Network Anomaly Detection

Anomaly Detection – Challenge 3

Team Pauliguel

Index

- Dataset description
- Undersampling
- Pre-processing
- Models and evaluation
- Tuning
- Kaggle submission

Dataset description

- Total of 82.332 instances.

- Unbalanced class representation
  - Positive class is 0.07% of the instances.

Undersampling

- To solve the unbalanced data.
- Random sample of the most common class.
- 80 instances in the training set.

Pre-processing

- Missing Values: inexistent.

- Categorical variables
    - dummy variables

Feature Extraction

Proto

	non attack	attack
arp	2859	1
icmp	15	0
igmp	18	0
eigr	0	1
gmtp	0	1
ipx-	0	1
ospf	64	2
pim	0	1
rvd	0	1
rtp	1	0
sctp	0	5
tcp	39121	18
udp	13922	8
unas	0	3

$$proto1 = arp + icmp + igmp$$

$$proto2 = eigrp + gmtp + ipxnip + \\ + pim + rvd + sctp + unas$$

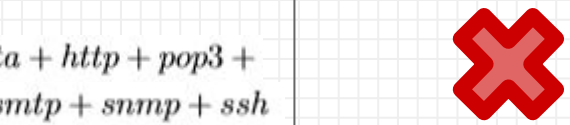
Service

	non attack	attack
dns	7493	5
ftp	1218	0
ftp-data	2552	0
http	5348	12
pop3	4	0
radius	2	0
smtp	1579	2
snmp	1	0
ssh	1291	0

$$service = ftp + ftpdata + http + pop3 + \\ + radius + smtp + snmp + ssh$$

State

	non attack	attack
CON	12099	5
ECO	12	0
FIN	37175	17
INT	5715	18
no	1	0
PAR	1	0
REQ	925	1
RST	71	0
URN	1	0



Feature Selection

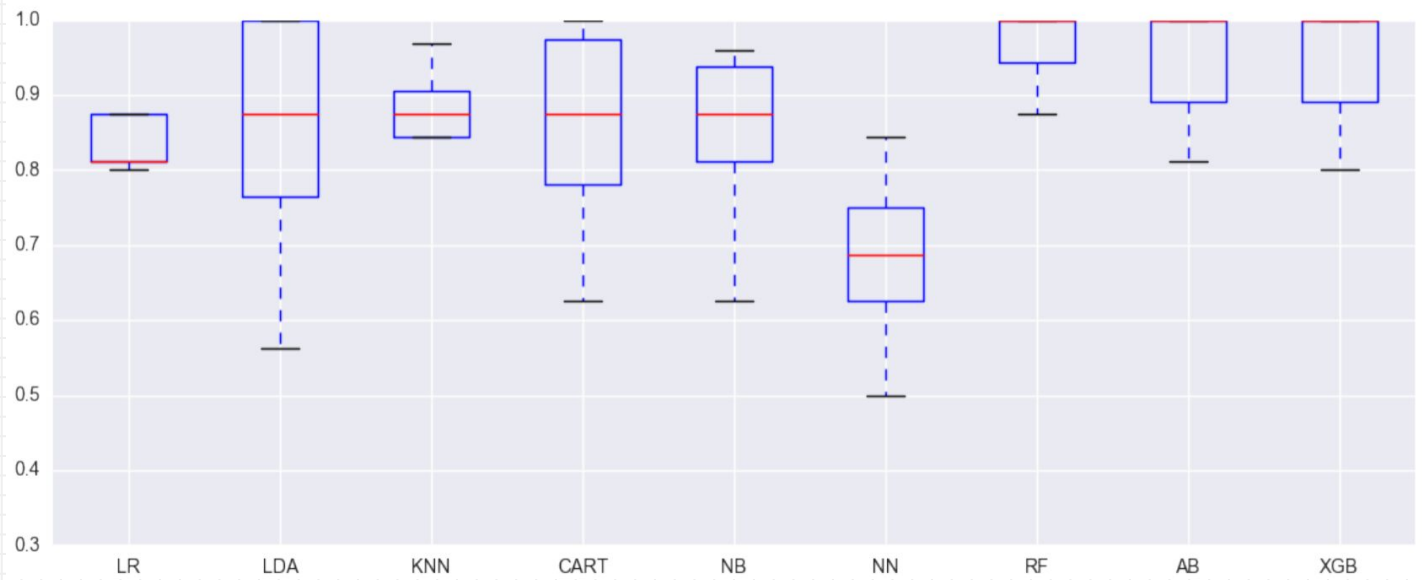
- PCA with 30 principal components
  - value based on the variances from the principal component
- Select only the variables that had non-zero importances
  - only applicable to the Random Forest algorithm

Models and evaluation

Models used:

- Logistic Regression
- Linear Discriminant Analysis
- K-Nearest Neighbors Classifier
- Decision Tree Classifier (CART)
- Naïve Bayes
- Multilayer Perceptron
- Random Forest
- AdaBoost
- Gradient Boosted Decision Trees (XGBoost)
    + One-class SVM
    + Isolation Forest

Results with AUC



Results with AUC

Feature Selection	Algorithm	Offline score (CV)	Public score
None	Random Forest	0.95850	0.78855
PCA (30)	Random Forest	0.87462	0.76687
Feature Importance	Random Forest	0.95850	0.85865

Tuning

- Random Forests parameters using grid search:
  - “number of estimators”,
  - “criterion”

Kaggle submission

0.85865

Public score - best submission

•••

Private score - best submission

Inspired with Glühwein!!!

THANK YOU!

Any questions?