**Abstract**

Our work concerns a model regression challenge, based on a Kaggle competition that uses data collected for houses sold in Ames, IA. Based on these provided data, we explored various data pre-processing strategies to maximize the effectiveness of a Scikit-learn gradient boosting regressor to create a predictive ensemble for the Kaggle test data set. We designed an algorithm capable of achieving an R-squared value of 0.96 using 10-fold cross-validation on the testing data set, thus presenting a promising approach to creating a reliable predictor based on the provided data features.