COMP 4448: DS TOOLS 2
Week 1 Live Session

NEBA NFONSANG

Welcome to Data Science Tools 2

Week 1 Objectives

- ▶ Self Introduction
- ▶ Syllabus
- ▶ Quick Python Review Resources
- ▶ Data Science, What really is it?
- ▶ Algorithms, why are they useful?
- ▶ Breakout Rooms: Week 1 activity
- ▶ Assignment 1
- ▶ Quiz 1 and 2

Activities for Breakout Rooms

▶ Write down 10 methods for solving problems in data science.

▶ How is data science related to machine learning and statistics.

▶ What is the difference between machine learning and statistics?

Some Methods used in Data Science

- Decision Trees
- Ensemble Methods
- KNN
- Naïve Bayes
- Linear Regression
- Logistic Regression
- SVM

- PCA
- LSA
- FA
- Clustering
- Neural Networks
- Deep Learning
- etc.

Quick Review of Assignment 1

Coming up next…!

Week 2: Quiz 1 Coverage

- What is data science?
- Variables and data types
- Datasets
- Supervised learning
- Unsupervised learning
- Algorithms
- Classification and regression

Week 2: Quiz 1 Coverage

- Data science workflow
- Data cleaning
- Data exploration
- Data quality
- Missing data mechanisms
- Handling missing data
- Outliers
- Data transformation

Coming up next…!

Week 3: Quiz 1 Coverage

- Model construction
- Model evaluation
- Model selection
- Model performance
- Bias vs variance trade off
- Overfitting, underfitting, optimal fit.
- Hyper parameters

Week 3: Quiz 2 Coverage

- Cross-validation
- Confusion matrix
- Sensitivity and specificity
- False and true negative rate
- False and true positive rates
- Overall accuracy
- Precision, recall, f-1 scores
- ROC curves