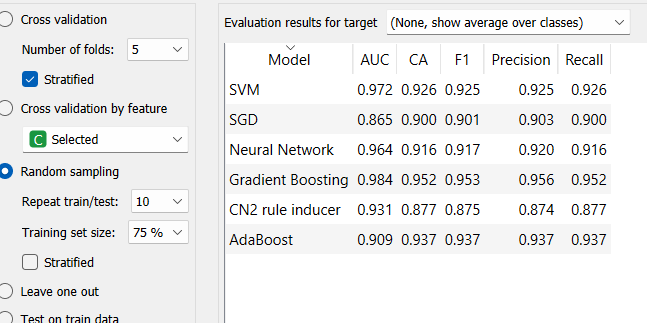
Title- Optimizing Campus Safety: A Cloud-Based Crime Reporting System with Machine Learning Insights

Dataset- <https://data.mendeley.com/datasets/fbs9mgmh4y/3>

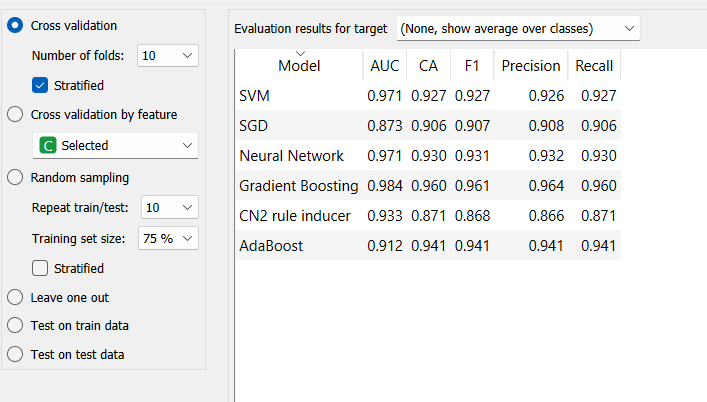
Description- For the current work we have used a system with Intel(R) Core (TM) i5-8265U CPU @ 1.60GHz (1.80 GHz) with 8 GB RAM & 2GB integrated graphics card with Windows 11 Home Insider Preview Single Language. We have used Orange Data Mining software version 3-3.38.0. Initially passing the dataset over the discussed machine learning yields the results as mentioned in figure 3. Gradient boosting shows higher values in the area under the curve, classification accuracy, F1-score, precision, and recall value with 5-fold and 10-fold cross-validation. Classification accuracy of around 81% is exhibited.

All the ML algorithms used here have 75:25 as the data split ratio.

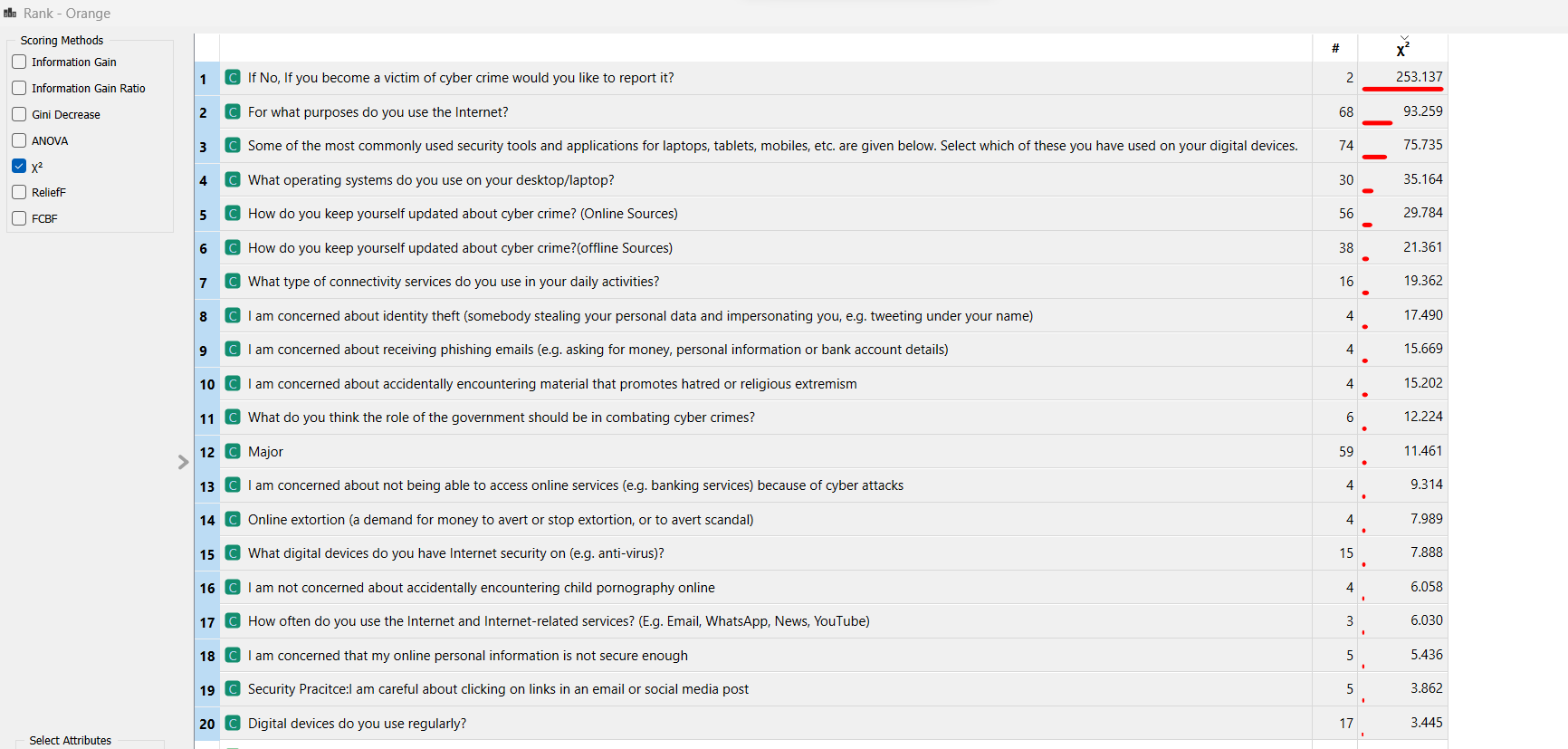
Cross validation: 5-fold; stratified



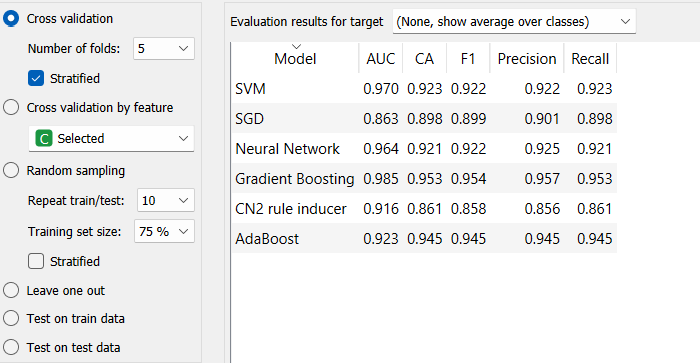
Cross validation: 10-fold; stratified



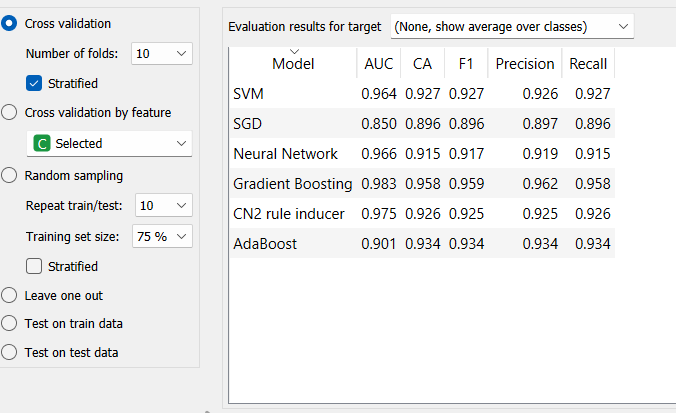
Chi square test showing 20 important features of the dataset

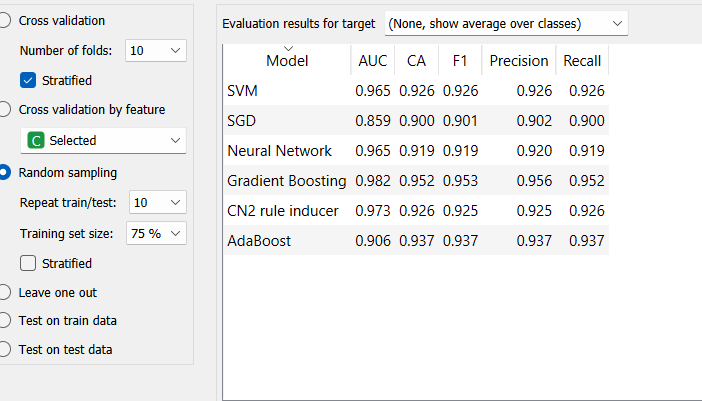


Cross validation: 5-fold; stratified

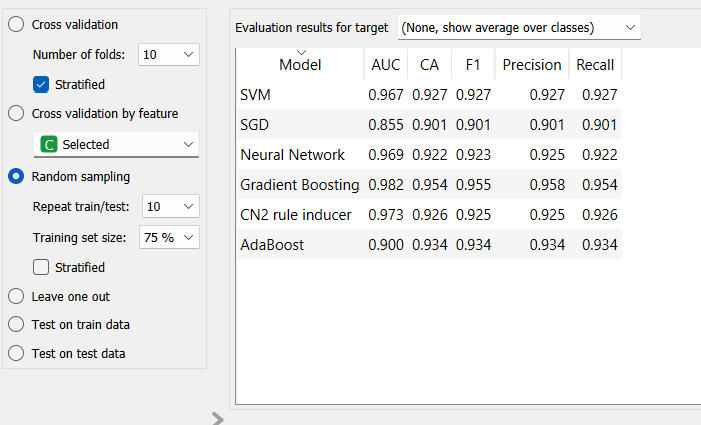


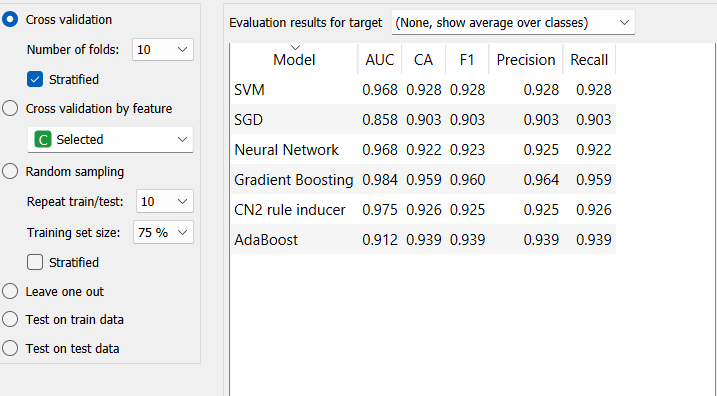
Cross validation: 10-fold; stratified



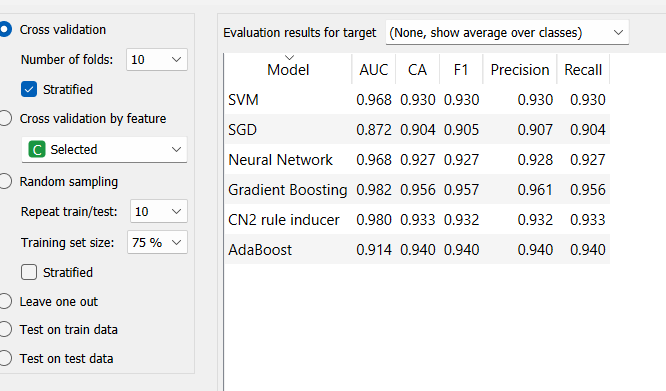
Cross validation: 10-fold; stratified 

With 25 features





With 40 features – Relieff



Methodology- The data preprocessing is done via Hot Deck Imputation followed by applying different ML algorithms namely CN2 Rule Induction (CN2); Stochastic Gradient Descent (SGD); Support Vector Machine (SVM); Random Forest (RF)l Gradient Boosting (GB); Stacking; Cross Validation