

a. Clear Instructions on how to run the script to train the model and test them

The cell with (third cell from the bottom) main execution block needs to be updated to train and test different models. In the data preprocessing section, there are three lines:

```
# preprocessor = TelcoChurnPreprocessor()
# preprocessor = AdultPreprocessor()
# preprocessor = CreditCardPreprocessor()
```

To train and test a specific model, uncomment the corresponding line. After running this cell, execute the two following cells to generate the violin plots and comparison table. If you want to see the results for another model, return to the main execution block, uncomment the desired model, and run all three cells again.

b. The table(s) and plot(s) mentioned in the performance evaluation section with my experimental results.

Dataset 1(Telco Customer Churn):

Violin plot:

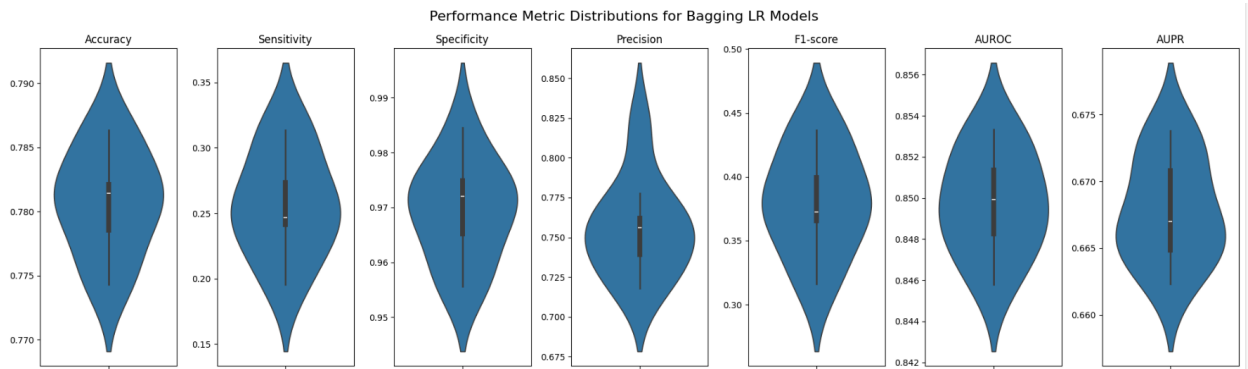


Table:

	LR*	Voting ensemble	Stacking ensemble
Metric			
Accuracy	0.7631	0.7779	0.7885
Sensitivity	0.2197	0.2440	0.3405
Specificity	0.9647	0.9701	0.9498
Precision	0.6979	0.7459	0.7095
F1-score	0.3342	0.3677	0.4601
AUROC	0.8338	0.6070	0.8363
AUPR	0.6340	0.3821	0.6529

Dataset 2(Adult):

Violin plot:

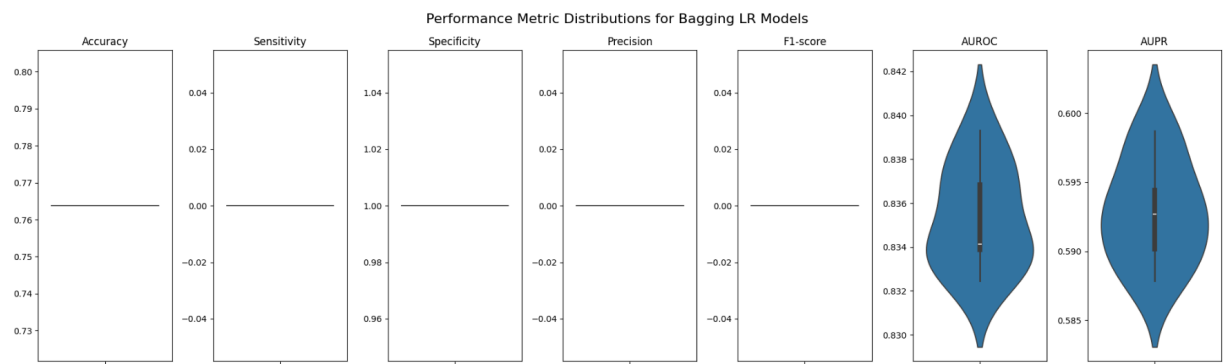


Table:

LR* Voting ensemble Stacking ensemble

Metric

Accuracy	0.7588	0.7638	0.7664
Sensitivity	0.0000	0.0000	0.0198
Specificity	1.0000	1.0000	0.9973
Precision	0.0000	0.0000	0.6972
F1-score	0.0000	0.0000	0.0384
AUROC	0.8282	0.5000	0.8379
AUPR	0.5994	0.2362	0.5954

Dataset 3(Credit Card Fraud Detection):

Violin plot:

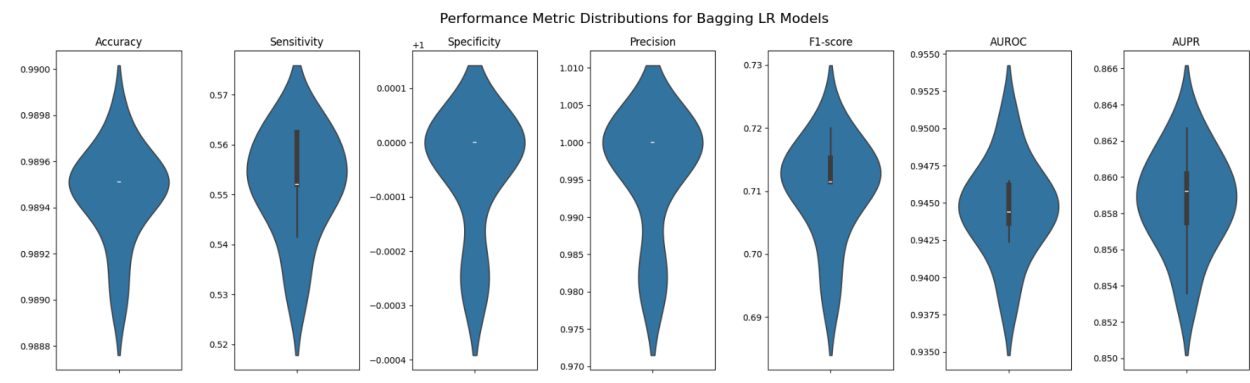


Table:

LR* Voting ensemble Stacking ensemble

Metric

Accuracy	0.9905	0.9895	0.9895
Sensitivity	0.5867	0.5521	0.5625
Specificity	1.0000	1.0000	0.9998
Precision	1.0000	1.0000	0.9818
F1-score	0.7395	0.7114	0.7152
AUROC	0.9729	0.7760	0.9380
AUPR	0.9071	0.5626	0.8569

c. Any observations

The violin plot of dataset 3 shows a single horizontal line instead of a violin shape for the first 5 metrics (Accuracy, Sensitivity, Specificity, Precision, and F1-score). This suggests there is no variation in these metrics across the bagging LR models being evaluated. Only the last two metrics (AUROC and AUPR) show typical violin plot shapes, indicating some variation in these measures across the models.