Data Science and You!

or why Wesley Crusher must DIE!

The term ‘Data Science’

- Overused
- Over Marketed
- Snippet from WikiPedia

  “Data science is a buzzword, often used interchangeably with analytics or big data, that is often abused for marketing anything involving data processing…”

Security Data Science...

Paul Braxton, founder of securitydatascience.org:

“Security data science is focused on advancing information security through practical applications of exploratory data analysis, statistics, machine learning and data visualization.”

Crisp and concise… our presentation will cover many of the themes in this definition.

Talk Outline

- Terminology Demystified
- Data Transformation Pipeline
- Use cases:
    - Machine learning on PE Files
    - Exploration of PCAPs

- Wrap Up (GitHub and You!)

Terminology Demystified

- Feature Vector: A ‘list’ of numerical features that represent an object/observation/sample.
- Data Frame: ‘Table’ where each row contains features for each observation. Does NOT have to be all numerical.

Checksum	Debug Size	Compile Date	IAT_RVA	Dorsey’s Mom
0	24	1384956112	5997	Fat
7134913	96	123410236	508732	XL
345979	96	239419032	13470	Blob
21397	28	2346192346	2134	Jabba

Terminology Demystified

- **X** Matrix: Rectangular array of numerical values arranged in rows and columns.
- **y** Vector: A label vector that often contain strings.

Checksum	Debug Size	Compile Date	IAT_RVA	Label
0	24	1384956112	5997	Bad
7134913	96	123410236	508732	Good
345979	96	239419032	13470	Good
21397	28	2346192346	2134	Bad

see: stattrek.com/matrix-algebra/matrix-notation.aspx

Data Transformation Pipeline

- Raw Data
- Data Frame (Pandas)
  - Gisting and Statistics
  - Visualization I
- X Matrix, y Vector (Scikit-Learn)
  - Machine Learning
  - Visualization II

Machine learning on PE Files

IPython notebook!

Exploration of PCAPs

IPython notebook!

Wrap Up

- Software:
  http://clicksecurity.github.io/data_hacking
- People:
  - Mike Sconzo (sconzo@clicksecurity.com)
  - Brian Wylie (bwylie@clicksecurity.com)
- Support:
  Click Security! Woot woot! http://clicksecurity.com