**DATA SCIENCE PROJECT 2 (MAST90107)   
Semester 2 2021**

**GROUP 17**

Meeting No. 10

**Date & Time:** 24th Sept 2021, 10 AM - 11 AM

**Agenda:** Progress with XGB and Report

**Attendees:** Brendan Ansell, Roberto Bonelli, Kartika Waluyo, Vrinda Rajendar Rajanahally

**Discussion:**

1. Error 502 - Sensor overload, memory overload while running xgboost with all genes.

The initial models list is too big, and added chunks to it at every stage slows down the process

1. Fixes to overcome the problem

Divide the set into chunks that are identifiable to know what we are training. Choose to save each model (each chunk) as we go, named after each gene.

1. How to report the models - report the problems we faced in brief and then explain how the final version of the model overcomes all those problems
2. Intro to the project - initial aim can be added into future direction.

**Resources/Next steps:**

1. Send pieces of the report to BA/RB so that they can review as we go.
2. The dataset has to strictly always be formatted as follows: Donor ID (rows) and Gene Ensembl ID (columns)