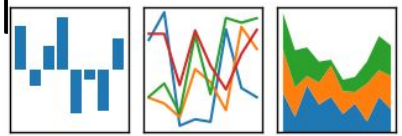
Autumn Session 2019

King's College London Health Science DTC

Our goals for today

- Learn about Data Science approaches.
- Use the time today to code together and exchange ideas and experiences.
- Using the group effort to solve data challenge.

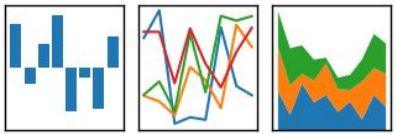
A group of experienced programmers (Mateusz, Matthew, and Paul) will be available to help out during today's session.



Our goals for today

Specifically:

- We will learn about machine learning algorithms for classification.
- How to a set up Logistic Regression and Random Forests models.
- Evaluate predictive models using different performance metrics.
- Fine-tune machine learning models.



Data set of interest

Mammographic Mass Data Set

Download: Data Folder, Data Set Description

Abstract: Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age.

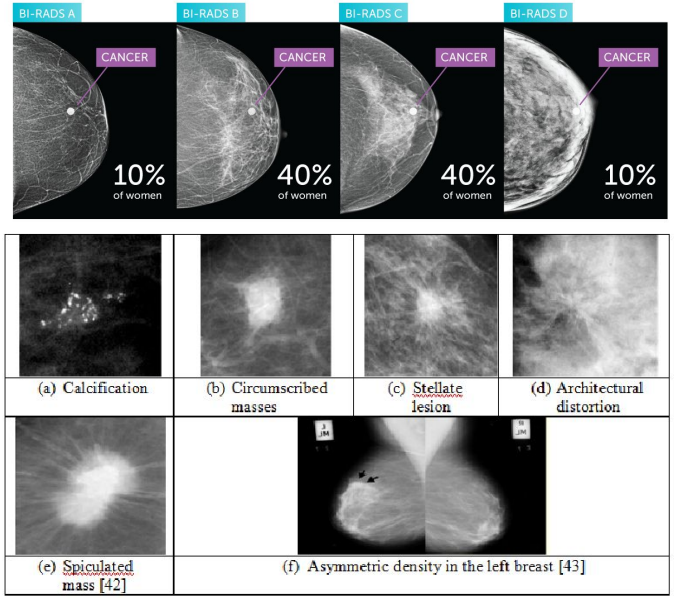
Data Set Characteristics:	Multivariate	**Number of Instances:**	961	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	6	**Date Donated**	2007-10-29
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	156390

Mammographic Mass data set:

- Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and patient's age.
- A supervised classification problem.

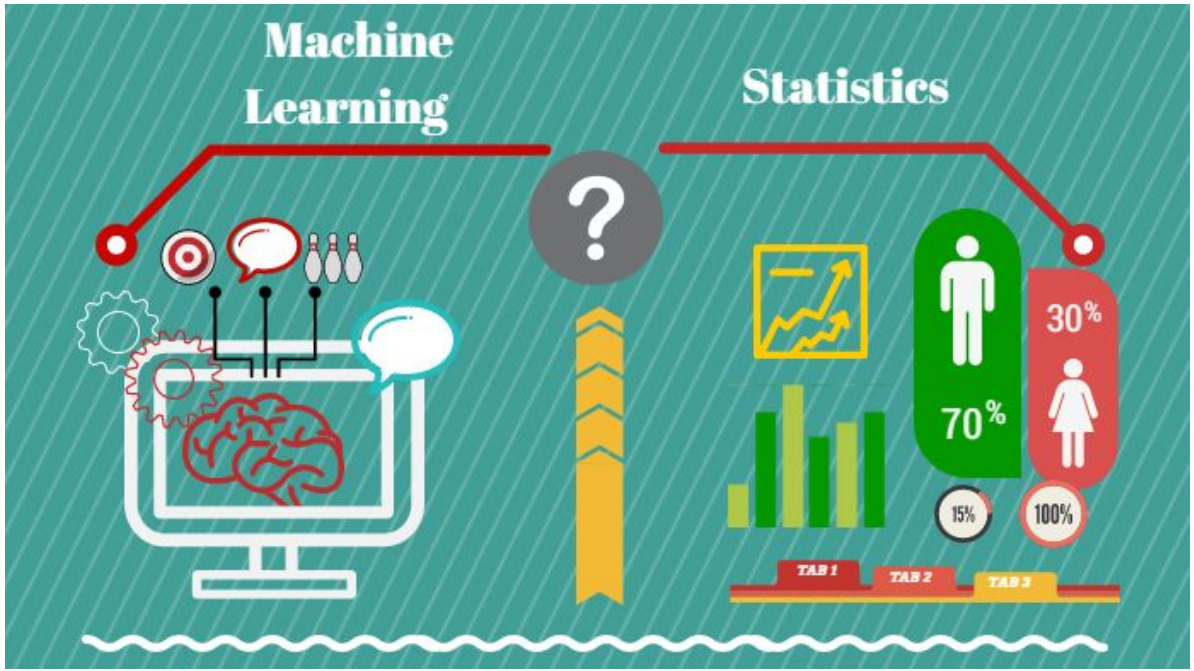
Mammography and BI-RADS

Low-energy X-rays for diagnosis and screening

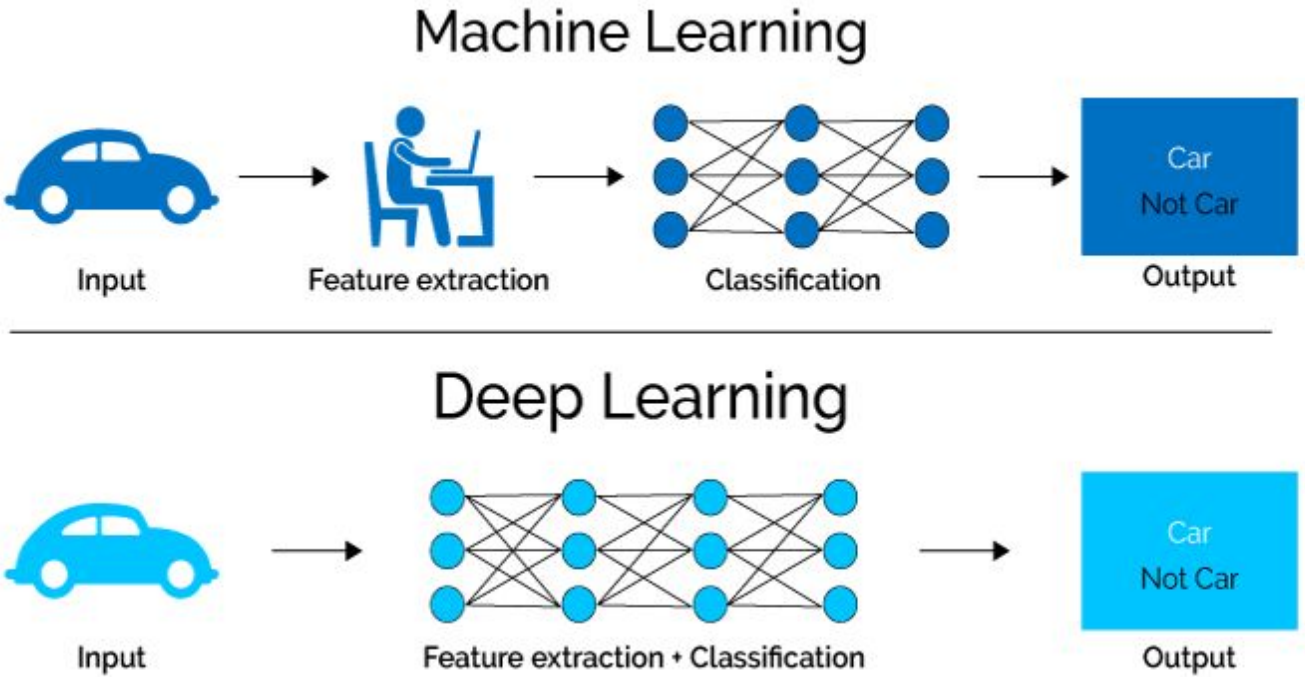


Breast Imaging Reporting and Data System (BI-RADS)

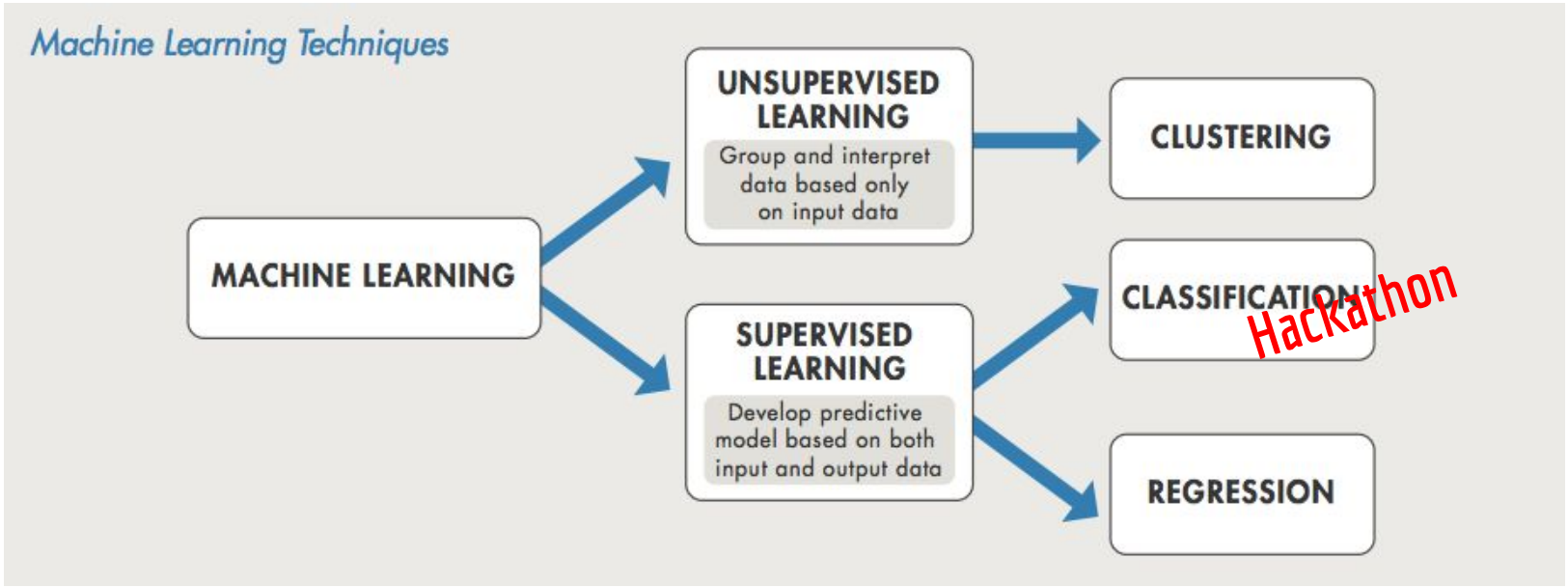
What is machine learning?



Machine vs deep learning



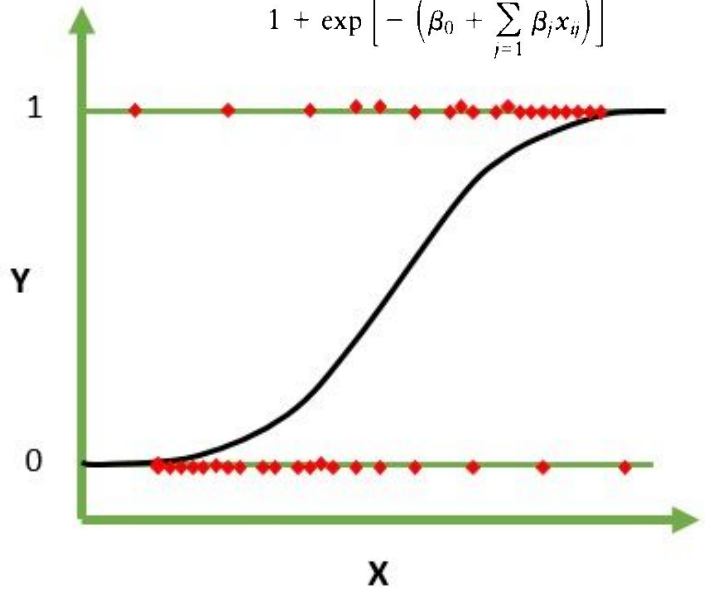
Supervised vs unsupervised learning



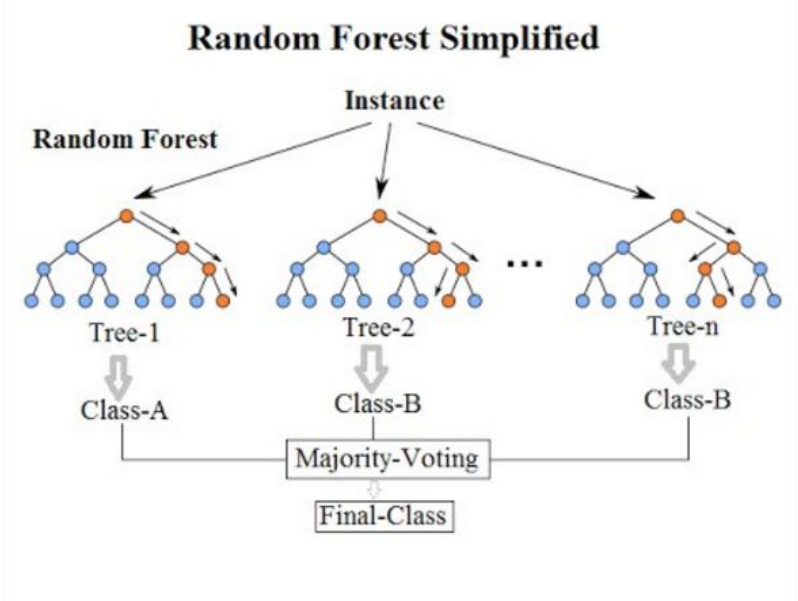
Learning models

Logistic regression

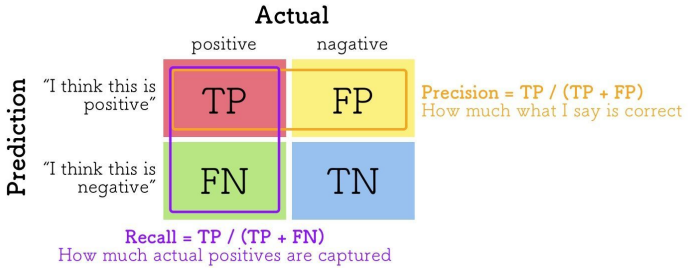
$$\theta_i = \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\right]}$$



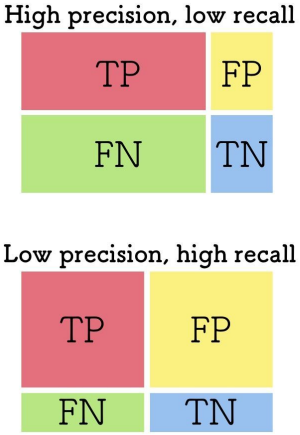
Random Forests



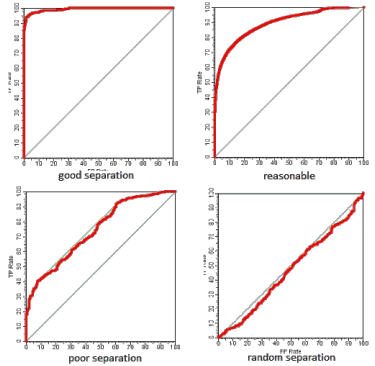
Metrics for evaluating models



Confusion matrices



Accuracy, precision, recall scores...



ROC curves

Acknowledgments

King's HSDTC:

Mr Matthew Coleman

Dr Fiona Watt

Fellow coders:

Mr Mateusz Bieniek

Mr Matthew Wai Heng Chung

Ms Lisa Grant

Dr Anna Laddach

Mr Paul Smith

Happy hacking!

Website:

https://khsdtc.github.io/Hackathon_Autumn2019

Challenge:

https://github.com/KHSDTC/Hackathon_Autumn2019_Challenge