**Regression Instruction Manual**

**Technology requirements**

Python 3.6

Pandas 0.24.2

Numpy 1.16.4

Scikit-learn 0.21.3

**Regression (Feature Reduction top 10 features)**:

To implement a python code, it requires 3 arguments, SETTING NUMBER, RATIO and NFEATURE.

Command

python ml\_reg\_top10\_wo\_NumPeers.py -cls CLASSIFIER -rat RATIO -fea NFEATURE

Note: reg\_reduction.sh file will execute all pairs required for the experiments.

The following are the information on the arguments.

* CLASSIFIER

GB: Gradient Boosting

RF: Random Forest

LR: Logistic Regression

* RATIO

20\_80: 20% training, 80% testing

50\_50: 50% training, 50% testing

80\_20: 80% training, 20% testing

* NFEATURE

10: uses top 10 features

9: uses top 9 features

….

2: uses top 2 features

1: uses top 1 feature

Command Example

python ml\_reg\_top10\_wo\_NumPeers.py -cls DNN -rat 20\_80 -fea 10

python ml\_reg\_top10\_wo\_NumPeers.py -cls GB -rat 50\_50 -fea 9

python ml\_reg\_top10\_wo\_NumPeers.py -cls LR -rat 80\_20 -fea 8

ml\_reg\_top10\_wo\_NumPeers.py creates regression.csv

* regression\_top10\_wo\_NumPeers.csv contains results of time, recall, RMSE

**Regression (feature comparison)**:

To implement a python code, it requires 3 arguments, SETTING NUMBER, RATIO and SET OF FEATURES.

Command

python ml\_reg.py -cls CLASSIFIER -rat RATIO -set SET OF FEATURES

Note: reg\_feature\_comparison.sh file will execute all pairs required for the experiments.

The following are the information on the arguments.

* CLASSIFIER

GB: Gradient Boosting

RF: Random Forest

LR: Logistic Regression

* RATIO

20\_80: 20% training, 80% testing

50\_50: 50% training, 50% testing

80\_20: 80% training, 20% testing

* SET OF FEATURES

1: top10 features

2: top 10 features w/o NumPeers

3: entire features

4: entire features w/o NumPeers

Command Example

python ml\_reg.py -cls DNN -rat 20\_80 -set 1

python ml\_reg.py -cls GB -rat 50\_50 -set 2

python ml\_reg.py -cls LR -rat 80\_20 -set 3

ml\_reg\_top10\_wo\_NumPeers creates regression.csv

* regression\_top10\_wo\_NumPeers.csv contains results of time, recall, RMSE