An Introduction to Machine Learning

With Python

Terry McCann | @SQLShark

Microsoft Partner
Microsoft
Gold Data Analytics
Gold Data Platform
Gold Cloud Platform

MVP Microsoft® Most Valuable Professional

adatis

adatis

*We **enable** our **clients** make **sense** of their **data***

DATA ARCHITECTURE

MANAGED SERVICES

DATA STRATEGY

DATA SCIENCE

DATA MANAGEMENT

DATA ANALYTICS

Questions?

Shout them out!!

Where to find the ***slides*** and ***demos***

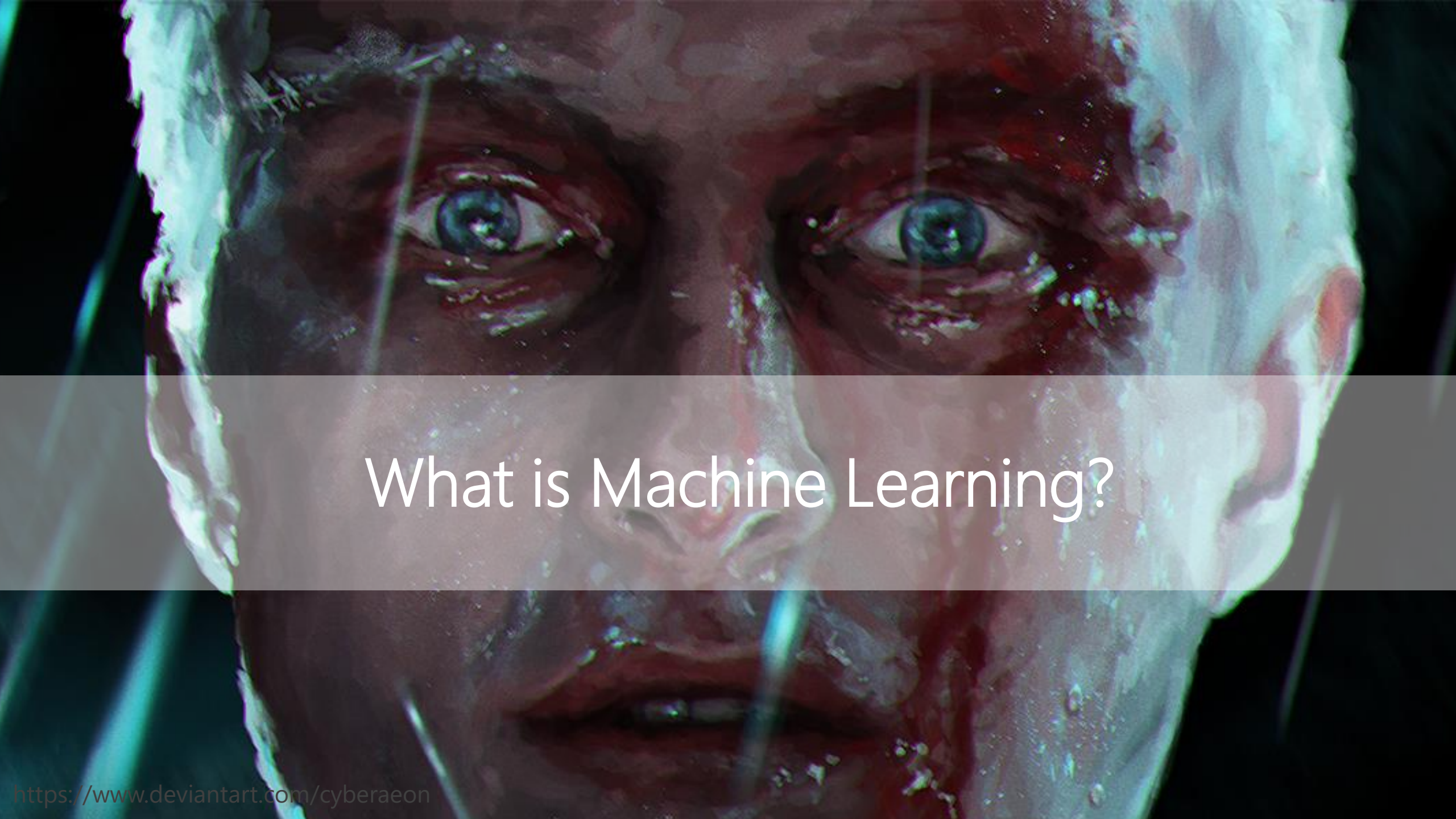
https://github.com/SQLShark

bit.ly/2uHQS74

In this session we will look to understand what is ***Machine Learning***?

We will explore ***data science***, the process, the use cases and how to get started.

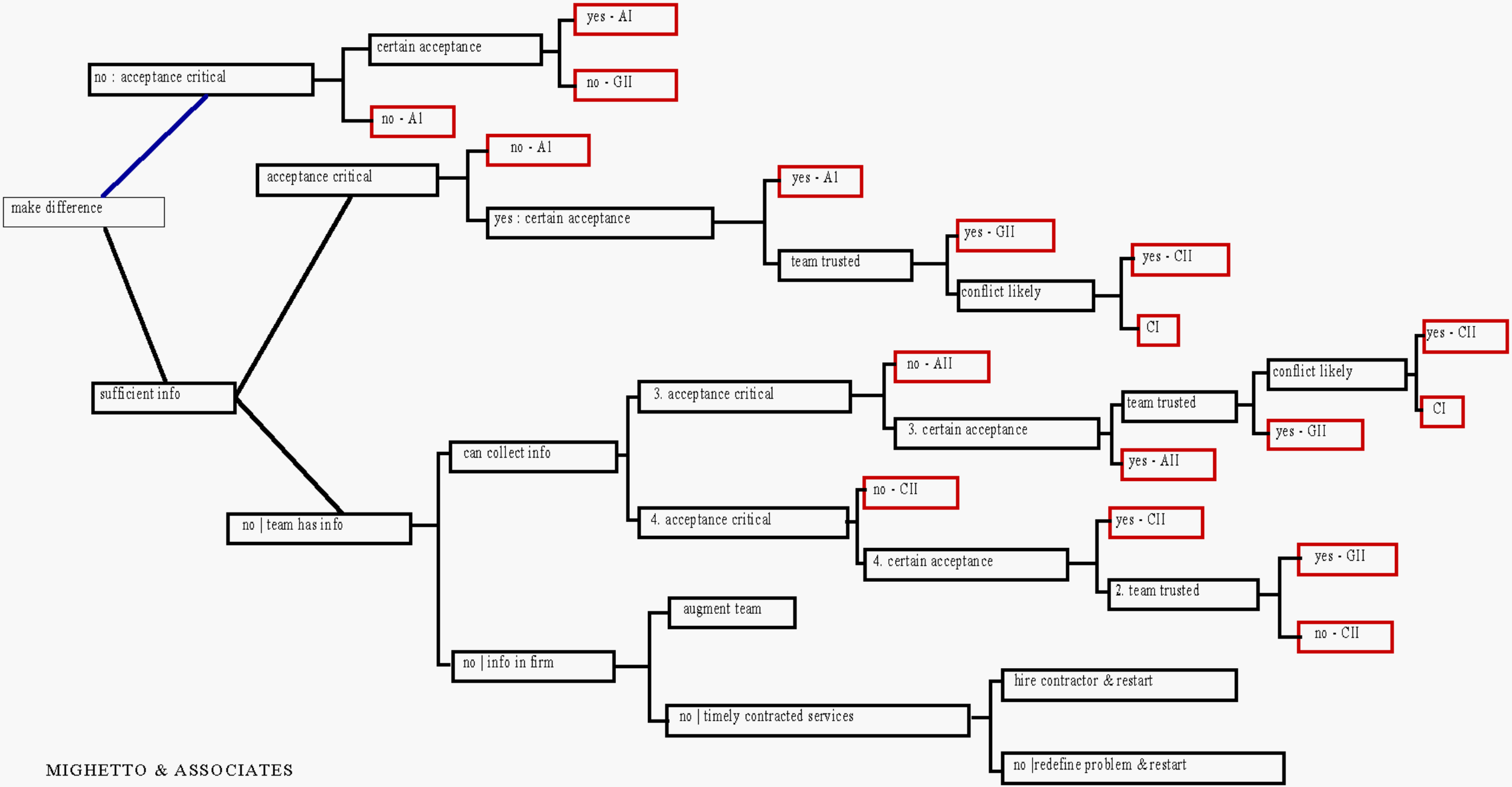
We will understand why ***Python*** is a great Language, and how to advance your skills in machine learning



What is Machine Learning?

https://www.deviantart.com/cyberaeon

Expert systems



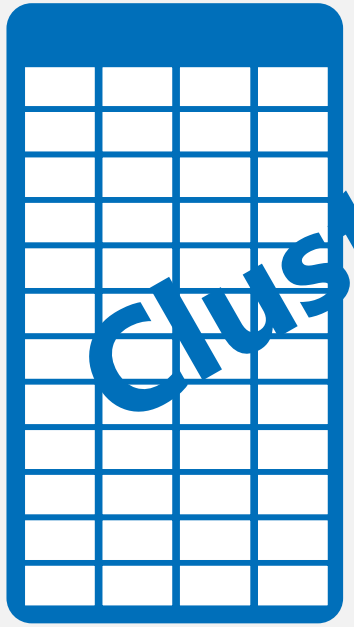
MIGHETTO & ASSOCIATES

A machine is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**

Its performance at tasks in **T**, as measured by **P**, improves with experience **E**.

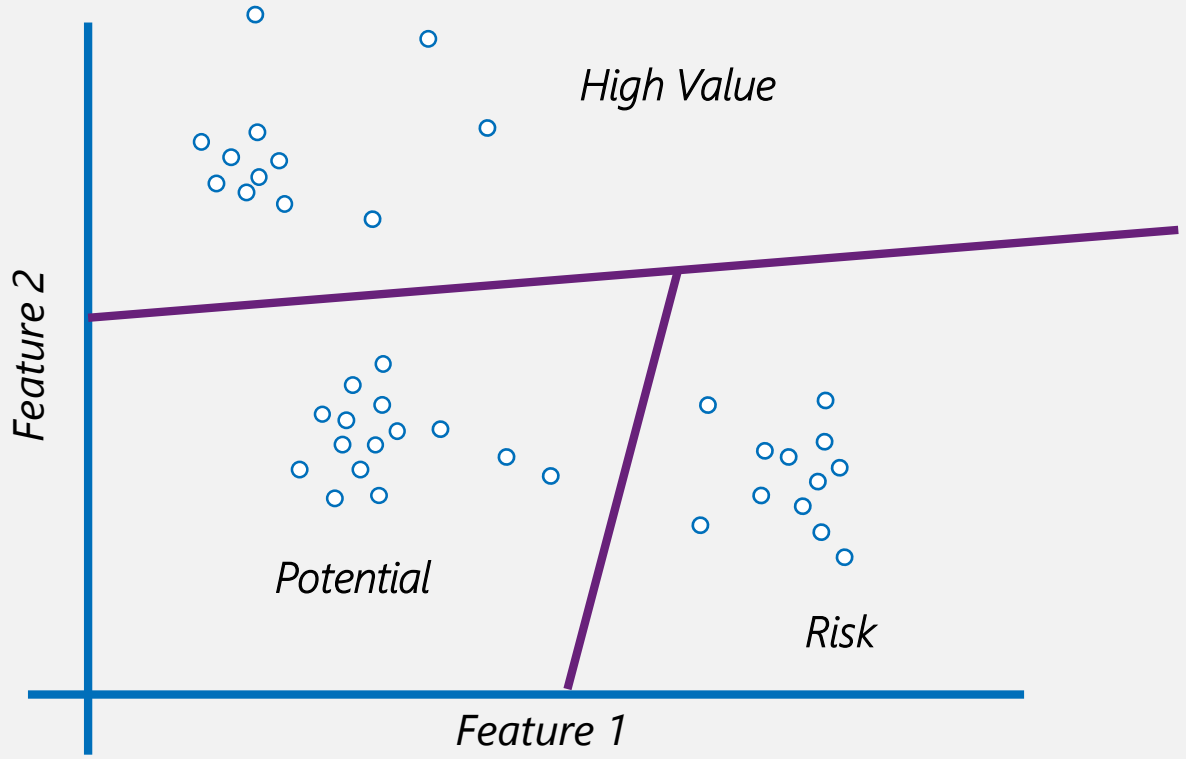
https://www.quora.com/What-is-the-best-definition-for-machine-learning

data

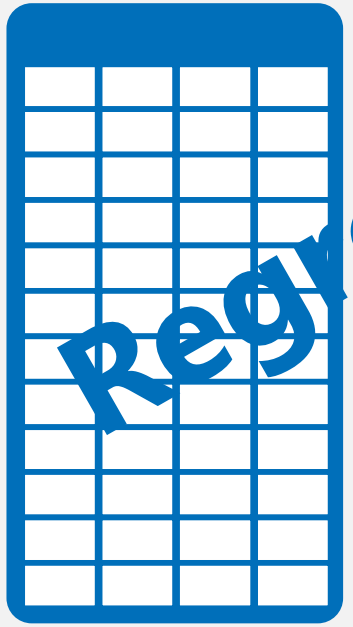


clustering

"Hey, I am trying to better **understand** my customers. I think there might be a **pattern** in their **behaviour**, but I am struggling to see it"

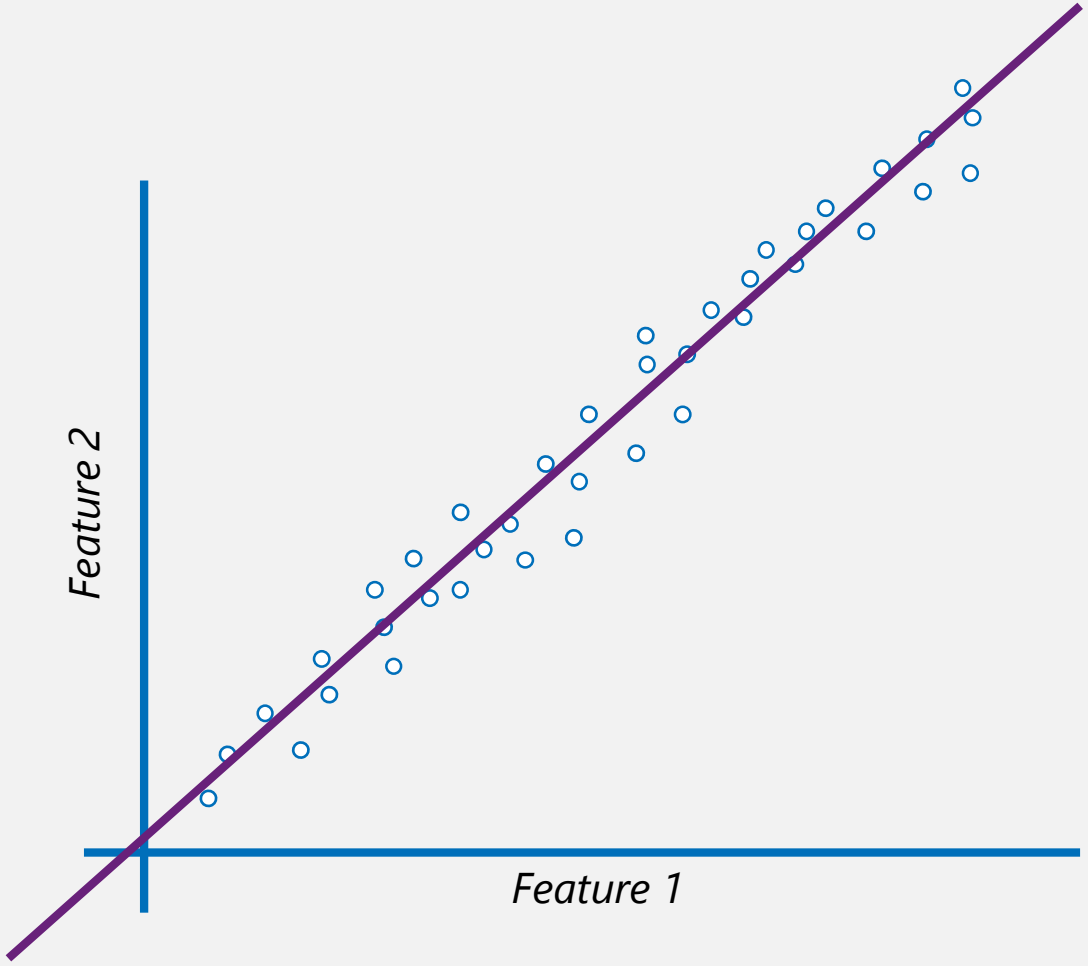


data

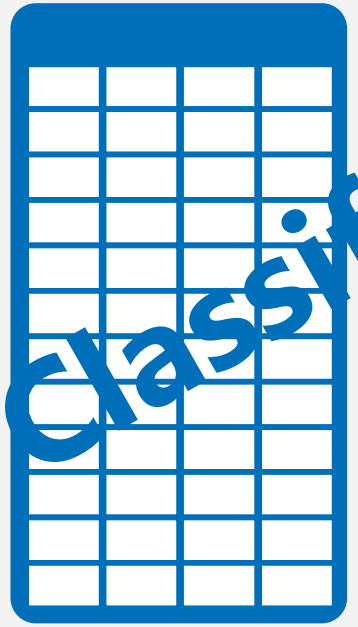


Regression

"Hey, I am trying to **predict** how much someone will **spend** next time they come to my store"

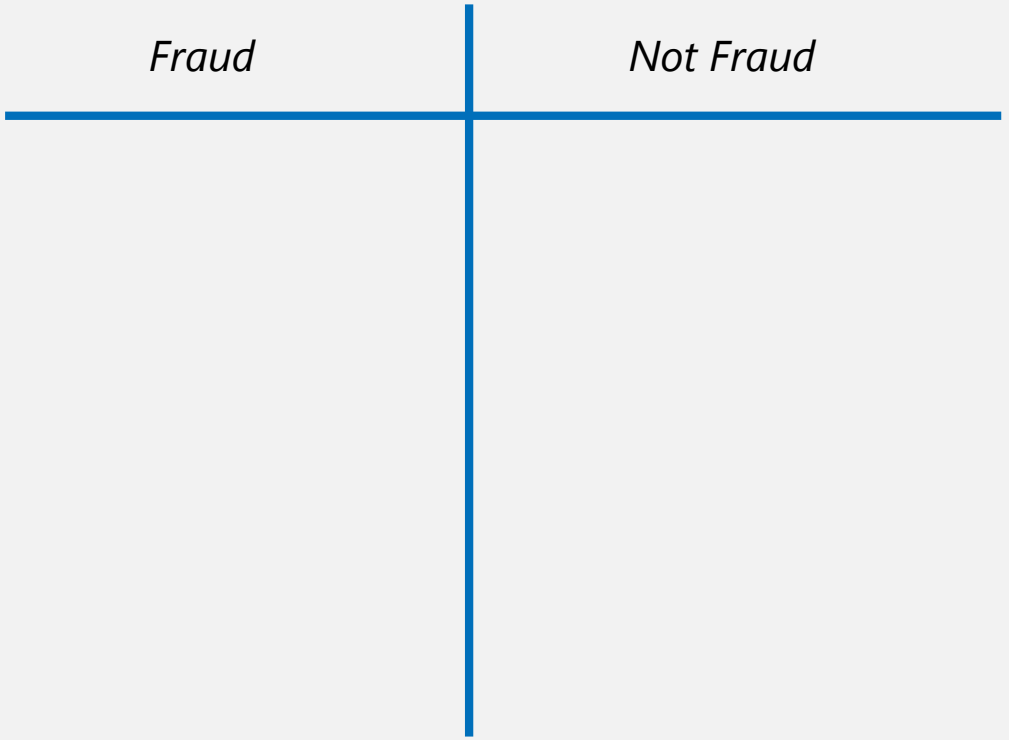


data

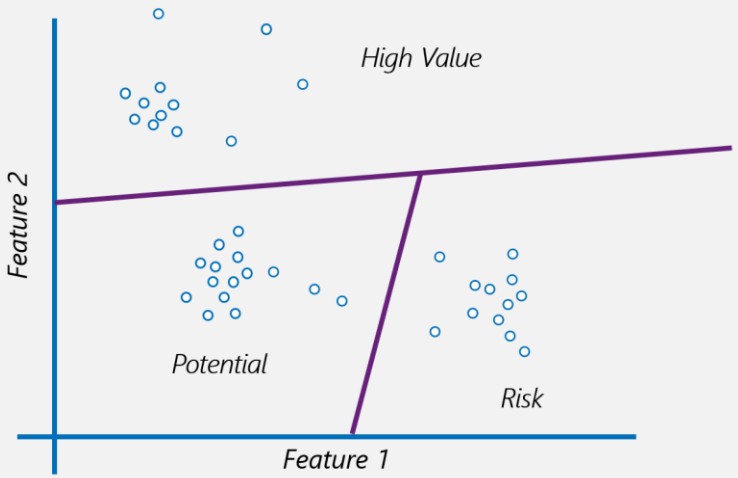


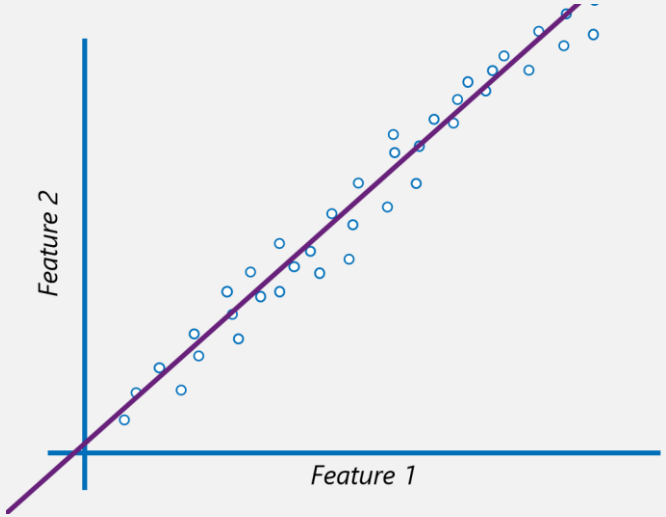
classification

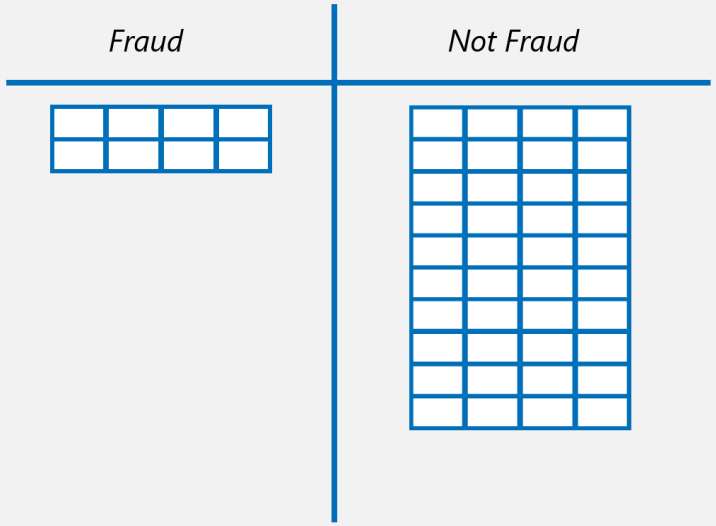
"Hey, I am trying to **predict** if an insurance claim is **fraudulent** or not"



Types of machine learning







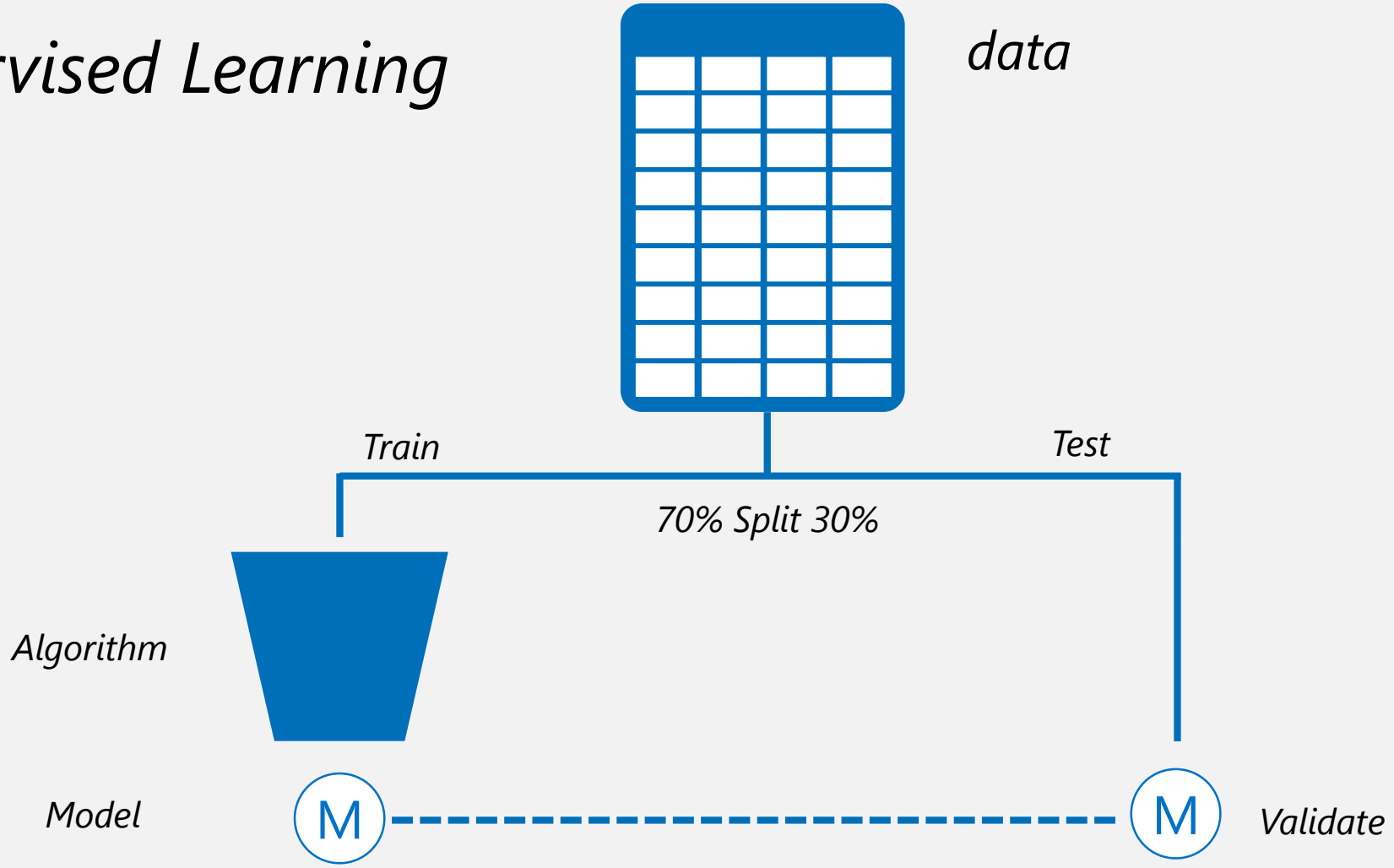
Clustering **Regression** **Classification**

Unsupervised

Supervised

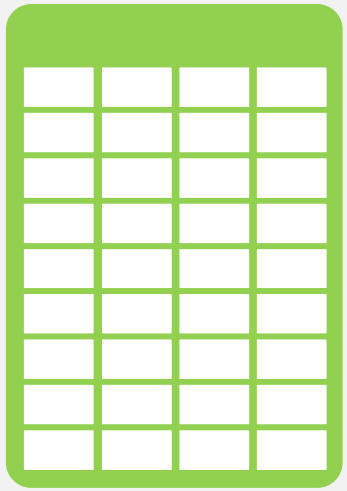
adatis

Supervised Learning



adatis

Supervised Learning



New data

Model

Prediction

adatis

R2
D3

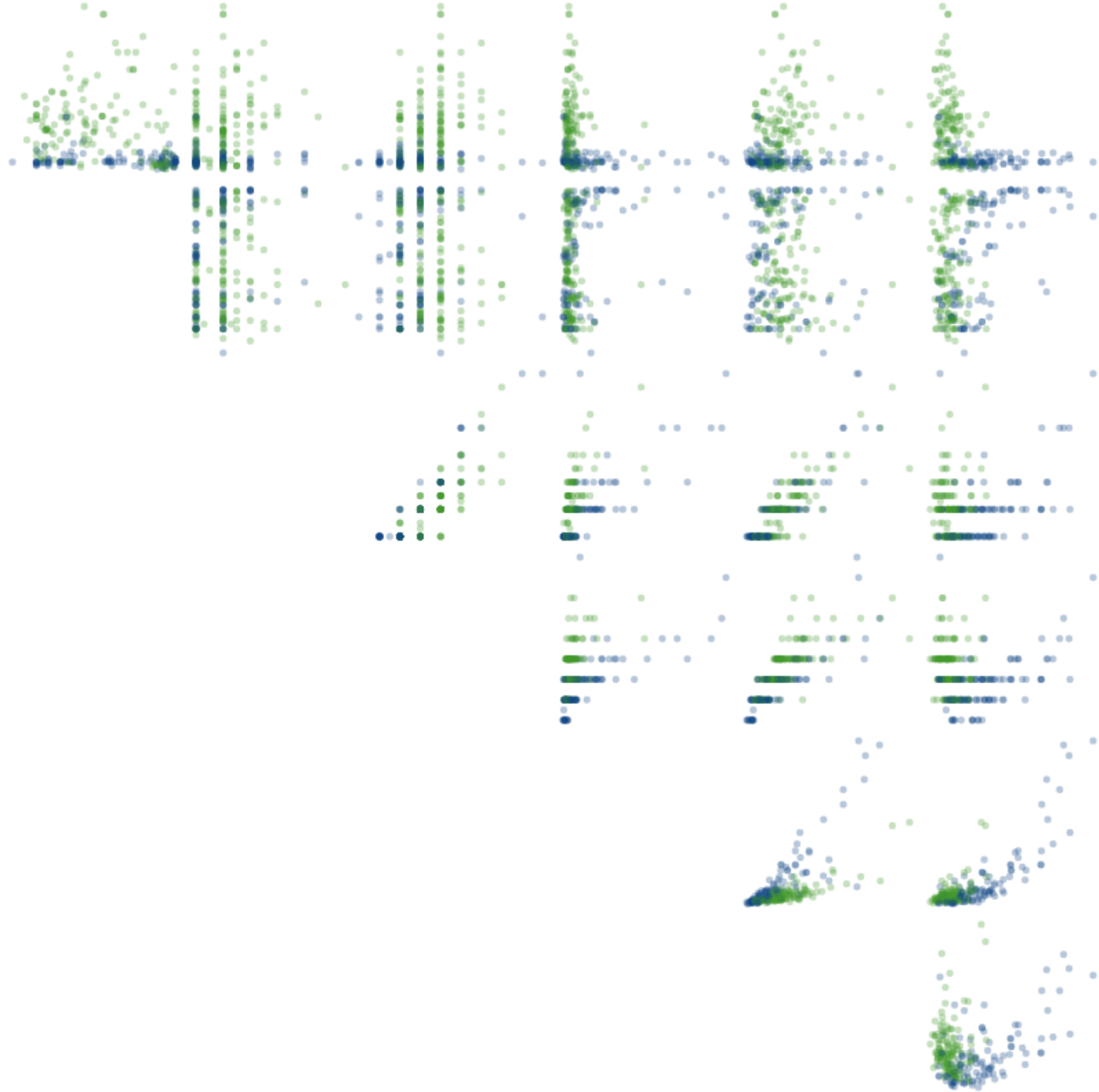
A visual introduction to machine learning

English

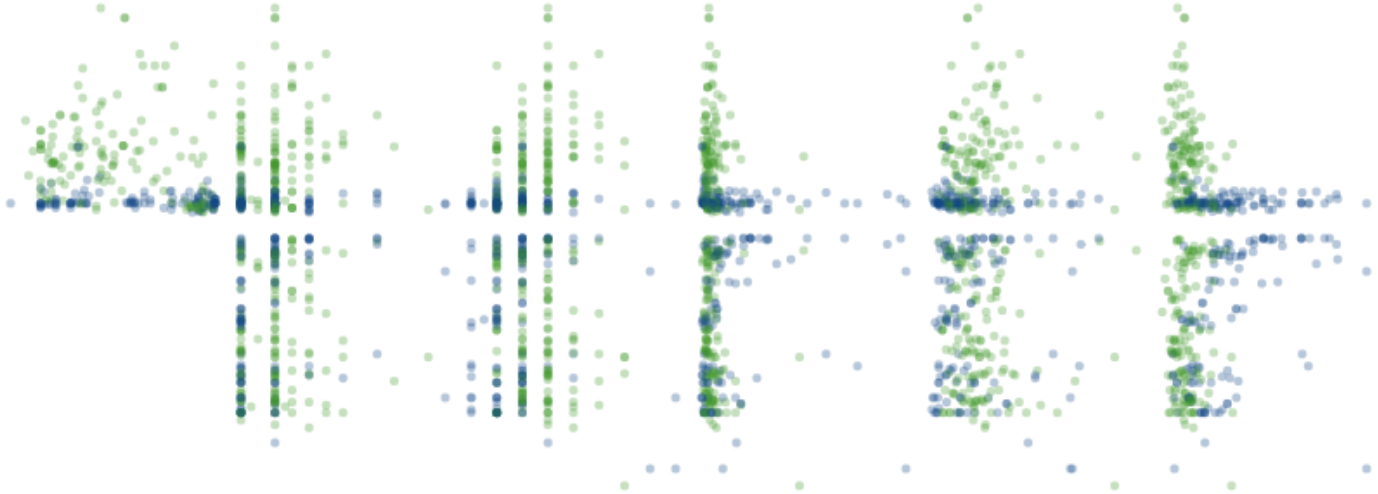
In machine learning, computers apply **statistical learning** techniques to automatically identify patterns in data. These techniques can be used to make highly accurate predictions.

Keep scrolling. Using a data set about homes, we will create a machine learning model to distinguish homes in New York from homes in San Francisco.

SCROLL



R2
D3

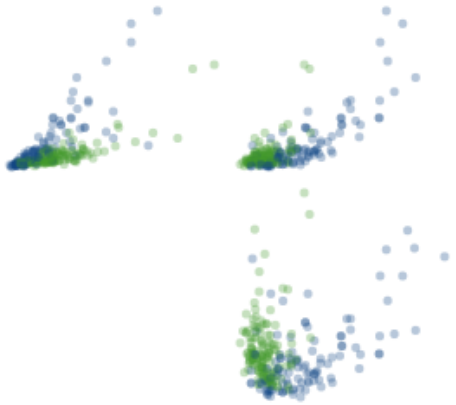


A visual introduction to

http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Keep scrolling. Using a data set about homes, we will create a machine learning model to distinguish homes in New York from homes in San Francisco.

SCROLL



Python: Fundamentals

```
print("Hello, Manchester!")
```

adatis

Beautiful is better than ugly
Explicit is better than implicit
Simple is better than complex
Complex is better than complicated
Readability counts

The Zen of Python (PEP 20)

adatis

Python is...

An interpreted language

It has a base language which is extended with modules

Indents are important! Indents over curly brackets

Supports: variables, lists, dictionaries and tuples

Duck typed

Awesome!!

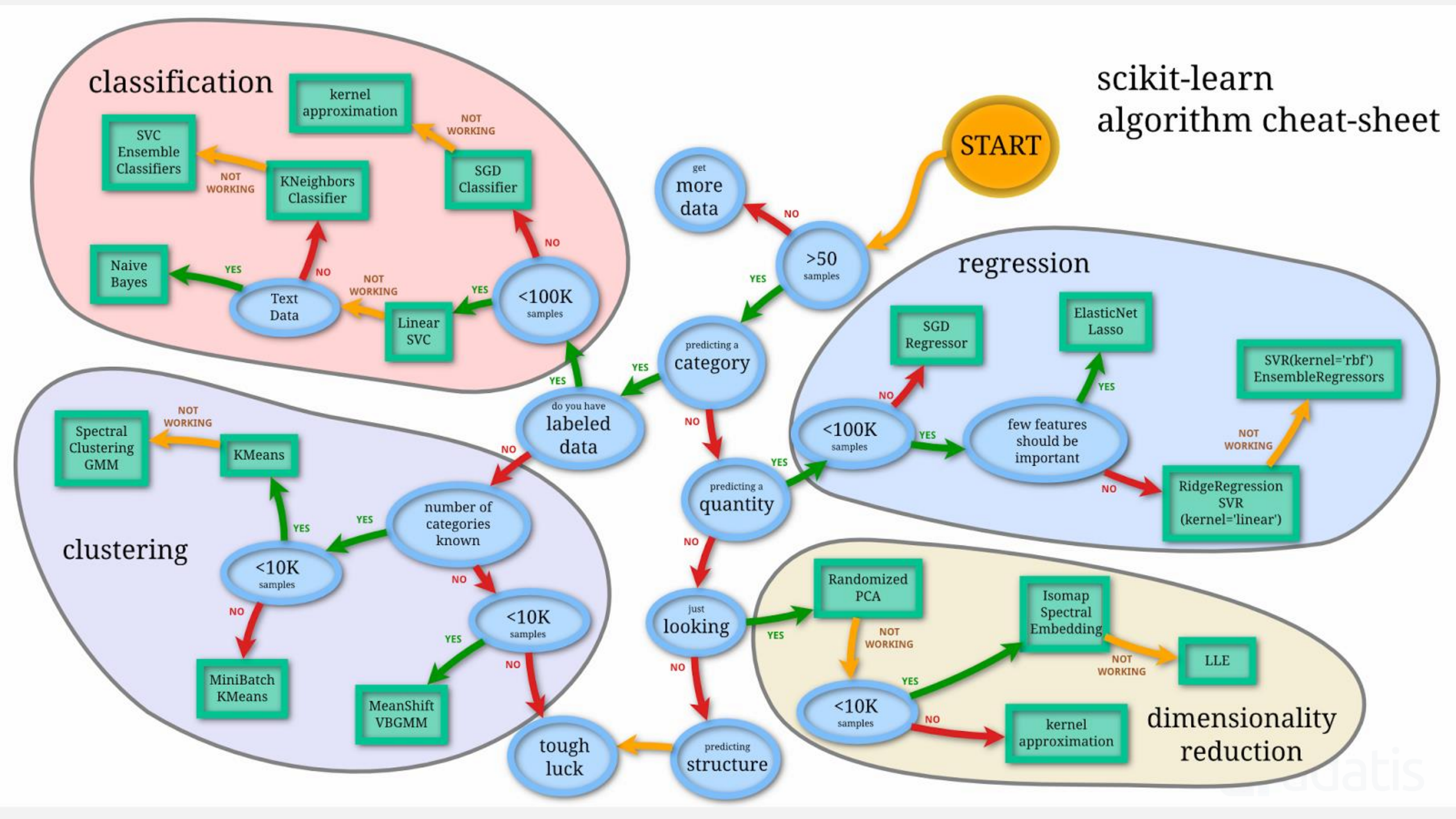
adatis

DS/ML Modules

Numpy	– Numerical Python
SciPy	– Scientific Python
Pandas	- Data wrangling
MatPlotLib	- Data Visualisation
Scikit-learn	- Machine Learning (shallow)
….	
TensorFlow	- Machine Learning (deep)
Keras	- High level DL
NLTK	- Natural Language processing

How to add modules? **PIP**

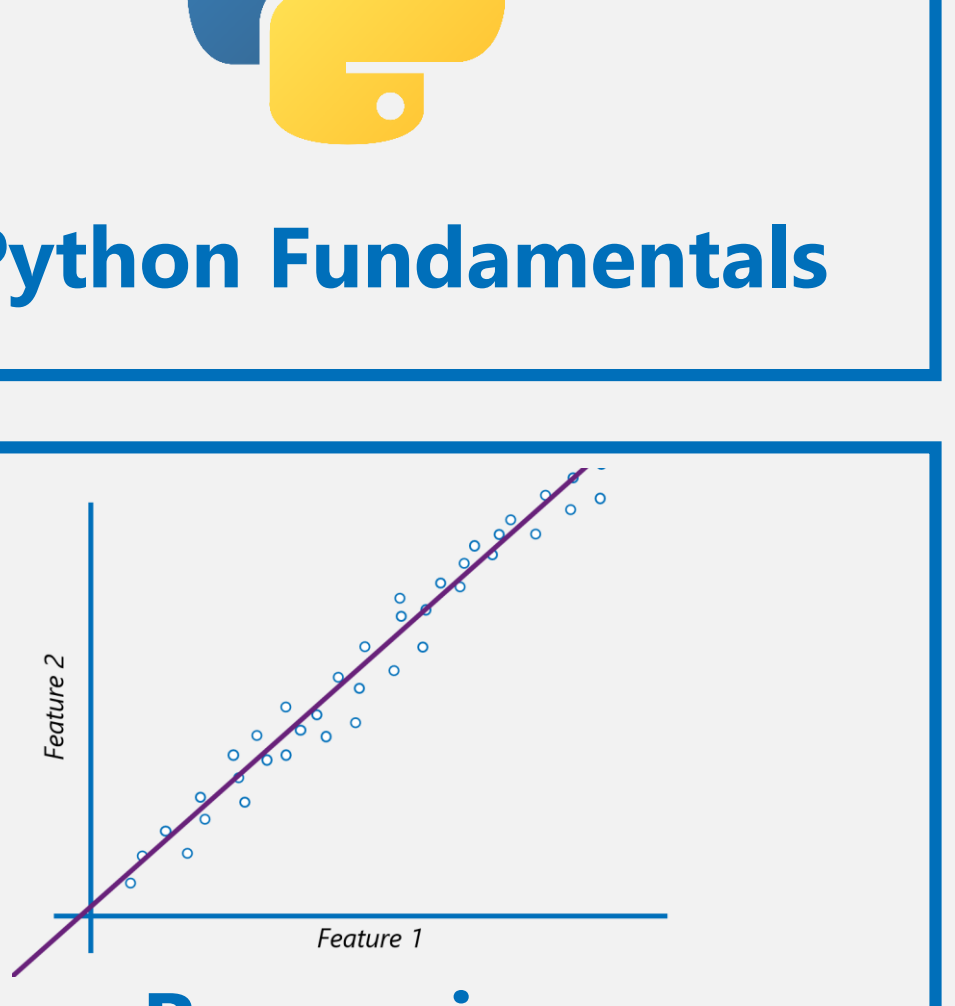
adatis


scikit-learn
algorithm cheat-sheet
START
get more data
NO
>50 samples
YES
predicting a category
YES
do you have labeled data
YES
<100K samples
NO
Naive Bayes
YES
Text Data
NO
KNeighbors Classifier
NO
SVC Ensemble Classifiers
NOT WORKING
kernel approximation
NOT WORKING
SGD Classifier
NOT WORKING
Linear SVC
NO
predicting a quantity
YES
<100K samples
NO
SGD Regressor
YES
few features should be important
YES
ElasticNet Lasso
NO
RidgeRegression SVR (kernel='linear')
NOT WORKING
SVR(kernel='rbf') EnsembleRegressors
NO
just looking
YES
Randomized PCA
NOT WORKING
Isomap Spectral Embedding
NOT WORKING
LLE
YES
<10K samples
NO
kernel approximation
NO
predicting structure
tough luck
NO
number of categories known
YES
<10K samples
NO
KMeans
NOT WORKING
Spectral Clustering GMM
YES
<10K samples
NO
MiniBatch KMeans
NO
<10K samples
YES
MeanShift VBGMM
NO
classification
clustering
regression
dimensionality reduction

Python: Demos

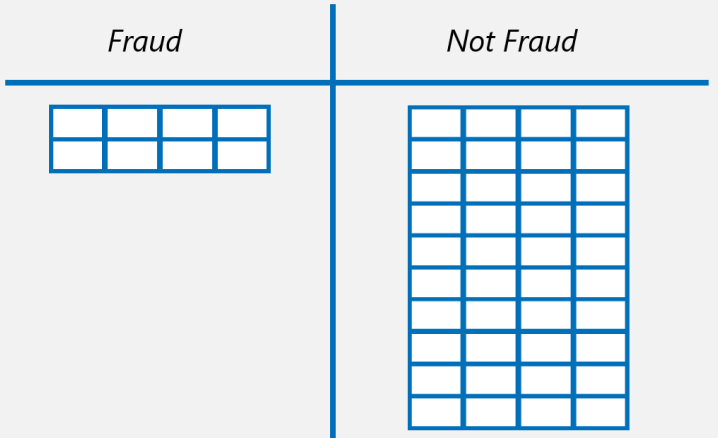
adatis

Python Fundamentals

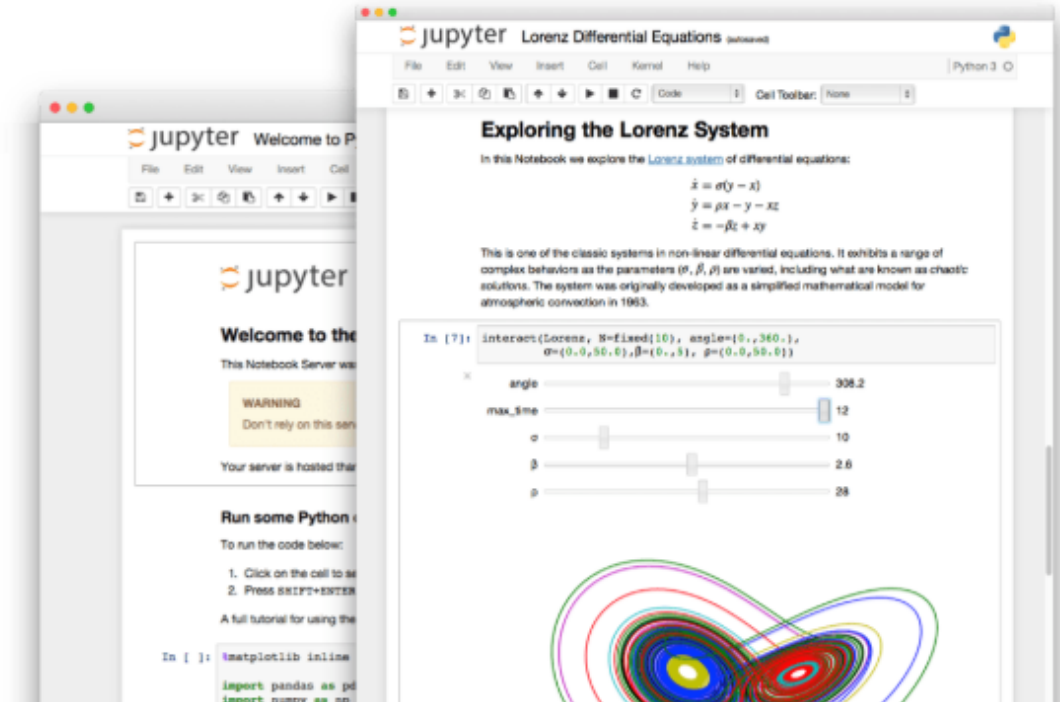


Clustering

Regression

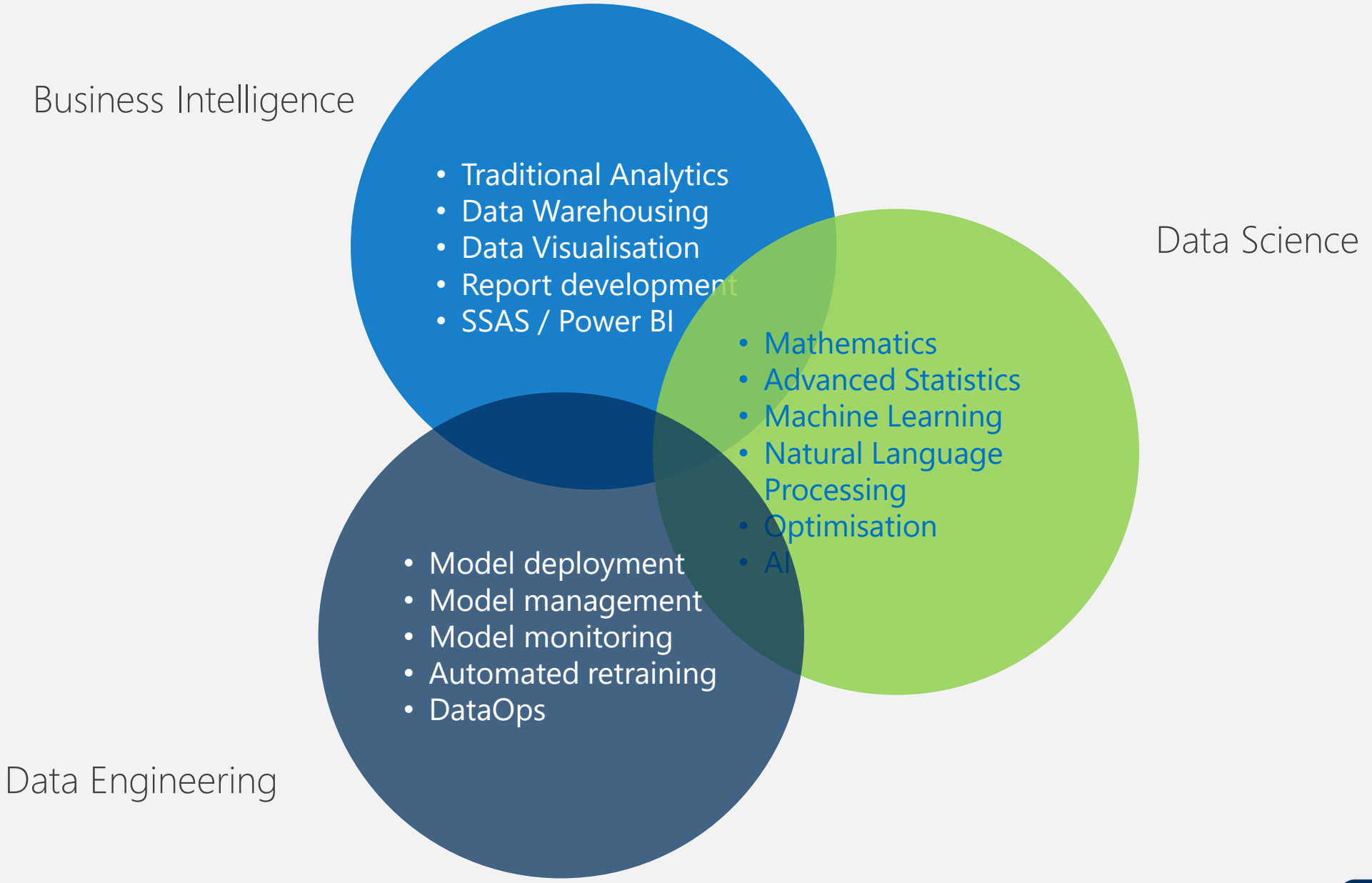


Classification



The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.


Business Intelligence
• Traditional Analytics
• Data Warehousing
• Data Visualisation
• Report development
• SSAS / Power BI
Data Science
• Mathematics
• Advanced Statistics
• Machine Learning
• Natural Language Processing
• Optimisation
• AI
• Model deployment
• Model management
• Model monitoring
• Automated retraining
• DataOps
Data Engineering

adatis

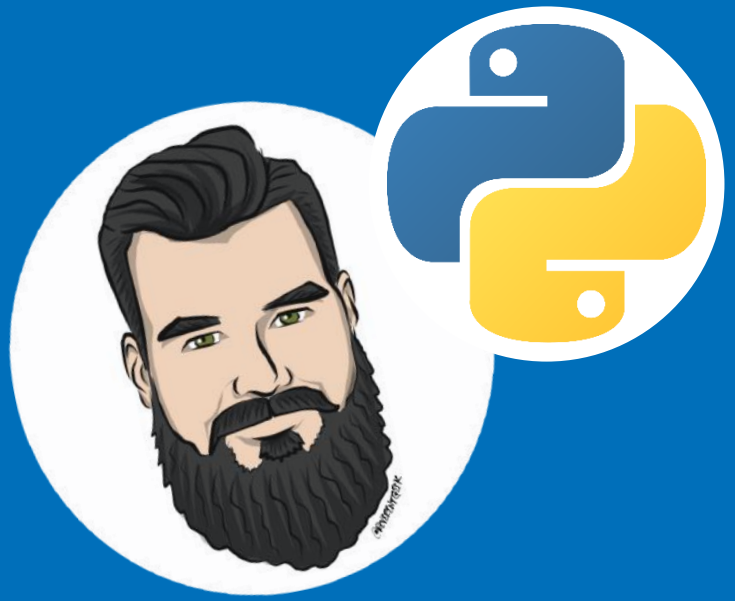
Python!

Python is amazing, so versatile.

If you are serious about Machine Learning then learn Python. Learn R too, but Python for production.

Start basic and build from there.

Thank you



@SQLShark
tpm@adatis.co.uk
www.adatis.co.uk
@AdatisBI