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

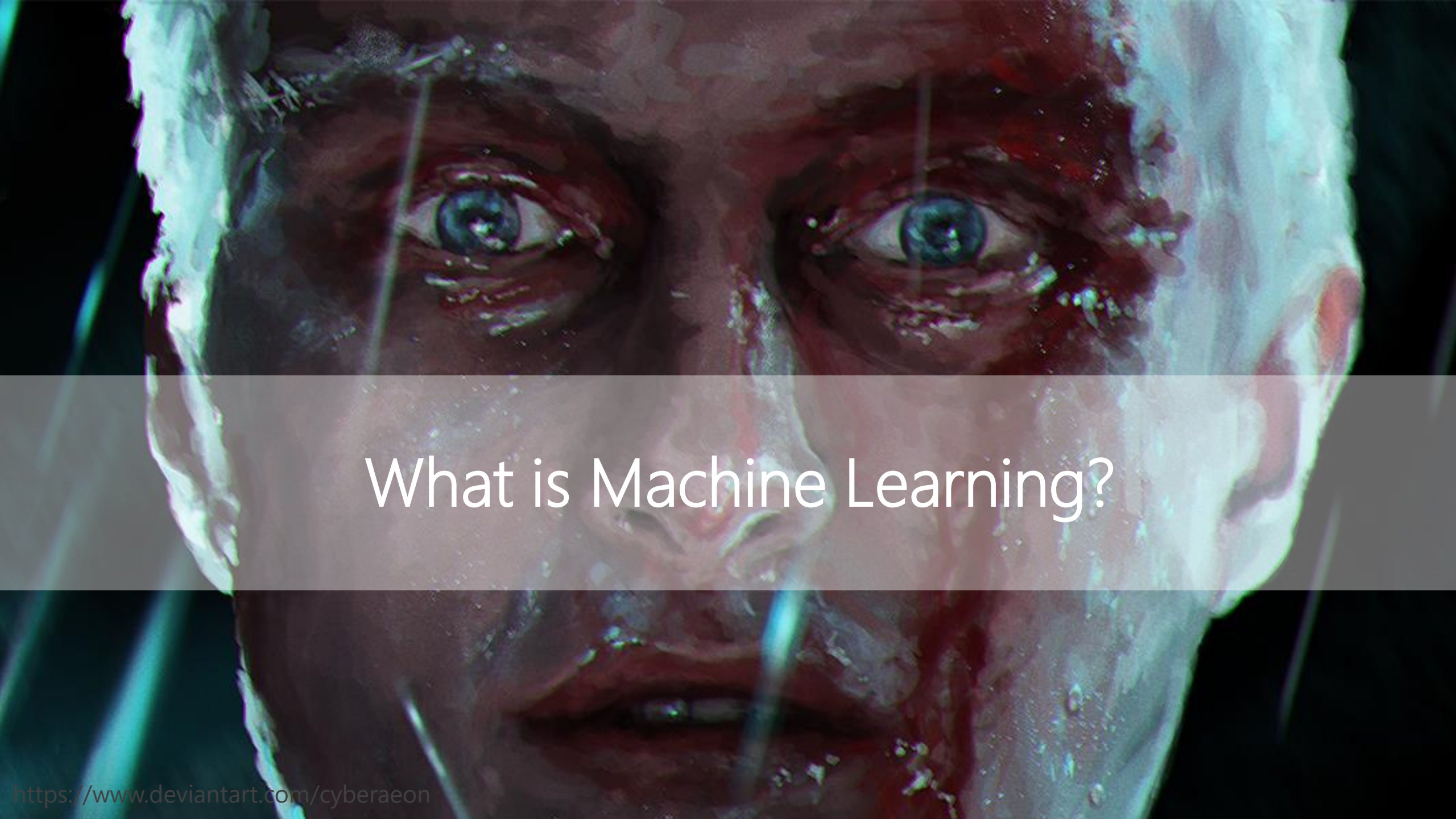
DATA MANAGEMENT

DATA ANALYTICS

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

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

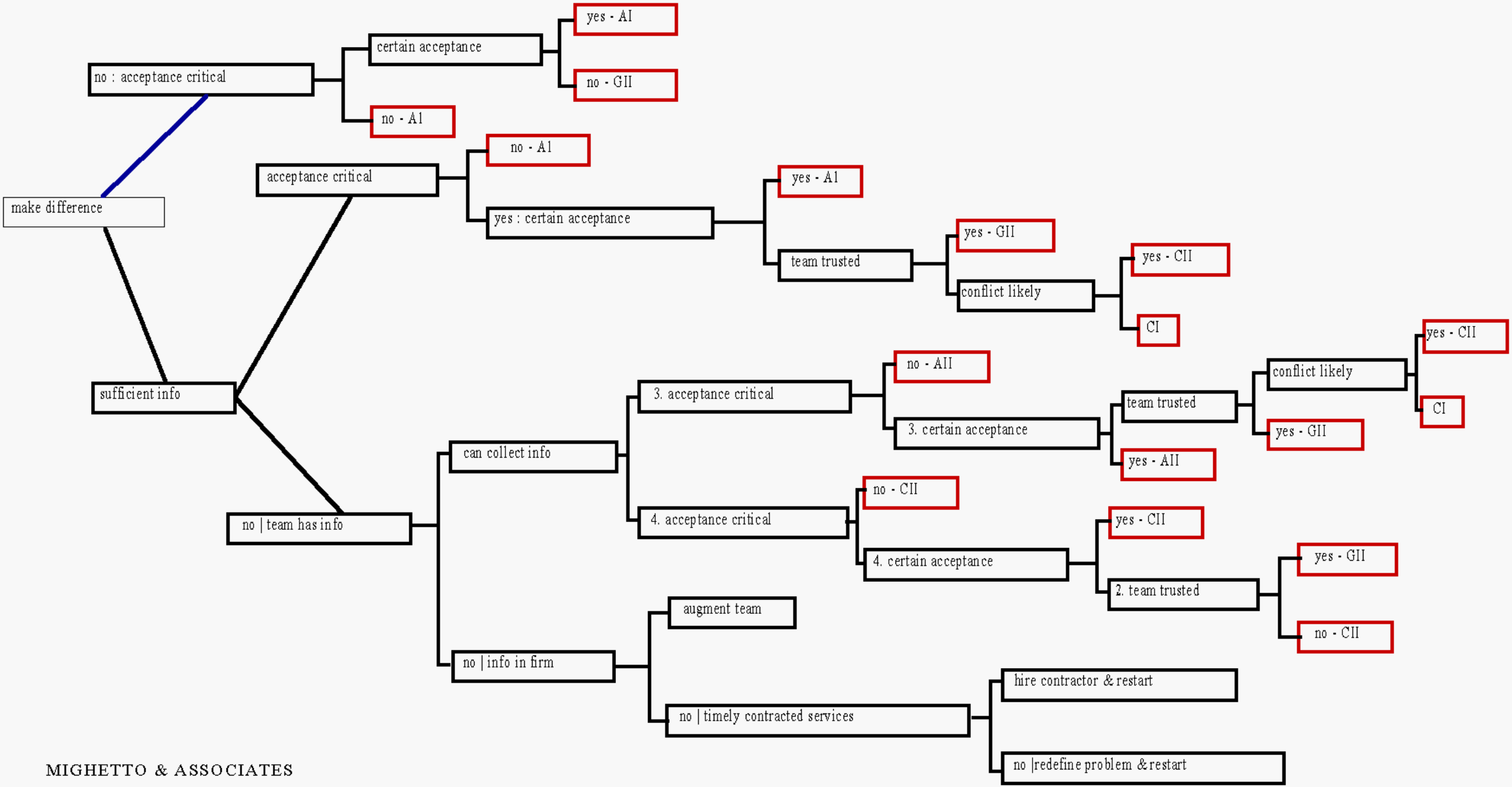
We will understand why ***Python*** is a great Language, and how to advance your skills in machine learning



What is Machine Learning?

https://www.deviantart.com/cyberaeon

Expert systems



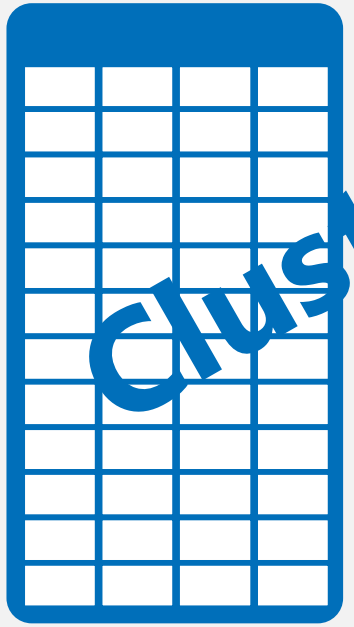
MIGHETTO & ASSOCIATES

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**.

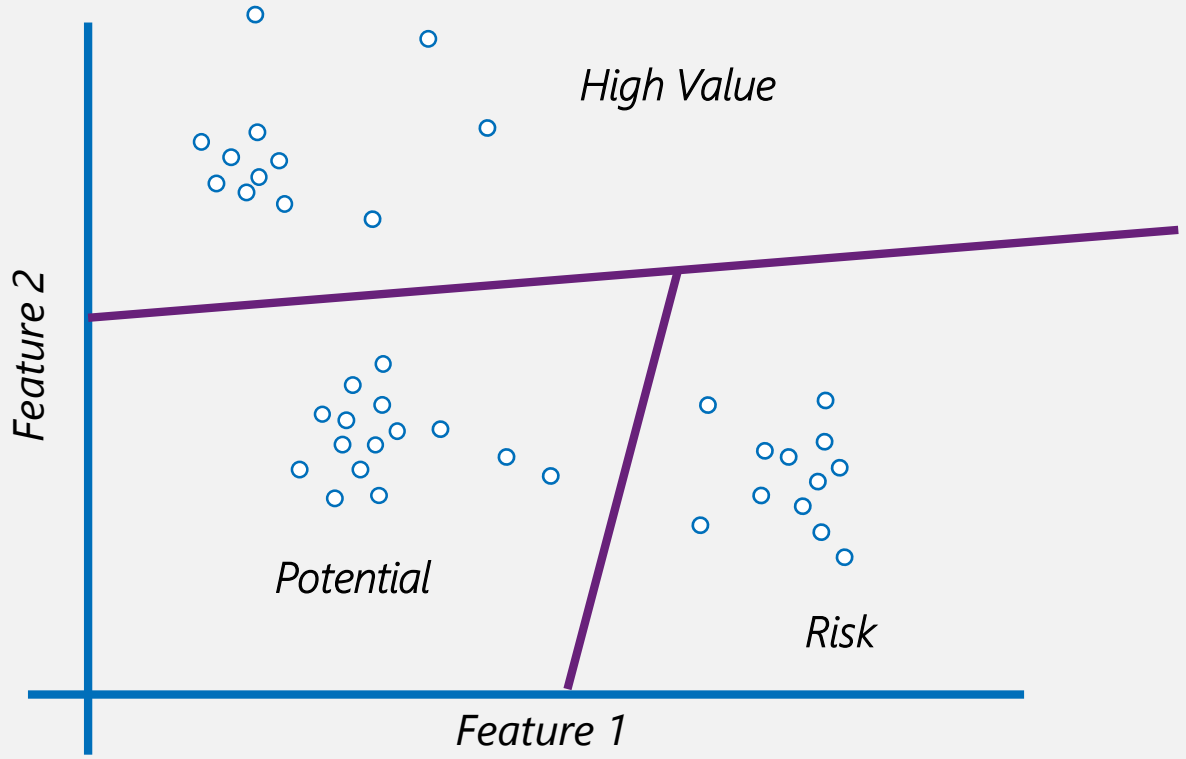
https://www.quora.com/What-is-the-best-definition-for-machine-learning

data

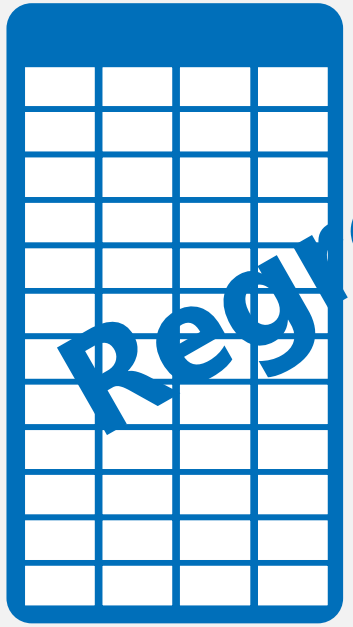


clustering

"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"

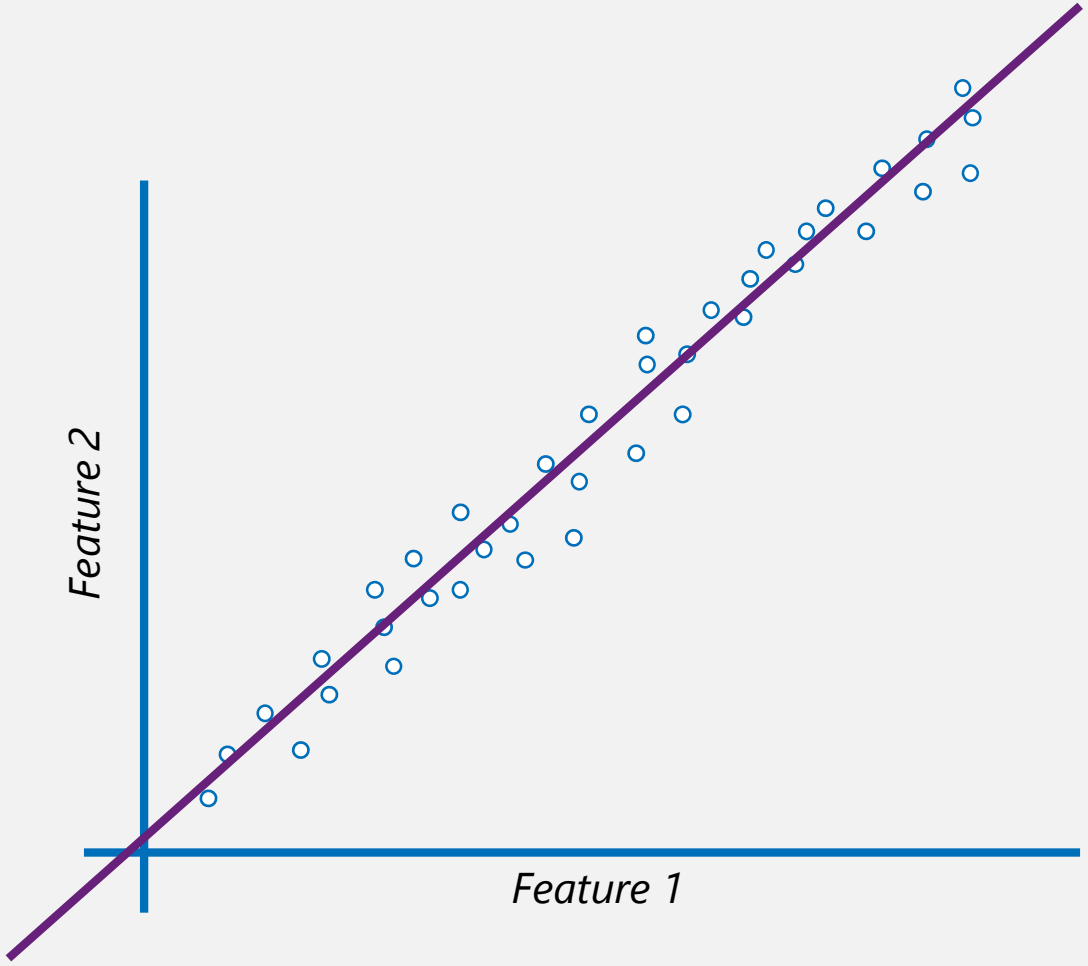


data

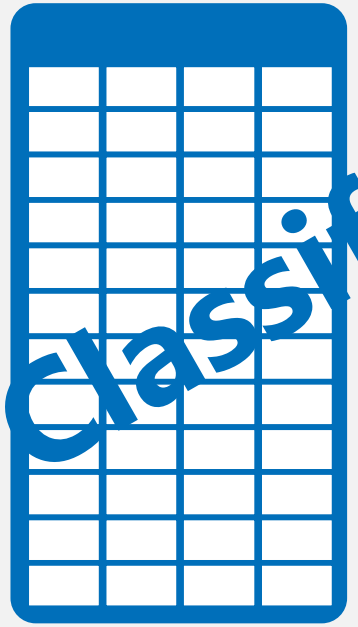


Regression

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

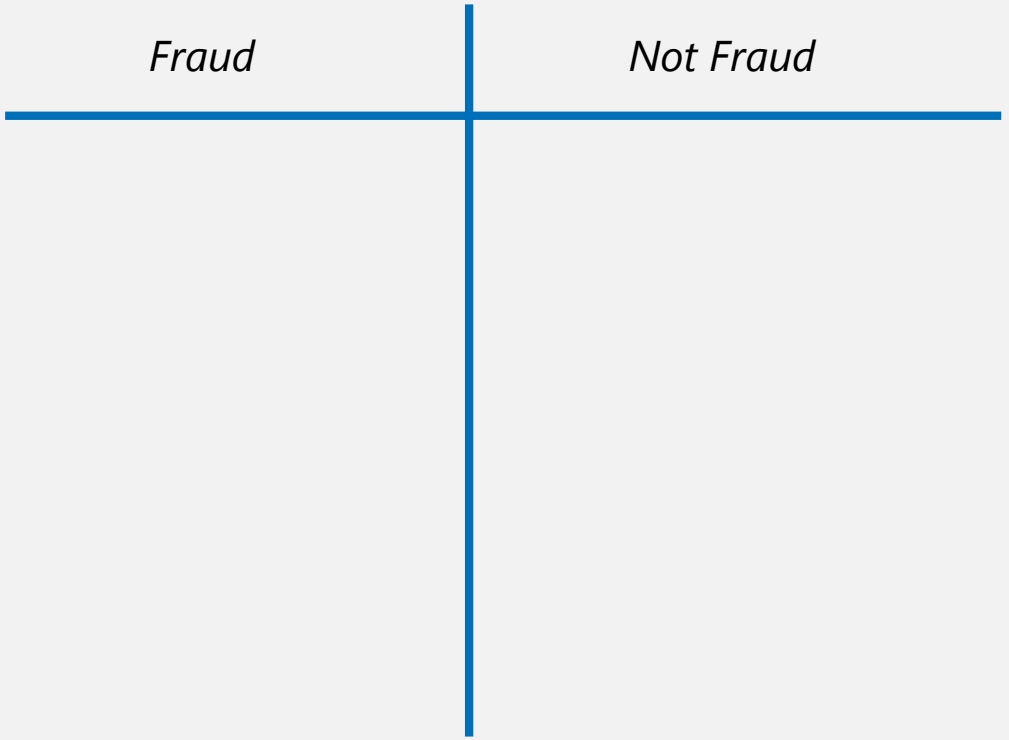


data

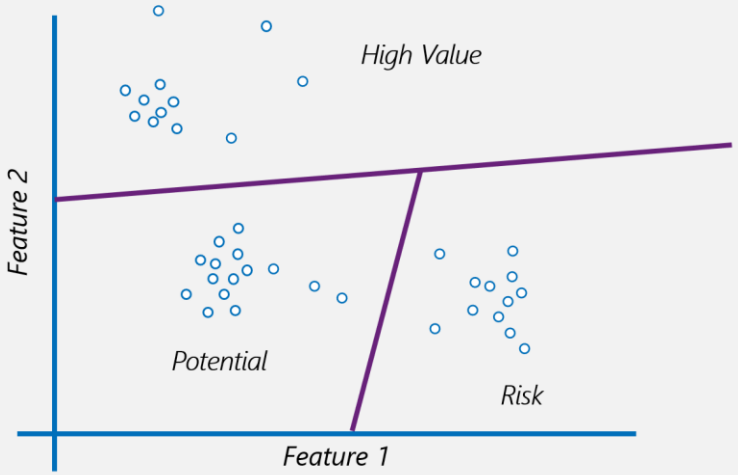


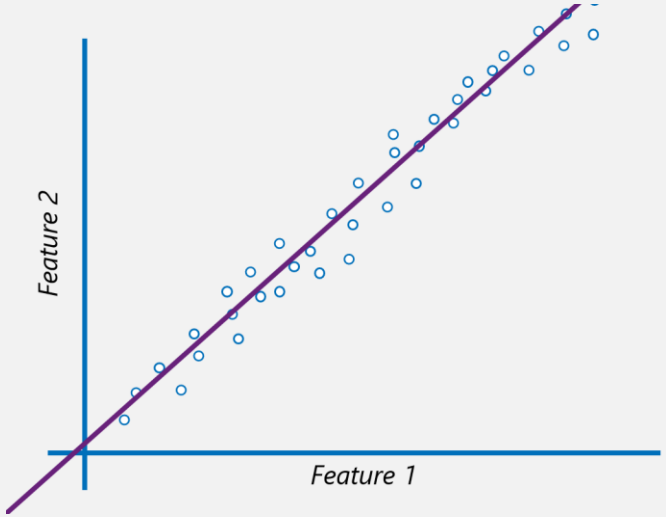
classification

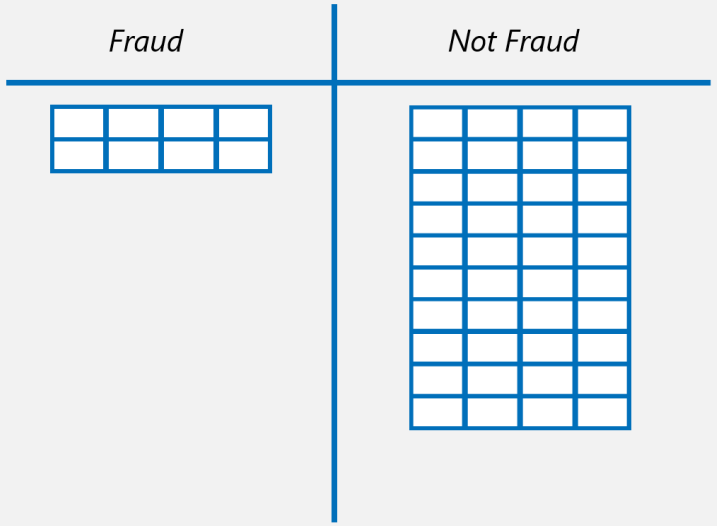
"Hey, I am trying to **predict** if an insurance claim is **fraudulent** or not"



Types of machine learning







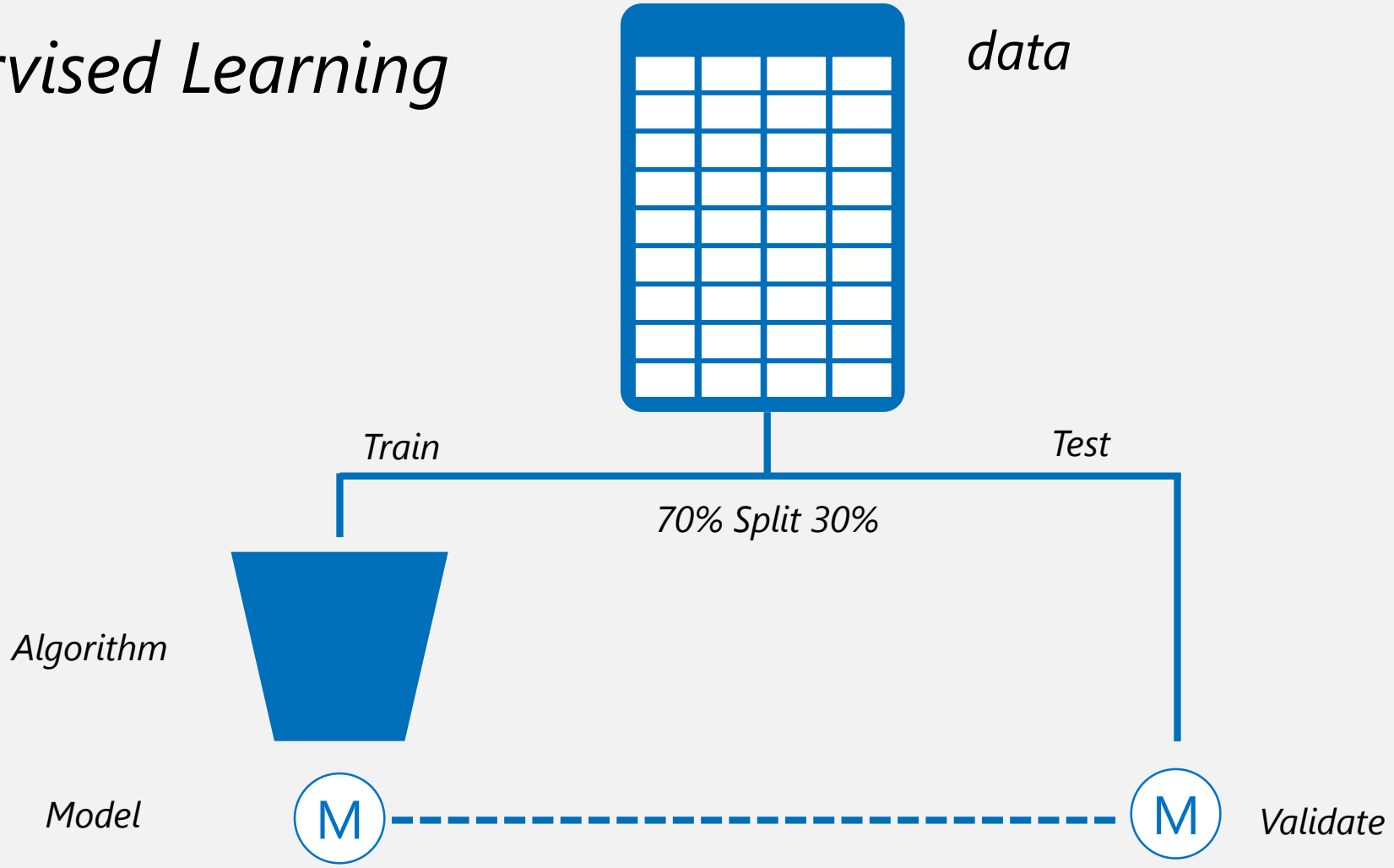
Clustering **Regression** **Classification**

Unsupervised

Supervised

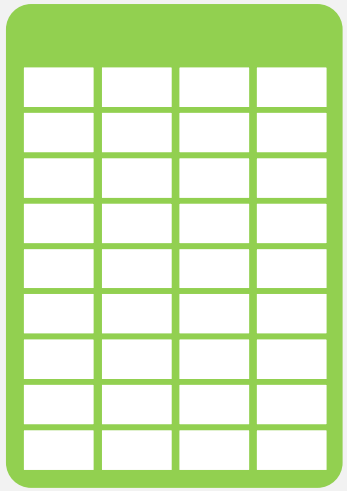
adatis

Supervised Learning



adatis

Supervised Learning



New data

Model

Prediction

adatis

R2
D3

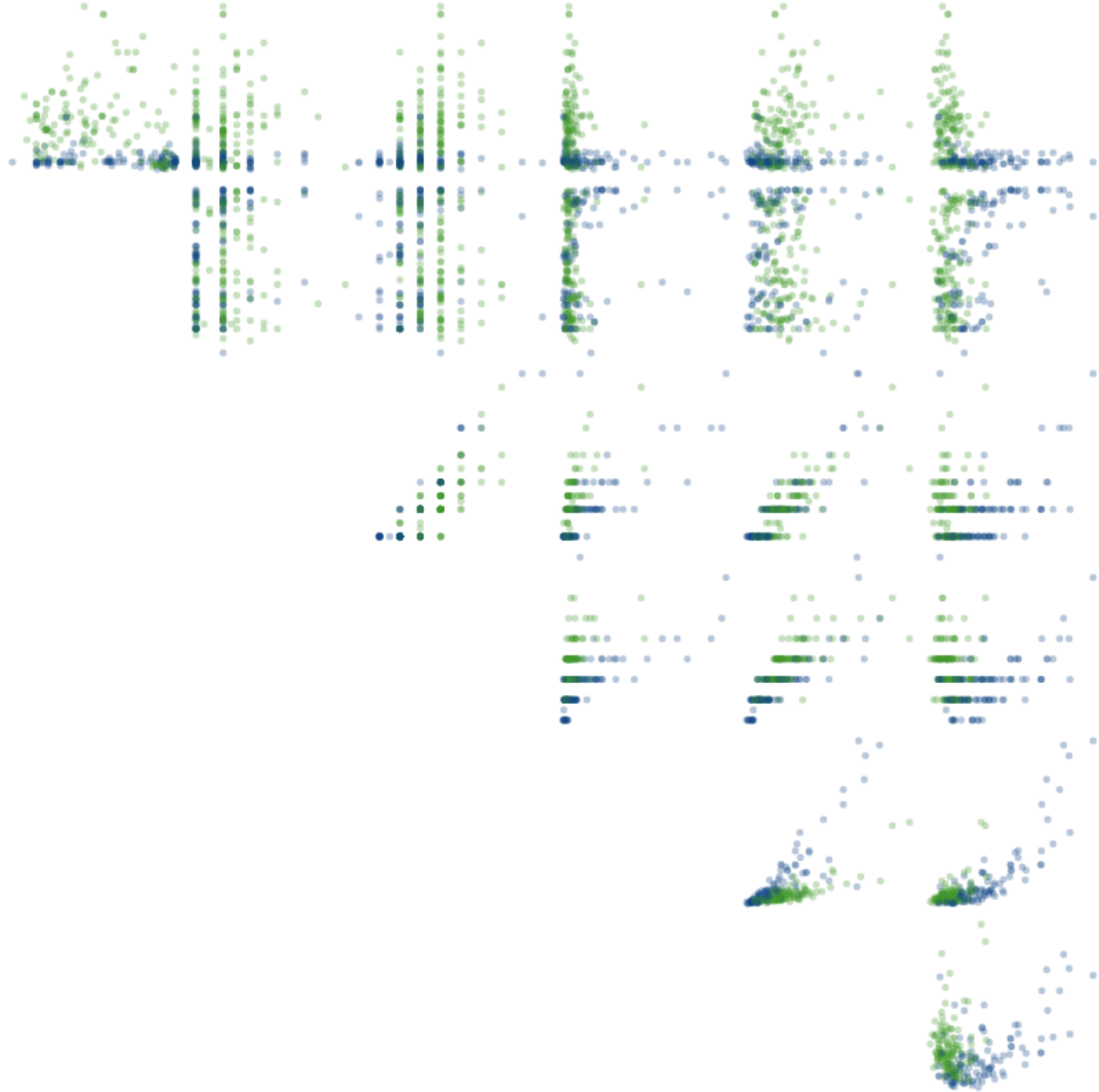
A visual introduction to machine learning

English

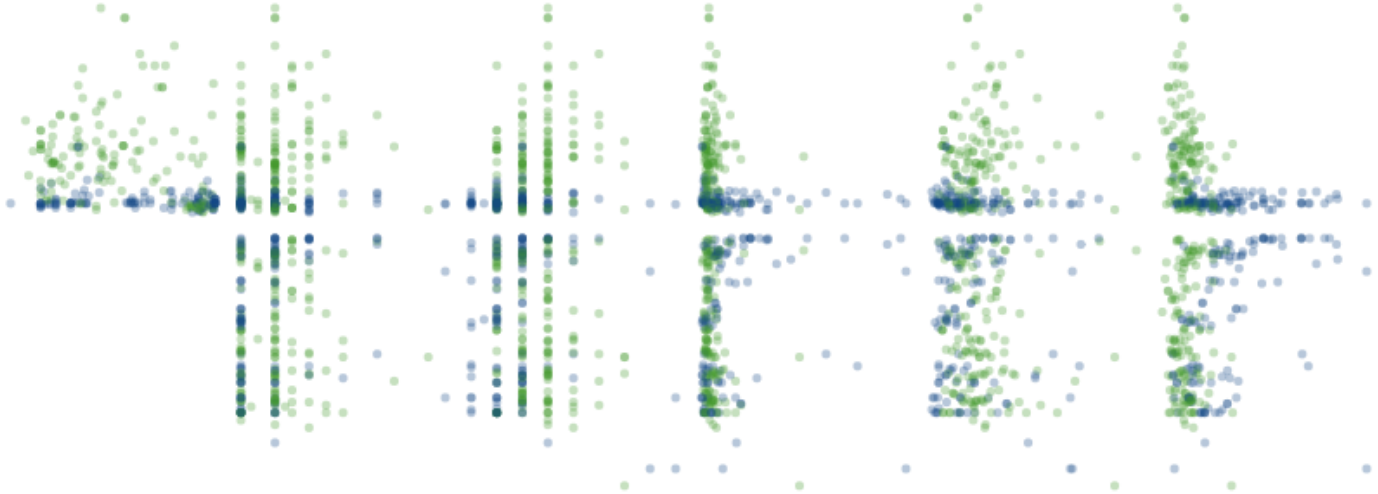
In machine learning, computers apply **statistical learning** techniques to automatically identify patterns in data. These techniques can be used to make highly accurate predictions.

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

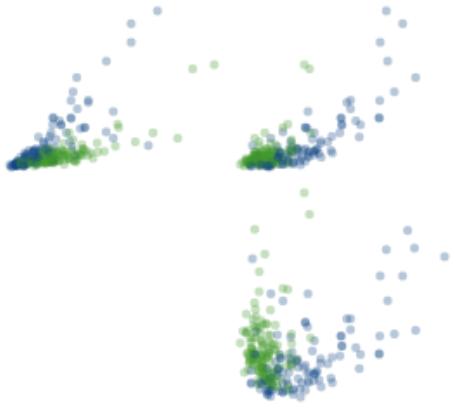


A visual introduction to

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, World!")
```

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

Download Anaconda Distribution

Version 5.2 | Release Date: May 30, 2018

Download For:

High-Performance Distribution

Easily install 1,000+ data science packages

Package Management

Manage packages, dependencies and environments with conda

Portal to Data Science

Uncover insights in your data and create interactive visualizations

Windows | macOS | Linux

Anaconda 5.2 For Windows Installer

Python 3.6 version *

Download

Python 2.7 version *

Download

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!!

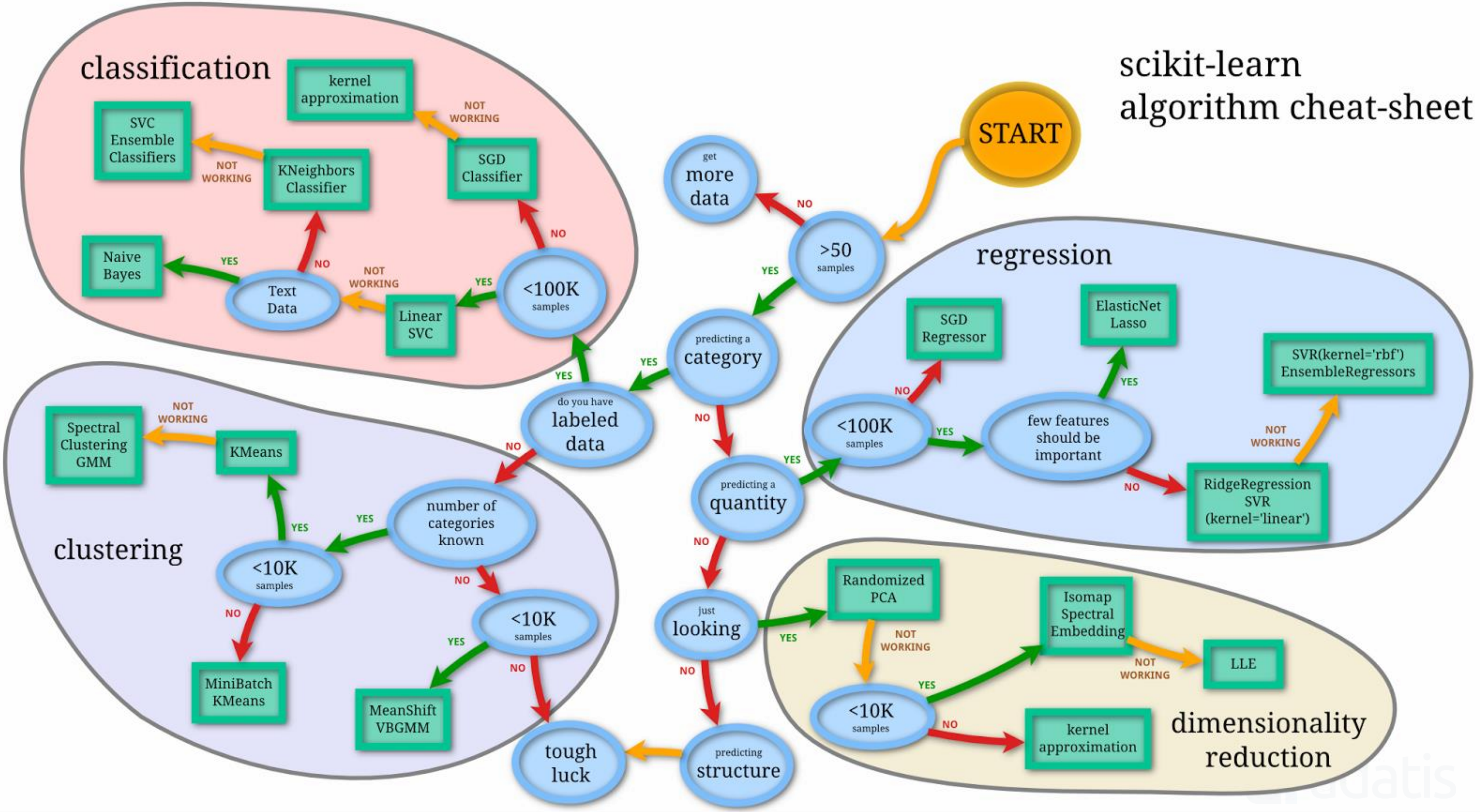
adatis

DS/ML Modules

Numpy	– Numerical Python
SciPy	– Scientific Python
Pandas	- Data wrangling
MatPlotLib	- Data visualisation
Scikit-learn	- Machine learning (shallow)
….	
TensorFlow	- Machine Learning (deep)
NLTK	- Natural Language processing

How to add modules? **PIP**

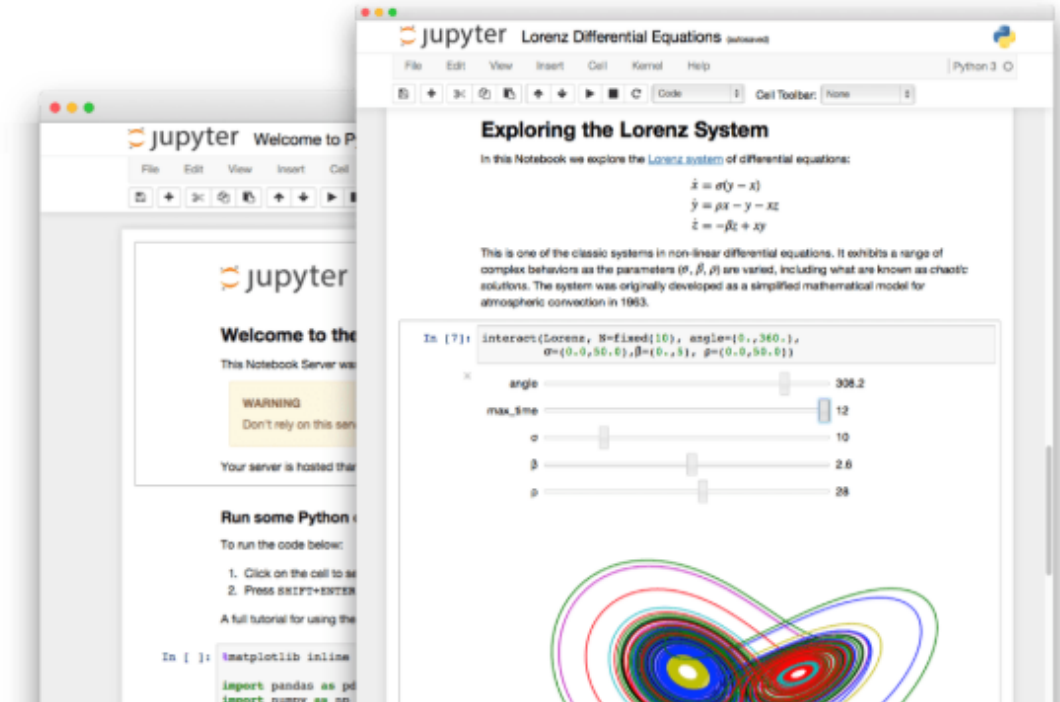
adatis



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

Python: Demos

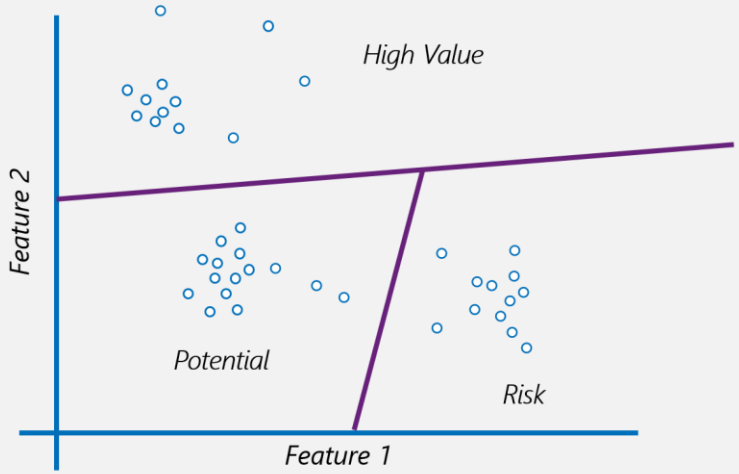
adatis



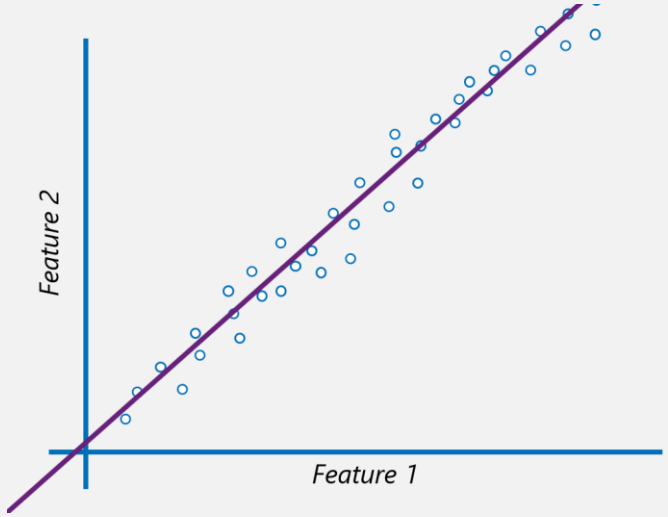
The Jupyter Notebook

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.

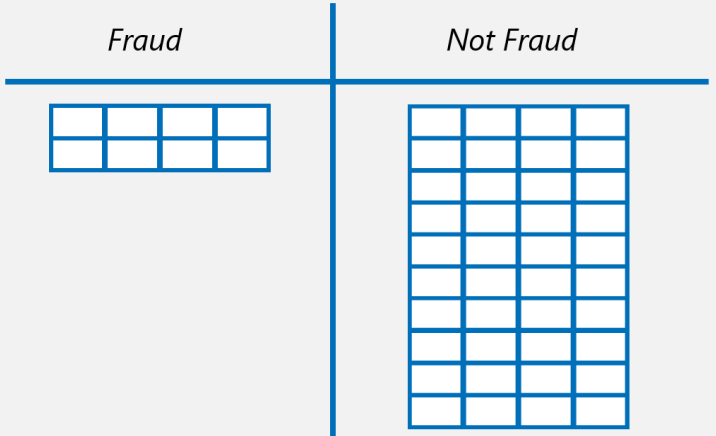
Python Fundamentals



Clustering



Regression



Classification