Title: Explainable AI for Customer Churn Prediction in the Energy Sector Using Ensemble Machine Learning Models

Author names: Abdullah Al Mamun, Md Fazla Saim Tanoor, Abdul Kadar Muhammad Masum, Touhid Bhuiyan, Md. Maruf Hassan

Abstract: Customer churn prediction is crucial for energy providers to preserve revenue and market share in a competitive setting. This research explores implementing explainable AI (XAI) in customer churn prediction with machine learning algorithms as well as interpretability methods. This research utilizes a dataset formed by combining client and price information from Kaggle's PowerCo dataset, making use of the 'id' column as a key. During the data preprocessing section, the data involves extensive preparation, such as dropping unnecessary columns, deleting duplicates, encoding category features, capping outlier values, and imputing missing values. The class imbalance problem was alleviated through making use of the Synthetic Minority Oversampling Technique (SMOTE) to enable robust training of models. The machine learning algorithms Random Forest, XGBoost, and LightGBM were built, implemented, and benchmarked with principal performance metrics like accuracy, precision, recall, F1-score, as well as ROC-AUC. Among the evaluated ensemble models, XGBoost demonstrated the most balanced trade-off between predictive accuracy and interpretability when combined with SHAP and LIME. SHAP (SHapley Additive exPlanations) as well as LIME (Local Interpretable Model-agnostic Explanations) are used for providing global as well as localinterpretability to identify key drivers of churn. This proposed approach exhibits potential to provide accurate predictions as well as salient insights, enabling energy providers to design targeted retention strategies.

Keywords: Customer Churn Prediction, Explainable AI (XAI), Machine Learning, Random Forest, XGBoost, LightGBM, SMOTE, SHAP, LIME, Performance Metrics, ROC-AUC, PR-AUC

DOI: https://doi.org/10.4108/airo.9813