Offline Assignment 2 Report

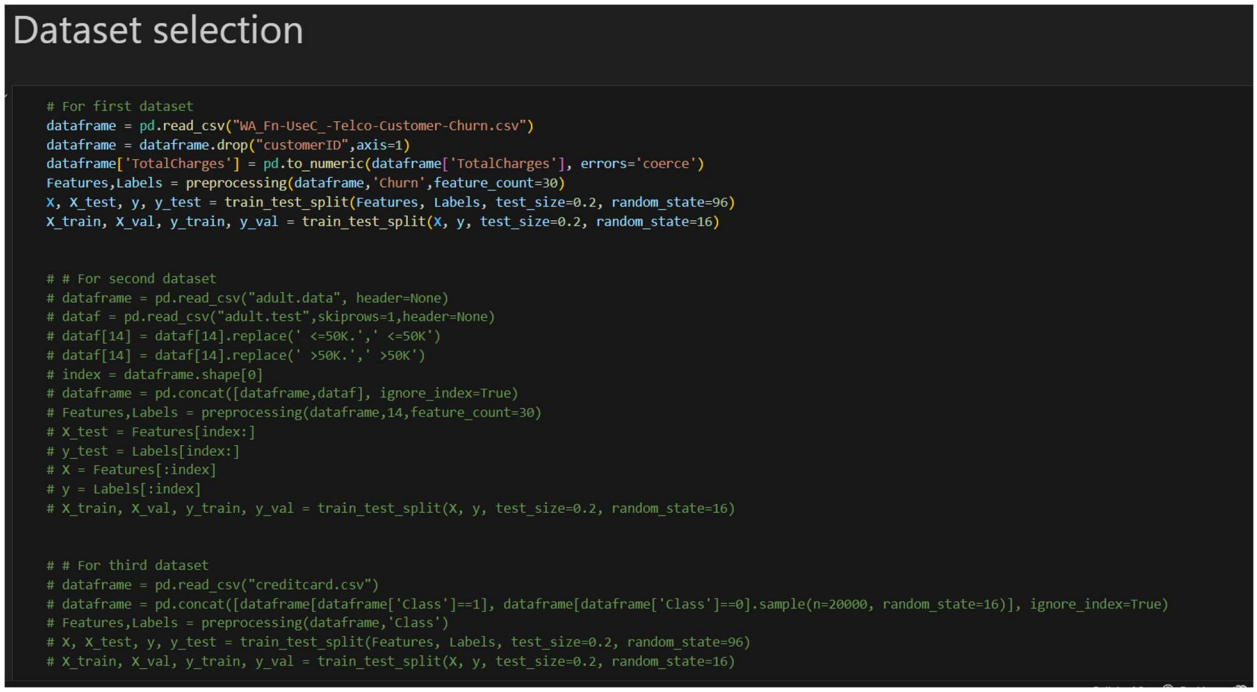
CSE 472 Sessional

Name: Md. Zulkar Naim

Student ID: 1905016

How to run script:

We just need to comment out dataset specific parts from **Dataset selection** code block (second code block in jupyter notebook).



- **For first dataset (https://www.kaggle.com/blastchar/telco-customer-churn):** Keep the first part of this code and comment the rest of this block.

- **For second dataset (https://archive.ics.uci.edu/ml/datasets/adult):** Keep the second part of this code and comment the rest of this block.

- **For third dataset (https://www.kaggle.com/mlg-ulb/creditcardfraud):** Keep the third part of this code and comment the rest of this block.

Performance Evaluation:

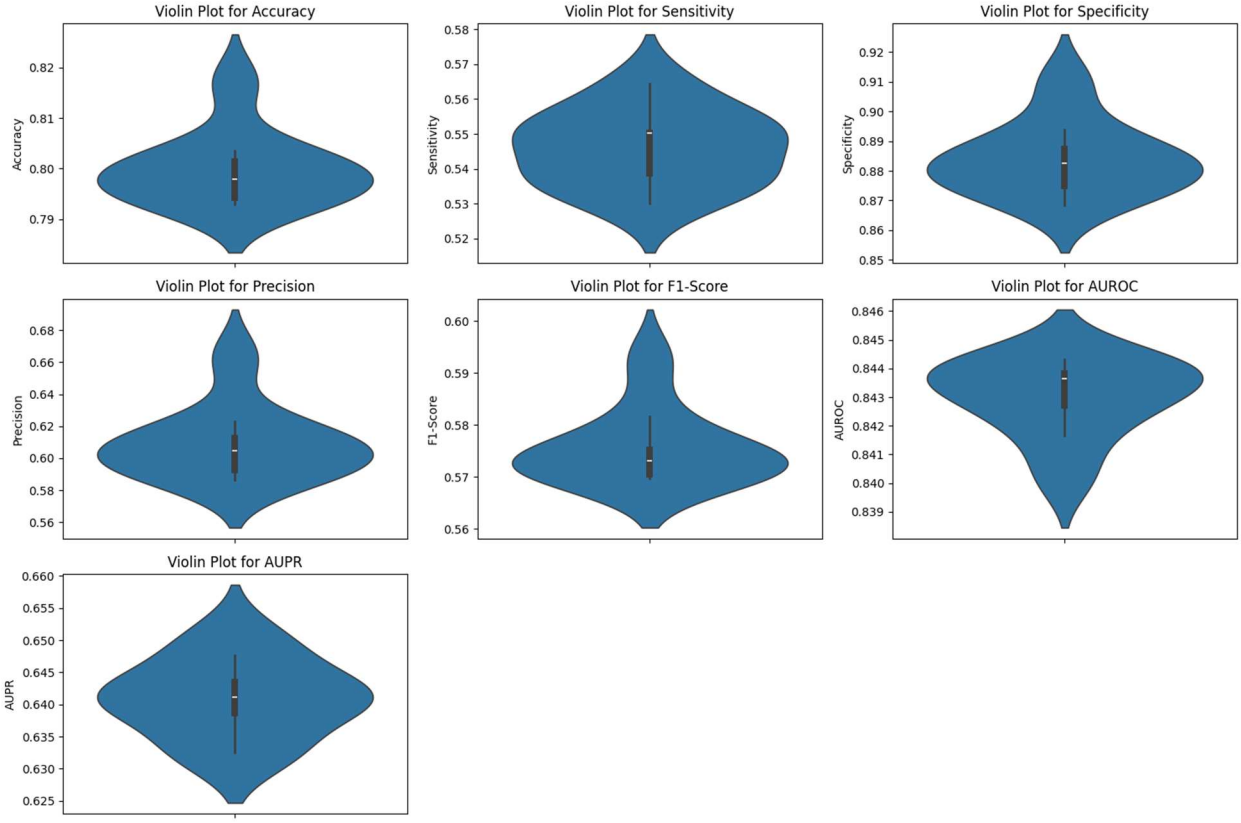
For first dataset (https://www.kaggle.com/blastchar/telco-customer-churn):

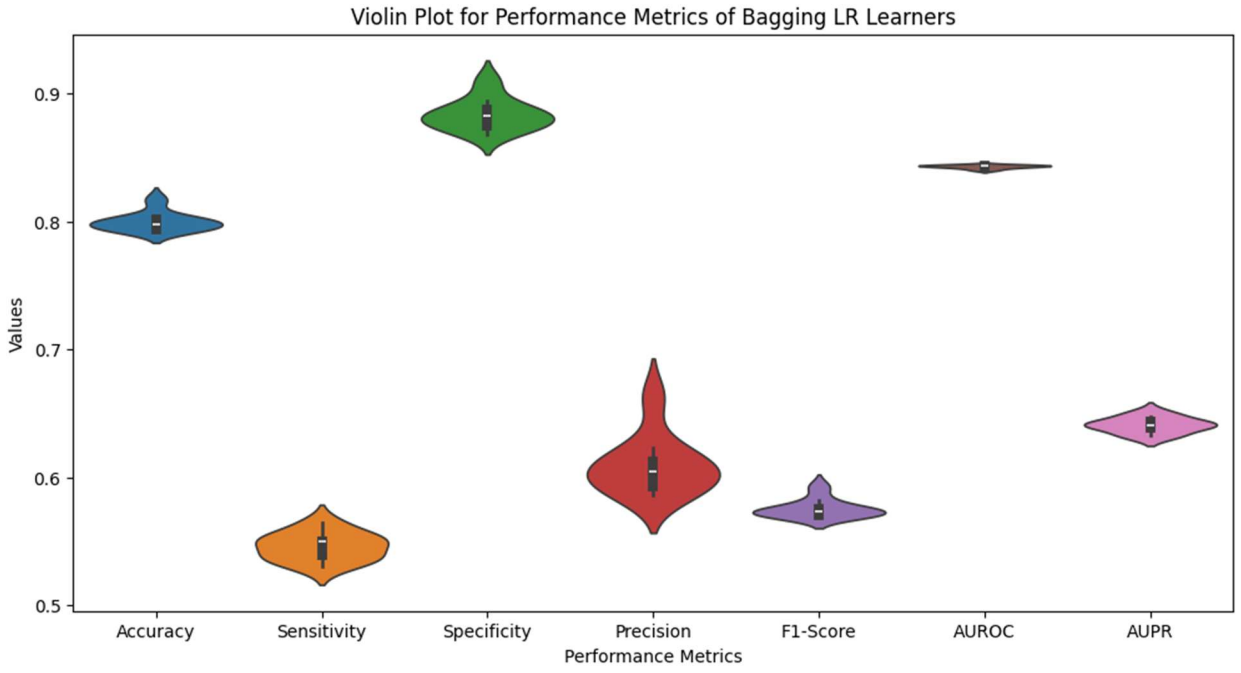
Learning rate = 0.0001, no regularization

Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUROC	AUPR
LR	0.800 ± 0.007	0.546 ± 0.010	0.884 ± 0.012	0.610 ± 0.022	0.576 ± 0.007	0.843 ± 0.001	0.641 ± 0.006
Voting ensemble	0.79573	0.538682	0.880682	0.598726	0.567119	0.844396	0.644381
Stacking ensemble	0.808541	0.515759	0.905303	0.642857	0.572337	0.748987	0.50329

Violin plots for each performance metric for the 9 bagging LR learners





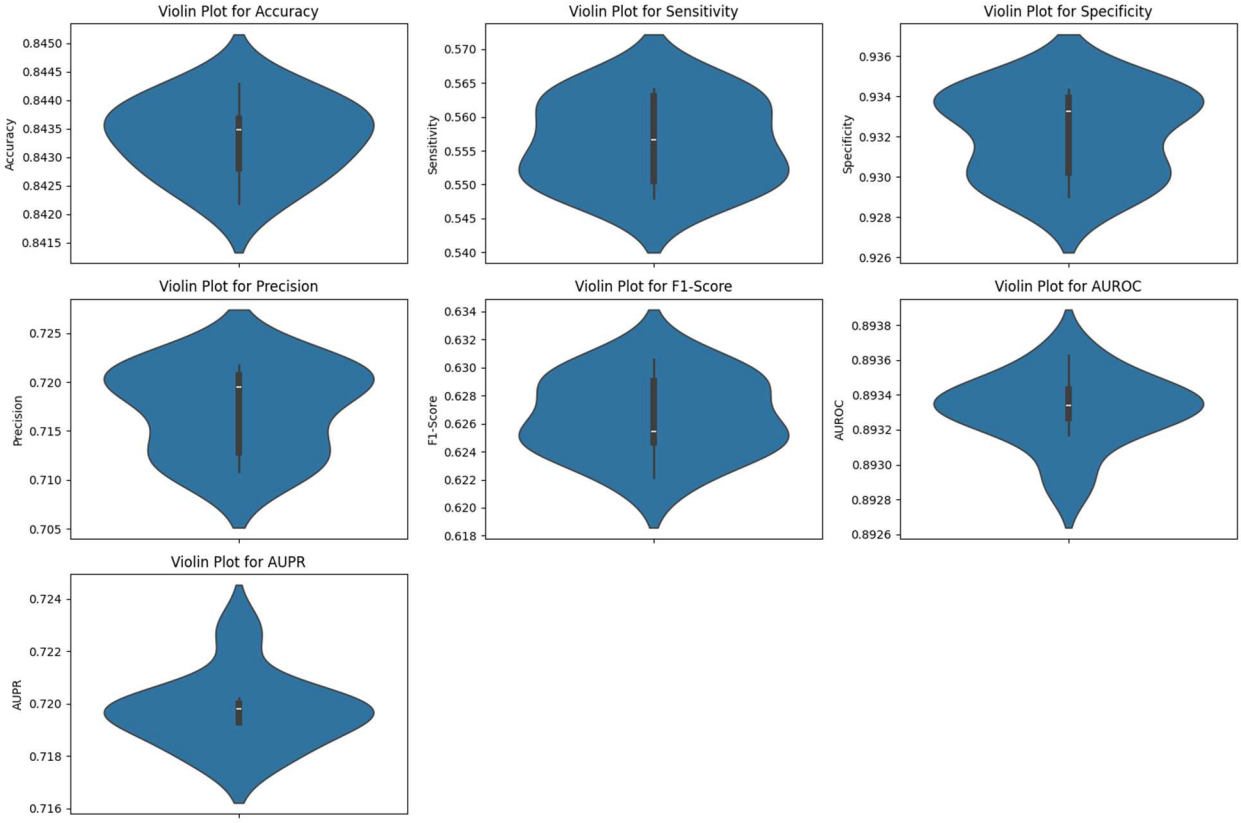
For second dataset (https://archive.ics.uci.edu/ml/datasets/adult):

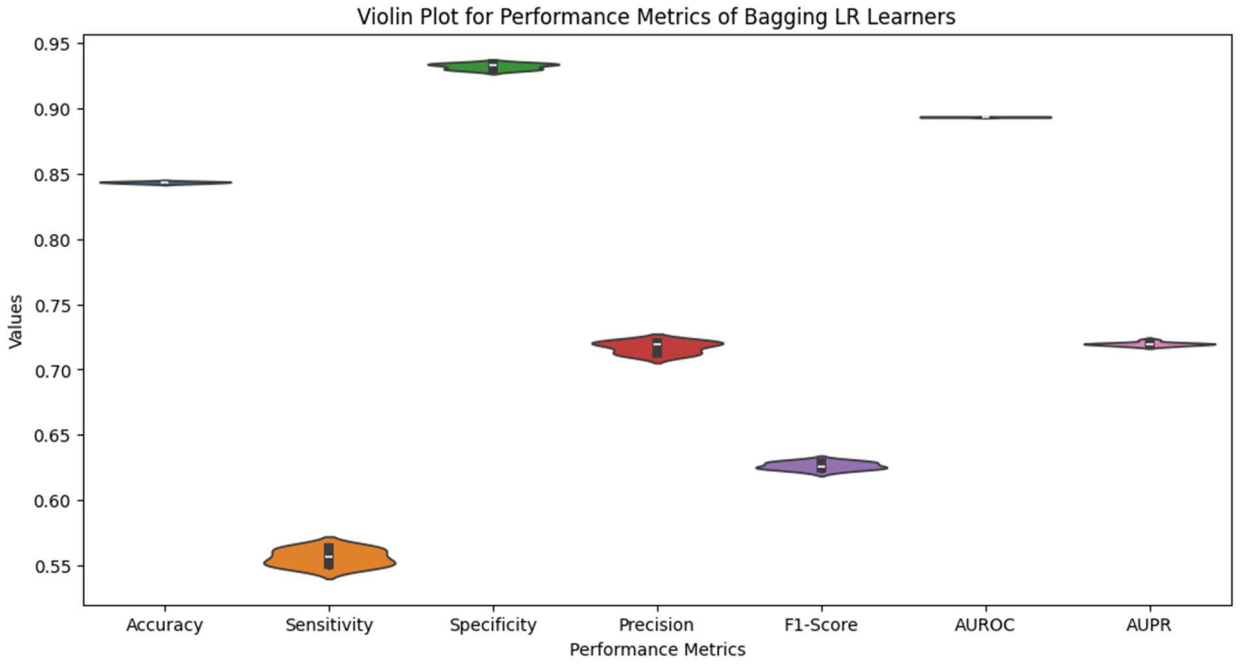
Learning rate = 0.0001, no regularization

Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUROC	AUPR
LR	0.843 ± 0.001	0.556 ± 0.006	0.932 ± 0.002	0.717 ± 0.004	0.627 ± 0.003	0.893 ± 0.000	0.720 ± 0.001
Voting ensemble	0.843552	0.555295	0.932715	0.718529	0.626453	0.893589	0.720353
Stacking ensemble	0.843552	0.556077	0.932473	0.718087	0.626782	0.755387	0.522309

Violin plots for each performance metric for the 9 bagging LR learners





For third dataset (https://www.kaggle.com/mlg-ulb/creditcardfraud):

Learning rate = 0.01, L2 regularization, Regularization strength = 0.1

Performance on Test set

	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUROC	AUPR
LR	0.993 ± 0.001	0.766 ± 0.009	0.999 ± 0.001	0.967 ± 0.027	0.855 ± 0.008	0.978 ± 0.001	0.872 ± 0.002
Voting ensemble	0.993405	0.767857	0.999749	0.988506	0.864322	0.978656	0.872587
Stacking ensemble	0.992672	0.776786	0.998744	0.945652	0.852941	0.888031	0.773377

Violin plots for each performance metric for the 9 bagging LR learners