Md. Raihan Sobhan

Student ID : 1905095

# Assignment 2: Logistic Regression with Bagging and Stacking

## How to run the models

In section ***6. Split the data into training, validation and testing sets*** of **1905095.ipynb**,

- Uncomment Features_Telco_TopK and label_Telco for Telco dataset (Dataset 1)
- Uncomment Features_adult_TopK and label_adult for adult dataset (Dataset 2)
- Uncomment Features_CreditCard_TopK and label_CreditCard for creditcard dataset (Dataset 3)

## For running with any new dataset:

In section ***6. Split the data into training, validation and testing sets*** of **1905095.ipynb**

1. Assign Features to the variable, Features_top_k_Scaled,
   *e.g. Features_top_k_Scaled = Features_New_Dataset*
2. Assign Label to the variable, label ,
   *e.g. label = label_New_Dataset*

---

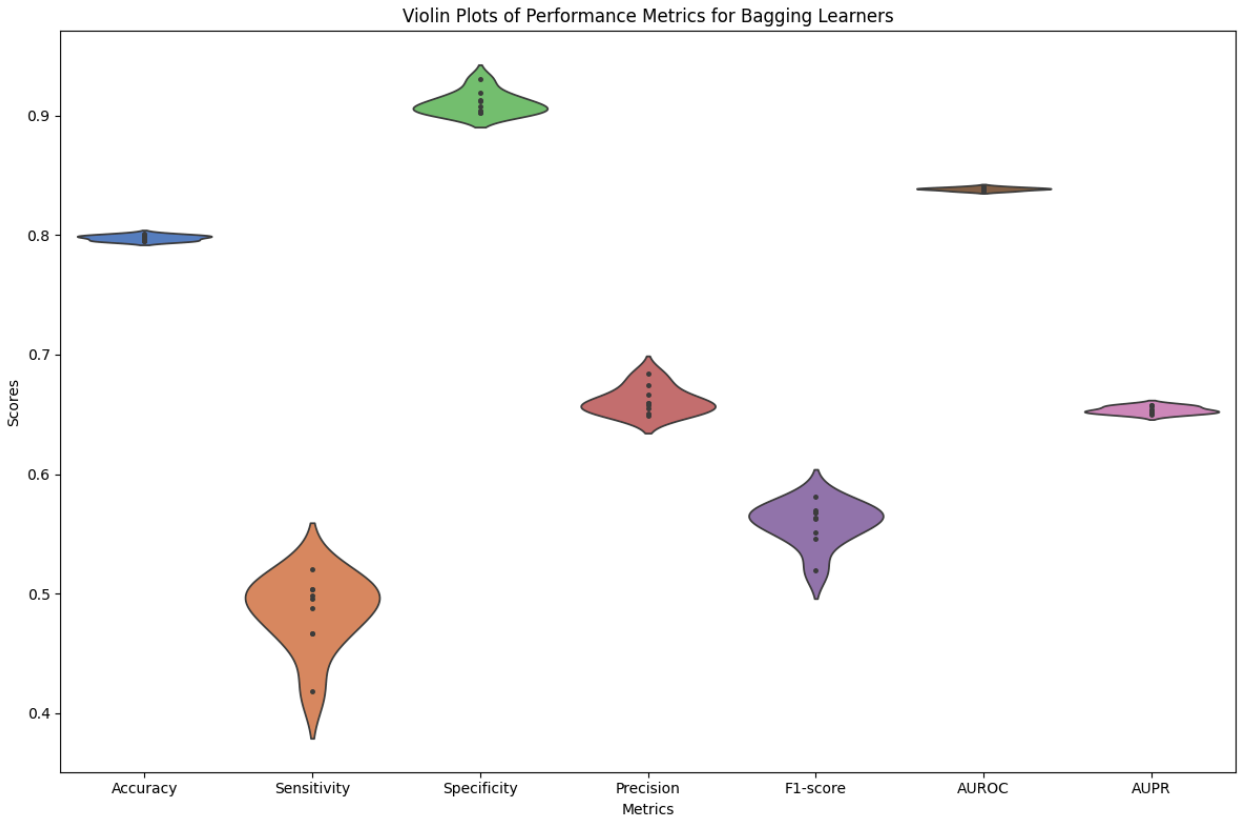
## Dataset - 1 : Telco

## Performance on Test set

| | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **F1-score** | **AUROC** | **AUPR** |
|---|---|---|---|---|---|---|---|
| LR* | 0.798 ± 0.002 | 0.485 ± 0.029 | 0.911 ± 0.009 | 0.662 ± 0.011 | 0.559 ± 0.017 | 0.839 ± 0.001 | 0.654 ± 0.003 |
| Voting ensemble | 0.798 | 0.493 | 0.908 | 0.659 | 0.564 | 0.839 | 0.655 |
| Stacking ensemble | 0.804 | 0.515 | 0.908 | 0.669 | 0.582 | 0.843 | 0.663 |

** For LR, report average ± stdev for the 9 bagging LR learners*
*learning_rate=0.01, num_iterations=1000, regularization='l2', lambda_param=0.1*

## Violin Plot :




---

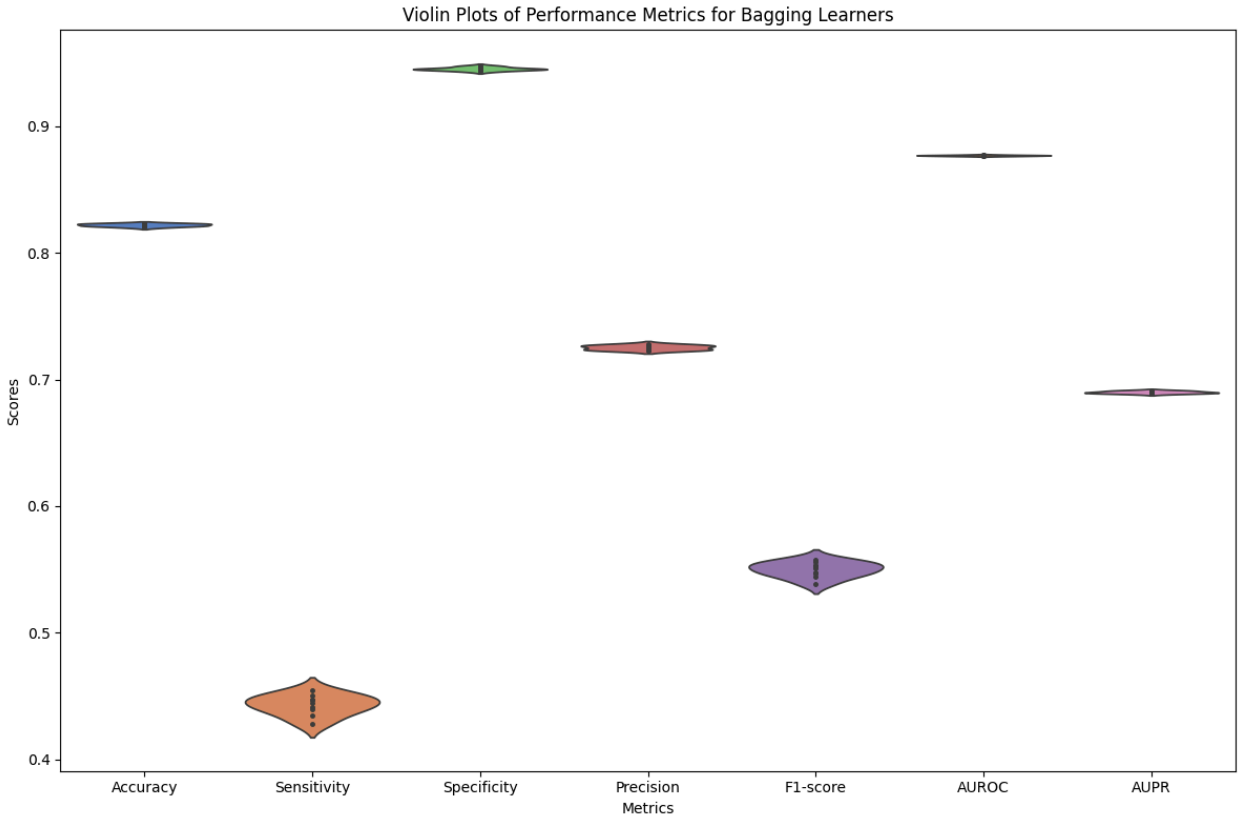
## Dataset - 2 : Adult

### Performance on Test set

| | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
|---|---|---|---|---|---|---|---|
| LR* | 0.822 ± 0.001 | 0.443 ± 0.008 | 0.945 ± 0.001 | 0.725 ± 0.002 | 0.550 ± 0.006 | 0.877 ± 0.000 | 0.690 ± 0.001 |
| Voting ensemble | 0.822 | 0.443 | 0.945 | 0.724 | 0.549 | 0.877 | 0.690 |
| Stacking ensemble | 0.825 | 0.474 | 0.939 | 0.717 | 0.571 | 0.876 | 0.689 |

** For LR, report average ± stdev for the 9 bagging LR learners*

## Violin Plot:




---

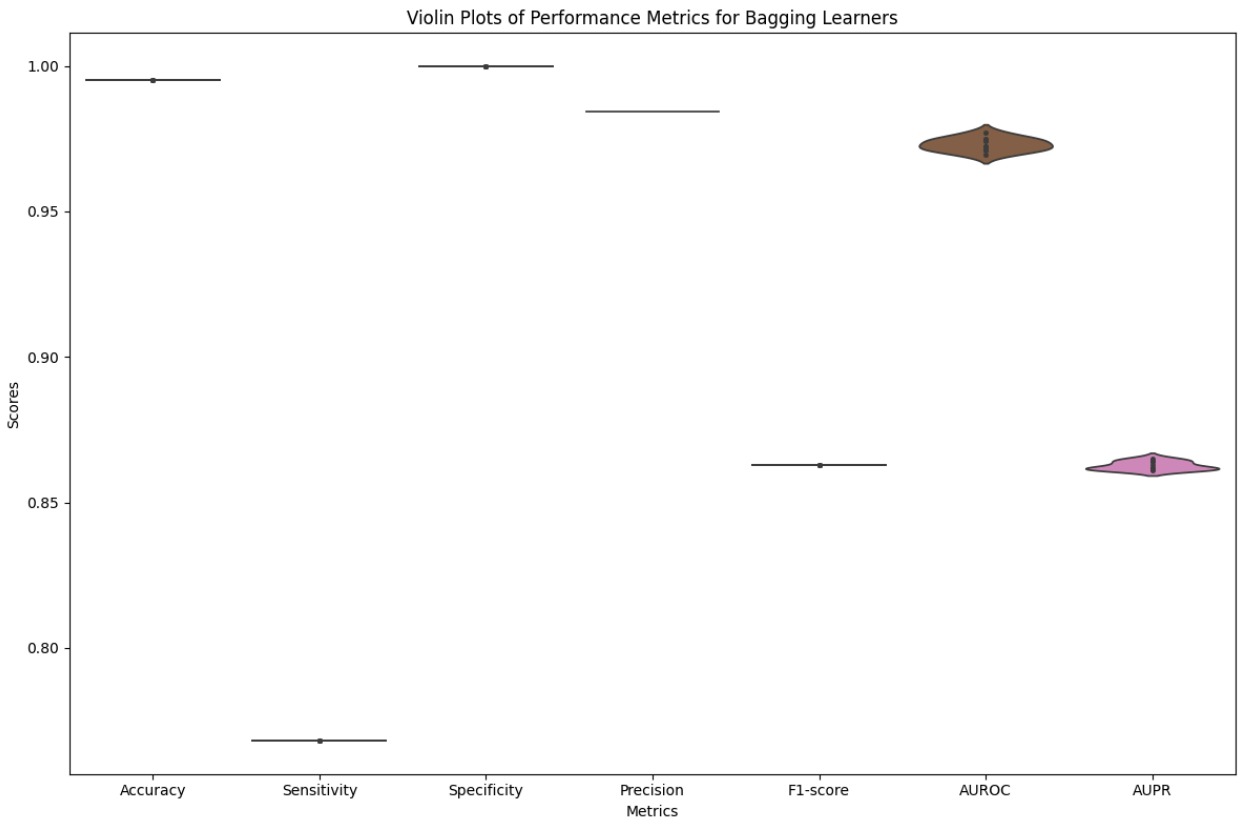
## Dataset - 3 : Credit Card Fraud Detection

### Performance on Test set

| | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
|---|---|---|---|---|---|---|---|
| LR* | 0.995 ± 0.000 | 0.768 ± 0.000 | 1.000 ± 0.000 | 0.984 ± 0.000 | 0.863 ± 0.000 | 0.973 ± 0.002 | 0.863 ± 0.001 |
| Voting ensemble | 0.995 | 0.768 | 1.000 | 0.984 | 0.863 | 0.974 | 0.864 |
| Stacking ensemble | 0.995 | 0.768 | 1.000 | 0.984 | 0.863 | 0.964 | 0.860 |

*learning_rate=0.01, num_iterations=1000, regularization='l2', lambda_param=0.1*

## Violin Plot:


Violin Plots of Performance Metrics for Bagging Learners
1.00
0.95
0.90
0.85
0.80
Scores
Accuracy
Sensitivity
Specificity
Precision
F1-score
AUROC
AUPR
Metrics