Co-ordinator: Ronny Mabokela  
  
- Understand your problem

Data needed:

1. Pictures
2. Text
3. Numbers (Source of data)

* How do you process the data (pre-processing) in order to get the machine working?

4 type machine learning

1. Supervised - Have the data
2. Unsupervised - Raw
3. Semi-supervised - Provide some data for some sections
4. Self-learning - They self-correct

Case scenario:

Email Validator: that that checks which email is fake and which is not real

* Annotate the data (give it a bunch of emails that are fake and not fake)
* Encode our data to be represented in numbers (tables)
* Machine works with 0’s and 1’s
* The machine needs training for it to know features (Differentiators). They need to be raw.
* After training, come up with a model.
* We need to split data into testing and training. Evaluate
* Evaluation: Accuracy below 50% is not accurate
* Confusion matrix, which part our AI can accurately extract features and which features confuse

Built machine learning algorithms that you can training your model:

Night base: Deals with probability – Given certain features, is the machine able to predict

Random forest

Decision tree

Active boost

Logistice regretions

Super vector

Binary clarification: True or False

Multi-class

Hosting machine learning

Axure

AWS

Horoco

Scikit-learn: Beginner website for machine learning

* Classification: Predicts which class a given sample of data is part off
* Regression: Predicts continuous values
* Clustering

<https://scikit-learn.org/stable/>

Install:

1. Python
2. Jupyter Notebook
3. Panda

Market basket system

Utf-8: Raw data

Download Cheat Sheet:  
Python

Data science

Install/import sklearn to get all libraries

* Pip install sklearn

What does Tfidf do?

We have Deep Learning Algorithm > Keras model