**Programming Assessment (Python/R)**

1. **Data description**:

The Garment Industry is one of the key examples of the industrial globalization of this modern era. It is a highly labour-intensive industry with lots of manual processes. Satisfying the huge global demand for garment products is mostly dependent on the production and delivery performance of the employees in the garment manufacturing companies. So, it is highly desirable among the decision makers in the garments industry to track, analyse and predict the productivity performance of the working teams in their factories. (UCI public dataset)

Description Link: https://archive.ics.uci.edu/ml/datasets/Productivity+Prediction+of+Garment+Employees

Raw Datasets:

https://archive.ics.uci.edu/ml/machine-learning-databases/00597/

**Target Condition:** predict the productivity range (0-1)

2. **Regression Algorithms** (pick 3 out of 11)

* Linear Regression
* Ridge Regression
* Stepwise Regression
* Neural Network Regression
* Lasso Regression
* Decision Tree Regression
* Random Forest
* KNN Model
* Support Vector Machines (SVM)
* Facebook Prophet
* Name your own

3. **Procedures** (include full scripts with necessary comments, citations and a final report is possible)

* Data cleaning and validation
  + Please include a data description after preprocessing (e.g. data dimensionality and etc.)
* Outlier and missing value detection (if possible)
  + Please explain what outlier detection methods/approaches you used
  + Please explain how you handle missing values
* Feature Selection and Model fitting
  + Please explain how you split the datasets
  + Any feature selection methods used
  + Any specific packages used to fit the models
* Regression prediction
  + Please explain why and how you applied each method
* Results comparison and visualization
  + How you calculate the accuracy? Explain the evaluation metrics you used
* Final report (doc, pdf, html, rmd or etc.)