



Introduction to Data Science (with Python)



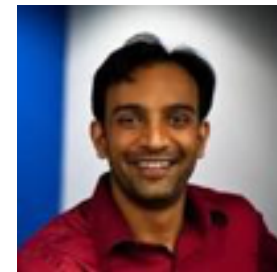
What, Why & Where

The Data Scientist

“the sexiest job in the next 10 years will be statisticians.”
— Hal Varian (Chief Economist @Google)



“The data age has arrived. From crowd-sourced product reviews to real-time traffic alerts, data science has become a regular part of our daily lives.” — Dr. D.J. Patil (first U.S. Chief Data Scientist)



What is Data Science?

The extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.

— Thomas H Davenport, Competing on Analytics



A **scientific art** of doing anything **meaningful** with data that makes business decision making more **accurate & easier.**

What is Data Science?

- Data Science is the application of: **computer technology, machine learning and domain knowledge** to solve problems in business and industry, to aid efficient and effective decision making
- Data Science is simply the **scientific process of converting raw data into knowledge** to support decision making
- Data Science involves **finding patterns in data**
- The goal of Data Science is to **improve business, society or personal performance** by gaining knowledge from data
- Data Science is **moving decision making from gut feel and guesstimates** to better, more informed ones driven by data

So why is Data Science so important all of a sudden?

Why is Data Science used?

Decision making is now fact and performance based

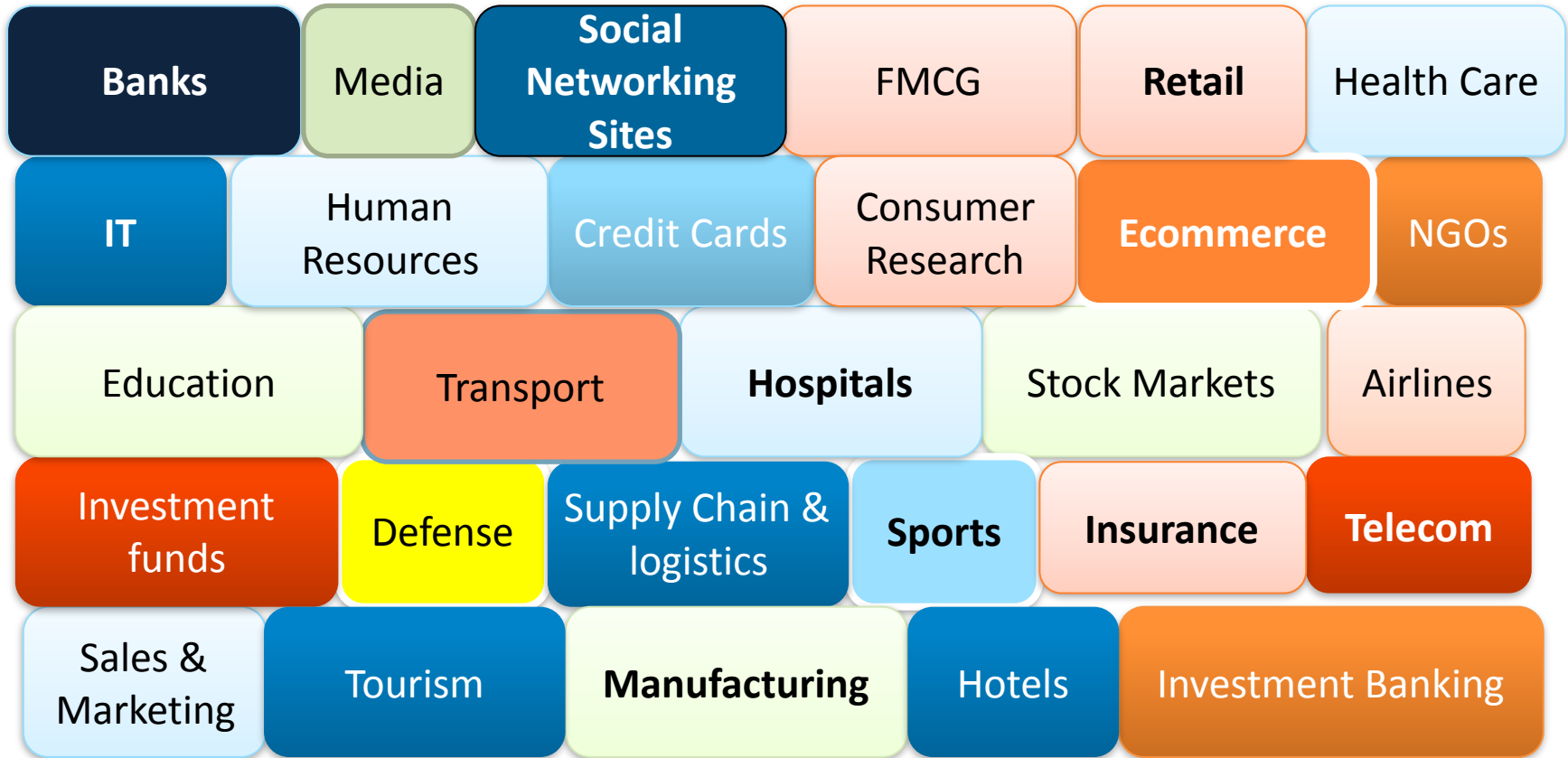
Intuition is out, metrics are in

Intense competition, shorter time-to-market, demanding customers

Make each and every dollar count and increase return on investment

Take real-time decisions

Industries using Data Science

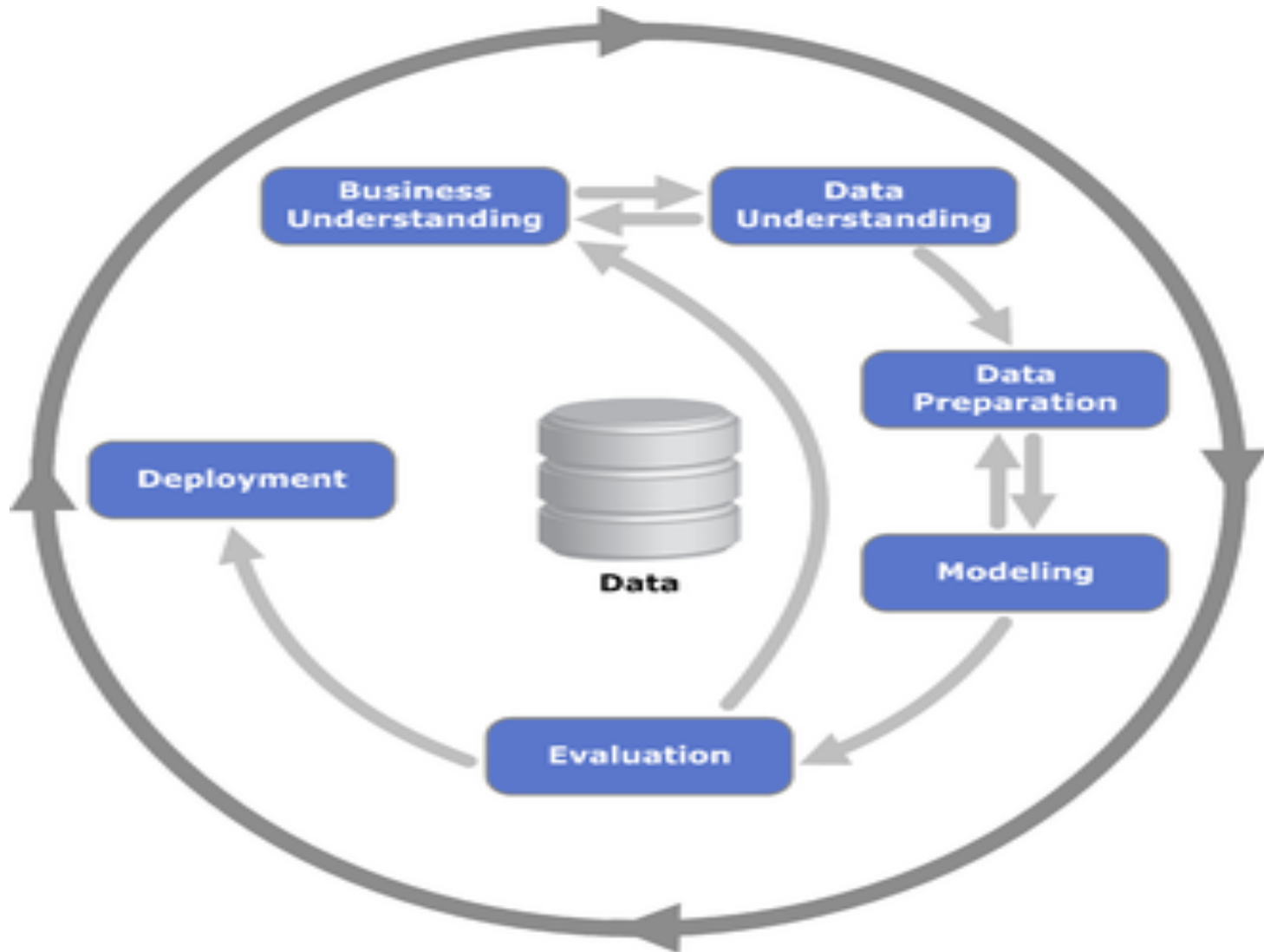


Insights → Actionable



How: Data Science Methodology

Data Science process



Data Scientist Toolbox

Capabilities of A data Scientist

- Conversant with at least one relevant programming language
- Can handle project level problems , version control
- In addition to traditional sources, knows how to collect data from web, social media platforms etc
- Conversant with multiple machine learning algorithms
- Knows how to build prototype end product , data driven solutions
- Is a good storyteller !!

What Is Python?

Python : The Programming Language



Python & Scikit-Learn: That Data-science Duo



+

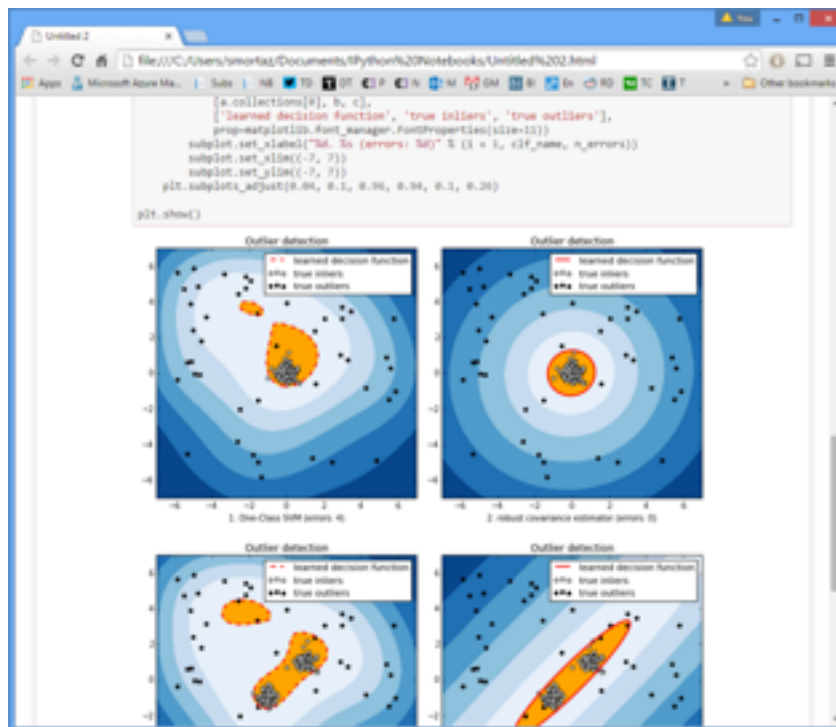


=



Course Components

Getting Ready with Python

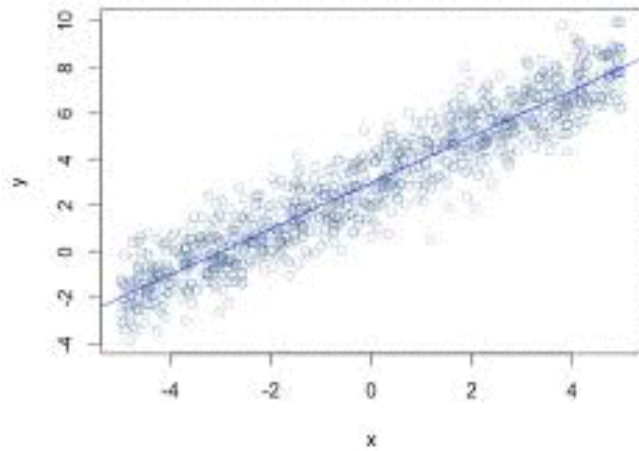


Version Control , Web Scraping and APIs

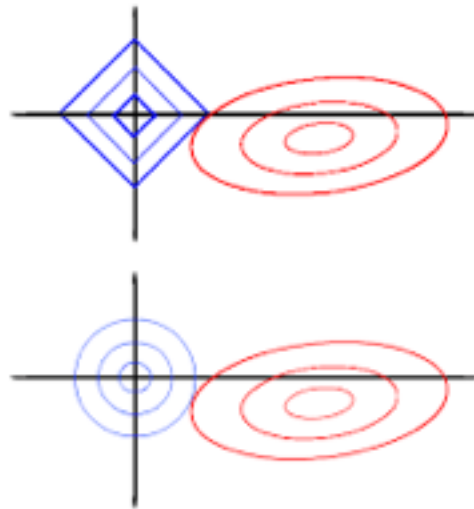
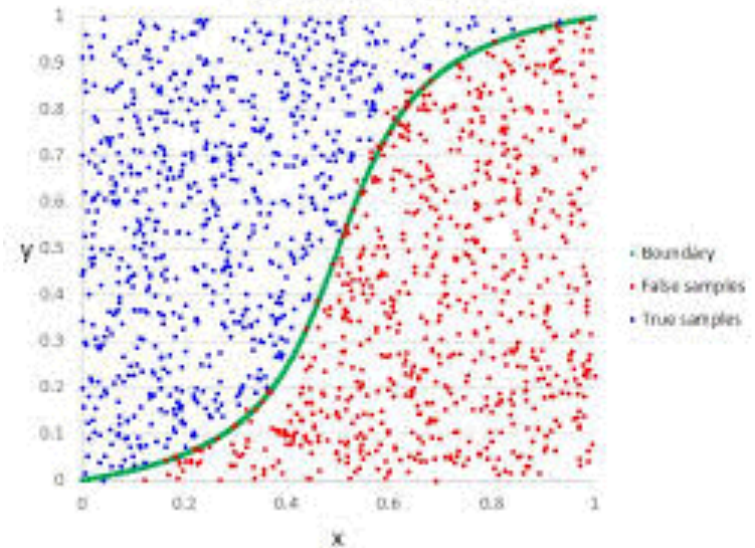


ML algorithms : Generalised Linear Models

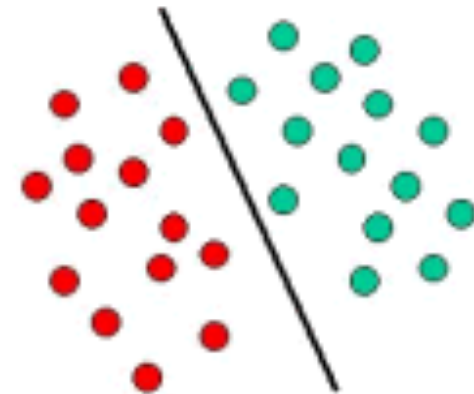
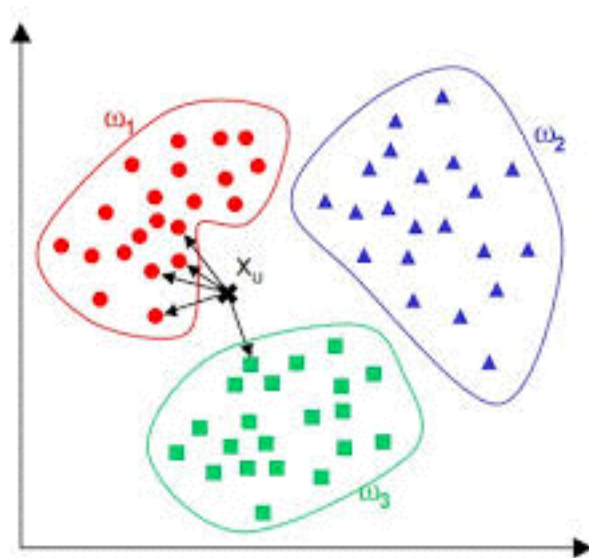
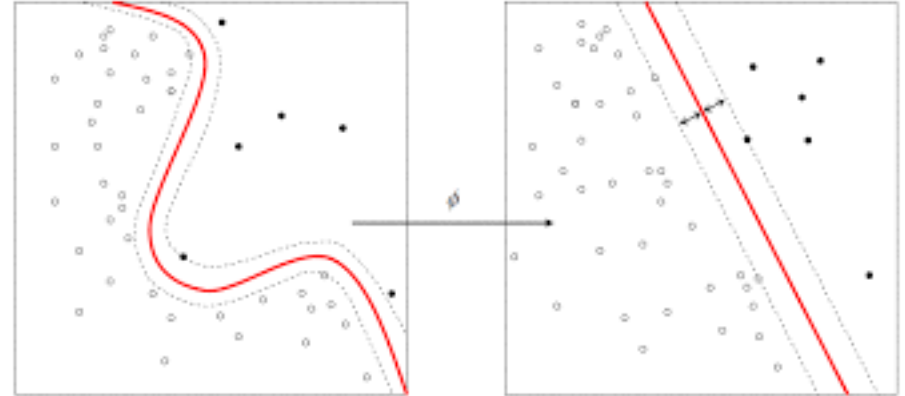
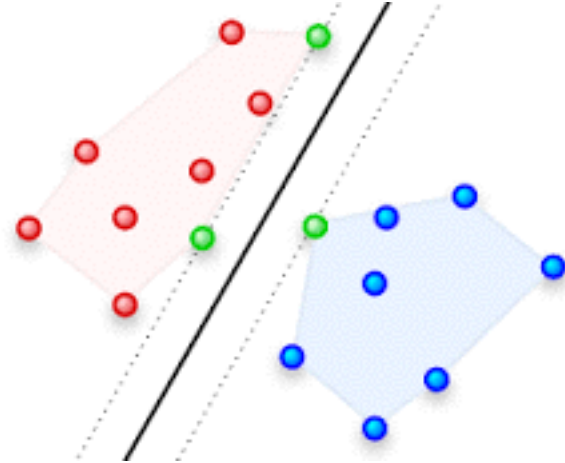
Linear regression



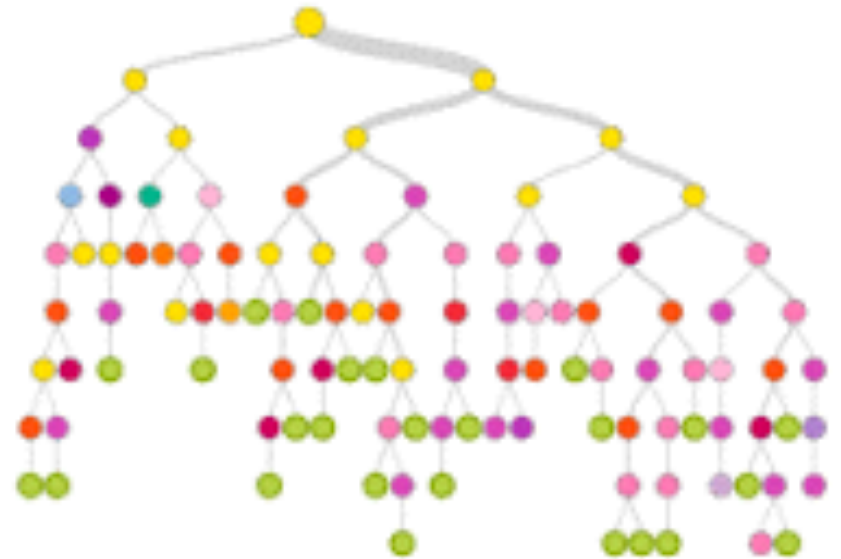
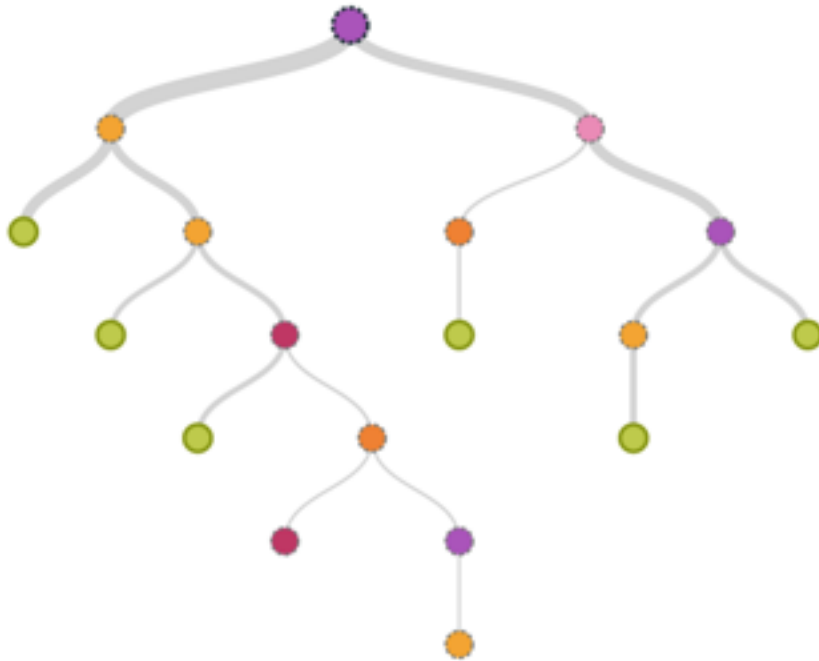
Logistic Regression Example



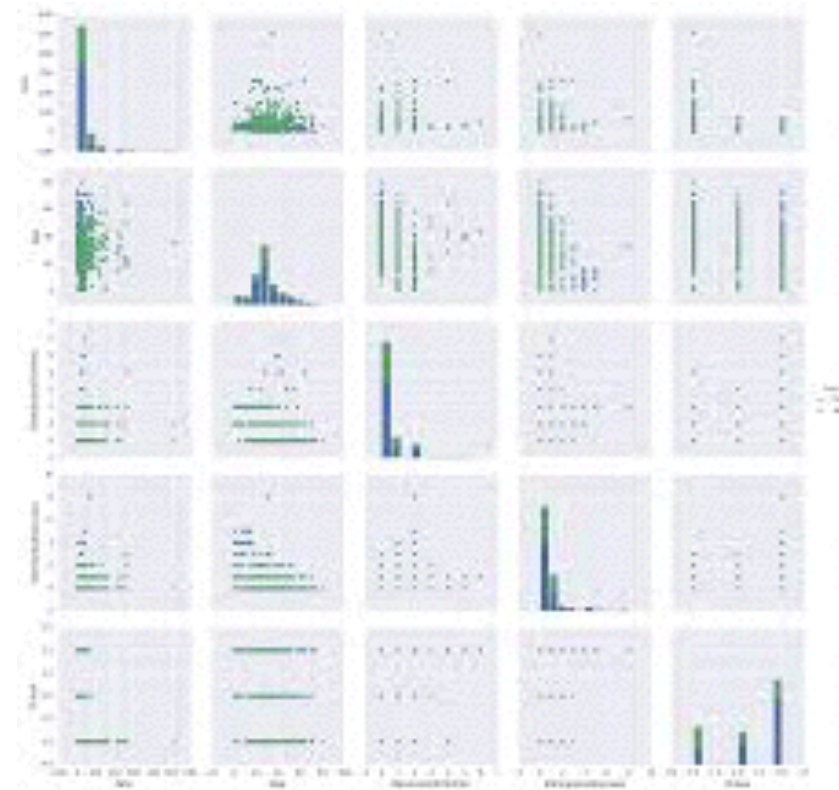
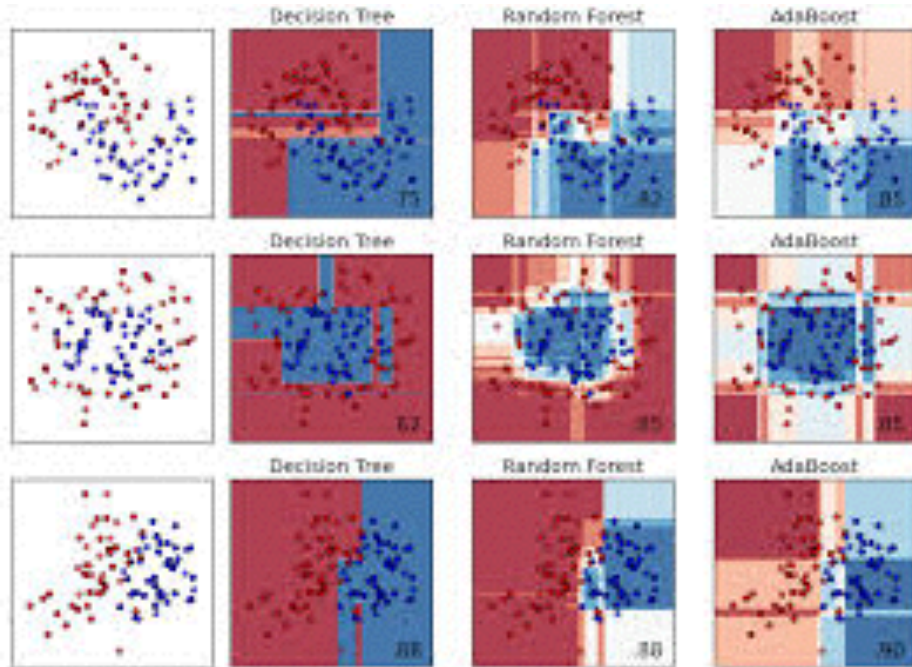
ML algorithms : Naive Bayes, KNN, SVM



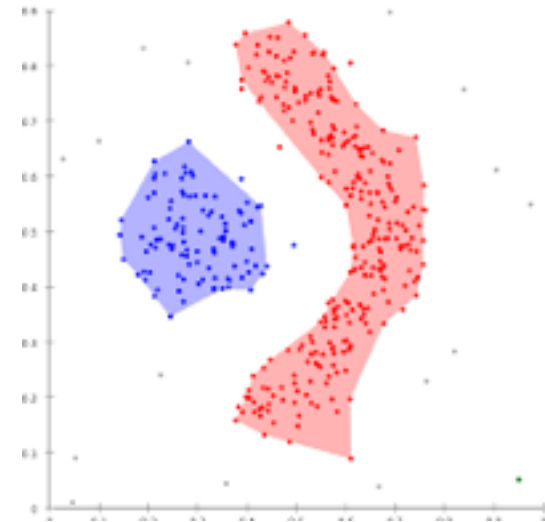
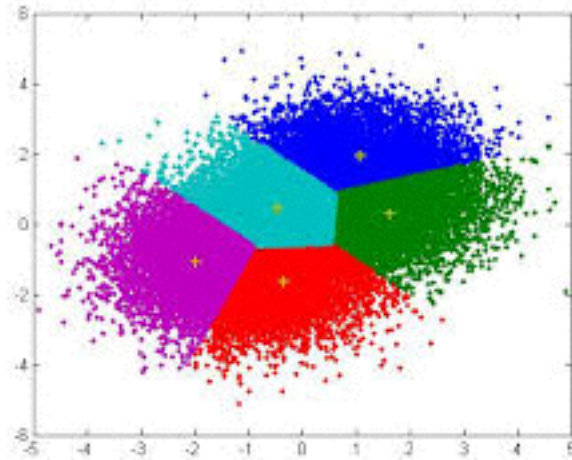
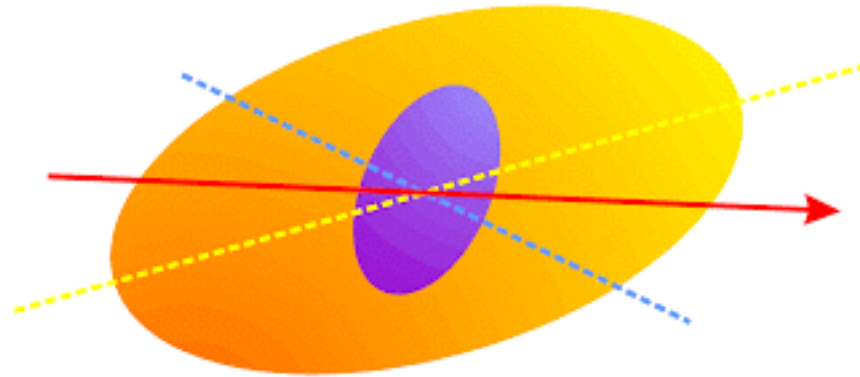
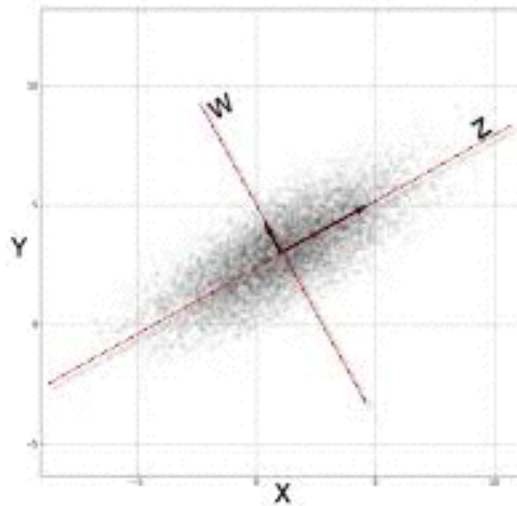
ML algorithms : D-trees & RandomForests



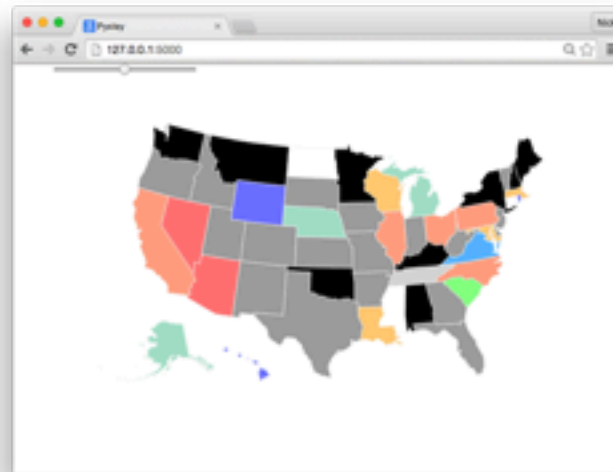
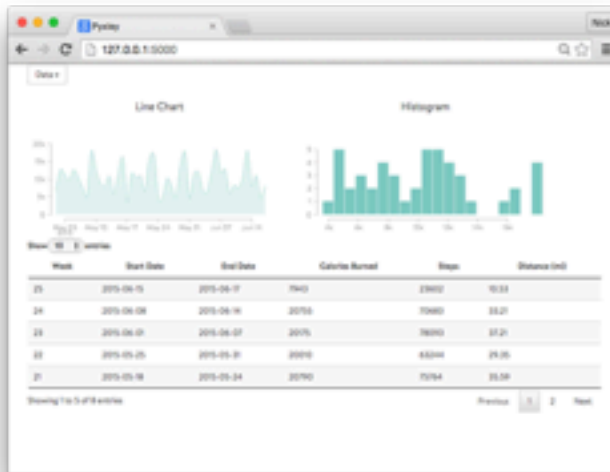
ML algorithms : Boosting Algorithm



ML algorithms : PCA,Factor, K-means, DBSCAN



Deploying your solutions



The Pyxley application window displays a table with columns: Position, Office, Start Date, and Salary Range. The table lists various roles and their associated details. The interface includes a search bar at the top and a 'Show' dropdown menu.

Position	Office	Start Date	Salary Range
System Architect	Edinburgh	2011-04-25	17k 10k
Accountant	Tokyo	2011-01-25	10k 10k
Junior Technical Author	San Francisco	2009-09-12	14k 17k
Senior Javascript Developer	Edinburgh	2012-03-29	10k 40k
Accountant	Tokyo	2008-11-28	10k 10k
Integration Specialist	New York	2012-12-02	30k 40k
Sales Assistant	San Francisco	2012-08-06	12k 15k
Integration Specialist	Tokyo	2010-10-16	30k 17k
Javascript Developer	San Francisco	2009-06-15	18k 18k

How Will we do it?

How.....

- This is one of the rare presentations that you will get to see
- Either white board or scripts
- Ask questions : Things need not make sense to you because they make sense to someone else
- I will not have all the answers , explore!
- Practice exercises at the end of every class ; except today.

Thank you