CSE 472
Offline 2
Model Training and Evaluation Report

Sabah Ahmed
1905118

September 20, 2024

Instructions for Running the Notebook

To train and evaluate the model using different datasets and scalers, follow these instructions:

Step 1: Choose the Scaler

In the script, you will find the following lines of code:

```
# choose the scaler to use
scaler = MinMaxScaler()
# scaler = StandardScaler()
# scaler = RobustScaler()
# scaler = PowerTransformer()
```

- To use the **MinMaxScaler**, keep the line 'scaler = MinMaxScaler()' uncommented and comment out all other 'scaler' lines.

- If you want to use **StandardScaler**, uncomment 'scaler = StandardScaler()' and comment out the others.

- Similarly, for **RobustScaler**, uncomment 'scaler = RobustScaler()' and comment out the others.

- For **PowerTransformer**, uncomment 'scaler = PowerTransformer()' and comment out the rest.

Step 2: Choose the Dataset

In the script, you will find the following lines:

```
# choose the dataset to use
dataset_id = '1'
# dataset_id = '2'
# dataset_id = '3'
```

- For **Dataset 1**, keep 'dataset_id = '1'' uncommented and comment out the other dataset lines.

- To use **Dataset 2**, uncomment 'dataset_id = '2'' and comment out the others.

- For **Dataset 3**, uncomment 'dataset_id = '3'' and comment out the others.

Step 3: Tune Hyperparameters

Here are the hyperparameters you can adjust in the script:

```
# tune hyperparameters
learning_rate = .5
epochs = 3000
decay_rate = 5e-3
threshold = 0.5
regularization = None
lambda_regularization = 0.0001
```

- **learning_rate**: The learning rate controls the step size during gradient descent. You can adjust this value based on your experimentation.

- **epochs**: This defines the number of iterations the model will train for. You can increase or decrease the number of epochs as needed.

- **decay_rate**: This parameter adjusts the learning rate over time. In this codebase, exponential decaying is used.

- **threshold**: The decision boundary for classification.

- **regularization**: You can set this to different types of regularization methods, such as 'L2', by modifying 'regularization = None' to something like 'regularization = 'l2''. Possible regularization values are 'l1' and 'l2' in this codebase.

- **lambda_regularization**: Set this to adjust the strength of regularization.

1 Dataset 1 : Telco Customer Churn

1.1 MinMaxScaler
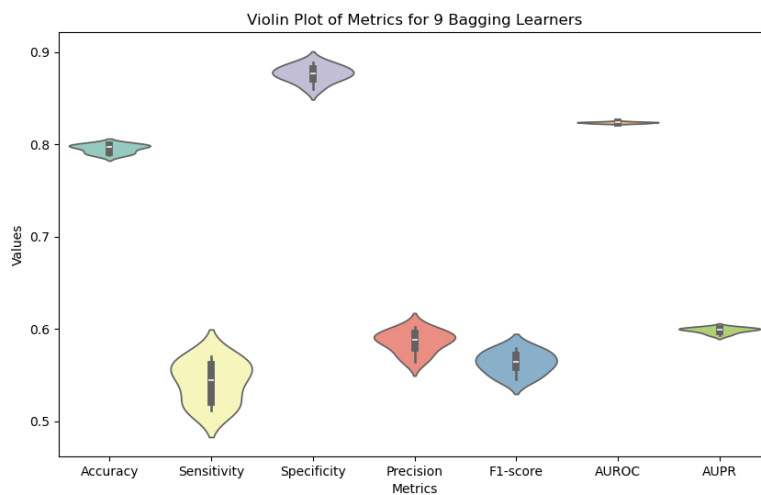
Table 1: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.7976 ± 0.0029	0.5346 ± 0.0174	0.8821 ± 0.0076	0.5930 ± 0.0095	0.5621 ± 0.0079	0.8244 ± 0.0007	0.5840 ± 0.0039
Voting Ensemble	0.8027	0.5353	0.8886	0.6067	0.5687	0.7288	0.4591
Stacking Ensemble	0.7984	0.5588	0.8754	0.5901	0.5740	0.8239	0.5844
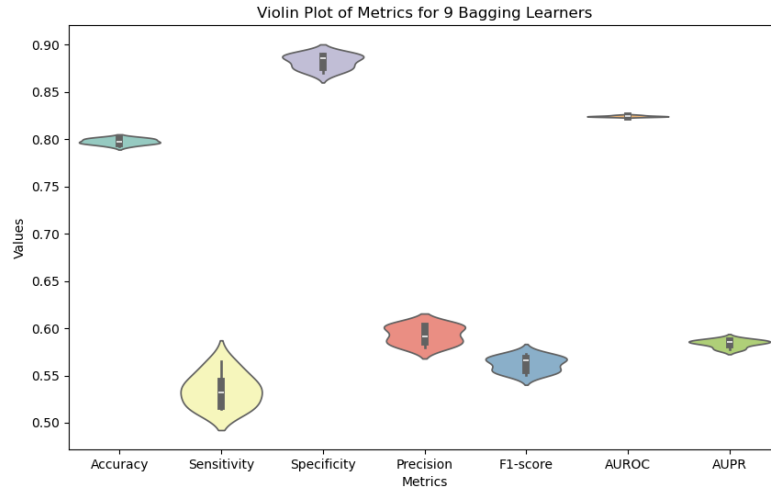
1.2 StandardScaler

Table 2: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.7956 ± 0.0045	0.5428 ± 0.0225	0.8768 ± 0.0089	0.5862 ± 0.0119	0.5634 ± 0.0121	0.8239 ± 0.0008	0.5988 ± 0.0030
Voting Ensemble	0.7999	0.5412	0.8829	0.5974	0.5679	0.7404	0.4634
Stacking Ensemble	0.7920	0.5324	0.8754	0.5783	0.5544	0.8237	0.5953

1.3 RobustScaler

Table 3: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.7949 ± 0.0031	0.5346 ± 0.0156	0.8785 ± 0.0073	0.5858 ± 0.0091	0.5589 ± 0.0072	0.8230 ± 0.0007	0.5939 ± 0.0025
Voting Ensemble	0.7991	0.5412	0.8820	0.5955	0.5670	0.7332	0.4613
Stacking Ensemble	0.7984	0.5412	0.8810	0.5935	0.5662	0.8237	0.5937
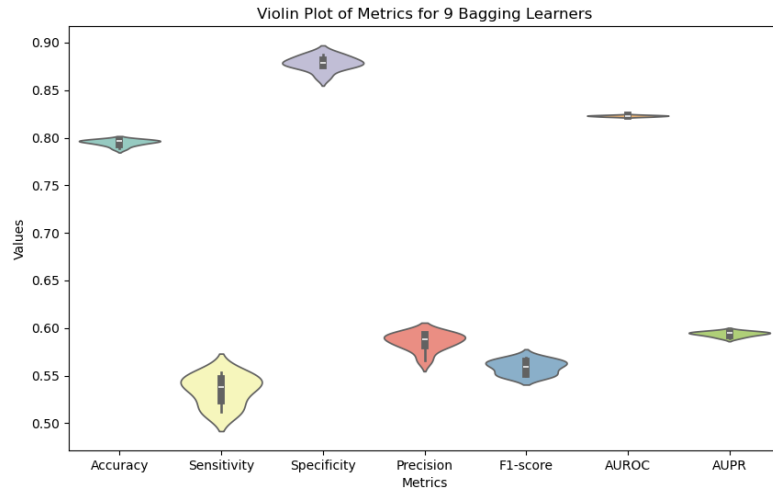
1.4 PowerTransformer

Table 4: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.8003 ± 0.0030	0.5376 ± 0.0195	0.8846 ± 0.0065	0.5994 ± 0.0080	0.5666 ± 0.0104	0.8319 ± 0.0006	0.6079 ± 0.0019
Voting Ensemble	0.8027	0.5412	0.8867	0.6053	0.5714	0.7428	0.4699
Stacking Ensemble	0.7977	0.5529	0.8763	0.5893	0.5706	0.8302	0.6070

Violin Plot of Metrics for 9 Bagging Learners
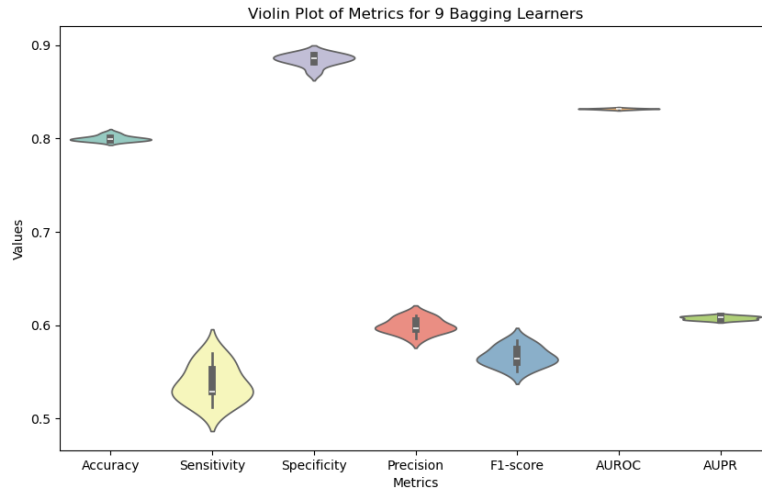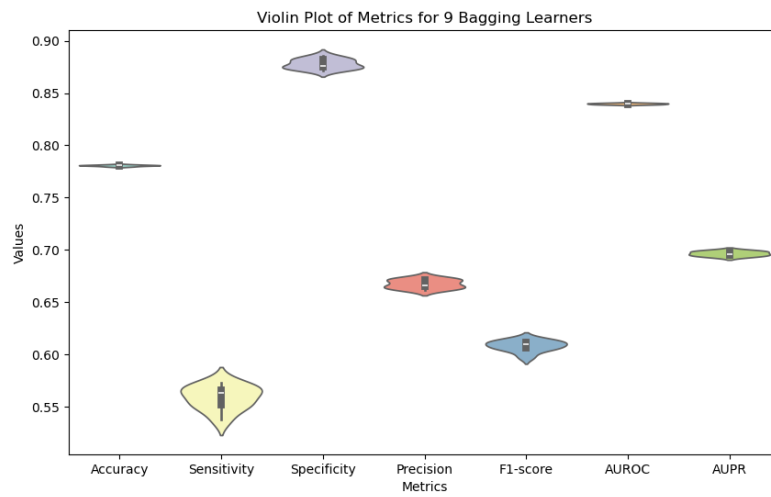
2   Dataset 2 : Adult UCI

2.1   MinMaxScaler

Table 5: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.7808 ± 0.0007	0.5592 ± 0.0113	0.8780 ± 0.0047	0.6677 ± 0.0043	0.6085 ± 0.0052	0.8398 ± 0.0006	0.6967 ± 0.0023
Voting Ensemble	0.7802	0.5576	0.8778	0.6667	0.6073	0.7287	0.5226
Stacking Ensemble	0.7820	0.5477	0.8848	0.6757	0.6050	0.8378	0.6901

2.2   StandardScaler

Table 6: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.8164 ± 0.0016	0.6245 ± 0.0089	0.9006 ± 0.0040	0.7336 ± 0.0058	0.6746 ± 0.0041	0.8756 ± 0.0004	0.7779 ± 0.0011
Voting Ensemble	0.8159	0.6232	0.9004	0.7329	0.6736	0.7754	0.5944
Stacking Ensemble	0.8121	0.5788	0.9144	0.7476	0.6525	0.8732	0.7709
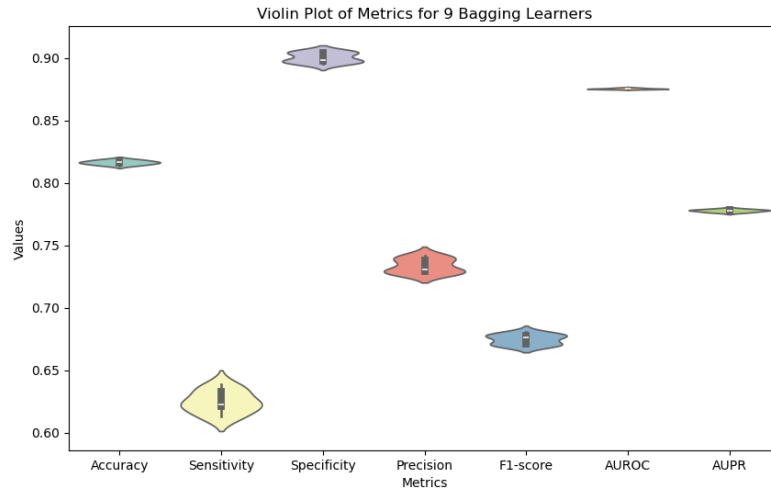
2.3 RobustScaler

Table 7: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.8141 ± 0.0013	0.6128 ± 0.0077	0.9023 ± 0.0027	0.7333 ± 0.0038	0.6676 ± 0.0040	0.8706 ± 0.0008	0.7701 ± 0.0011
Voting Ensemble	0.8141	0.6119	0.9028	0.7339	0.6674	0.7694	0.5869
Stacking Ensemble	0.8050	0.5285	0.9263	0.7586	0.6230	0.8684	0.7643

Violin Plot of Metrics for 9 Bagging Learners
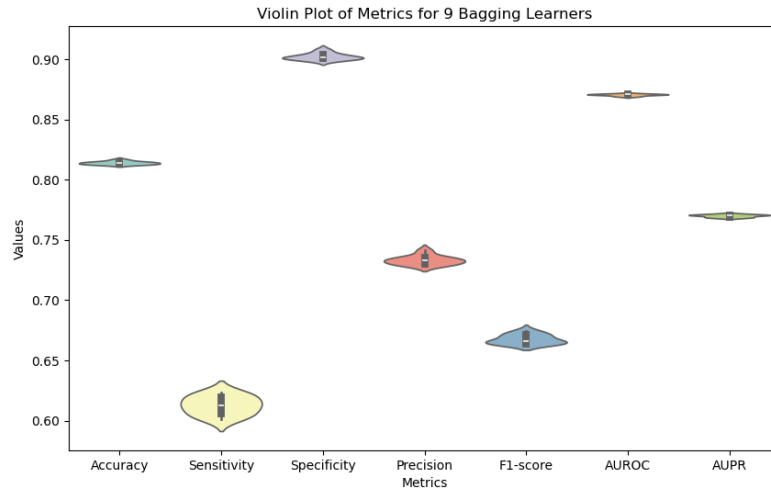
2.4   PowerTransformer

Table 8: Performance on Test set

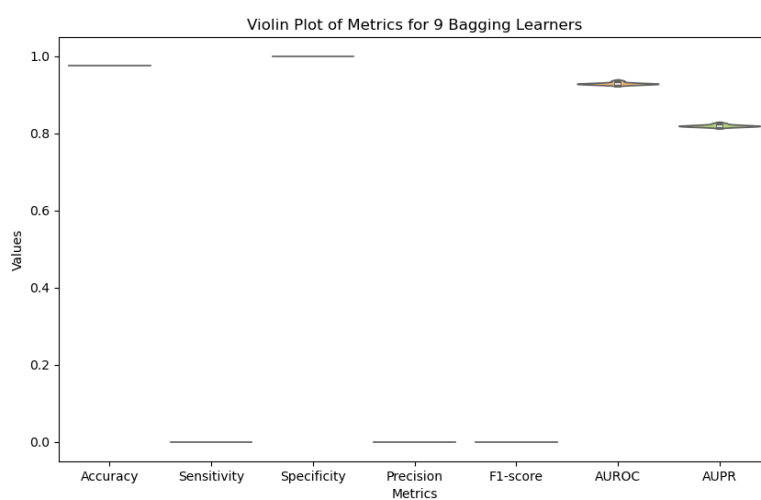|	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.8006 ± 0.0010	0.6122 ± 0.0071	0.8832 ± 0.0032	0.6968 ± 0.0037	0.6517 ± 0.0030	0.8593 ± 0.0005	0.7304 ± 0.0013
Voting Ensemble	0.8014	0.6139	0.8836	0.6980	0.6533	0.7612	0.5650
Stacking Ensemble	0.7891	0.5258	0.9045	0.7070	0.6031	0.8559	0.7213
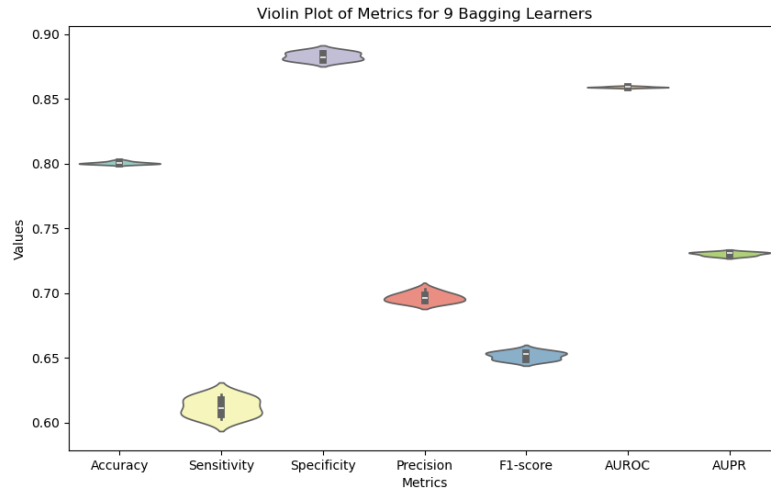
3 Dataset 3 : Credit Card Fraud Detection

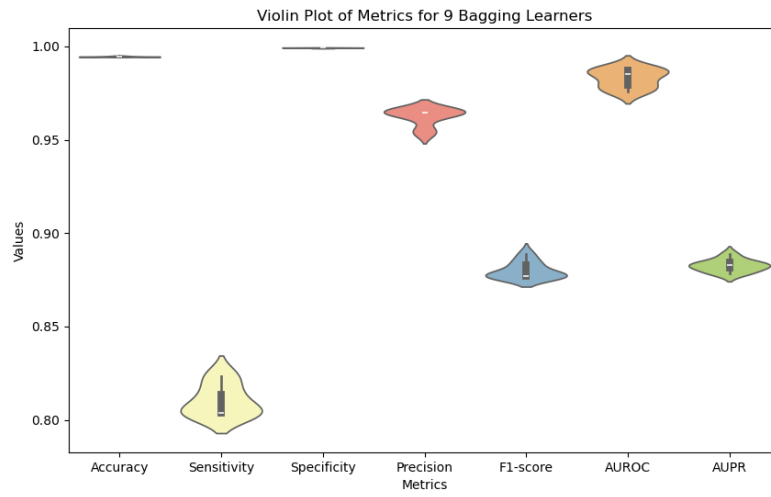## 3.1 MinMaxScaler

Table 9: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.9751 ± 0.0000	0.0000 ± 0.0000	1.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	0.9298 ± 0.0030	0.8205 ± 0.0029
Voting Ensemble	0.9751	0.0000	1.0000	0.0000	0.0000	0.5000	0.0249
Stacking Ensemble	0.9751	0.0000	1.0000	0.0000	0.0000	0.9319	0.8209

3.2   StandardScaler

Table 10: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.9945 ± 0.0002	0.8105 ± 0.0085	0.9992 ± 0.0001	0.9625 ± 0.0047	0.8799 ± 0.0044	0.9829 ± 0.0050	0.8827 ± 0.0032
Voting Ensemble	0.9944	0.8039	0.9992	0.9647	0.8770	0.9114	0.7988
Stacking Ensemble	0.9941	0.8039	0.9990	0.9535	0.8723	0.9781	0.8792

3.3 RobustScaler

Table 11: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.9943 ± 0.0003	0.7952 ± 0.0134	0.9993 ± 0.0001	0.9682 ± 0.0055	0.8732 ± 0.0071	0.9505 ± 0.0078	0.8631 ± 0.0077
Voting Ensemble	0.9941	0.7941	0.9992	0.9643	0.8710	0.9114	0.8068
Stacking Ensemble	0.9944	0.8039	0.9992	0.9647	0.8770	0.9479	0.8678
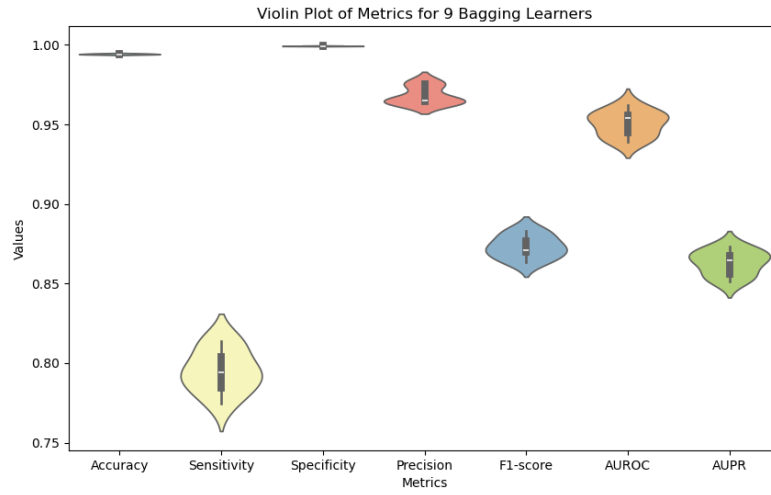
3.4   PowerTransformer

Table 12: Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR*	0.9943 ± 0.0001	0.8028 ± 0.0059	0.9992 ± 0.0001	0.9622 ± 0.0048	0.8753 ± 0.0030	0.9781 ± 0.0065	0.8808 ± 0.0036
Voting Ensemble	0.9944	0.8039	0.9992	0.9647	0.8770	0.9065	0.7892
Stacking Ensemble	0.9944	0.8039	0.9992	0.9647	0.8770	0.9724	0.8793