**Cyber Security-Network Security: Network Intrusions Detection Case-Study**

* For Binary Classification Models all the Datasets have been imported and appended.
* All the EDA Process is done like checking Null Values, Missing Value Treatment, Outlier Capping.
* For Feature Engineering: I have performed Univariate Regression by taking Somerce D Values of each column and taking Top 15 Variables, RFE Done for 15 variables, Select K Best done for 15 variables, and in the end taking Set of all three functions I get my Important Variables.
* Selecting my Final Variables after deducting Variables having VIF Values <=5.
* Algorithms used for Binary Classification:

1. Simple Logistic Regression
2. KNN Classifier
3. Ada Boost Classifier

**A] Simple Logistic Regression:**

1. **Gini Index Score for Train and Test are 99%.**
2. **ROC-AUC Scores for Train and Test are 99%.**
3. **Predicted Probability for separating 1 and 0, Calculated Cut-off, Specificity, Sensitivity, Accuracy.**

**and choosing the cut-off based on Highest Accuracy.**

1. **The Metrics Classification Report gives Accuracy of 99% for Both Train and Test.**
2. **This is a Good Stable Model.**

**B] KNN Classifier:**

1. **Grid Search done for getting Best Parameters.**
2. **ROC-AUC Scores for Train and Test are 99%.**
3. **Metrics Classification Report gives Accuracy of 100% for Train and Test.**
4. **The other Metrics such as Precision, Recall, F1-Score are all 100% for Train and Test.**
5. **It’s a Highly Stable Model.**

**C] Ada Boost Classifier:**

1. **Grid Search done for getting Best Parameters.**
2. **ROC-AUC Scores for Train and Test are 99%.**
3. **Metrics Classification Report gives Accuracy of 100% for Train and Test.**
4. **The other Metrics such as Precision, Recall, F1-Score are all 100% for Train and Test.**
5. **It’s a Highly Stable Model.**

**To Conclude:**

**All the Three Algorithms are Giving Good Accuracies and are Highly Stable, but if 100% Accuracy needed for Binary Classification - Ada Boost and KNN Models have 1% higher accuracy than Logistic Regression.**

**Multinomial Classification**

* **Same Datasets are appended but this time each attack has its own Classification Number.**
* **Same Variables chosen as chosen in Binary Classification Model.**
* **Algorithms used for Multinomial Classification are:**

1. **Simple Logistic Regression**
2. **Random Forest Classifier**
3. **Gradient Boost Classifier**

**A] Simple Logistic Regression:**

1. **Accuracy Score for Train and Test Data is 98%.**
2. **Metric Classification Report Gives 99% Accuracy for both Train and Test Data but other Metrics such as Precision, Recall and F1-Score are not that Good.**

**B] Random Forest Classifier:**

1. **With help of Grid Search CV I have got best Parameters for Random Forest Classifier.**
2. **Accuracy Score for Train and Test Data is 99%.**
3. **Metric Classification Report Gives 100% Accuracy for both Train and Test Data And other Metrics such as Precision, Recall and F1-Score are Better than Simple Logistic Regression.**

**C] Gradient Boost Classifier:**

1. **Tuning Parameters taken from Past Files because Best Parameters were not giving output even after setting up Grid. Due to this Tuning Parameters were given based on Past Files.**
2. **Sample taken from Train Dataset as the Model were not fitting on whole Train Data, and Based on Sample Data the Model is Fitted.**
3. **Accuracy of Sample Dataset of Train and Test is 99%.**
4. **Metrics Classification Report gives Normal Precision, Recall, F1-Score.**
5. **If the same sample model is Tested on Whole Train and Test Dataset the Accuracy is 70% for Train and Test.**
6. **If the whole Train data was fitted with Grid Search we would have got best parameters and the Accuracy would also get Increased.**
7. **But with Normal Parameters the Accuracy of Sample Train and Test is 99% and same model on whole Train Data is 70% for Train and Test, which is Good Stable Model.**

**To Conclude: Random Forest Classifier is a very Good and Stable Model as it gives a bit higher Metric Scores than Logistic Regression and Gradient Boosting Technique in this Scenario!**