documentation

Abstract

We can analyze hacker attacks through our use of data and use models to predict penetration rates

Design

This project originates from the bootcame sdaia. The data is downlod from kaggel

Data

The dataset contains 22544 rows & 41 columns . consists of a wide variety of intrusions simulated in a military network environment. The LAN was focused like a real environment and blasted with multiple attacks. A connection is a sequence of TCP packets starting and ending at some time duration between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Also, each connection is labelled as either normal or as an attack with exactly one specific attack type. Each connection record consists of about 100 bytes.

For each TCP/IP connection, 41 quantitative and qualitative features are obtained from normal and attack data (3 qualitative and 38 quantitative features) .The class variable has two :categories

- Normal
- .Anomalous

documentation

Algorithms

Mapping latitude and longitude to 3-dimensional coordinates so nearby continuous values would also be close in reality
Converting categorical features to binary dummy variables
Combining particular dummies and ranges of numeric features to highlight strong signals and illogical values status identified during EDA

has been used
Logistic regression, k nearest neighbors, and decision tree

We discovered that the test data does not contain a class so we used the train data of the full training dataset and divided it into train and test

Tools

Numpy and Pandas for data manipulation
Scikit-learn for modeling
Matplotlib and Seaborn for plotting