Appendices

Parameters for Classification Models

All the following parameters are chosen for Scikit-Learn classification models by parameter searching specifically.

Dataset	leaf_size	n_neighbors
Adult	5	5
Compas	5	5
Titanic	50	5
Communities	5	2
German	5	2
Bank	5	2

Table 1: K-Nearest Neighbors

Dataset	C	max_iter	tol
Adult	0.01	1000	0.01
Compas	0.1	1000	0.01
Titanic	1	1000	1e-06
Communities	10	1000	0.001
German	0.1	1000	1e-06
Bank	0.1	1000	0.001

Table 2: Support Vector Machine

Dataset	max_depth	min_ samples_ leaf	n_ estimators
Adult	50	1	200
Compas	200	5	500
Titanic	10	5	50
Communities	null	1	200
German	200	1	100
Bank	200	50	50

Table 3: Random Forest

Dataset	C	max_iter	tol
Adult	0.01	100	1e-06
Compas	0.1	100	1e-06
Titanic	1	100	1e-06
Communities	10.0	100	1e-06
German	100.0	100	1e-06
Bank	100.0	100	1e-06

Table 4: Logistic Regression

Dataset	max_depth	max_ leaf_ nodes	min_ samples_ leaf
Adult	200	1000	1
Compas	10	100	1
Titanic	200	100	1
Communities	10	100	5
German	10	1000	10
Bank	10	1000	5

Table 5: Decision Tree

Dataset	alpha	hidden_ layer_sizes
Adult	1e-05	[500]
Compas	1e-05	[100]
Titanic	0.0001	[500]
Communities	1e-05	[500]
German	1e-05	[100]
Bank	0.0001	[500]

Table 6: Multi-layer Perceptron (Part 1)

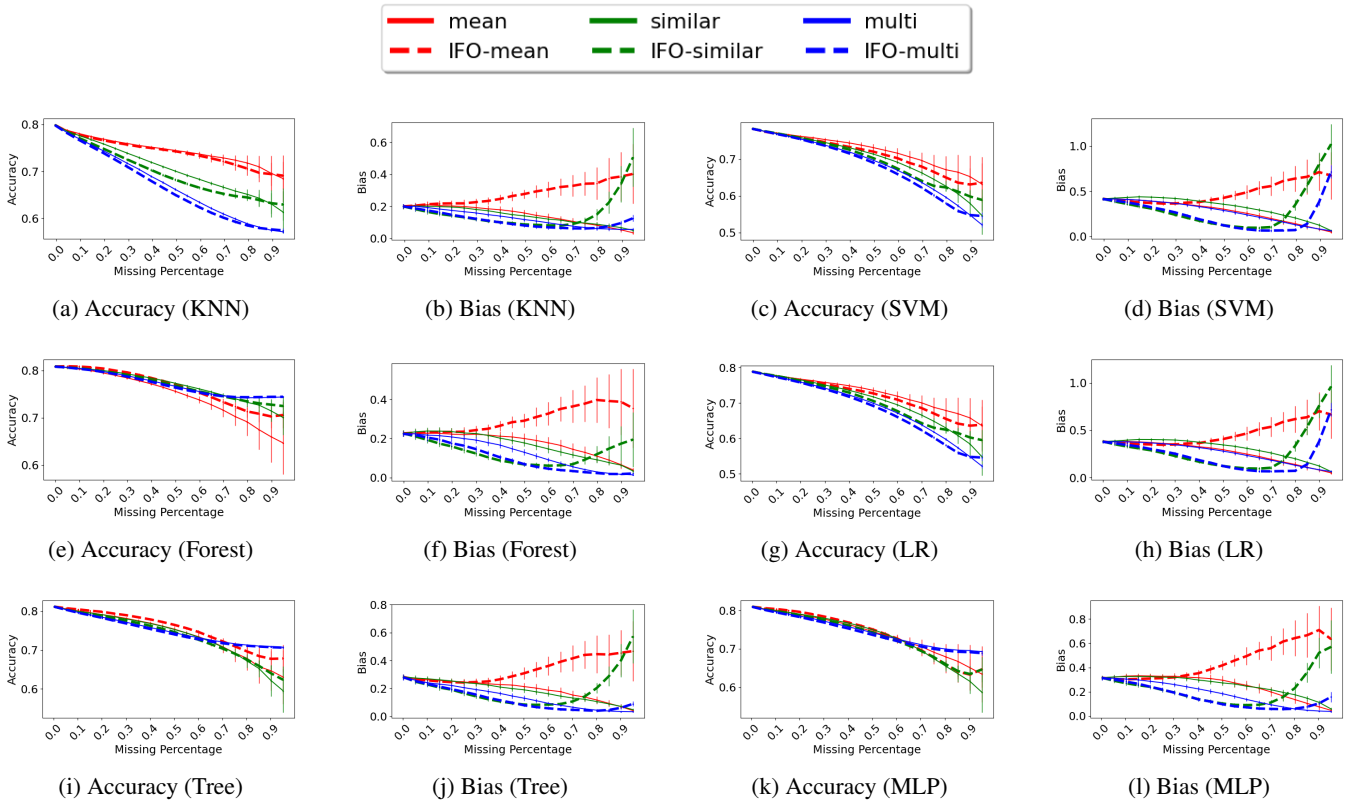
Dataset	learning_ rate_init	max_iter
Adult	0.0001	500
Compas	0.001	200
Titanic	0.001	500
Communities	0.0001	1000
German	0.01	500
Bank	0.0001	200

Table 7: Multi-layer Perceptron (Part 2)

Experimental Results (MCAR)

The following figures are the results from experiments that induce missingness with MCAR. Each figure consists of 2 sub-plots: accuracy plot and bias plot. X-axis represents the amount of missingness induced by MCAR, Y-axis shows the respective values. Different colors of lines show different ML algorithms used at that point. The length of error bar at each point shows the standard deviation computed from 200 trials.

Since we split dataset to 9:1 for cross validation, test data may contain only 1 target instead of all, and because our bias definition contains FPR and FNR, which will cause division by zero error, so we dropped these data points. Therefore, we may see some empty data in communities dataset.

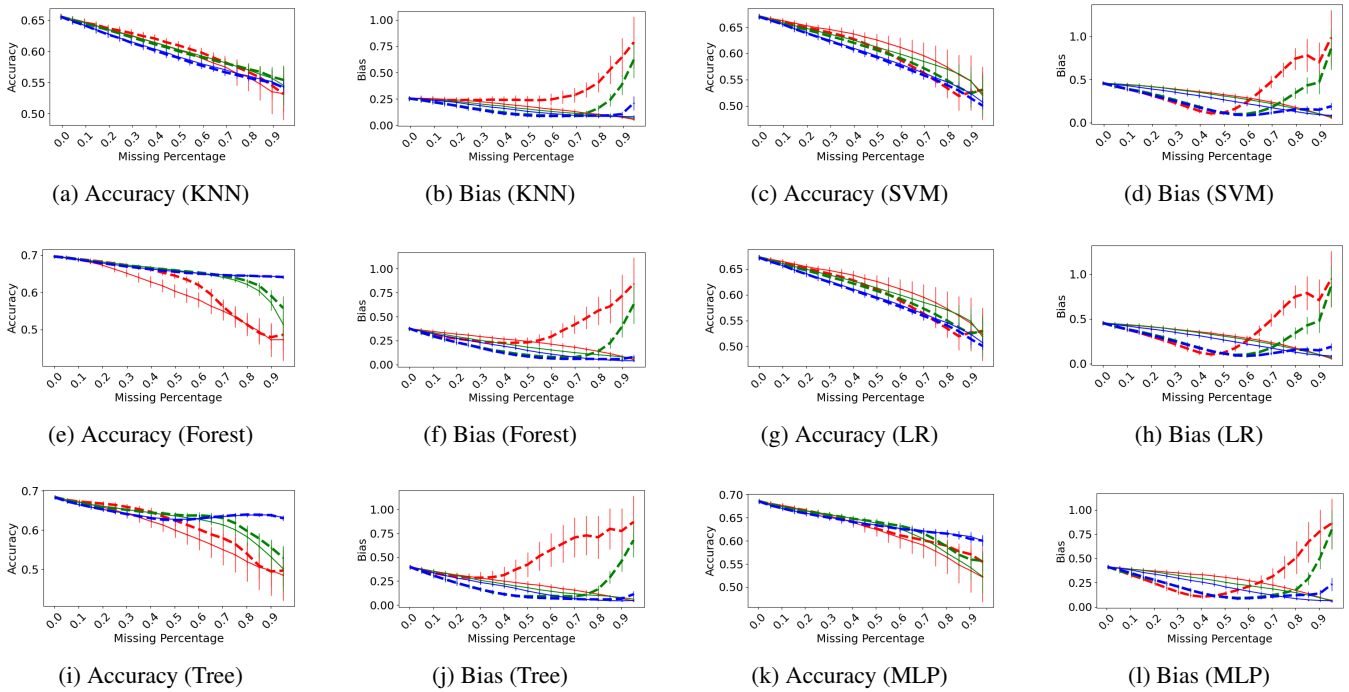


(a) Accuracy (KNN) (b) Bias (KNN) (c) Accuracy (SVM) (d) Bias (SVM)

(e) Accuracy (Forest) (f) Bias (Forest) (g) Accuracy (LR) (h) Bias (LR)

(i) Accuracy (Tree) (j) Bias (Tree) (k) Accuracy (MLP) (l) Bias (MLP)

Figure 1: Adult Dataset (Combined Imputation Methods)

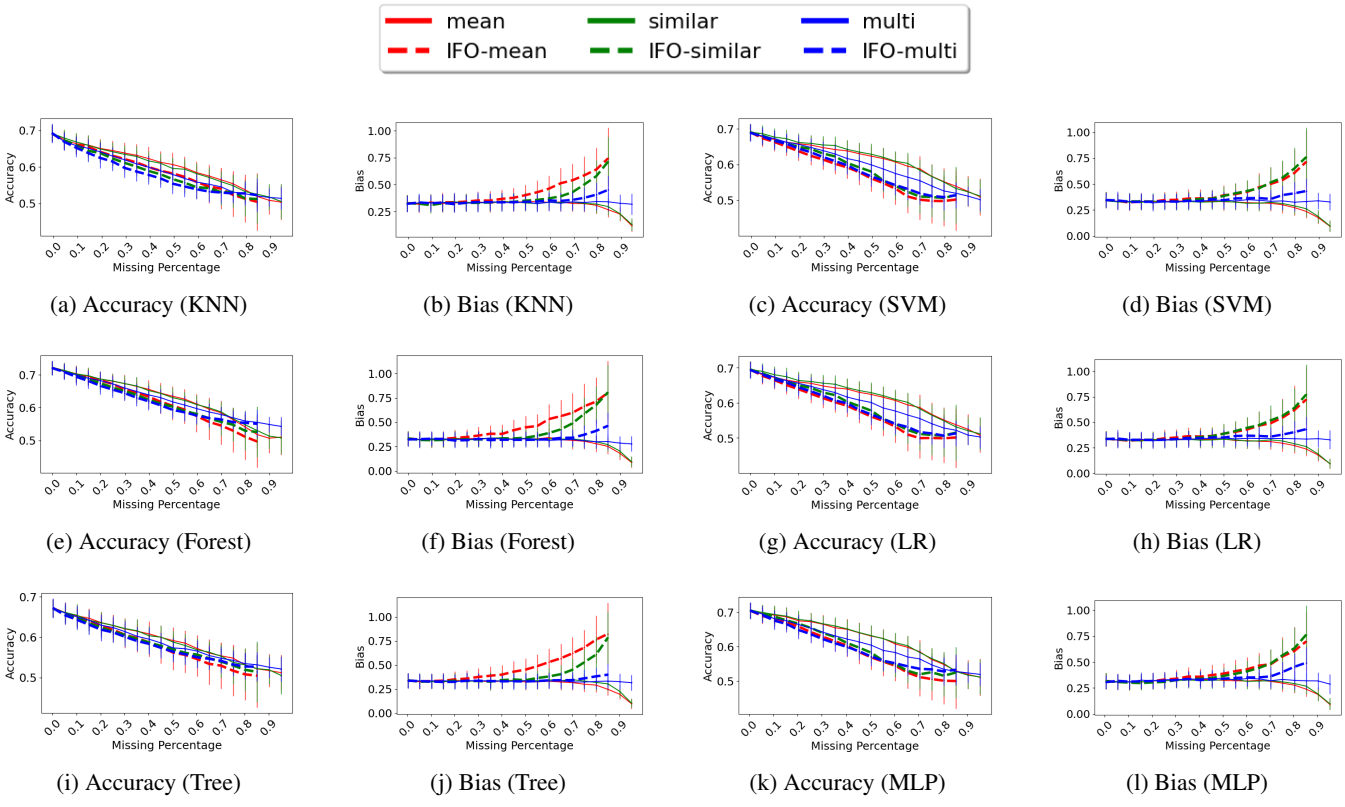


(a) Accuracy (KNN) (b) Bias (KNN) (c) Accuracy (SVM) (d) Bias (SVM)

(e) Accuracy (Forest) (f) Bias (Forest) (g) Accuracy (LR) (h) Bias (LR)

(i) Accuracy (Tree) (j) Bias (Tree) (k) Accuracy (MLP) (l) Bias (MLP)

Figure 2: Compas Dataset (Combined Imputation Methods)

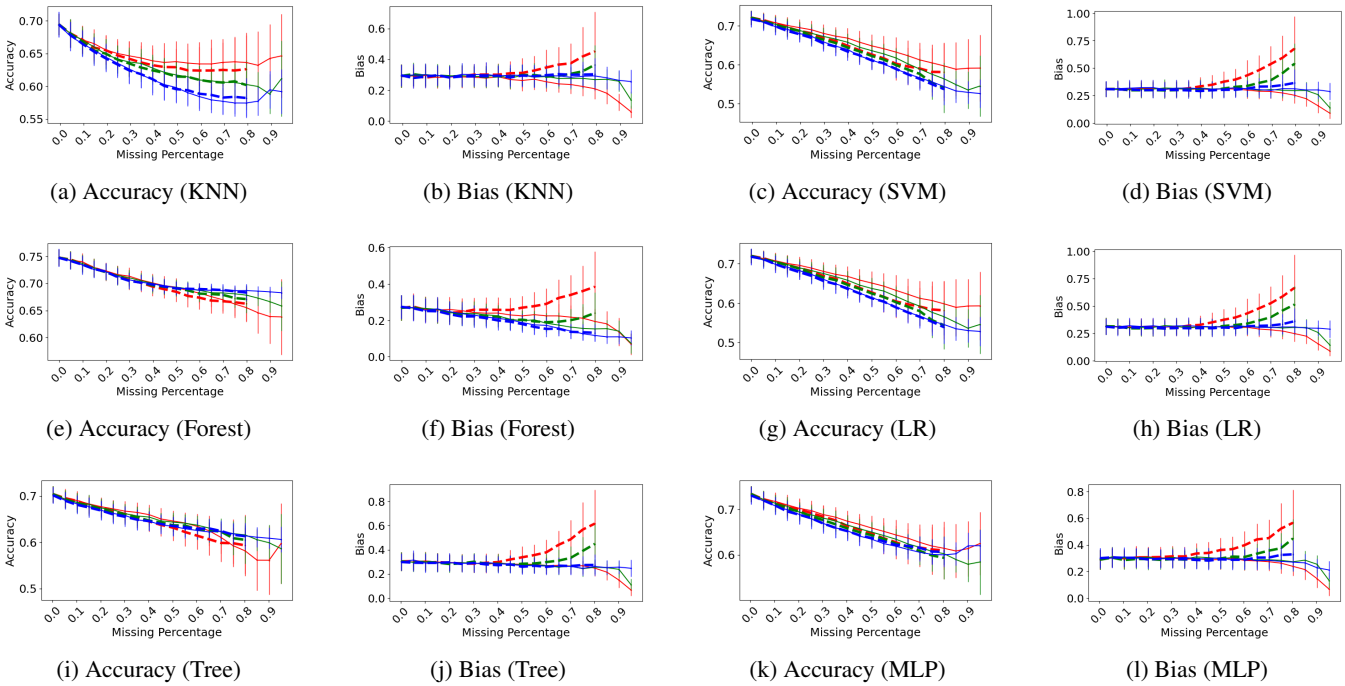


(a) Accuracy (KNN) (b) Bias (KNN) (c) Accuracy (SVM) (d) Bias (SVM)

(e) Accuracy (Forest) (f) Bias (Forest) (g) Accuracy (LR) (h) Bias (LR)

(i) Accuracy (Tree) (j) Bias (Tree) (k) Accuracy (MLP) (l) Bias (MLP)

Figure 3: Titanic Dataset (Combined Imputation Methods)

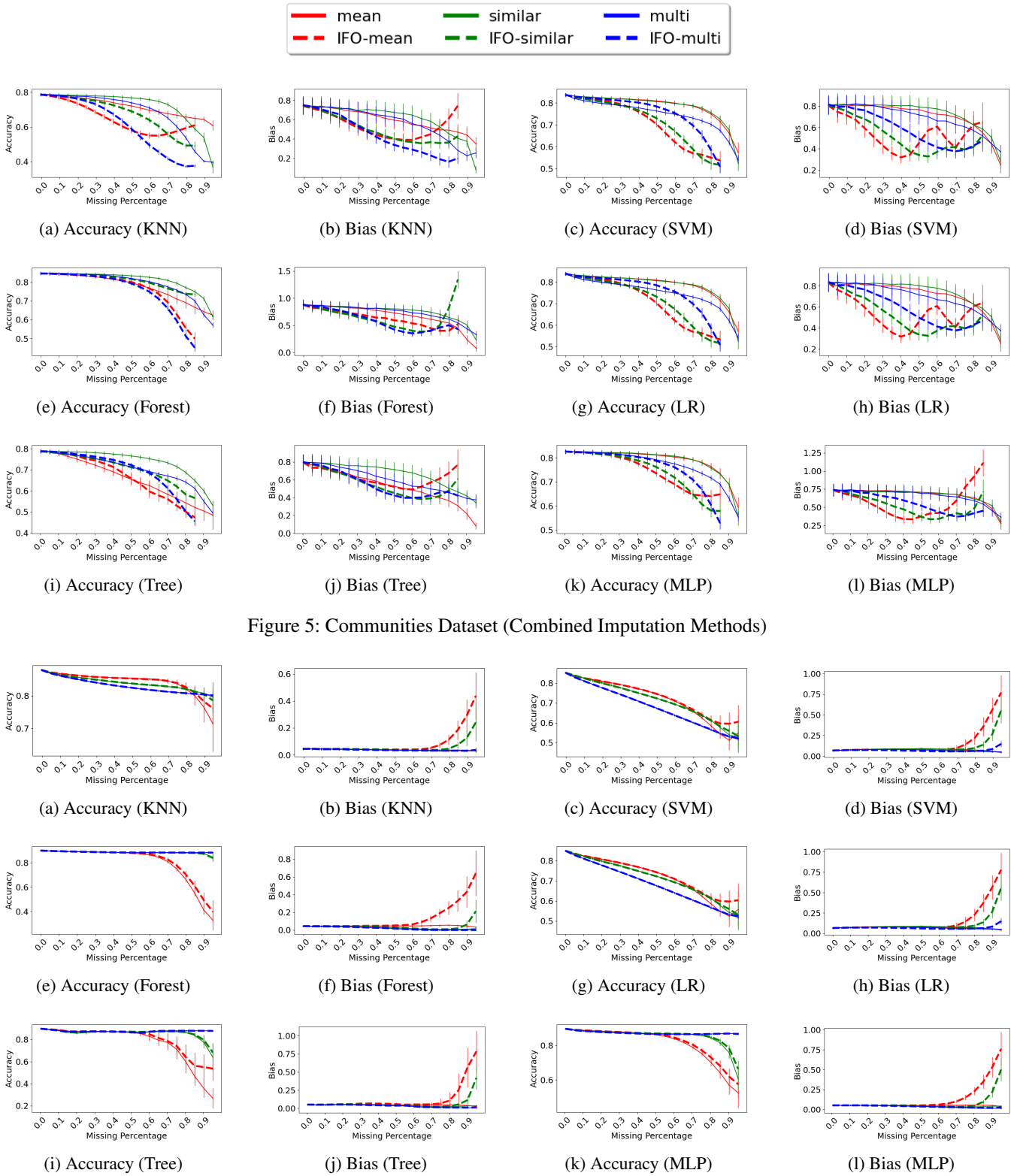


(a) Accuracy (KNN) (b) Bias (KNN) (c) Accuracy (SVM) (d) Bias (SVM)

(e) Accuracy (Forest) (f) Bias (Forest) (g) Accuracy (LR) (h) Bias (LR)

(i) Accuracy (Tree) (j) Bias (Tree) (k) Accuracy (MLP) (l) Bias (MLP)

Figure 4: German Credit Dataset (Combined Imputation Methods)



(a) Accuracy (KNN) (b) Bias (KNN) (c) Accuracy (SVM) (d) Bias (SVM)

(e) Accuracy (Forest) (f) Bias (Forest) (g) Accuracy (LR) (h) Bias (LR)

(i) Accuracy (Tree) (j) Bias (Tree) (k) Accuracy (MLP) (l) Bias (MLP)

Figure 5: Communities Dataset (Combined Imputation Methods)

(a) Accuracy (KNN) (b) Bias (KNN) (c) Accuracy (SVM) (d) Bias (SVM)

(e) Accuracy (Forest) (f) Bias (Forest) (g) Accuracy (LR) (h) Bias (LR)

(i) Accuracy (Tree) (j) Bias (Tree) (k) Accuracy (MLP) (l) Bias (MLP)

Figure 6: Bank Dataset (Combined Imputation Methods)