

#### Volume

- · Terabite, petabite
- · Records
- · Transaction data
- · Tables and files

3V of Big Data

- · Batch (Intermittent piling)
- · Near real-time
- · Real-time
- · Continuously streaming data

- Structured
- Unstructured
- · Semi-structured
- · SNS data, Logs, Sensor data
- Text, images and other media

Velocity Variety

## **Big Data Business Models**

Access

Opening up new data sources

Commodify

Assessment and commodification

Crossuse

New uses for existing data

Combine

Fusion of data sources

Pattern recognition

**Testing** 

Modeling

**Predictions** 

**Visualisation** 

**Efficiency** 

**Planning** 

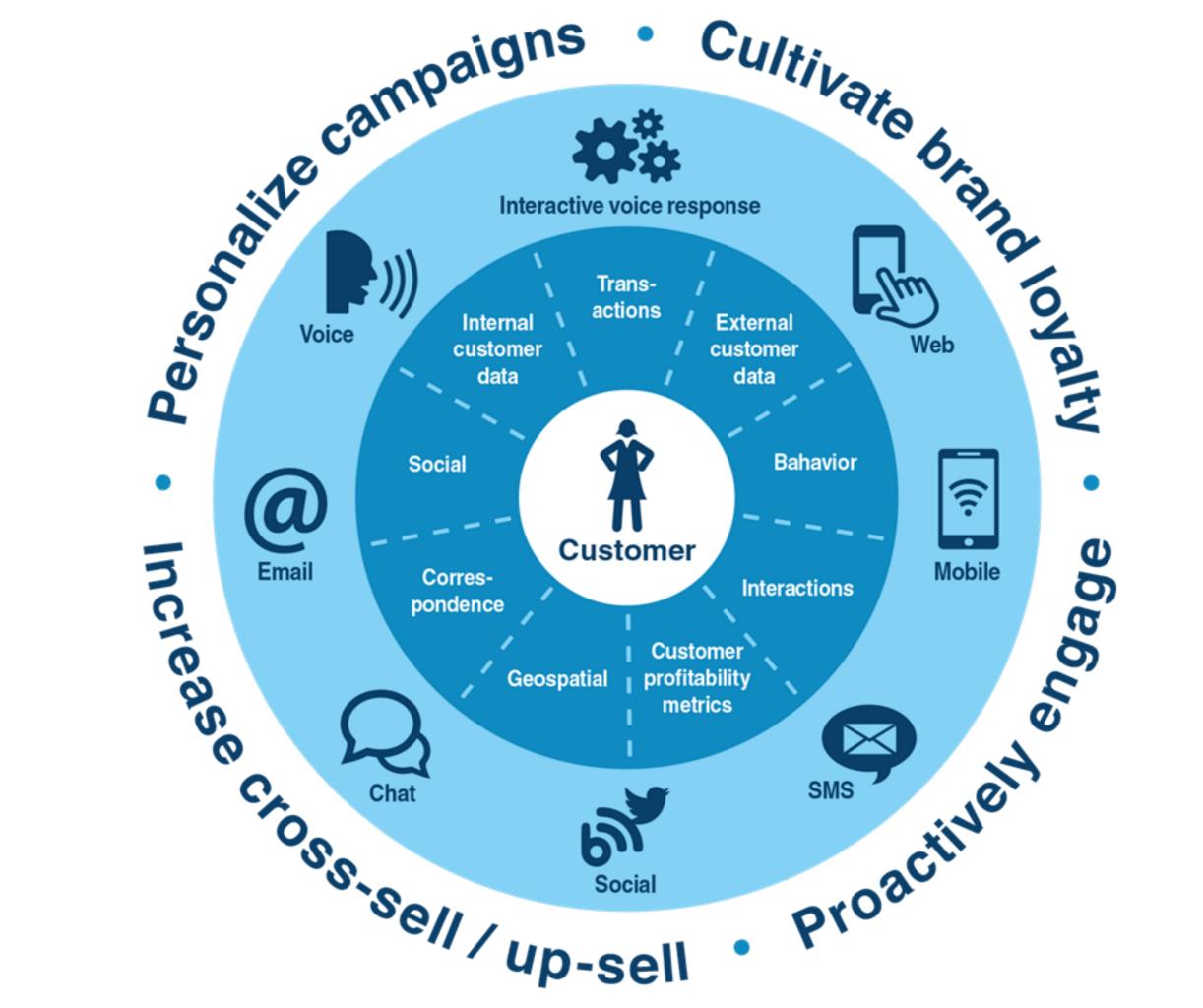
Innovation

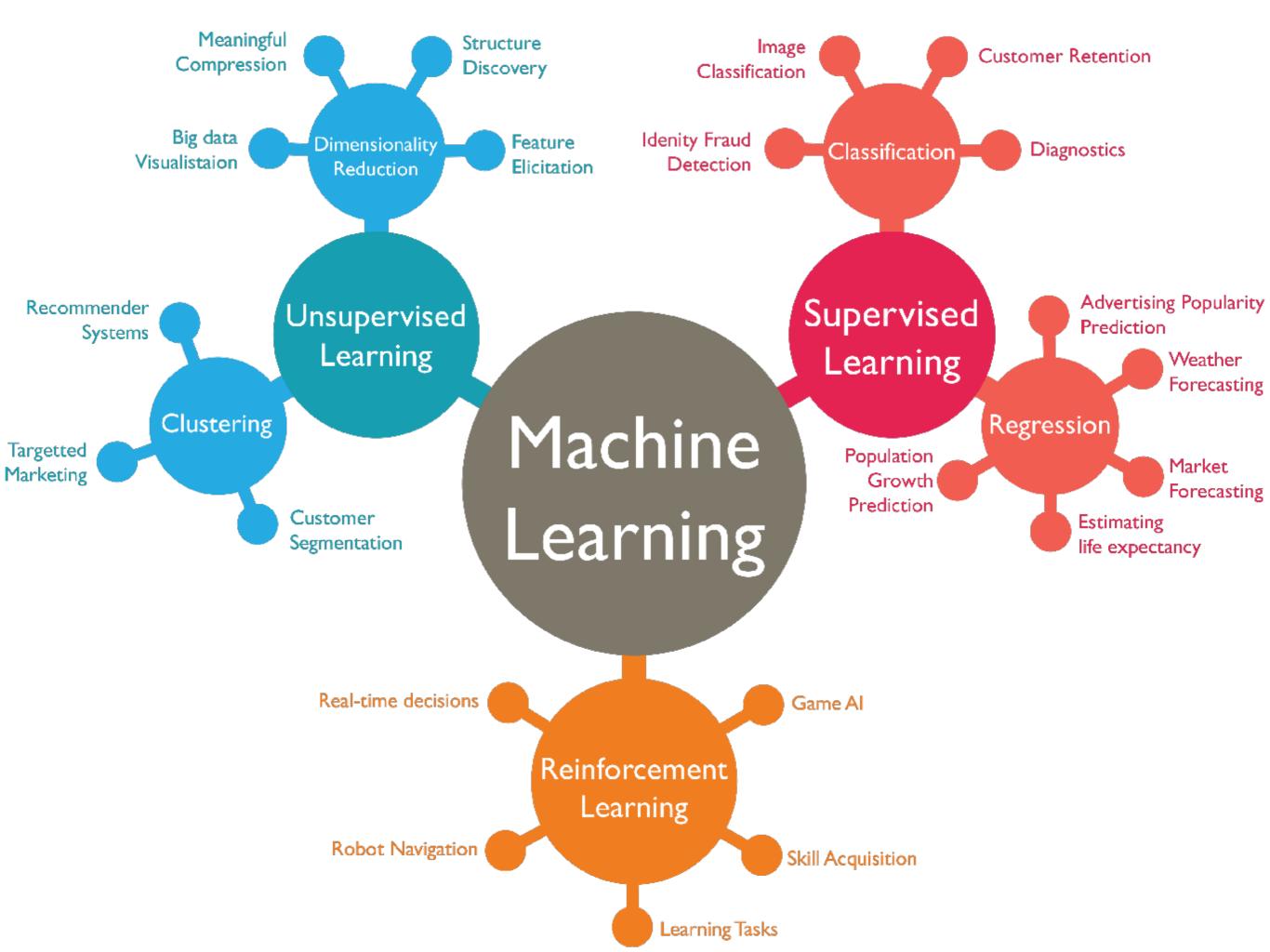
**Customer Insights** 

Personalization

Quality

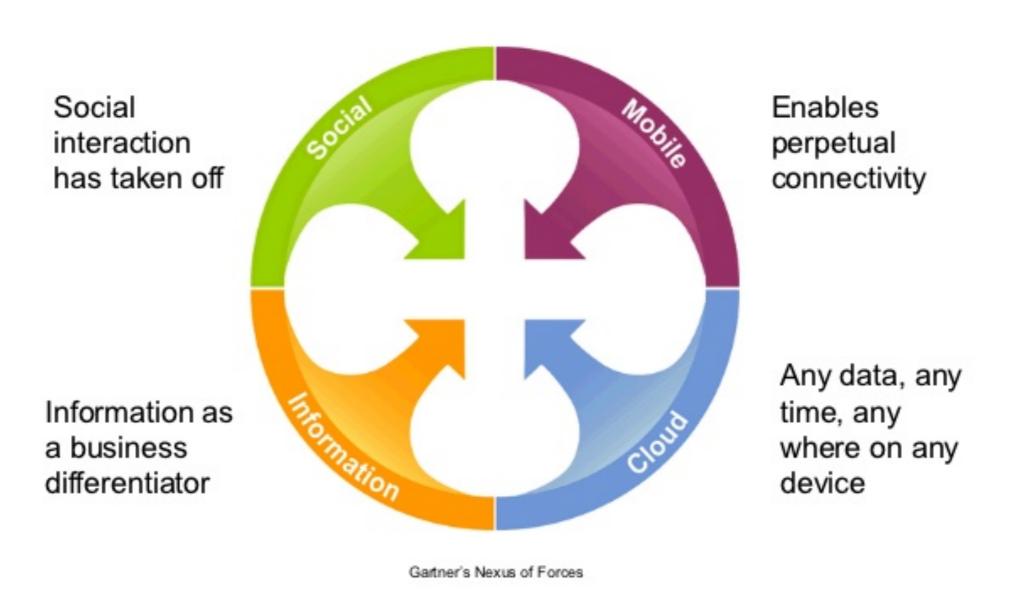
Intelligence





#### Why Big Data Now?

Confluence of 4 forces - Mobile, Social, Cloud and Information



#### A mostly complete chart of

#### Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org



Backfed Input Cell

Input Cell

Noisy Input Cell

Hidden Cell

Probablistic Hidden Cell

Spiking Hidden Cell

Output Cell

Match Input Output Cell

Recurrent Cell

Memory Cell

Different Memory Cell

Kernel

Convolution or Pool

Perceptron (P)



Feed Forward (FF)

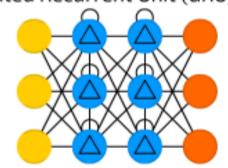


Radial Basis Network (RBF)





Gated Recurrent Unit (GRU)



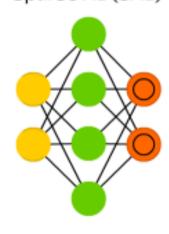
Recurrent Neural Network (RNN)



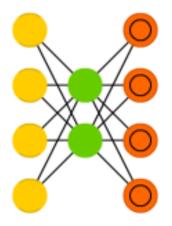
Long / Short Term Memory (LSTM)



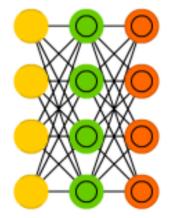
Sparse AE (SAE)



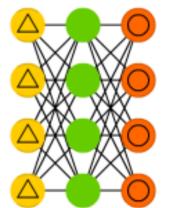
Auto Encoder (AE)



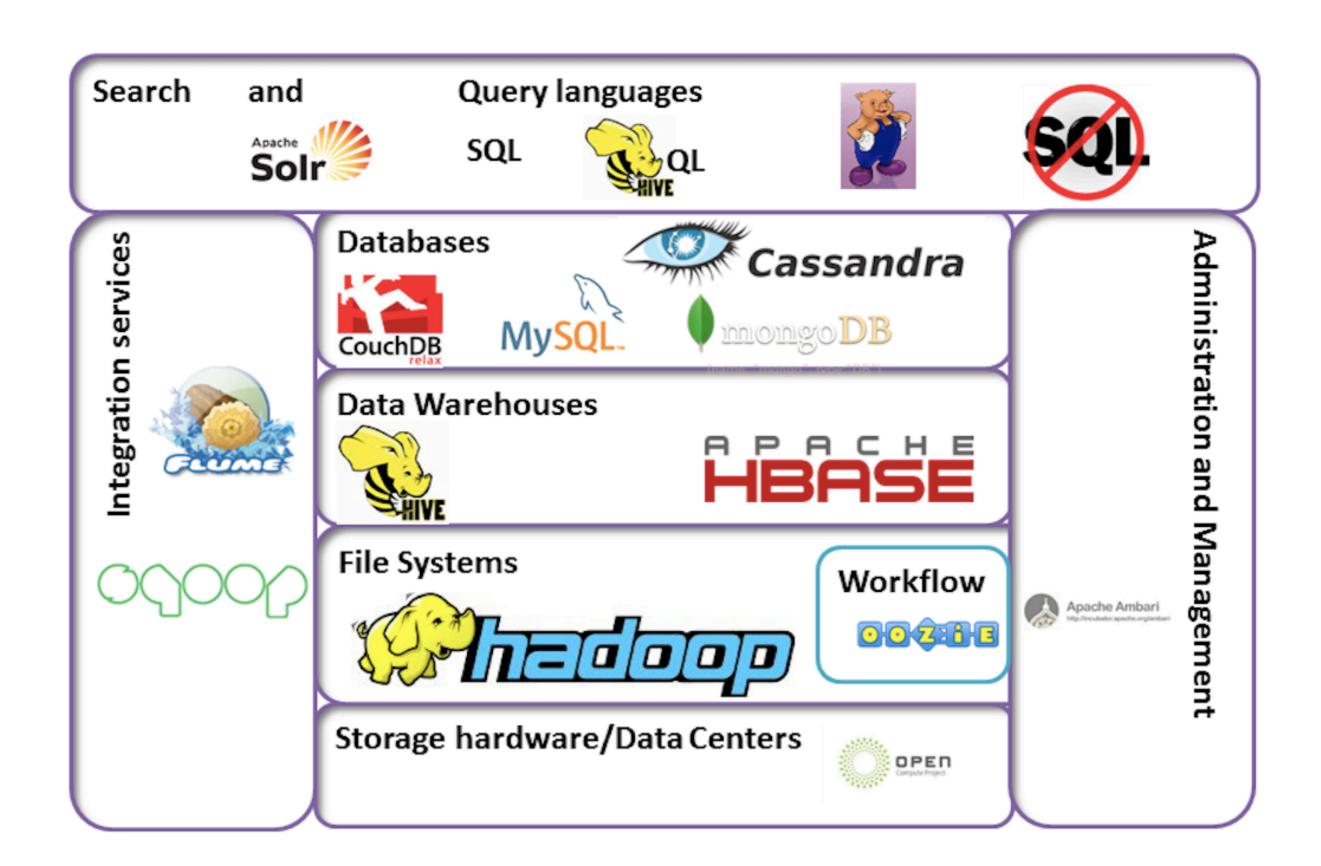
Variational AE (VAE)



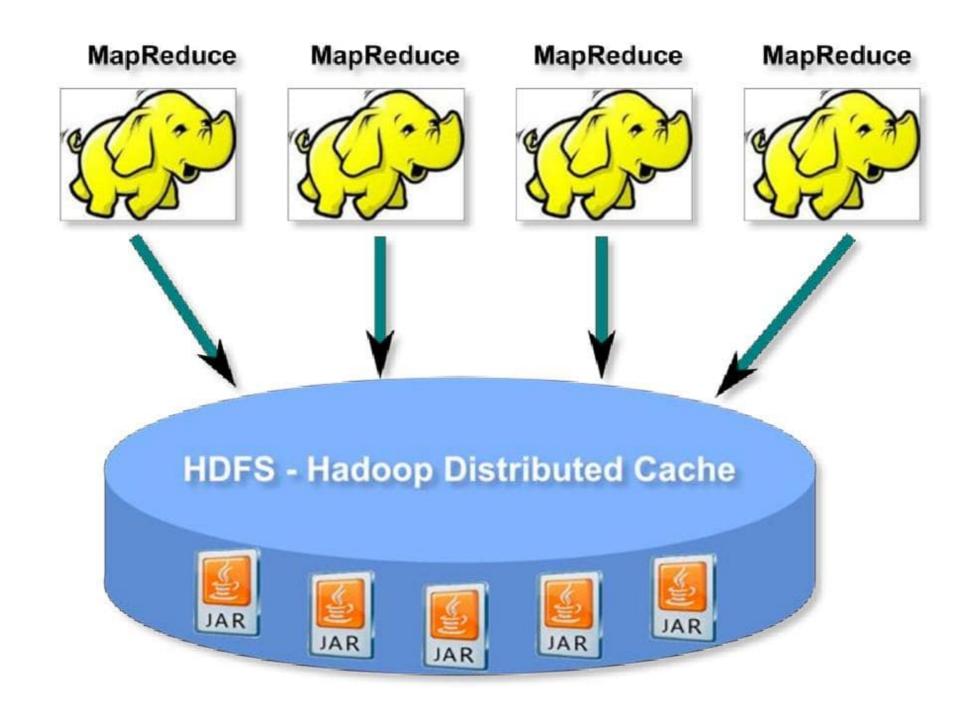
Denoising AE (DAE)



### The Big Data Open Source Technology Stack



#### **Distributed Cache**

















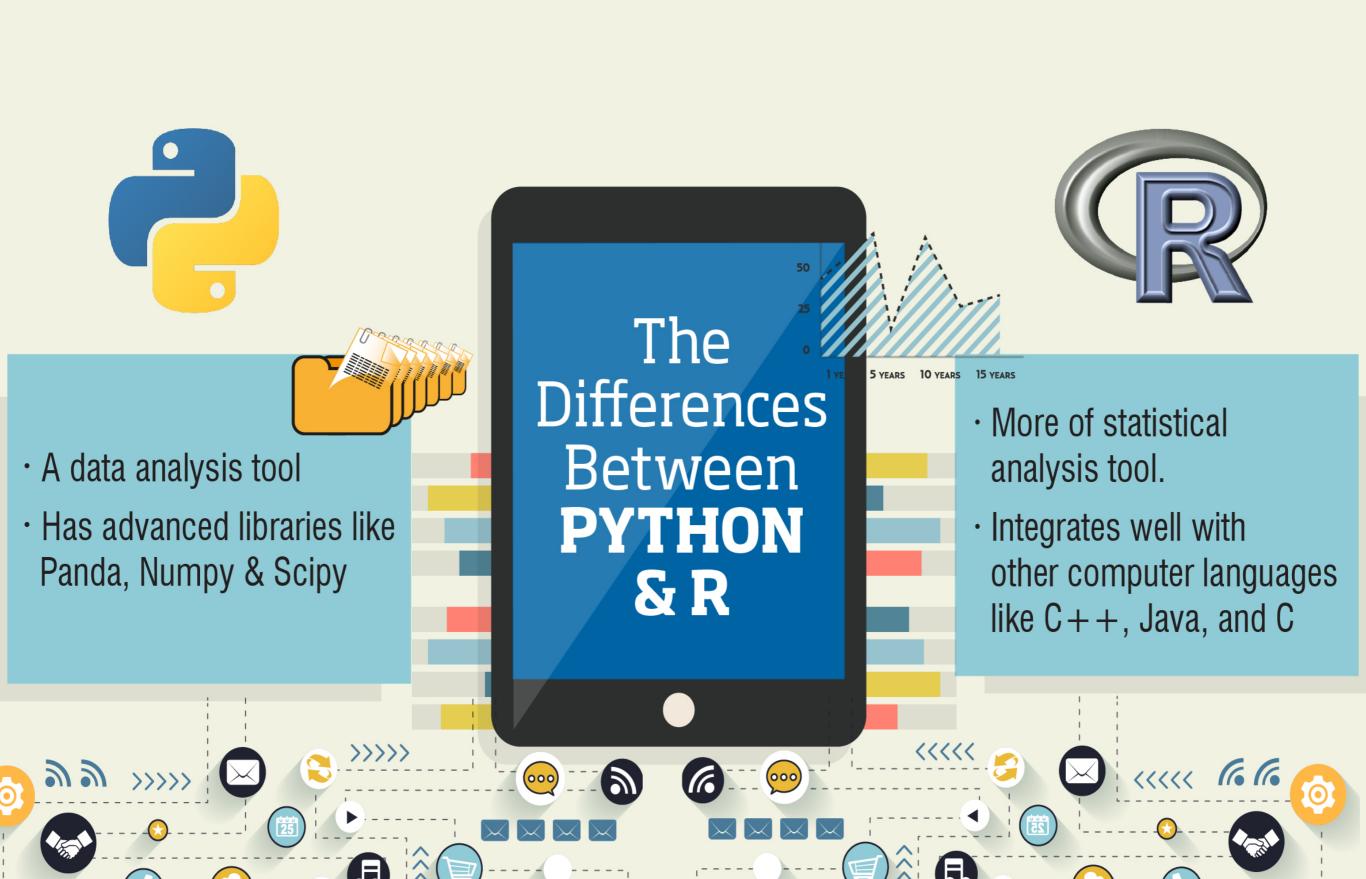
**HYPERTABLE** INC



Neo4j

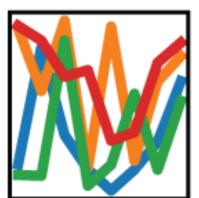


redis

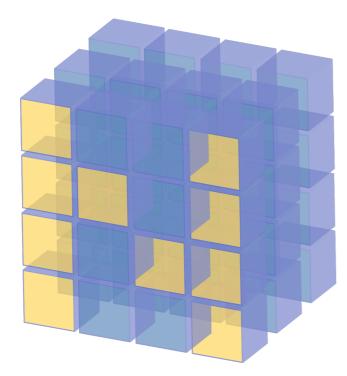


# $\underset{y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}}{\mathsf{pandas}} \left[ \mathbf{y}_{it} \right]$

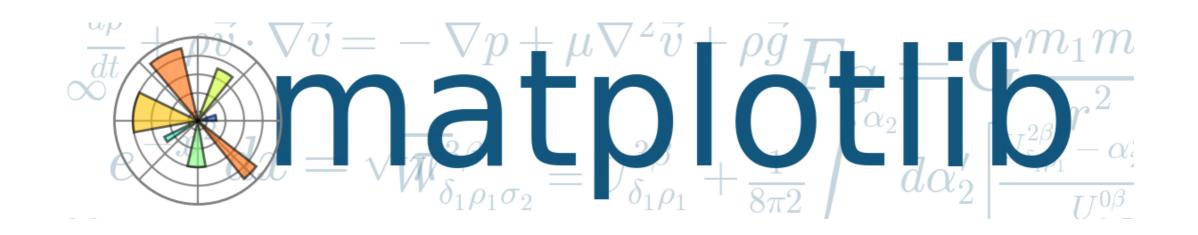


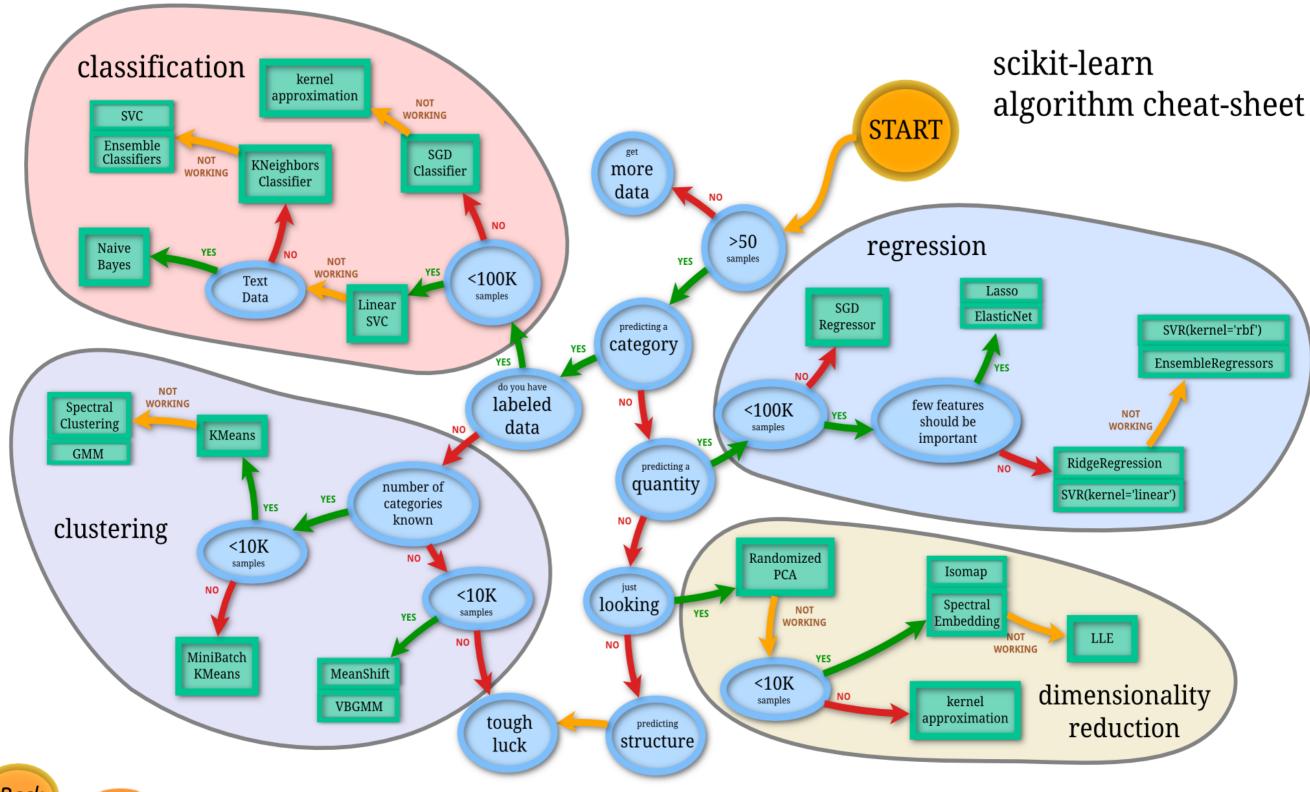






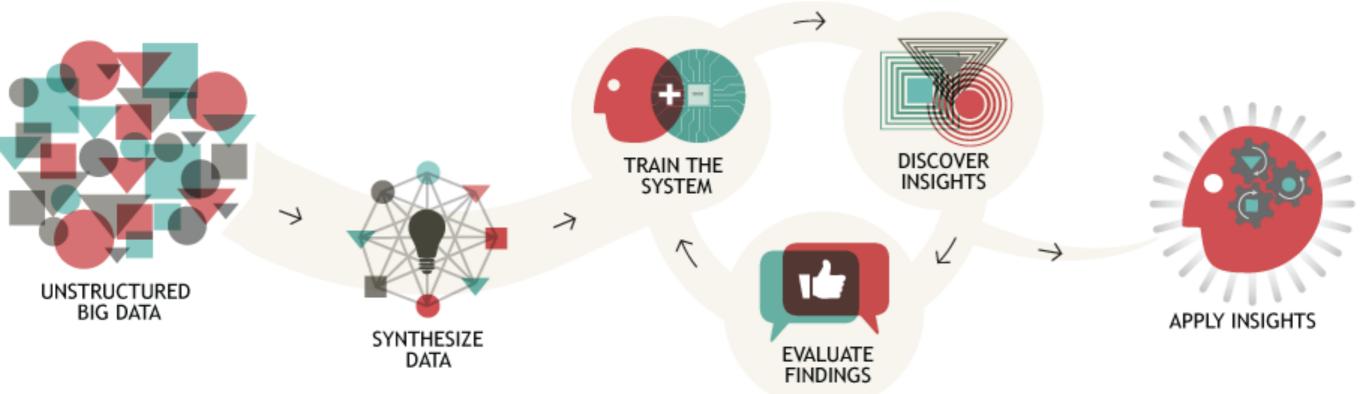
## NumPy











© DISCOVIX 2012