gambar 6. Error untuk 227.163 Pelanggan dengan Back Month 12

Dataset sangat besar ini menghasilkan grafik dengan error yang stabil dan konsisten setelah mencapai sekitar 100–200 pohon. Penambahan pohon lebih lanjut tidak memberikan peningkatan signifikan dalam performa model, karena dataset telah cukup besar untuk menangkap pola dengan baik.

Perbandingan Evaluasi Model pada Random Forest

Pemodelan menggunakan fitur Moving Average dengan periode 3, 6, dan 12 bulan memberikan gambaran mengenai pola keterlambatan pembayaran pelanggan dalam jangka waktu yang berbeda. Setiap periode Moving Average menawarkan tingkat akurasi yang berbeda berdasarkan kemampuan model dalam mengenali pola pembayaran dari data historis. Pada subbab ini, dibandingkan hasil error dari ketiga periode Moving Average untuk menentukan periode yang menghasilkan performa terbaik. Perbandingan evaluasi model menggunakan algoritma Random Forest dengan fitur Moving Average pada periode 3, 6, dan 12 bulan memberikan wawasan penting mengenai efektivitas setiap periode dalam memprediksi keterlambatan pembayaran pelanggan. Setiap periode Moving Average menawarkan tingkat akurasi yang berbeda, tergantung pada kemampuan model dalam mengenali pola pembayaran dari data historis. Moving Average 3 bulan, meskipun lebih cepat dalam menangkap perubahan jangka pendek, tidak cukup untuk menggambarkan pola jangka panjang, sehingga menghasilkan tingkat error yang lebih tinggi dibandingkan dengan periode lainnya. Moving Average 6 bulan sedikit lebih baik, tetapi masih menunjukkan fluktuasi error yang cukup besar pada jumlah pohon yang lebih tinggi. Sebaliknya, Moving Average 12 bulan menghasilkan performa terbaik

dengan error yang paling rendah dan stabil, mencerminkan kemampuan model untuk mengenali pola jangka panjang dengan lebih akurat.

Penelitian oleh [7] menemukan bahwa penggunaan Moving Average dengan periode yang lebih panjang dapat meningkatkan stabilitas model, terutama dalam memprediksi tren yang lebih lama dan variabel dengan fluktuasi musiman. Selain itu, penelitian oleh [8] menunjukkan bahwa Moving Average dengan periode lebih panjang mengurangi fluktuasi error dalam model prediksi pembayaran, yang pada gilirannya meningkatkan akurasi dan efisiensi model dalam memprediksi pola keterlambatan. Penelitian internasional oleh [9] juga mendukung temuan ini, di mana mereka melaporkan bahwa Moving Average 12 bulan memberikan hasil yang lebih baik dalam pemodelan data yang memiliki pola musiman atau tren jangka panjang, dengan error yang lebih kecil dan lebih konsisten dibandingkan periode yang lebih pendek. Secara keseluruhan, hasil perbandingan ini menunjukkan bahwa Moving Average 12 bulan merupakan pilihan terbaik untuk memodelkan keterlambatan pembayaran, terutama ketika dataset yang digunakan mencakup variabilitas musiman atau pola jangka panjang yang perlu dikenali oleh model predik.

Analisis Performa Model

Dari hasil pengamatan grafik dan nilai **Mean Absolute Error (MAE)**:

1. **Moving Average 3 Bulan** hanya cocok untuk memodelkan pola jangka pendek, tetapi gagal menangkap tren perilaku pelanggan dalam jangka panjang. Hal ini menyebabkan nilai MAE sebesar **0.0044**, yang lebih tinggi dibandingkan periode lainnya.
2. **Moving Average 6 Bulan** memberikan hasil yang lebih baik dibandingkan 3 bulan, dengan nilai MAE sebesar **0.0042**. Namun, fluktuasi kecil tetap terlihat pada jumlah pohon yang besar.
3. **Moving Average 12 Bulan** memberikan hasil terbaik di antara ketiga periode, dengan nilai MAE yang paling kecil, yaitu **0.00387**, serta performa yang paling stabil. Informasi historis yang lebih panjang membantu model menangkap pola jangka panjang dengan lebih baik, menghasilkan prediksi yang lebih akurat.

Berdasarkan analisis grafik dan nilai **Mean Absolute Error (MAE)**, model dengan fitur **Moving Average 12 Bulan** menunjukkan performa terbaik di antara ketiga periode. Hal ini terlihat dari nilai MAE sebesar **0.00387**. Grafik Moving Average 12 Bulan juga menunjukkan error yang paling stabil, bahkan pada jumlah pohon yang rendah, dengan penurunan error yang signifikan hingga mencapai stabilitas pada sekitar 100 pohon. Informasi historis yang lebih panjang dalam Moving Average 12 Bulan membantu model menangkap pola jangka panjang dengan lebih baik, sehingga menghasilkan prediksi yang lebih akurat dan konsisten. Oleh karena itu, fitur Moving Average 12 Bulan direkomendasikan sebagai pilihan utama untuk memodelkan keterlambatan pembayaran, terutama pada dataset dengan skala besar. Fitur tambahan seperti Lag Features dan Month juga memberikan kontribusi signifikan, meskipun fitur bulanan (Month) memiliki pengaruh yang lebih kecil dibandingkan fitur lainnya dalam memprediksi keterlambatan pembayaran. Tidak ada fitur yang secara signifikan mengurangi performa model, tetapi hasil ini mengindikasikan bahwa menambah lebih banyak fitur historis atau kategori tambahan mungkin dapat meningkatkan akurasi model. MAE sebesar 0.0044 pada Moving Average 3 bulan dan 0.00387 pada 12 bulan menunjukkan bahwa rata-rata kesalahan prediksi model terhadap nilai aktual sangat rendah. Dalam konteks keterlambatan pembayaran, nilai MAE ini mengindikasikan bahwa model dapat memprediksi keterlambatan pembayaran pelanggan dengan rata-rata selisih waktu kurang dari 1 hari. Hasil analisis menggunakan grafik dan nilai Mean Absolute Error (MAE) menunjukkan perbedaan signifikan dalam performa model dengan berbagai periode Moving Average (MA). Model dengan Moving Average 3 bulan hanya mampu memodelkan pola jangka pendek, namun gagal menangkap tren perilaku pelanggan dalam jangka panjang. Hal ini tercermin dari nilai MAE yang relatif lebih tinggi, yaitu 0.0044, dibandingkan dengan periode lainnya. Moving Average 6 bulan memberikan hasil yang lebih baik dengan nilai MAE sebesar 0.0042, meskipun fluktuasi kecil tetap muncul pada jumlah pohon yang besar. Namun, Moving Average 12 bulan menghasilkan performa terbaik, dengan nilai MAE

terendah yaitu 0.00387 dan performa yang paling stabil. Penggunaan informasi historis lebih panjang dalam Moving Average 12 bulan memungkinkan model untuk menangkap pola jangka panjang dengan lebih baik, menghasilkan prediksi yang lebih akurat dan konsisten. Penelitian oleh [10] menunjukkan bahwa Moving Average dengan periode yang lebih panjang lebih efektif dalam memprediksi variabel yang memiliki pola musiman atau tren jangka panjang, terutama dalam konteks prediksi keuangan. Penelitian lainnya oleh [11] menambahkan bahwa penggunaan Moving Average yang lebih panjang membantu model dalam menangkap fluktuasi yang lebih luas dalam data historis, sehingga memperbaiki akurasi prediksi. Selain itu, studi internasional oleh [12] mengungkapkan bahwa pendekatan berbasis Moving Average dengan informasi lebih panjang dapat mengurangi volatilitas error dalam prediksi, memberikan hasil yang lebih stabil dan dapat diandalkan dalam prediksi jangka panjang.

Berdasarkan temuan ini, fitur Moving Average 12 Bulan sangat direkomendasikan untuk memodelkan keterlambatan pembayaran, terutama dalam dataset dengan skala besar. Hasil ini juga menunjukkan bahwa penambahan fitur lain seperti Lag Features dan Month dapat memberikan kontribusi positif, meskipun pengaruh fitur bulanan lebih kecil dibandingkan fitur lainnya. Meskipun demikian, penambahan lebih banyak fitur historis atau kategori tambahan mungkin dapat lebih meningkatkan akurasi model, tanpa secara signifikan mengurangi performanya.

Validasi Data dengan Data Test

Pada hal ini, model **Random Forest Regressor** diuji menggunakan data test untuk mengevaluasi kinerja prediksinya terhadap keterlambatan pembayaran listrik. Model dilatih menggunakan data historis keterlambatan dengan variabel tambahan yang merepresentasikan pola musiman dan dilengkapi dengan lag periode 12 bulan. Validasi dilakukan dengan membandingkan prediksi model terhadap nilai aktual pada data test. Model **Random Forest Regressor** diuji menggunakan data test untuk mengevaluasi kinerjanya dalam memprediksi keterlambatan pembayaran listrik. Model ini dilatih menggunakan data historis yang mencakup keterlambatan pembayaran sebelumnya, dengan variabel tambahan berupa pola musiman dan lag periode selama 12 bulan. Validasi dilakukan dengan membandingkan hasil prediksi model terhadap nilai aktual menggunakan data test, yang memberikan gambaran tentang kemampuan model dalam menangkap pola keterlambatan pembayaran berdasarkan data historis. Hasil ini sejalan dengan penelitian [13] yang menemukan bahwa penggunaan variabel musiman dan lag waktu dalam algoritma Random Forest secara signifikan meningkatkan akurasi prediksi dalam analisis perilaku konsumen pada data transaksi listrik Prabayar. Penelitian lain oleh [14] menegaskan bahwa model berbasis Random Forest memiliki keunggulan dalam memproses data dengan dimensi temporal yang kompleks, dibandingkan algoritma prediksi lainnya, seperti Support Vector Regression. Sementara itu, studi internasional oleh [15] mengungkapkan bahwa penerapan Random Forest pada dataset berbasis waktu menghasilkan prediksi yang lebih konsisten dalam kondisi variasi data yang tinggi, khususnya dalam sektor energi. Dengan pendekatan ini, validasi model membuktikan keandalan algoritma Random Forest dalam memprediksi keterlambatan pembayaran. Model ini tidak hanya memberikan wawasan yang bermanfaat untuk pengelolaan operasional, tetapi juga berpotensi mendukung kebijakan proaktif, seperti pemberian notifikasi dini atau opsi pembayaran fleksibel bagi pelanggan berisiko tinggi.

Proses Validasi

Proses validasi model prediktif menggunakan dataset yang diolah dalam format jendela waktu (window) selama 12 bulan untuk memprediksi nilai keterlambatan pada bulan ke-13 menunjukkan pendekatan berbasis data temporal yang efektif. Penambahan variabel musiman berupa bulan berjalan membantu menangkap pola periodik keterlambatan pembayaran. Model prediksi menggunakan algoritma Random Forest Regressor dengan parameter utama seperti jumlah estimator sebanyak 220, porsi data uji sebesar 10% dari total dataset, dan pengaturan `random_state` sebesar 42 untuk memastikan hasil konsisten. Evaluasi performa model dilakukan dengan membandingkan hasil prediksi dan nilai aktual menggunakan Mean Squared Error (MSE) sebagai metrik utama, dengan

MSE rata-rata sebagai indikator kualitas model. Hasil ini sejalan dengan penelitian [16] yang menunjukkan bahwa metode Random Forest memiliki keunggulan dalam memprediksi perilaku pelanggan dengan pola temporal karena kemampuan algoritma menangkap interaksi antar fitur yang kompleks. Penelitian lain oleh [17] juga menekankan pentingnya penggunaan variabel musiman dalam model prediksi, terutama dalam konteks analisis keuangan, untuk meningkatkan akurasi hasil. Selain itu, studi terbaru oleh [18] mengungkap bahwa penerapan Random Forest Regressor dalam analisis data berbasis waktu menghasilkan prediksi yang lebih stabil dibandingkan metode lain, seperti Support Vector Regression (SVR), terutama pada dataset dengan pola yang tidak teratur. Dengan pendekatan ini, model prediktif mampu memberikan hasil yang dapat diandalkan untuk mendukung pengelolaan keterlambatan pembayaran. Integrasi model ini berpotensi meningkatkan efisiensi operasional dan mengurangi risiko keterlambatan pembayaran melalui implementasi strategi berbasis data. Adopsi model prediksi ini juga membuka peluang untuk pengembangan kebijakan baru yang lebih responsif terhadap pola keterlambatan pelanggan.

Hasil Validasi

Rata-rata MSE pada data pengujian adalah 3.9089, yang bervariasi berdasarkan pola keterlambatan pelanggan. Pada pelanggan dengan pola keterlambatan rendah, model menghasilkan prediksi yang sangat akurat (MSE mendekati 0). Namun, untuk pelanggan dengan variasi keterlambatan yang signifikan, MSE menjadi lebih tinggi, menunjukkan bahwa pola yang tidak teratur membutuhkan fitur tambahan atau pendekatan lain untuk meningkatkan akurasi. Model prediktif yang dikembangkan memiliki potensi implementasi langsung dalam operasi PLN. Temuan ini dapat digunakan untuk mendukung pengelolaan pelanggan dengan mengidentifikasi pelanggan berisiko tinggi keterlambatan pembayaran, sehingga memungkinkan prioritas pengiriman invoice atau pemberian notifikasi dini. Selain itu, hasil penelitian ini juga dapat mendorong pembuatan kebijakan baru, seperti pengadopsian sistem pengelolaan tagihan berbasis risiko untuk mengurangi jumlah tunggakan dan menekan biaya operasional.

Strategi lain yang dapat diimplementasikan adalah memberikan opsi pembayaran yang lebih fleksibel kepada pelanggan berisiko tinggi, misalnya melalui penjadwalan ulang tagihan atau pengingat pembayaran yang lebih personal. Namun, implementasi model ini dihadapkan pada beberapa tantangan, seperti integrasi model prediksi ke dalam sistem PLN yang sudah ada, yang memerlukan waktu dan sumber daya tambahan. Selain itu, ketersediaan data secara real-time menjadi faktor penting untuk menjaga akurasi prediksi, dan penerimaan pelanggan terhadap pemberitahuan dini atau penjadwalan ulang perlu diperhatikan agar strategi ini dapat berjalan dengan optimal. Penelitian ini menunjukkan bahwa algoritma Random Forest, terutama dengan fitur Moving Average 12 bulan, merupakan pendekatan yang efektif dalam memprediksi keterlambatan pembayaran tagihan listrik. Dengan penerapan model ini, PLN diharapkan dapat meningkatkan efisiensi operasional, mengurangi tunggakan, serta menjaga stabilitas keuangan perusahaan. Langkah selanjutnya adalah mengatasi tantangan implementasi yang ada dan memastikan adaptasi sistem ini untuk digunakan dalam skala yang lebih luas. Analisis prediksi keterlambatan pembayaran tagihan listrik menggunakan algoritma Random Forest menunjukkan bahwa model ini efektif dalam mengidentifikasi pelanggan yang berisiko tinggi mengalami keterlambatan pembayaran. Hal ini sejalan dengan penelitian sebelumnya yang membandingkan metode klasifikasi Decision Tree dan Support Vector Machine dalam memprediksi keterlambatan pembayaran listrik. Hasil penelitian tersebut menunjukkan bahwa metode Decision Tree memiliki akurasi yang lebih tinggi dibandingkan Support Vector Machine dalam memprediksi keterlambatan pembayaran listrik [19]. Selain itu, penelitian lain menggunakan metode K-Means dan XGBoost untuk menganalisis perilaku pelanggan listrik. Hasilnya menunjukkan bahwa kombinasi kedua metode tersebut mampu mengelompokkan pelanggan berdasarkan perilaku pembayaran dan memprediksi keterlambatan pembayaran dengan akurasi yang baik [20]. Penelitian lain juga membahas prediksi keterlambatan pembayaran rekening listrik menggunakan metode Random Forest. Hasil penelitian tersebut menunjukkan bahwa metode Random Forest memiliki akurasi yang tinggi dalam memprediksi keterlambatan pembayaran rekening listrik [21]. Dengan demikian, implementasi model prediksi keterlambatan pembayaran

tagihan listrik menggunakan algoritma Random Forest dapat membantu PLN dalam mengidentifikasi pelanggan yang berisiko tinggi mengalami keterlambatan pembayaran. Hal ini memungkinkan PLN untuk mengambil tindakan preventif, seperti memberikan notifikasi dini atau menawarkan opsi pembayaran yang lebih fleksibel, guna meningkatkan efisiensi operasional dan mengurangi jumlah tunggakan.

IV. KESIMPULAN

Berdasarkan hasil penelitian mengenai prediksi keterlambatan pembayaran listrik di PLN UP3 Makassar Utara menggunakan algoritma Random Forest, dapat disimpulkan beberapa hal. Algoritma Random Forest menunjukkan kinerja yang sangat baik dalam memprediksi keterlambatan pembayaran, terutama dengan menggunakan fitur Moving Average 12 bulan. Model ini menghasilkan nilai Mean Absolute Error (MAE) terendah sebesar 0.00387, yang menandakan kemampuan model dalam mengenali pola jangka panjang pelanggan secara akurat. Jika dibandingkan dengan penelitian serupa yang menggunakan metode lain seperti C4.5 atau Support Vector Machine, nilai MAE dari Random Forest ini lebih rendah, menunjukkan peningkatan signifikan dalam akurasi prediksi keterlambatan pembayaran. Selain akurasi, model ini juga efisien dalam hal waktu komputasi, dengan jumlah pohon optimal berada pada kisaran 100–200 pohon. Stabilitas model dalam kisaran ini memastikan prediksi yang konsisten tanpa peningkatan beban komputasi yang signifikan.

Model prediktif ini memiliki potensi besar untuk membantu PLN dalam mengelola risiko keterlambatan pembayaran. Hasil prediksi dapat diterapkan dalam operasional dengan langkah-langkah praktis seperti mengidentifikasi pelanggan berisiko tinggi dan memprioritaskan pengiriman invoice atau pemberian notifikasi dini kepada mereka. Selain itu, PLN dapat mengembangkan strategi mitigasi seperti pemberitahuan lebih awal atau memberikan opsi penjadwalan ulang pembayaran yang lebih fleksibel untuk pelanggan yang berpotensi menunggak. Dengan demikian, sistem prediksi ini dapat mengurangi jumlah tunggakan pembayaran secara signifikan, meningkatkan arus kas, dan mendukung stabilitas keuangan perusahaan.

Namun, penelitian ini juga memiliki beberapa keterbatasan yang perlu diperhatikan. Model tidak mempertimbangkan faktor-faktor eksternal seperti kondisi ekonomi makro atau perubahan kebijakan tarif listrik yang dapat memengaruhi perilaku pembayaran pelanggan. Selain itu, model hanya menggunakan data historis pelanggan dari PLN UP3 Makassar Utara, sehingga hasilnya mungkin tidak sepenuhnya berlaku untuk wilayah lain dengan karakteristik pelanggan yang berbeda. Pembaruan data secara berkala juga diperlukan untuk menjaga akurasi prediksi, yang dapat menjadi tantangan logistik bagi PLN. Ke depan, penelitian ini dapat diperluas dengan mengintegrasikan variabel tambahan yang lebih kompleks atau membandingkan algoritma lainnya untuk memastikan model yang paling sesuai digunakan di berbagai wilayah operasional PLN. Dengan adaptasi dan pengembangan lebih lanjut, model ini dapat menjadi alat strategis dalam pengelolaan pelanggan dan pengurangan risiko keterlambatan pembayaran di sektor kelistrikan.

REFERENSI

- [1] S. Sutialisna, P. N. Sabrina, dan H. Ashaury, "Prediksi Tagihan Listrik Menggunakan Metode Weighted Exponential Moving Average (WEMA)," *JATI J. Mhs. Tek. Inform.*, vol. 7, no. 5, Art. no. 5, 2023, doi: 10.36040/jati.v7i5.7478.
- [2] E. Riduwan, D. Purwitasari, dan A. B. Raharjo, "Identifikasi Fitur untuk Prediksi Penerimaan Program Listrik Prabayar: Kasus di PLN Tahuna.," *Techno Com*, vol. 21, no. 3, 2022, Diakses: 13 Januari 2025. [Daring]. Tersedia pada: <https://www.academia.edu/download/103557529/3095.pdf>
- [3] D. N. Batubara, A. P. Windarto, dan E. Irawan, "Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine," *JURIKOM J. Ris. Komput.*, vol. 9, no. 1, hlm. 102–108, 2022.

- [4] Y. P. Dewi, "Prosedur Penanganan Tunggakan Rekening Listrik Pascabayar pada PT PLN (Persero) Unit Layanan Pelanggan Sanur," diploma, Politeknik Negeri Bali, 2024. Diakses: 13 Januari 2025. [Daring]. Tersedia pada: <https://repository.pnb.ac.id/15405/>
- [5] R. H. Nugraha, D. Purwitasari, dan A. B. Raharjo, "K-Means And Xgboost For Customer Electricity Account Payment Behavior Analysis (Case Study: Pln Ulp Panakkukang)," *JUTI J. Ilm. Teknol. Inf.*, doi: <http://dx.doi.org/10.12962/j24068535.v20i2.a1132>.
- [6] T. A. Y. Siswa dan R. P. Wibowo, "Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah," *Teknika*, vol. 12, no. 1, Art. no. 1, Mar 2023, doi: 10.34148/teknika.v12i1.601.
- [7] A. Nurlifa dan S. Kusumadewi, "Sistem Peramalan Jumlah Penjualan Menggunakan Metode Moving Average Pada Rumah Jilbab Zaky," *INOVTEK Polbeng - Seri Inform.*, vol. 2, no. 1, Art. no. 1, Jun 2017, doi: 10.35314/isi.v2i1.112.
- [8] Tedyyana, Agus, Osman Ghazali, and Onno W. Purbo. "Machine learning for network defense: automated DDoS detection with telegram notification integration." *Indonesian Journal of Electrical Engineering and Computer Science* 34.2 (2024): 1102..
- [9] S. Putri, F. Badruzzaman, dan E. Harahap, "Perbandingan Metode Single Moving Average dan Single Exponential Smoothing dalam Peramalan Jumlah Pengguna Pospay pada PT Pos Indonesia KCU Bandung," *Mat. J. Teori Dan Terap. Mat.*, vol. 22, no. 1, Art. no. 1, Mei 2023.
- [10] Y. Astuti, B. Novianti, T. Hidayat, dan D. Maulina, "Penerapan Metode Single Moving Average Untuk Peramalan Penjualan Mainan Anak," *SENSITIf Semin. Nas. Sist. Inf. Dan Teknol. Inf.*, hlm. 253–261, Des 2019.
- [11] A. N. Nurfadhilah, W. Budi, E. Kurniati, dan D. Suhaedi, "Penerapan Metode Moving Average untuk Prediksi Indeks Harga Konsumen," *Mat. J. Teori Dan Terap. Mat.*, vol. 21, no. 1, Art. no. 1, Jun 2022.
- [12] F. Sembiring, D. Gustian, A. Erfina, dan Y. Vikriansyah, "Analisis Tingkat Akurasi Algoritma Moving Average dalam Prediksi Pergerakan Uang Elektronik Bitcoin," *Jutisi J. Ilm. Tek. Inform. Dan Sist. Inf.*, vol. 10, no. 1, Art. no. 1, Apr 2021, doi: 10.35889/jutisi.v10i1.577.
- [13] M. S. Simamora dan S. Siregar, "Analysis of Factors Affecting Electricity Sales at PLN ULP Medan Timur Analisis Faktor-Faktor Yang Mempengaruhi Penjualan Listrik Pada PLN ULP Medan Timur", Diakses: 13 Januari 2025. [Daring]. Tersedia pada: <https://pdfs.semanticscholar.org/bb5c/d86a55ce53dee91f54741c22c8c96102cfc8.pdf>
- [14] E. Riduwan, D. Purwitasari, dan A. B. Raharjo, "Identifikasi Fitur untuk Prediksi Penerimaan Program Listrik Prabayar: Kasus di PLN Tahuna," *Techno.Com*, vol. 21, no. 3, hlm. 434–444, Agu 2022, doi: 10.33633/tc.v21i3.6451.
- [15] R. A. Saputra, "Analisis Pengaruh Kinerja Layanan Terhadap Kepuasan Konsumen Pelanggan Listrik Pintar di Kota Pontianak," *Pros. CORISINDO 2023*, Agu 2023, Diakses: 13 Januari 2025. [Daring]. Tersedia pada: <https://ojs.stmikpontianak.ac.id/index.php/corisindo/article/view/273>
- [16] A. A. Putri, "Pengaruh Siklus Pendapatan Listrik Prabayar, Pelanggan Pascabayar, Penjualan Listrik Terhadap Penerapan Sistem Informasi Akuntansi," masters, Universitas Indo Global Mandiri, 2024. Diakses: 13 Januari 2025. [Daring]. Tersedia pada: <http://repository.uigm.ac.id/id/eprint/1034/>
- [17] D. R. Hartono, "Prototipe Sistem Monitoring Daya Listrik Berbasis Internet of Things (IoT) Menggunakan Metode Exponential Smoothing Untuk Prediksi Kebutuhan Energi," Master's Thesis, Universitas Islam Sultan Agung (Indonesia), 2023. Diakses: 13 Januari 2025.
- [18] B. Darmawan, "Analisis Kinerja Keuangan PT PLN Setelah Ditetapkannya Payment Point Online Bank (PPOB) pada Tahun 2014-2019," undergraduate, IAIN Metro, 2022.
- [19] D. N. Batubara, A. P. Windarto, dan E. Irawan, "Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine," *JURIKOM J. Ris. Komput.*, vol. 9, no. 1, Art. no. 1, Feb 2022, doi: 10.30865/jurikom.v9i1.3833.
- [20] M. N. Adiwiranto, C. B. Waluyo, dan B. Sudibya, "Prototipe Sistem Monitoring Konsumsi Energi Listrik Serta Estimasi Biaya Pada Peralatan Rumah Tangga Berbasis Internet Of Things," *J. Edukasi Elektro*, vol. 6, no. 1, Art. no. 1, Mei 2022, doi: 10.21831/jee.v6i1.44272.
- [21] D. R. Hartono, "Monitoring Daya Listrik Berbasis Internet of Things Menggunakan Metode Simple Exponential Smoothing untuk Prediksi Kebutuhan Energi," *CYCLOTRON*, vol. 6, no. 2, Art. no. 2, Agu 2023, doi: 10.30651/cl.v6i2.17948.