

Predicting and Explaining Privacy Exposure in Mobility Data - Hyperparameters settings

1 Hyperparameters settings

In order to learn the machine learning models used for the evaluation of EXPERT, we performed hyper-parameter tuning by grid search in the parameter space. In this section we provide the details about the hyperparameters settings we found, for each value of adversary background knowledge configuration $h = 2, 3, 4, 5$. The machine learning models we used, are the following: Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and GCForest (GC). Table 1 reports the setting for the `istat` dataset, while Table 2 show the setting for the `voronoi` dataset.

Model	Parameters
LG	h=2,3,4,5 $C = 0.001$, penalty = 12
DT	h=2 criterion = gini, max depth = 200, max features = sqrt, min samples leaf = 30, min samples split = 50; h=3 criterion = gini, max depth = 300, max features = sqrt, min samples leaf = 20, min samples split = 30; h=4 criterion = entropy, max depth = 400, max features = auto, min samples leaf = 20, min samples split = 10; h=5 criterion = gini, max depth = 500, max features = 5, min samples leaf = 10, min samples split = 30
RF	h=2 bootstrap = True, criterion = gini, max depth = 100, max features = sqrt, min samples leaf = 5, min samples split = 25, estimators = 30; h=3 bootstrap = True, criterion = entropy, max depth = 500, max features = auto, min samples leaf = 3, min samples split = 10, estimators = 200; h=4 bootstrap = False, criterion = entropy, max depth = 500, max features = auto, min samples leaf = 5, min samples split = 30, estimators = 200; h=5 bootstrap = False, criterion = entropy, max depth = 40, max features = auto, min samples leaf = 3, min samples split = 5, estimators = 50
GC	h=2 XGB: bootstrap = True, eta = 0, gamma = 8, learning rate = 0.01, max depth = 20, n estimators = 300, tree method = approx; RF: bootstrap = True, criterion = gini, max depth = 100, max features = sqrt, min samples leaf = 5, min samples split = 25, n estimators = 30; h=3 XGB: bootstrap = True, tree method = hist, learning rate = 0.1, max depth = 20, n estimators = 500; RF: bootstrap = True, criterion = entropy, max depth = 500, max features = auto, min samples leaf = 3, min samples split = 10, n estimators = 200; h=4 XGB: bootstrap = True, tree method = auto, max depth = 20, learning rate = 0.1, n estimators = 400; EXTRA: bootstrap = False, criterion = entropy, max depth = 350, max features = auto, min samples leaf = 3, min samples split = 5, n estimators = 350; h=5 XGB: bootstrap = True, learning rate = 0.05, max depth = 20, n estimators = 400, tree method = hist; EXTRA: bootstrap = False, criterion = gini, max depth = 20, max features = sqrt, min samples leaf = 3, min samples split = 10, n estimators = 50

Table 1: Parameter setting **istat** dataset.

Model	Parameters
LG	h=2 $C = 0.001$, penalty = 12 h=3 $C = 0.09$, penalty = 12 h = 4 $C = 10$, penalty = 12 h = 5 $C = 0.001$, penalty = 12
DT	h=2 criterion = gini, max depth = 60, max features = 5, min samples leaf = 40, min samples split = 5; h=3 criterion = entropy, max depth = 500, max features = 5, min samples leaf = 15, min samples split = 5; h=4 criterion = gini, max depth = 60, max features = auto, min samples leaf = 40, min samples split = 25; h=5 criterion = gini, max depth = 300, max features = 4, min samples leaf = 20, min samples split = 5
RF	h=2 bootstrap = True, criterion = entropy, max depth = 100, max features = 5, min samples leaf = 10, min samples split = 5, estimators = 30; h=3 bootstrap = True, criterion = entropy, max depth = 40, max features = auto, min samples leaf = 5, min samples split = 10, estimators = 200; h=4 bootstrap = True, criterion = entropy, max depth = 500, max features = log2, min samples leaf = 3, min samples split = 5, estimators = 200; h=5 bootstrap = True, criterion = gini, max depth = 20, max features = sqrt, min samples leaf = 3, min samples split = 10, estimators = 50
GC	h=2 XGB: bootstrap = True, eta = 0, gamma = 8, learning rate = 0.05, max depth = 20, n estimators = 500, tree method = auto; RF: bootstrap = True, criterion = entropy, max depth = 100, max features = 5, min samples leaf = 10, min samples split = 5, n estimators = 30; h=3 EXTRA: bootstrap = False, criterion = entropy, max depth = 20, max features = auto, min samples leaf = 3, min samples split = 10, estimators = 20; RF: bootstrap = True, criterion = entropy, max depth = 40, max features = auto, min samples leaf = 5, min samples split = 10, n estimators = 200; h=4 EXTRA: bootstrap = False, criterion = entropy, max depth = 500, max features = auto, min samples leaf = 5, min samples split = 10, n estimators = 50; RF: bootstrap = True, criterion = entropy, max depth = 500, max features = log2, min samples leaf = 3, min samples split = 5, estimators 200; h=5 RF: bootstrap = True, criterion = gini, max depth = 20, max features = sqrt, min samples leaf = 3, min samples split = 10, estimators = 50; EXTRA: bootstrap = False, criterion = entropy, max depth = 100, max features = 5, min samples leaf = 3, min samples split = 5, n estimators = 50

Table 2: Parameter setting for voronoidataset.