# DBA3803/DSC3216: PREDICTIVE ANALYTICS IN BUSINESS

NUS Business School      Department of Analytics & Operations (DAO)

Location: BIZ1-0202
Time: Wednesday noon - 3:00 p.m. (**A1**) & 6:30 p.m. - 9:30 p.m. (**A2**)

## Administrative Information

- Instructor: Dr. Long ZHAO
    - Office: BIZ1 8-61
    - Office Hours: Thursday 3:00 p.m. - 5:00 p.m. (online and offline)
    - Email: longzhao@nus.edu.sg

- TA: Atlanta CHAKRABORTY
    - Email: atlanta@u.nus.edu

- Prerequisite: DAO2702
    - I will review linear algebra (matrix multiplication) and normal distribution.

- Evaluation:
    - Individual Assignments: 20%
    - Four Group Projects: 10% + 10% + 10% + 20%
        * Final (last) project requires both report and e-presentation. Others only require the report.
    - Two Quizzes: 7.5% + 12.5%
    - Participation: 10%

- Coding: R or Python
    - No prerequisite on the skill level. Individual assignments will be challenging but manageable for a rookie coder.
    - It is acceptable to use one of R or Python mainly.
    - One small catch. You will learn how to read and write simple code in the other.

## Course Outline and Schedule

This course aims to develop an understanding of forecasting methods from data science for analyzing complex issues and solving business problems. We will make productive use of analytics tools available in R and Python. Because these packages are mature and convenient, we will instead focus on the thinking behind the methodology, which also applies to more advanced tools. Moreover, we will also learn the limitations of forecasting methods and common illusions in predictive analytics. Although the class focuses on simplified models, it aims to bridge the classroom knowledge and business applications, such as portfolio construction, e-commerce, and smart city operations.

| Week | Date | Topic | Remark |
|---|---|---|---|
| 1 | Aug. 12th | Introduction to Data Science & Tips for Coding | - |
| 2 | Aug. 19th | Linear Regression and Model Complexity | Assignment 1 Due |
| 3 | Aug. 26th | Model Selection and Regularization | Assignment 2 Due |
| 4 | Sep. 2nd | Project 1: Portfolio Optimization | Assignment 3 Due |
| 5 | Sep. 9th | Decision Trees and Enhancement | Project 1 Due |
| 6 | Sep. 16th | Smart Choice of Objective & Project 2 | Assignment 4 Due |
| - | Sep. 23rd | Recess Week | - |
| 7 | Sep. 30th | Linear Classification | Project 2 Due & In-class Quiz 1 |
| 8 | Oct. 7th | Tree Based Methods for Classification | Assignment 5 Due |
| 9 | Oct. 14th | SVM & Project 3 | Assignment 6 Due |
| 10 | Oct. 21st | Case Study or Extended Topic 1 | Project 3 Due |
| 11 | Oct. 28th | Introduction of Final Project | TBD |
| 12 | Nov. 4th | Case Study or Extended Topic 2 | TBD & In-class Quiz 2 |
| 13 | Nov. 11th | Final Project e-Presentation | Final Project Due |

Extended topics include the following:

1. Brief introduction to causal inference.
2. Time-series prediction.
3. Brief introduction to unsupervised learning.
4. Brief introduction to reinforcement learning. (The intelligent behind AlphaGo)

## Logistics

Due to uncertainty brought by COVID 19, we have two plans. Plan A is the default, while plan B is the back-up.

- Plan A: Hybrid.
    - Based on the enrollment prediction, there will be two cohorts.
    - Each week, only one cohort could (**not required**) come to the classroom.
    - For example, cohort 1 could come in the first week while cohort 2 in second.
    - Accommodation is possible. Please let me know ASAP.
    - The students in the classroom are **required** to join the online zoom session. They should also ask questions via their microphones to let the online students hear.
    - The zoom meeting will be recorded.

I will send out the cohorts information on **Aug 10th**. Please

- Plan B: E-learning.
    - I will teach **synchronized** online at the class time. Namely, you are expected to show up online.
    - The zoom meeting will be recorded.

- Zoom related:
    - You shall see the zoom link for each class in LumiNUS.
    - I highly recommend **not** show your face. This is due to privacy reasons because the meeting will be recorded.
    - I might not be able to notice the questions online in time. To compensate this, I will stop every 15-30 minutes to check zoom.

- Because I am awful at reading questions in zoom chat, I hope that you could raise your hand in zoom first. Then with my permission, you could unmute yourself and ask the question directly.
- If you are too shy to ask, as I used to be, we could have a one-on-one meeting after class or use emails.

## Resources

- Datacamp.com may be the best way to start learning data science coding.
  - Here is the invitation link which is valid starting August 2nd. You should be able to use DataCamp for free with your **NUS email**.
  - If you cannot join it, please let me know ASAP. All our individual assignments are from DataCamp.

*Optional Books & Videos:*

- [ISLA] An Introduction to Statistical Learning with Applications in R
  - Here are the videos, Youtube: StatsLearning.

- [DSB] Data Science for Business

- [LFD] Learning From Data
  - Here are the videos, Youtube: Caltech's Machine Learning Course.
  - The videos are fantastic, but they are too high-level for our course.

## Individual Assignments

All of the individual assignments come from DataCamp. The majority of the time, there will be two parallel tracks for R and Python. That is to say, if you use Python, you could ignore the corresponding R courses. Meanwhile, when it is only available in one language, you have no choice but to finish it.

Each assignment contains **2-3** mini-courses or projects which take a rookie who is not familiar with the **coding basics** 6-8 hours while a guru 2-3 hours. The grade of individual assignments (in this course) is defined as

$$\text{grade} = 70\% \times \text{completion} + 30\% \times \text{accuracy}.$$

For example, if you finished one mini project with 80% accuracy. Then your grade will be $70\% \times 100\% + 30\% \times 80\% = 94\%$.

## Four Group Projects

Because you will mainly work with strangers in a company, to enhance your abilities in such cooperation, **I will 'dictate' the groups.** The size of a group is 3 (preferred) or 4. Each group will be within the same cohort such that it is possible to discuss face-to-face after class. I have an algorithm to generate groups such that one will work with different people across projects.

I will first introduce each project during class and then provide a brief discussion. I expect all of them are challenging for the following three reasons.

- All have the flavor of open-ended questions. Just copying what I have done in the class will not be enough. You need to think and analyze the problems.
- All have unique characteristics that make it special. That is to say, you need to have a customized approach.
- It is always hard to write code from scratch.

## Two Quizzes

Quizzes will focus on conceptual arguments. They will be available on LumiNUS during the class time.

## Participation

I will use zoom attendance records. I agree that it is quite extreme because if you come to the classroom without joining zoom, you did not get a score. However, this might lead to a better online experience. In my opinion, helping the disadvantaged is very valuable.

## Questions

For questions regarding course materials, please post on LumiNUS Forum. For questions in emails, I might re-post them and answer them on the Forum. I might also create a FAQ on Google Docs as a more organized approach.

For issues in course administration, please email with the subject "DBA3803/DSC3216+Name".