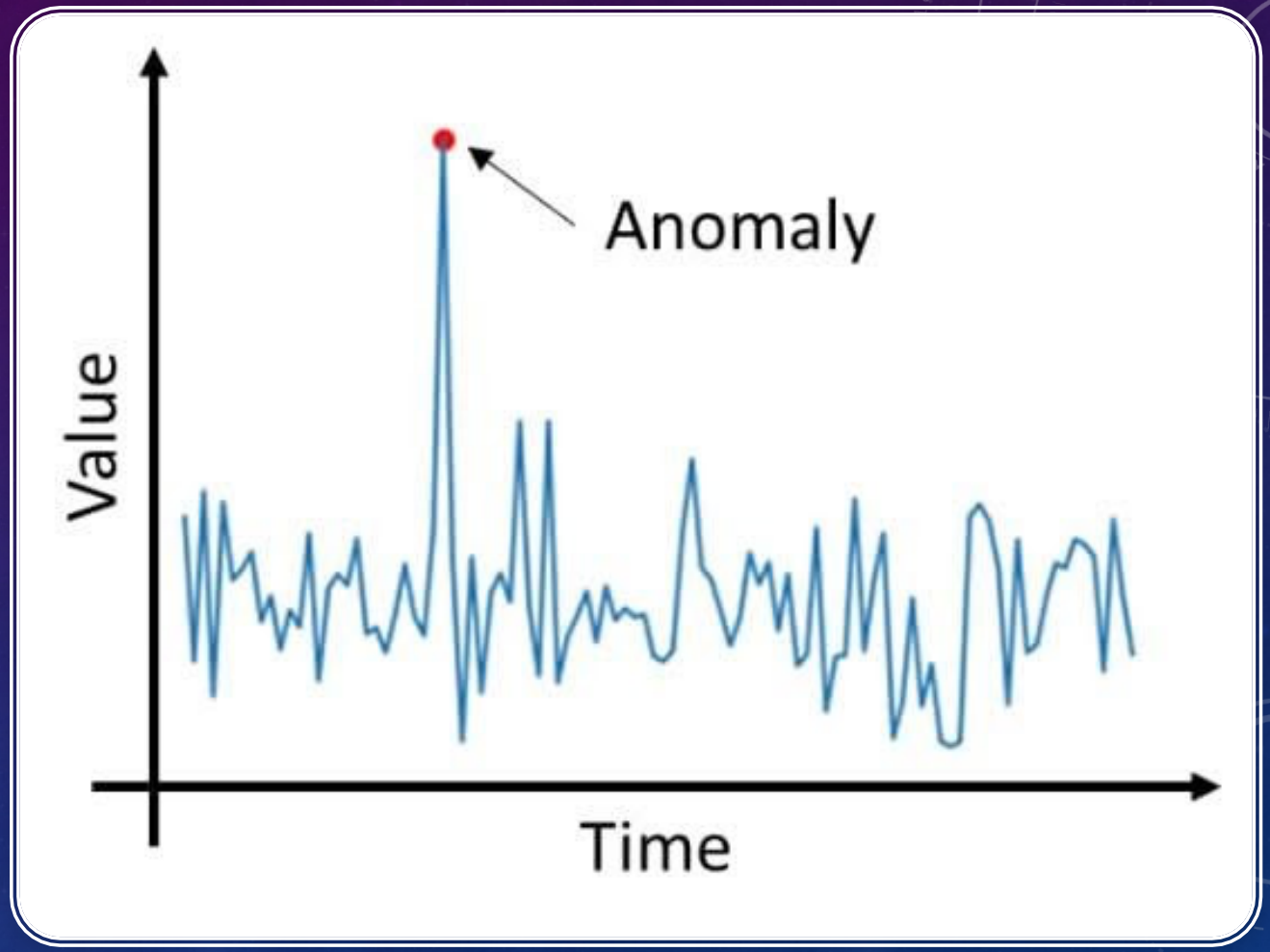
ANOMALY DETECTION

BY MARCOS DOMINGUEZ

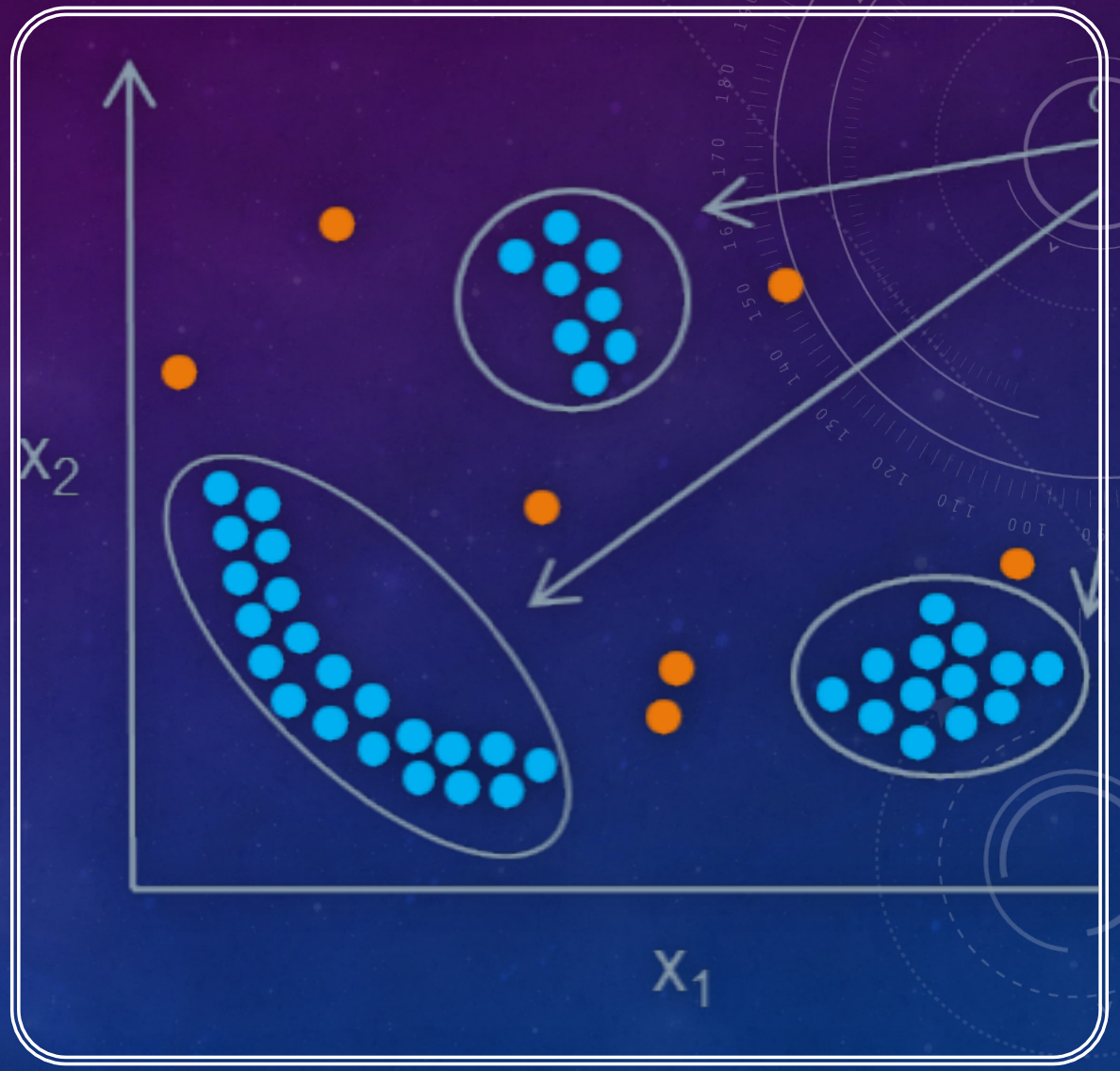
WHAT IS IT?

- Detecting unexpected items
- Unsupervised Learning
- 2 basic assumptions:
  - Anomalies ONLY occur rarely
  - Their features differ significantly from normal data



HOW IS IT USED?

- Detecting performance outliers
- Fraud detection
- Extreme price changes (stock trading)
- Detecting a cyber breach



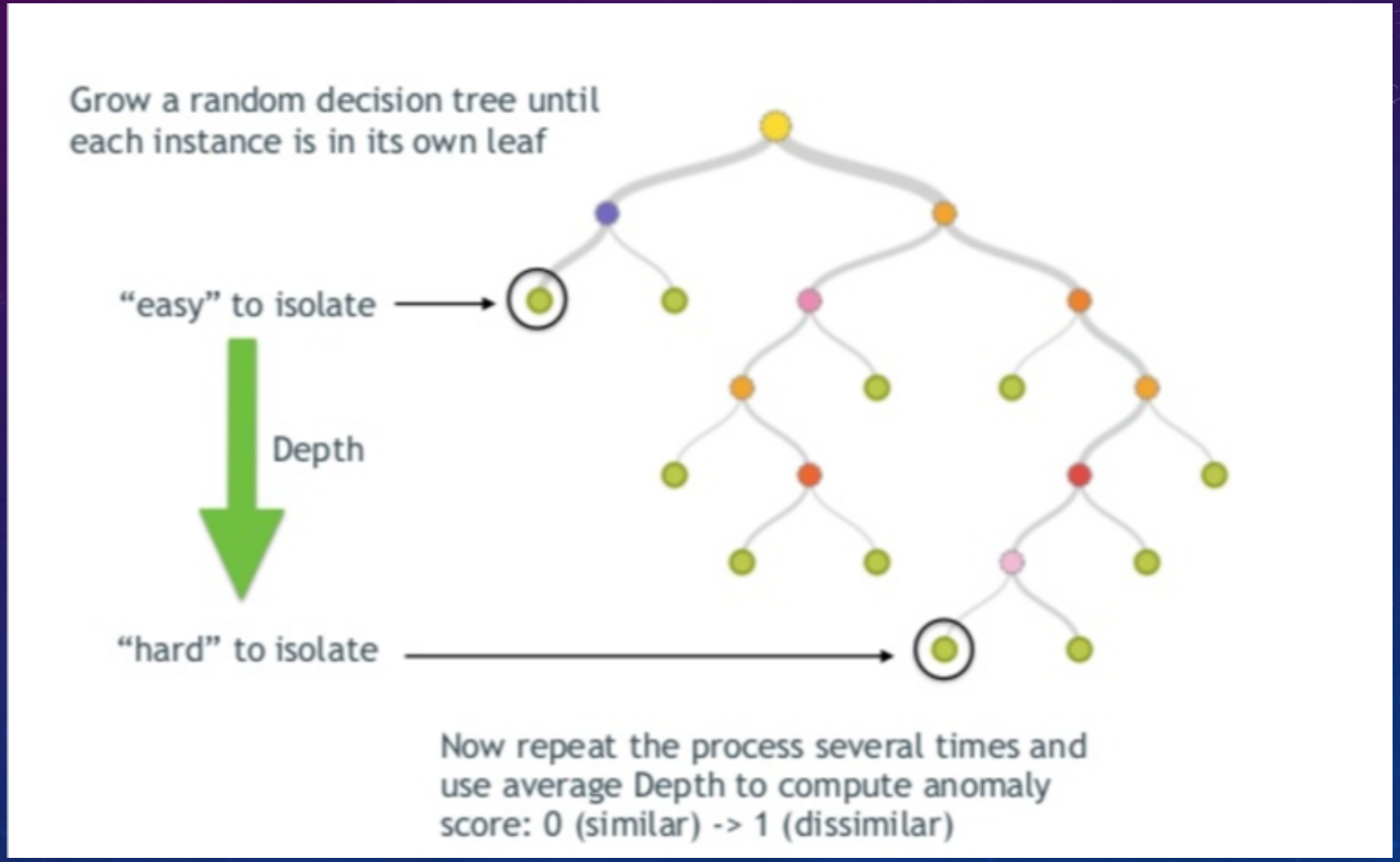
TYPES OF ALGORITHMS

- Univariate
  - Not very informative
- Multivariate
  - More informative
- Algorithms:
  - Cluster-based Local Outlier Factor (CBLOF)
  - Isolation Forest
  - K - Nearest Neighbors (KNN)

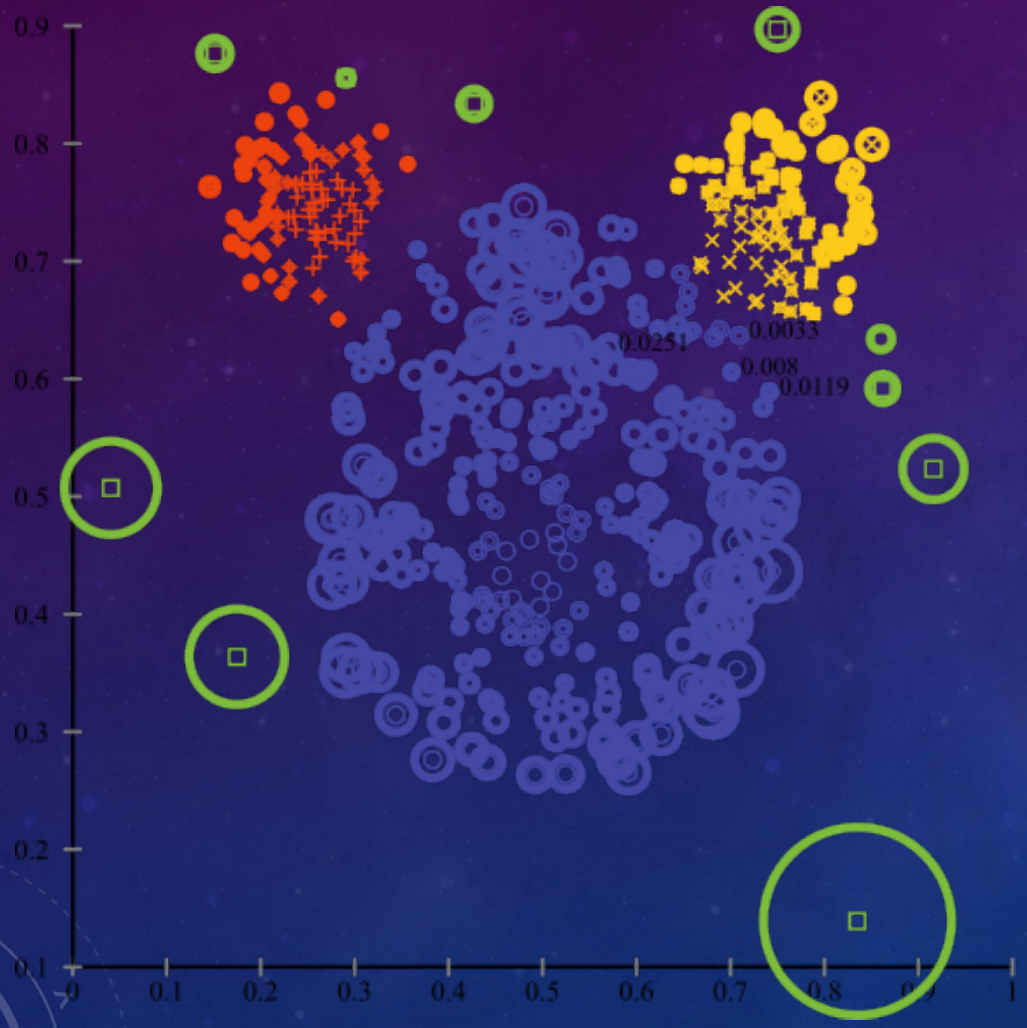
HOW DOES IT WORK?

- Entirely based on statistical outliers
- Assigning an anomaly score
  - Compare to threshold

ISOLATION FOREST



CLUSTER-BASED OUTLIER FACTOR



- Typically uses K-means clustering
- Calc anomaly score based on distance from clusters

ACKNOWLEDGEMENTS

- Susan Li, Towards Data Science
- Lakshay Arora, AnalyticsVidhya
- Surbhi S, KeyDifferences
- Tina Kavacova, Medium article
- Stack Exchange
- Open-source:
  - Scikit-learn
  - Seaborn, Matplotlib
  - Python, PyOD
- Full code and presentation here