PREDIKSI TAGIHAN LISTRIK BERDASARKAN KARAKTERISTIK SEGMEN PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING DAN RANDOM FOREST REGRESSION

Desi Fitriani Ramadan, 202231032

Dibawah bimbingan Dine Tiara Kusuma, S.T., M.Kom.

ABSTRAK

Meningkatnya kebutuhan energi listrik serta beragamnya karakteristik rumah tangga menyebabkan prediksi tagihan listrik menjadi penting untuk mendukung pengelolaan energi yang lebih efisien dan akurat. Penelitian ini menerapkan StandardScaler, K-Means Clustering untuk segmentasi pelanggan, serta Random Forest Regression dengan hasil segmentasi sebagai fitur tambahan dalam memprediksi tagihan listrik rumah tangga berdasarkan karakteristik pelanggan, sedangkan kualitas clustering dievaluasi menggunakan Elbow Method, Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index, serta performa model dievaluasi menggunakan Mean Absolute Error (MAE) dan Weighted Mean Absolute Percentage Error (WMAPE). Hasil penelitian menunjukkan bahwa pelanggan berhasil dikelompokkan ke dalam tiga segmen berdasarkan karakteristik rumah tangga, sehingga menghasilkan kelompok dengan karakteristik dan tingkat tagihan listrik yang berbeda. Model Random Forest Regression mampu menghasilkan prediksi tagihan listrik dengan nilai MAE sebesar 1.953.234 dan WMAPE sebesar 35,61%, serta analisis *feature importance* menunjukkan bahwa karakteristik rumah tangga dan informasi segmentasi memberikan kontribusi dalam proses prediksi. Penelitian selanjutnya disarankan membandingkan performa model dengan algoritma lain, menggunakan dataset yang lebih besar dan beragam, serta menambahkan variabel yang berpotensi memengaruhi tagihan listrik agar akurasi prediksi dapat ditingkatkan.

Kata kunci: K-Means Clustering, Random Forest Regression, Prediksi Tagihan Listrik, Segmentasi Pelanggan, Machine Learning.

PREDICTING ELECTRICITY BILLS BASED ON CUSTOMER SEGMENT CHARACTERISTICS USING K-MEANS CLUSTERING AND RANDOM FOREST REGRESSION

Desi Fitriani Ramadan, 202231032

Under the Guidance of Dine Tiara Kusuma, S.T., M.Kom.

ABSTRACT

The increasing demand for electricity and the diverse characteristics of households make electricity bill prediction essential for supporting more efficient and accurate energy management. This study applies StandardScaler, K-Means Clustering for customer segmentation, and Random Forest Regression using cluster labels as additional features to predict household electricity bills based on customer characteristics. Clustering quality was evaluated using the Elbow Method, Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index, while model performance was assessed using Mean Absolute Error (MAE) and Weighted Mean Absolute Percentage Error (WMAPE). The results show that customers were successfully grouped into three segments with distinct household characteristics and electricity bill patterns. The Random Forest Regression model achieved an MAE of 1,953,234 and a WMAPE of 35.61%, while feature importance analysis indicated that household characteristics and customer segmentation contributed to electricity bill prediction. Future research is recommended to compare the proposed model with other machine learning algorithms, use larger and more diverse datasets, and include additional variables that may influence electricity bills to improve prediction accuracy.

Keywords: K-Means Clustering, Random Forest Regression, Electricity Bill Prediction, Customer Segmentation, Machine Learning.