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

**GROUP 17**

## Meeting No. 3

**Date & Time:** 6th August 2021, 10AM - 11AM

**Agenda:**

1. Progress and challenges with autoencoders and autokeras
2. Progress and challenges with XG Boost

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

**Discussion:**

1. Model using Auto encoders:

* To first focus on Autokeras
* To find the difference between autokeras and autoencoders

1. Model using XG Boost:

* Figure out how to use Thresholds
* We need to use regression to predict (many variables from many dependent variables)
* Look into Multi-labelled data

**Resources/Next steps:**

* <https://cran.r-project.org/web/packages/xgboost/vignettes/discoverYourData.html>
* <https://xgboost.readthedocs.io/en/latest/R-package/xgboostPresentation.html>
* Regression: <https://www.datatechnotes.com/2020/08/regression-example-with-xgboost-in-r.html>
* Using toy data (about 100 entries) and try fitting a regression prediction XGB model on the training data → then use data that has missing values and see how the model performs
* <https://blogs.rstudio.com/ai/posts/2019-04-16-autokeras/>
* A simple structure to try:

Input -> 1000 -> 500 -> 1000 -> outputModels for correlated input and correlated output:

Auto encoders & Recurrent Neural NetFind Hyper-parameters

AutoML -> automates the architecture of models -> Autokeras