**Machine Learning Assignment – 1**

**Load the data set winequality-white.csv and perform the following tasks.**

* **Separate the target feature[‘quality], split data in 7:3 proportion (30% form a holdout set, use random\_state=17), and preprocess data with StandardScaler.**
* **Train a simple linear regression model using sci-kit learn and statsmodel.**
* **What are mean squared errors of model predictions on train and holdout sets?**
* **Create a data frame to display coefficients of each features.**
* **Which feature this linear regression model treats as the most influential on wine quality?**
* **Train a LASSO model with α=0.01and scaled data. Again, set random\_state=17.**
* **Which feature is the least informative in predicting wine quality, according to this LASSO model?**
* **Train LassoCV with random\_state=17 to choose the best value of α- alpha in 5-fold cross-validation.(use LassoCV instead of Gridsearch). The list of alphas to be passed are [0.01,0.001,0.1,0.2,0.02,0.002].**
* **Which feature is the least informative in predicting wine quality, according to the tuned LASSO model?**
* **What are mean squared errors of tuned LASSO predictions on train and holdout sets?**
* **Train a Random Forest, setting only random\_state to be 17.**
* **What are mean squared errors of tuned randomforest predictions on train and holdout sets?**
* **Tune the max\_features and max\_depth hyperparameters with GridSearchCV and again check mean cross-validation MSE and MSE on holdout set. Parameters to tune**

**forest\_params = {'max\_depth': list(range(10, 25)),**

**'max\_features': list(range(6,12))}**

* **Output RF's feature importance. Again, it's nice to present it as a DataFrame. What is the most important feature, according to the Random Forest model?**
* **Finally try different regression algorithms and say which is best and why? And do the diagnostics tests for OLS which was build first.**

**Notes:**

* **Submit the file in ipynb and html, make a zip and keep both the files.**
* **File name should be like this yourname\_ML\_A1**
* **Give inference where ever required.**