# Instructions:

1. The script will ask for a user input with the prompt: “Enter 1, 2 or 3: “
2. The inputs are associated with the three datasets
   1. input=1 is for Telco Customer Churn dataset
   2. input=2 processes the Adult dataset
   3. input=3 works with Credit Card Fraud Detection dataset.
3. Feature selection is carried out by default by correlation analysis in the function correlationAnalysis(n\_feature\_selection, features, target). To enable feature selection by information gain, look for the comment ‘# comment out the following 2 lines for information gain’ in the second last block and comment out as it says. At the same time, comment out the two lines followed by the comment ‘# comment out the following 2 lines for correlation analysis’ to turn off information gain.

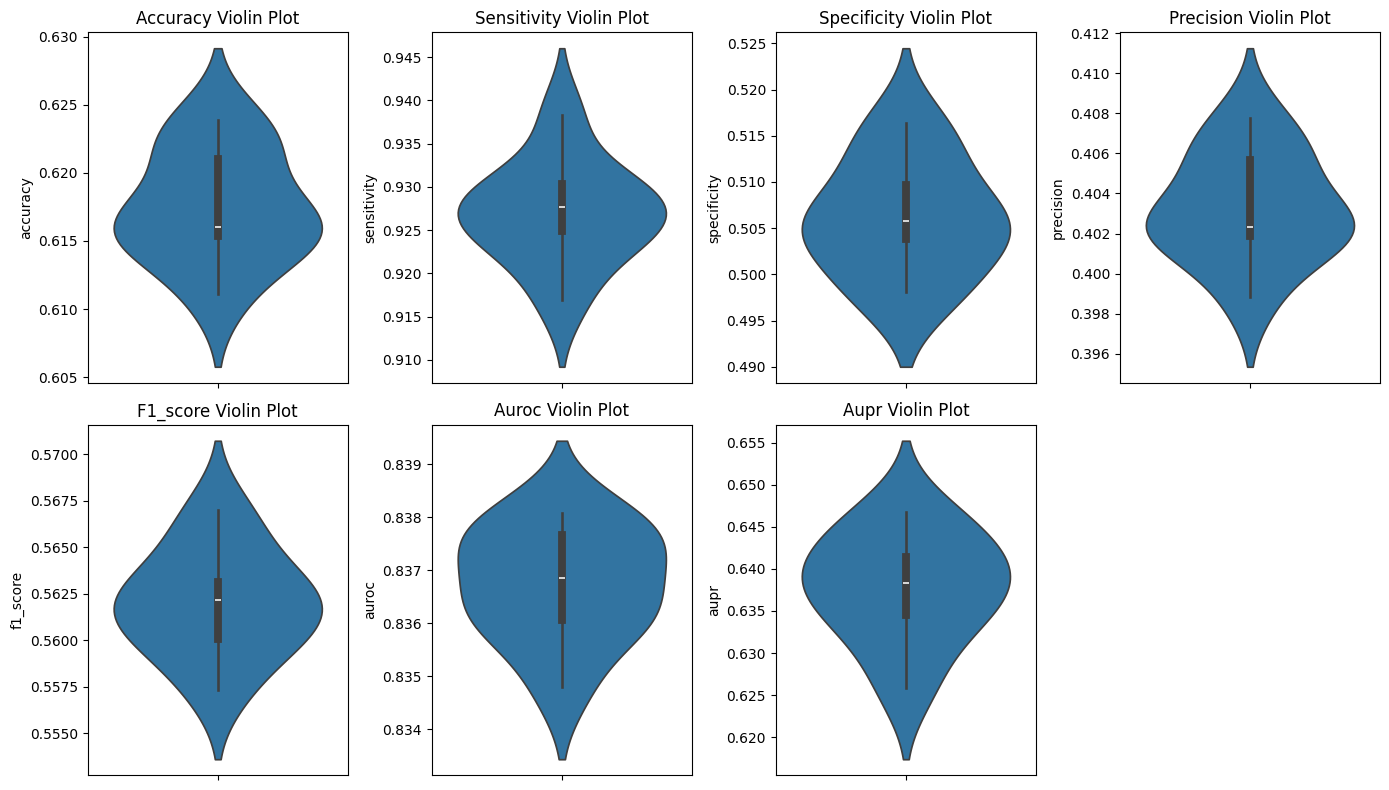
# Points to be noted:

1. The target column names are hardcoded for each dataset. Though coincidentally in every dataset the last column is the target, we are not assuming anything.

# Dataset 1:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
| LR\* | 0.6175 ± 0.0039 | 0.9270 ± 0.0056 | 0.5061 ± 0.0059 | 0.4033 ± 0.0026 | 0.5620 ± 0.0027 | 0.8367 ± 0.0010 | 0.6378 ± 0.0062 |
| Voting ensemble | 0.6188786373314408 | 0.9276139410187667 | 0.5077220077220077 | 0.40420560747663553 | 0.5630593978844589 | 0.8379568768308715 | 0.6367674534437616 |
| Stacking ensemble | 0.6139105748757985 | 0.9168900804289544 | 0.5048262548262549 | 0.4 | 0.5570032573289903 | 0.8333130622004616 | 0.6315303326143523 |

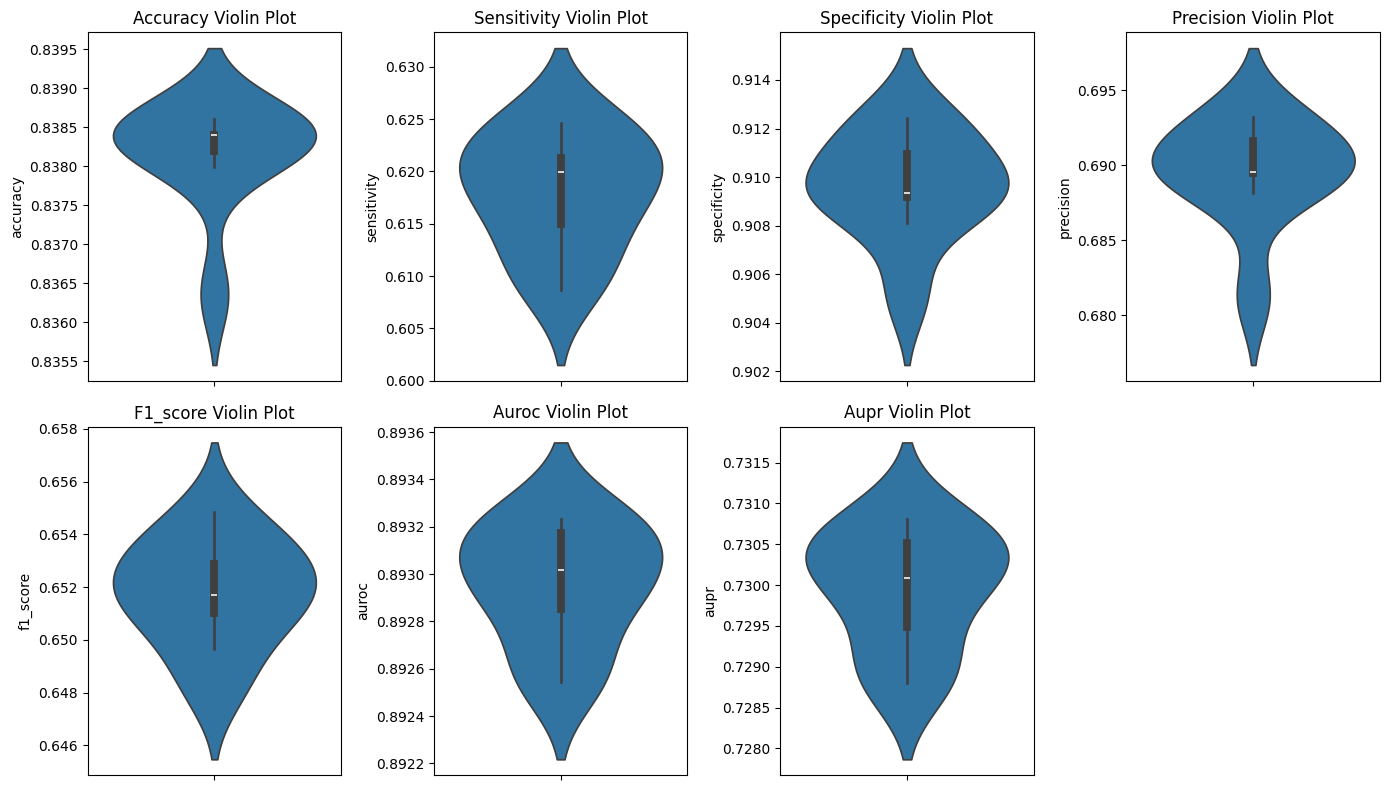
# Violin Plot:



# Dataset 2:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
| LR\* | 0.8382 ± 0.0007 | 0.6180 ± 0.0053 | 0.9096 ± 0.0021 | 0.6895 ± 0.0033 | 0.6518 ± 0.0019 | 0.8930 ± 0.0002 | 0.7300 ± 0.0007 |
| Voting ensemble | 0.8384055743416333 | 0.6178929765886287 | 0.910004072213927 | 0.6903316207379729 | 0.6521067725568056 | 0.893429889142261 | 0.7310326513560472 |
| Stacking ensemble | 0.8386105133722718 | 0.6199832775919732 | 0.9095968508212298 | 0.690088413215449 | 0.653160096894957 | 0.8910580628700799 | 0.7258806944006679 |

# Violin plot:



# Dataset 3:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
| LR\* | 0.9990 ± 0.0000 | 0.4086 ± 0.0172 | 0.9999 ± 0.0000 | 0.8802 ± 0.0044 | 0.5580 ± 0.0169 | 0.9620 ± 0.0085 | 0.7085 ± 0.0017 |
| Voting ensemble | 0.9989779015260988 | 0.4111111111111111 | 0.9999117480937588 | 0.8809523809523809 | 0.5606060606060606 | 0.9710219962973422 | 0.710583977465309 |
| Stacking ensemble | 0.9989602791386177 | 0.4 | 0.9999117480937588 | 0.8780487804878049 | 0.549618320610687 | 0.9564125796228309 | 0.6960891212287458 |

# Violin plot:

