

Machine learning classifier for URLs

Nicolas Ricardo Enciso



UNSECURE



Data set

- 25065 attacks (anomalous) marked as 1
- 36000 normal (normal) marked as 0
- Total 61065 cases in the dataset.



UNSECURE



Splitted data

- 70% for training
- 30% for testing
- Random election of the cases from the data
- Dataset :
 - Original : 8 features with no changes



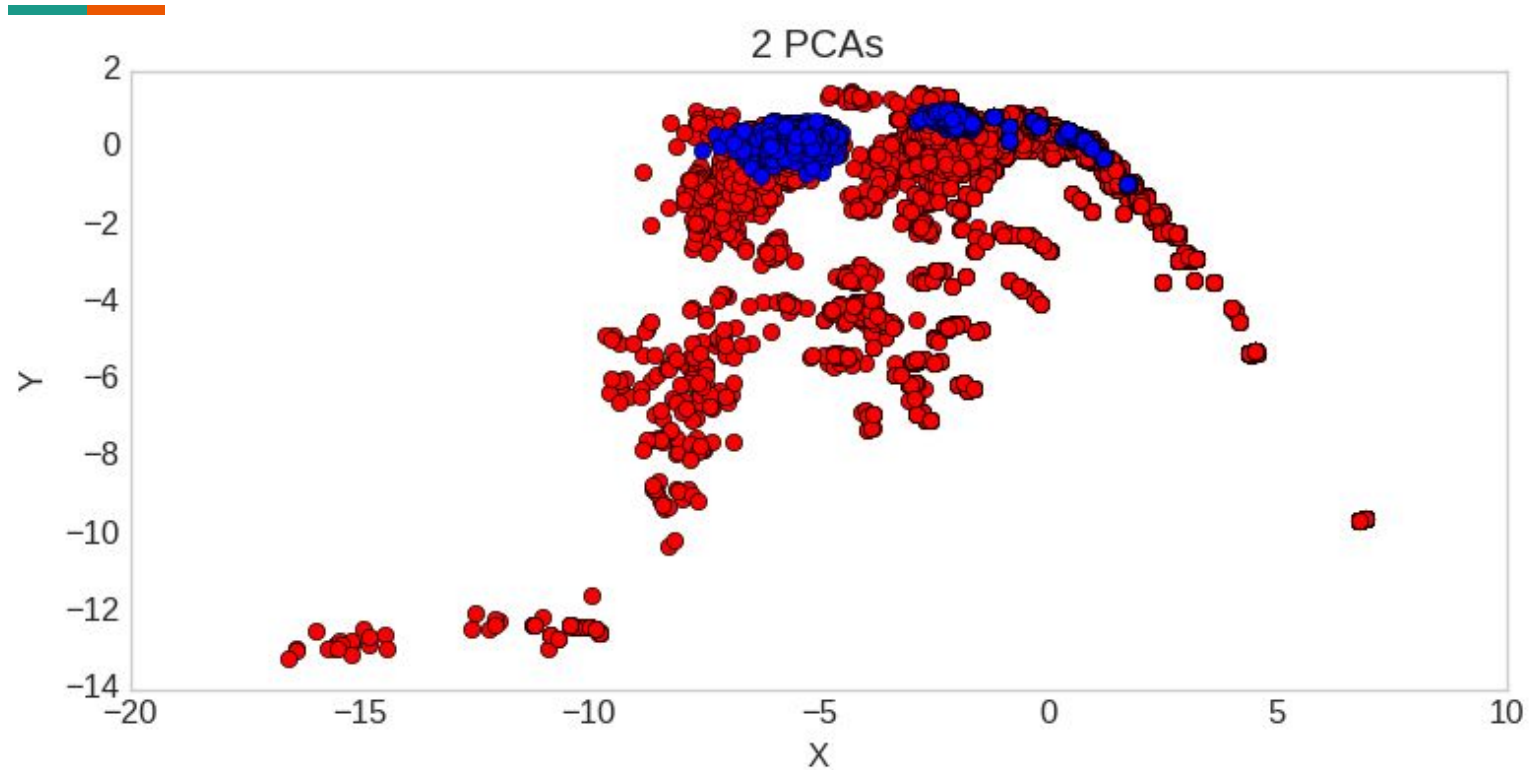
U N S E C U R E

Classifier algorithms

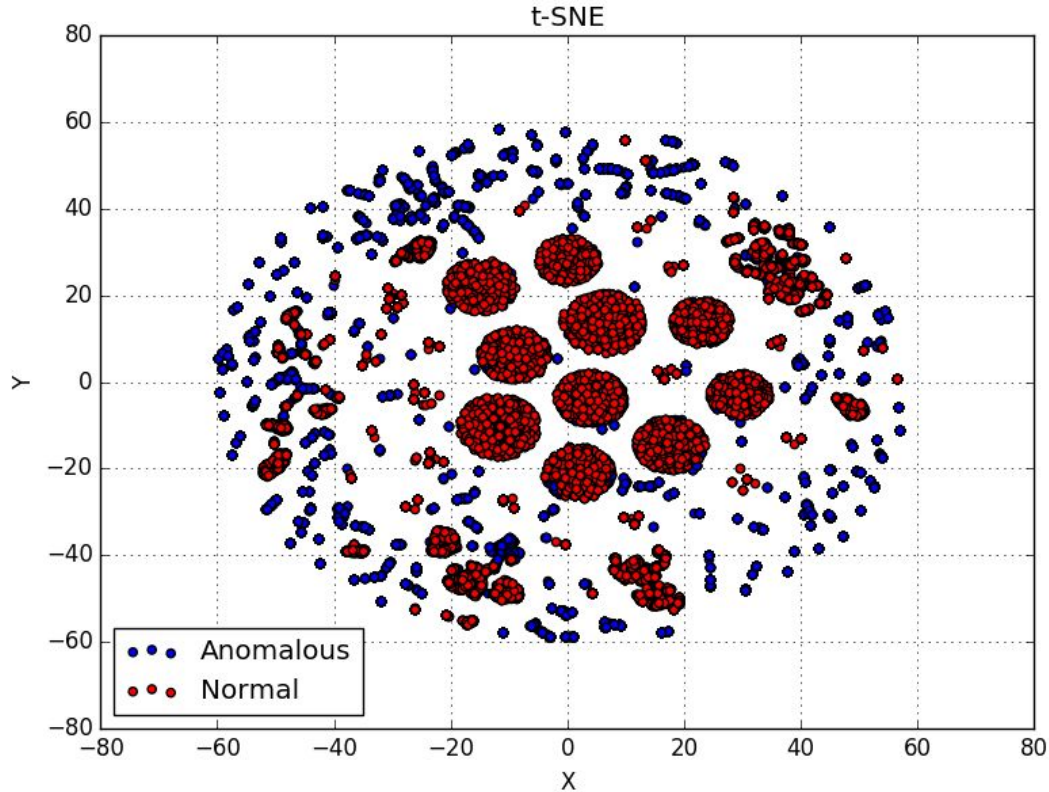


- Naive Bayes classifier:
 - Gaussian classifier
 - Multinomial classifier
- Support Vector Machine:
 - Linear
 - Gaussian ($C = 1.11$ and $\gamma = 0.09$)
 - Sigmoid
- Logistic Regression
- Random Forest (estimators = 100)

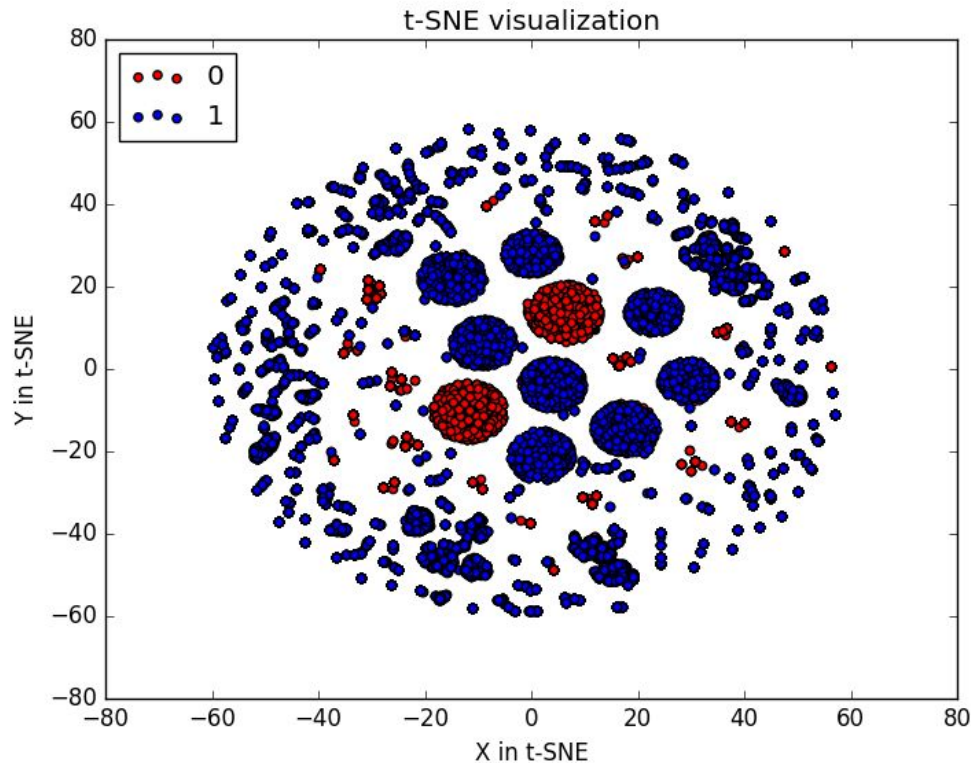
Visualization data: PCA (Anomalous in blue)



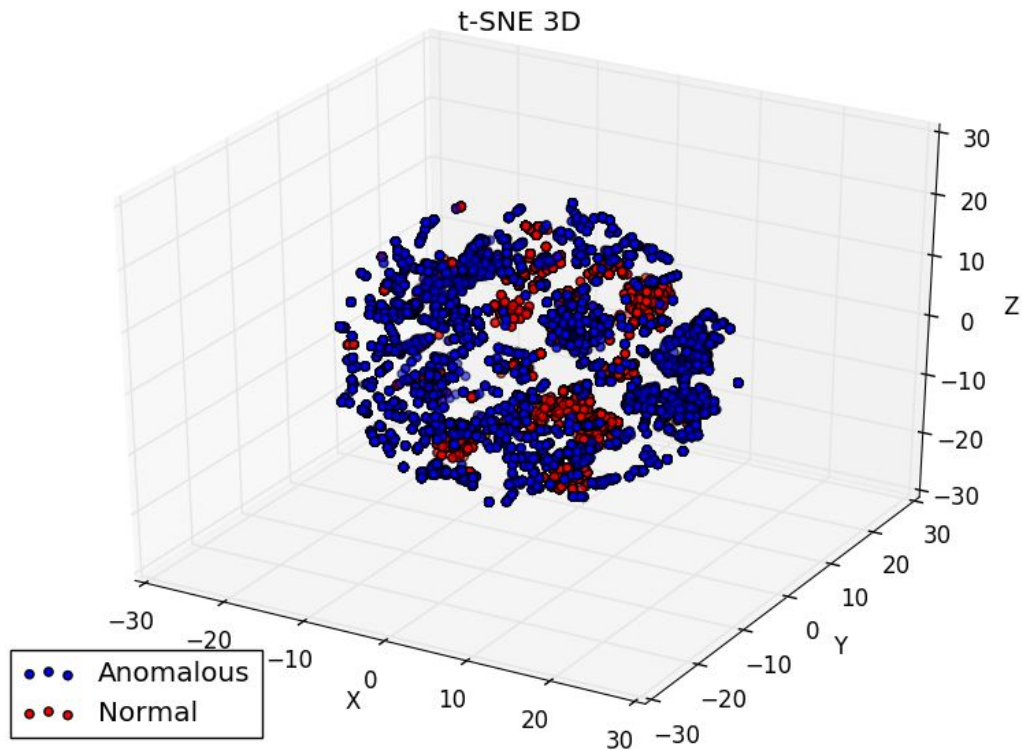
Visualization data: t-SNE 2D



Visualization data: t-SNE (1 anomalous - 0 normal)



Visualization data: t-SNE 3D





UNSECURE

Results original data

Results performance classification

		Anomalous predictions Original data			
	Precision	Recall	F1-Score	Accuracy	AUC
NB + Gaussian	0,720	0,200	0,320	0,643504366812	0,657966634589
NB + Multinomial	0,750	0,230	0,350	0,653056768559	0,650886405340
SVM + Sigmoid	0,000	0,000	0,000	0,592139737991	0,390236405688
SVM + linear	0,810	0,260	0,390	0,674672489083	0,693580346245
SVM + Gaussian	0,830	0,620	0,710	0,794596069869	0,870937300964
Logistic Regression	0,700	0,410	0,520	0,688373362445	0,707206599566
Random Forest	0,750	0,880	0,810	0,832860262009	0,933937629906

Results performance classification

	Normal cases predictions Original data				
	Precision	Recall	F1-Score	Accuracy	AUC
NB + Gaussian	0,630	0,950	0,760	0,643504366812	0,657966634589
NB + Multinomial	0,640	0,950	0,760	0,653056768559	0,650886405340
SVM + Sigmoid	0,590	1,000	0,740	0,592139737991	0,390236405688
SVM + linear	0,650	0,960	0,780	0,674672489083	0,693580346245
SVM + Gaussian	0,780	0,910	0,840	0,794596069869	0,870937300964
Logistic Regression	0,680	0,880	0,770	0,688373362445	0,707206599566
Random Forest	0,910	0,800	0,850	0,832860262009	0,933937629906

ROC

