Use Data-driven framework for the detection of non-technical losses in distribution grids

Abstract

Non-technical losses (NTL) occurring in the electric grid, particularly at the distribution level may cause a negative impact on utilities' interest, paying consumers and states. Reducing NTL can increase revenue, profit, reliability, among other aspects of the power system. Therefore, this subject brings for a critical concern to utilities and authorities. This study proposes the recognition of NTL using several machine learning models. The dataset was provided by a distributor system operator (DSO) in the coastal region in Colombia. Nine (9) models were trained and tested, considering not only aspects related to energy consumption but socio-demographics also. Three feature selection methods were used to reduce the number of final predictors. The models were evaluated through the accuracy and the F1 score using a 10-fold cross-validation algorithm. Results showed that the final subsets provided enough overall performance. However, the best subset correspond to the Tree-based subset. A gradient boosting machine was the model outperformed the rest, giving a mean accuracy of 74.3% and an F1 score of 83.1. These results represent great insights to local DSO and utilities to join artificial intelligence to their energy meters to reduce NTL significantly and therefore increase profit and reliability.