

## WEEKLY REPORT

**Dear Mohammed Shbaikat**

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

