CSE 472 : Machine Learning Sessional
Offline-2 Report

Name : Protoy Barai
Student ID : 1905068
Date : 20 September, 2024

Instructions :

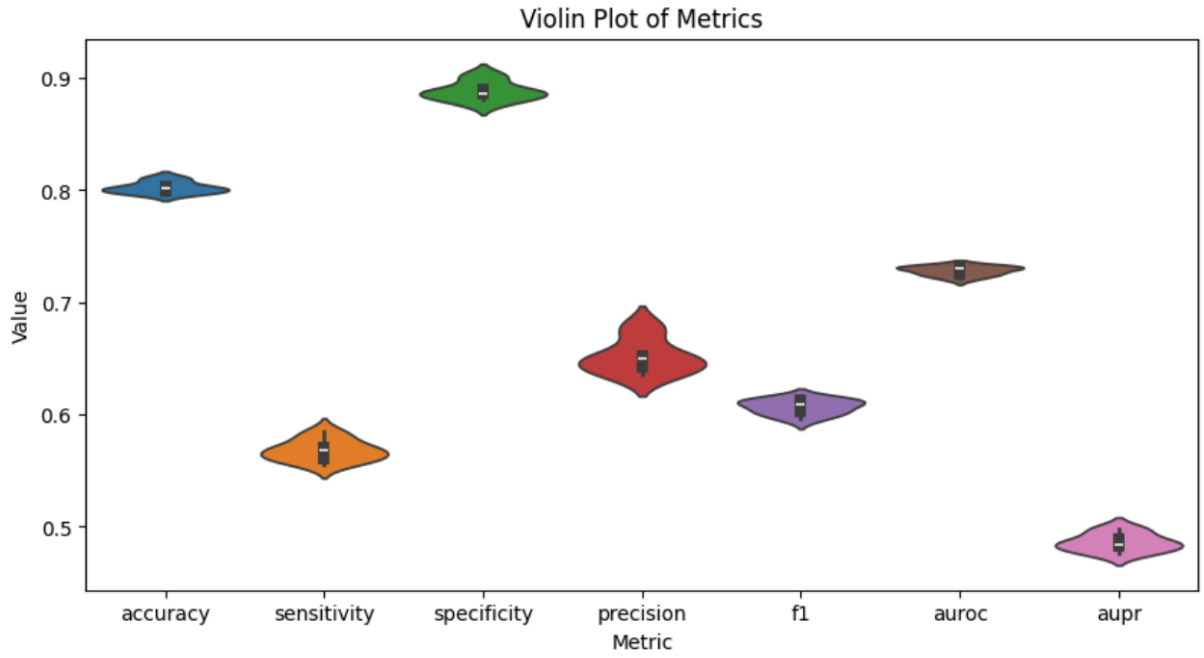
1. Create a folder named ‘datasets’ in the same folder where the 1905068.ipynb file is located.
2. Now download all the datasets. Unzip the adult dataset.
3. Now open the 1905068.ipynb file and execute all the cells till ‘Preprocessing Dataset 1’. From here, you can scroll down and choose a dataset to preprocess.
4. After preprocessing a dataset, scroll down below to the last cell with the markdown ‘Testing Any Dataset’ and execute the cell to get the scores.

Performance Evaluation :

Dataset 1 : learning rate : 0.01, number of iterations : 1000

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR	0.8026 ± 0.0047	0.5682 ± 0.0088	0.8886 ± 0.0077	0.6519 ± 0.0143	0.6070 ± 0.0062	0.7284 ± 0.0038	0.4862 ± 0.0074
Voting Ensembler	0.8027	0.5661	0.8894	0.6524	0.6062	0.7278	0.4858
Stacking Ensembler	0.8020	0.5952	0.8778	0.6410	0.6173	0.7365	0.4902

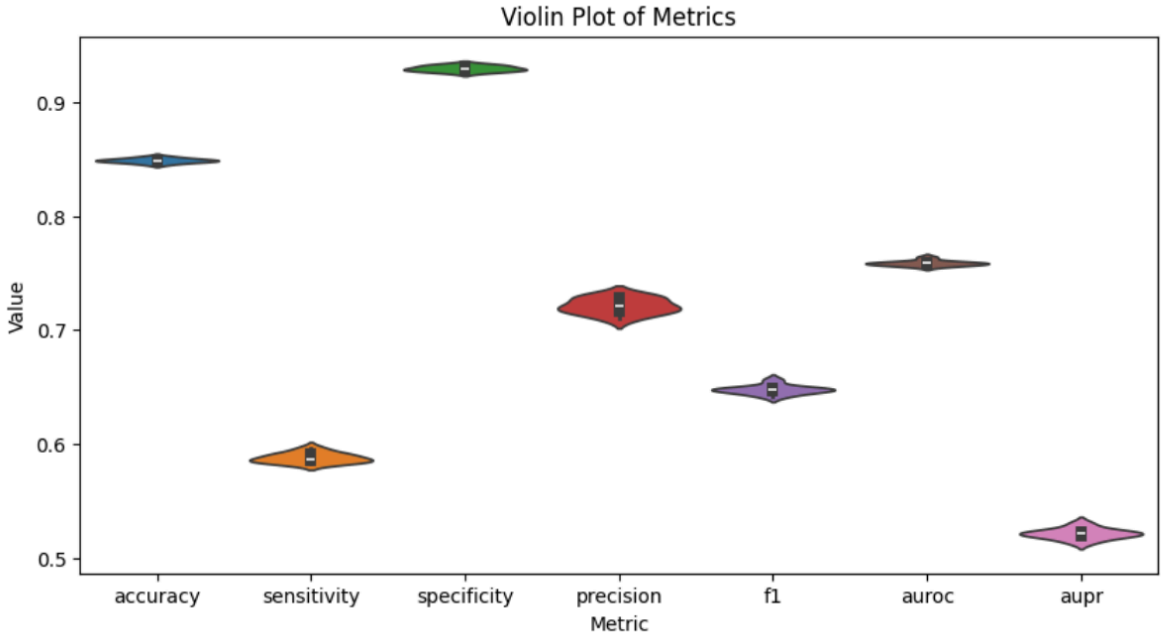
Violin Plot for Dataset 1 :



Dataset 2 : learning rate : 0.1, number of iterations : 1000

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR	0.8490 ± 0.0018	0.5881 ± 0.0040	0.9297 ± 0.0023	0.7214 ± 0.0064	0.6479 ± 0.0036	0.7589 ± 0.0021	0.5216 ± 0.0043
Voting Ensembler	0.8510	0.5913	0.9313	0.7271	0.6522	0.7613	0.5265
Stacking Ensembler	0.8498	0.5936	0.9291	0.7214	0.6513	0.7614	0.5242

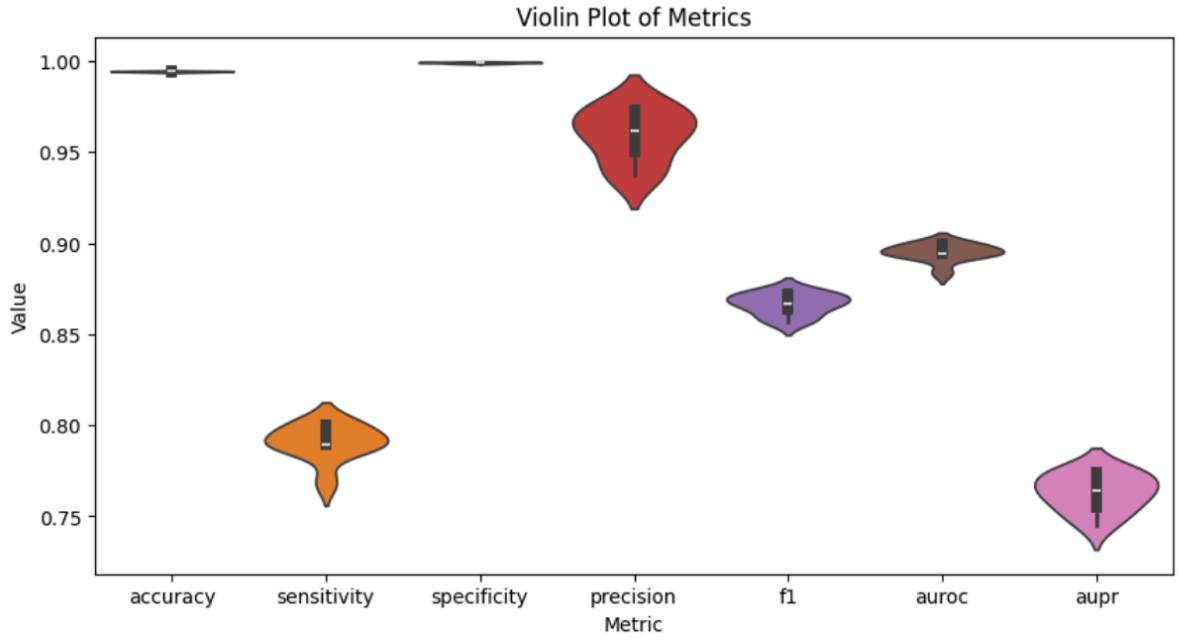
Violin Plot for Dataset 2



Dataset 3 : learning rate : 0.01, number of iterations : 1000

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUROC	AUPR
LR	0.9944 ± 0.0002	0.7906 ± 0.0092	0.9992 ± 0.0003	0.9591 ± 0.0135	0.8667 ± 0.0055	0.8949 ± 0.0045	0.7631 ± 0.0097
Voting Ensembler	0.9946	0.8000	0.9992	0.9620	0.8736	0.8996	0.7743
Stacking Ensembler	0.9934	0.8000	0.9980	0.9048	0.8492	0.8990	0.7284

Violin Plot for Dataset 3 :



Observations :

1. The performance scores vary slightly with the change of learning rate and number of iterations.
2. In most cases, The Majority Voting Classifier performs the best.
3. Data preprocessing plays the most important role in getting good performances overall.