

## WEEKLY REPORT

Dear Aya Helmy

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

