

WEEKLY REPORT

Dear Amro Haggag

In the below Table you will see a specified table that includes the good, the bad & the improvements that can be done on your Classification Task

Correctly Done	Can Be Improved
<ul style="list-style-type: none"> • Encoding • Scaling • Handle Outlier • Check correlation • Check and handle by vif • Check p-value • Applying Algorithms 	<ul style="list-style-type: none"> • Handle multicollinearity using drop cols iteratively with high vif or pca (not necessary to increase performance, the integrity will be) • You can apply other techniques for outlier as transformation (log), capping, fill by median (try and determine the good one) • You can try encoding using one-hot, label or manual encoder as type1:1, type2:2,... • You can apply robust algorithm as random forest, xgboost,... • You can apply cross-validation to know genera performance of model on all data • Split data into train and test before any preprocessing • Handle Imbalanced • You can add more features as total_members, total_night, percent_can((p_c)/(p_c + p_not_c

