**DATA SCIENCE PROJECT 1 (MAST90106)   
Semester 1 2021**

**GROUP 17**

Meeting No. 11

**Date & Time:** 12th May 2021

**Agenda:** Project progress with Supervisor

**Attendees:** Kartika Waluyo, Ziad Al Bkhetan, Vrinda Rajendar Rajanahally

**Discussion:**

1. Project report - directions given by Michael and our plan
2. Training, Testing and Validation methods we can use
3. Discussions with clients - to always talk about the big picture and how each task is related to the the end goal of the project

**Next steps:**

1. Start discussing the project report outline

**Resources:**

1. Both cross validation and bootstrapping are resampling methods.  
     
   Bootstrap resamples with replacement (and usually produces new "surrogate" data sets with the same number of cases as the original data set). Due to the drawing with replacement, a bootstrapped data set may contain multiple instances of the same original cases, and may completely omit other original cases.   
     
   Cross validation resamples without replacement and thus produces surrogate data sets that are smaller than the original. These data sets are produced in a systematic way so that after a pre-specified number 𝑘 of surrogate data sets, each of the 𝑛 original cases has been left out exactly once. This is called k-fold cross validation or leave-x-out cross validation with 𝑥=𝑛𝑘, e.g. leave-one-out cross validation omits 1 case for each surrogate set, i.e. 𝑘=𝑛. As the name cross validation suggests, its primary purpose is measuring (generalization) performance of a model. On contrast, bootstrapping is primarily used