|  |  |
| --- | --- |
|  | Report |
|  |  |
|  | Sumaiya Saeha  Student ID: 1905033 |

# 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 correlation analysis in the function correlationAnalysis(n\_feature\_selection, features, target) by default. 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.

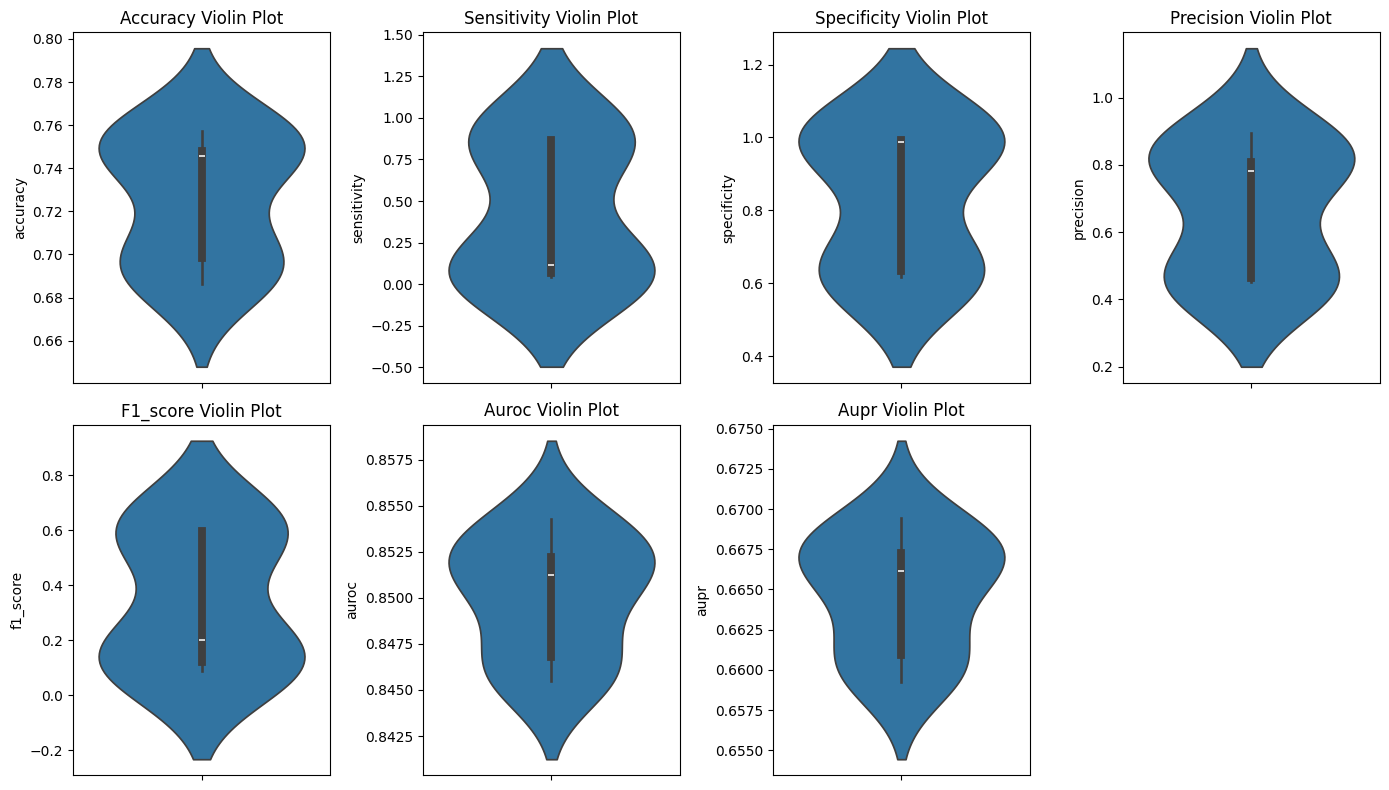
# Points to be noted:

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\* | 0.7257 ± 0.0281 | 0.4284 ± 0.3966 | 0.8328 ± 0.1804 | 0.6633 ± 0.1845 | 0.3422 ± 0.2343 | 0.8501 ± 0.0031 | 0.6647 ± 0.0035 |
| Voting ensemble | 0.759403832505323 | 0.12332439678284182 | 0.9884169884169884 | 0.7931034482758621 | 0.21345707656612528 | 0.849567060357945 | 0.6629838100705672 |
| Stacking ensemble | 0.6983676366217175 | 0.8820375335120644 | 0.6322393822393823 | 0.46338028169014084 | 0.6075715604801477 | 0.8531434057573468 | 0.6565204642312056 |

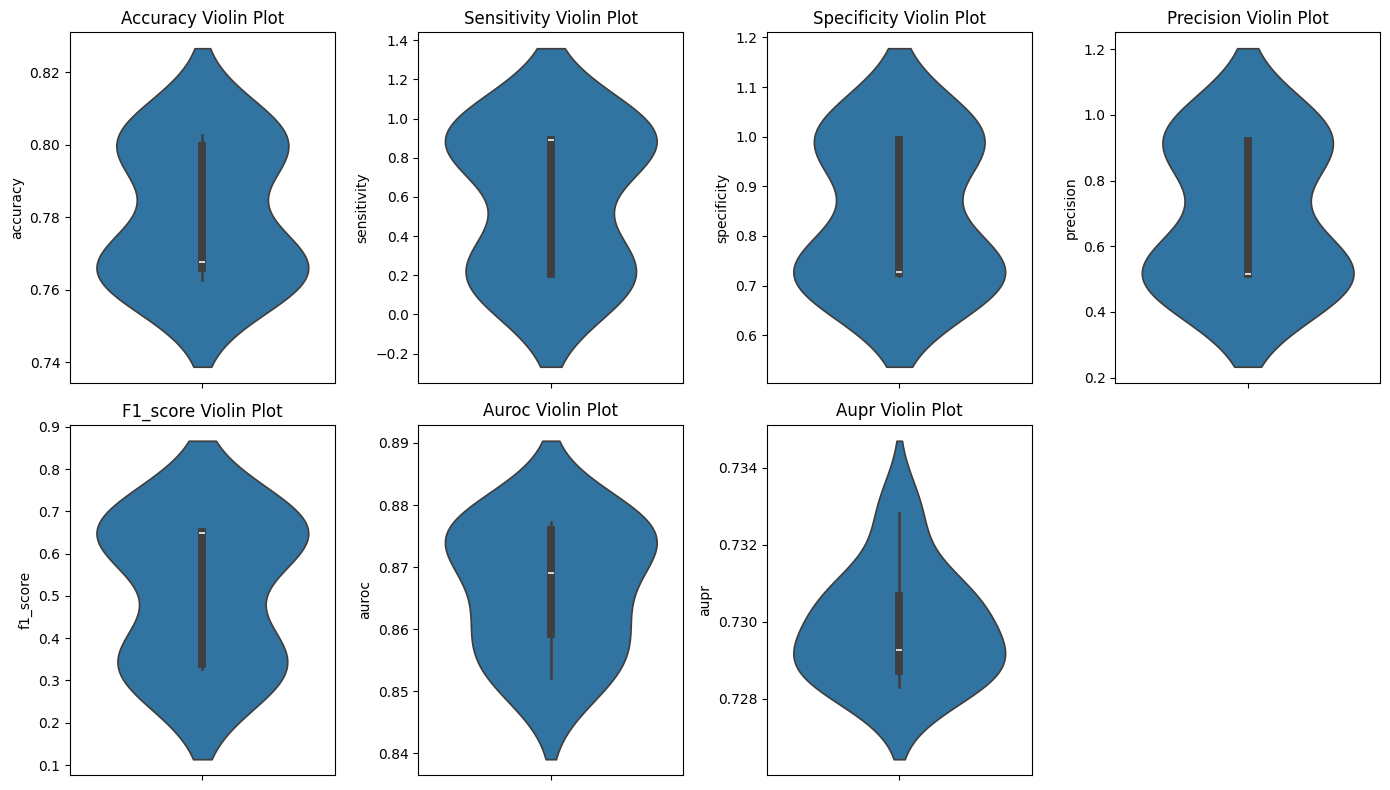
# Violin Plot:



# Dataset 2:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
| LR\* | 0.7811 ± 0.0174 | 0.5876 ± 0.3403 | 0.8443 ± 0.1341 | 0.6949 ± 0.2026 | 0.5125 ± 0.1561 | 0.8667 ± 0.0096 | 0.7299 ± 0.0014 |
| Voting ensemble | 0.7728018036482885 | 0.880932556203164 | 0.7374932028276238 | 0.5228564368668149 | 0.6562257714374322 | 0.8777249813120207 | 0.7310830178442778 |
| Stacking ensemble | 0.8088747694199631 | 0.7477102414654455 | 0.828847199564981 | 0.5878887070376432 | 0.6582371266263515 | 0.8825342213009912 | 0.7326262716534329 |

# Violin plot:



# Dataset 3:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | Precision | F1-score | AUROC | AUPR |
| LR\* | 0.9938 ± 0.0002 | 0.7183 ± 0.0143 | 0.9997 ± 0.0002 | 0.9791 ± 0.0102 | 0.8285 ± 0.0077 | 0.9742 ± 0.0018 | 0.8407 ± 0.0060 |
| Voting ensemble | 0.9936476911800636 | 0.7093023255813954 | 0.9997504367357125 | 0.9838709677419355 | 0.8243243243243243 | 0.9773767999024963 | 0.8419291601895765 |
| Stacking ensemble | 0.9943806498900561 | 0.7790697674418605 | 0.99900174694285 | 0.9436619718309859 | 0.8535031847133758 | 0.9631342824475773 | 0.7840230181844635 |

# Violin plot:

