



ABSTRAK

Mengidentifikasi anomali dalam data rekening pelanggan, seperti kesalahan pencatatan meter, pola konsumsi listrik yang tidak wajar, atau indikasi kecurangan merupakan langkah krusial untuk menjaga keakuratan tagihan listrik dan meningkatkan kepuasan pelanggan. Namun, sebagian besar penelitian terdahulu lebih berfokus pada pola konsumsi abnormal dan deteksi pencurian listrik, bukan pada aspek keakuratan verifikasi angka tagihan pelanggan, khususnya verifikasi pembacaan kWh-meter pascabayar. Padahal, kesalahan baca meter yang tidak terdeteksi dapat mengakibatkan ketidaksesuaian tagihan dan menurunkan kepercayaan pelanggan.

Metode konvensional seperti analisis statistik dasar memiliki keterbatasan dalam menangani kompleksitas pola konsumsi pelanggan dan skala besar data rekening listrik, sehingga kurang efektif untuk mendeteksi anomali secara akurat. Hal ini mendorong kebutuhan akan pendekatan yang lebih adaptif dan *scalable*. Pada penelitian ini, studi kasus difokuskan pada data Aplikasi Catat Meter Terpusat (ACMT) milik PT PLN (Persero) Unit Induk Distribusi Jawa Barat, yang mencakup lebih dari 5 juta baris historis pembacaan kWh-meter selama enam bulan. Setelah melalui tahap pra-pemrosesan dan pelabelan berbasis hasil verifikasi pengawas baca meter, data dibagi menjadi set pelatihan dan pengujian menggunakan *stratified sampling* untuk menjaga proporsi kelas. Untuk meningkatkan representasi kelas minoritas (anomali) pada data latih, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE).

Dalam penelitian ini digunakan tiga *base learners*, yaitu *Random Forest*, *Extra Trees*, dan *AdaBoost*, yang kemudian dikombinasikan ke dalam *Voting Classifier* untuk meningkatkan akurasi dan stabilitas prediksi. Pemilihan ketiga algoritma tersebut didasarkan pada hasil evaluasi terhadap 11 kandidat algoritma *machine learning* yang umum digunakan untuk klasifikasi. Evaluasi dilakukan menggunakan metrik *Precision*, *Recall*, dan *F1 Score*, di mana ketiga algoritma ini menunjukkan kinerja terbaik dengan nilai *F1 Score* masing-masing sebesar 0,9130, 0,9149, dan 0,8936. Selanjutnya, kombinasi ketiga model tersebut dalam *Voting Classifier* menghasilkan performa yang lebih unggul, dengan *F1 Score* sebesar 0,9286, yang merepresentasikan keseimbangan optimal antara sensitivitas terhadap anomali dan ketepatan prediksi. Hasil ini membuktikan bahwa pendekatan *Voting Classifier* efektif untuk sistem verifikasi tagihan listrik berskala besar dan berpotensi meningkatkan transparansi, efisiensi operasional, serta kepercayaan pelanggan terhadap PLN.

Kata kunci—Pendeteksian Anomali, *Voting Classifier*, *Machine Learning*, Verifikasi Tagihan, Data Konsumsi Energi



ABSTRACT

Identifying anomalies in customer billing data, such as meter reading errors, unusual electricity consumption patterns, or indications of fraud is a crucial step to ensure billing accuracy and enhance customer satisfaction. However, most previous studies have focused on abnormal consumption patterns and electricity theft detection, rather than the accuracy of billing verification, particularly the verification of postpaid kWh-meter readings. Undetected meter reading errors can lead to billing discrepancies and reduce customer trust.

Conventional methods, such as basic statistical analysis, have limitations in handling the complexity of customer consumption patterns and the large scale of billing data, making them less effective for accurate anomaly detection. This underscores the need for a more adaptive and scalable approach. In this study, the case is focused on data from the Centralized Meter Reading Application (ACMT) of PT PLN (Persero) West Java Distribution Unit, comprising more than 5 million rows of historical kWh-meter readings over six months. After undergoing preprocessing and labeling based on supervisor verification, the data were split into training and testing sets using stratified sampling to maintain class proportions. To improve the representation of minority classes (anomalies) in the training data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied.

This research employs three base learners, Random Forest, Extra Trees, and AdaBoost which are then combined into a Voting Classifier to improve prediction accuracy and stability. The selection of these three algorithms was based on evaluations of 11 candidate machine learning classifiers commonly used for classification tasks. The evaluation used Precision, Recall, and F1 Score metrics, where the three algorithms achieved the best F1 Scores of 0,9130, 0,9149, and 0,8936, respectively. Furthermore, their combination in the Voting Classifier achieved superior performance with an F1 Score of 0,9286, representing an optimal balance between sensitivity to anomalies (recall) and prediction accuracy (precision). These findings demonstrate that the Voting Classifier is effective for large-scale electricity billing verification systems and has the potential to enhance transparency, operational efficiency, and customer trust in PLN.

Keywords—Anomaly detection, Voting Classifier, Machine Learning, Billing Verification, Energy Consumption Data