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

**GROUP 17**

## Meeting No. 12

**Date & Time:** 8th October 2021, 10 AM - 11 AM

**Agenda:** Discussion of the report, corrections on the presentation

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

**Discussion:**

1. Biological Validation
2. Conclusion
3. Results - flow and structure of this section
4. Slides:

* New image for Gtex slide.
* Changes for the training, test and validation graphic.
* Include a sample size for each tissue in the cross validation slide.
* Modify the images in R, do not stretch them in ppt.
* Drop the legend at the bottom for the Correlation slides.
* Arrow in the neural network for the back propagation
* Drop Slide 13 for the Uni presentation
* Less text → the better
* Use text that is relevant to the project. (expression of gene x for tissue y)
* Slide 16 - explaining the training of the dataset over the best hyperparameters. The image shows the training that has occurred over 200 epochs, and shows that the MAE reduces over the epochs - for this slide, remove the legend and create your own. Add the formula for MSE for this slide
* TIP: present the xgboost first, then neural network.

**Resources/Next steps:**

1. Image for GTEX (use image B): <https://www.researchgate.net/figure/Sample-and-data-types-in-the-GTEx-v8-study-A-Illustration-of-the-54-tissue-types_fig1_336246636>
2. MSE derivation picture example<https://pascal-martin.netlify.app/post/2018-11-14-nicer-scatterplot-in-gggally_files/figure-html/unnamed-chunk-10-1.png>