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

**GROUP 17**

Meeting No. 17

**Date & Time:** 21st June 2021 Monday, 11:30AM - 12:40PM

**Agenda:**

1. Discussion on Mid-Year Report and Improvements
2. Discussion on presenting to WEHI and CSL teams
3. Our experience with swapping responsibilities
4. Modelling and Analysis pipeline

**Attendees:** Kartika Waluyo, Vrinda Raj, Roberto Bonelli, Brendan Ansell

**Discussion:**

1. Discussion on Mid-Year Report and Improvements

* Provide new dates of next report submission in Sem2
* Reference images, by citing the URL   
  E.g.: ‘Image from: URL’
* Tables: All R tables need to be converted to properly formatted tables
* Nos. and Titles of every image must be correctly added and formatted to be visible.
* Front page needs editing

1. Discussion on presenting to WEHI and CSL teams

* Presenting to WEHI and CSL will entail: Covering what we’ve done so far and what we plan to do next
* Open the floor for Q&A and feedback towards the end
* Presentation: add more content and structure it better
* Plan and schedule a mock run for the presentation

1. Modelling and Analysis pipeline

* New data files have been added → where the cause of death have been cleaned out
* XGB modelling (Vrinda)
* How can we fix genes that are not captured by blood? Is there a way in RFs that can account for this missing feature(s)?

1. Neural Network modelling

* Describe Model1, Model2 used in the code
* Create another Model3 → compare it with the others later
* Does the model work if the batch size is big? Does it slow down the network?
* How do we tune a model such that the training set and validation set start to converge?
* Potential idea: should we make about 15-17 models with different parameters and layers and compare their performance?
* Grid search → algorithm in R that can be used to choose the best/optimal parameter. (Research this and see if it can be used)
* Tidy Models package

**Next steps:**

1. Work on presentation and send for feedback
2. Schedule a time for presentation to WEHI and CSL
3. Correct the report