DBA3803/DSC3216: PREDICTIVE ANALYTICS IN BUSINESS

NUS Business School   Department of Analytics & Operations (DAO)

08 June, 2021

Administrative Information

- Instructor: Dr. Long ZHAO
    - Office: BIZ1 8-61
    - Office Hours: By Appointment
    - Email: longzhao@nus.edu.sg

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

- Evaluation:
    - Individual Assignments: 20%
    - Three Group Projects: 12.5% + 12.5% + 25%
        * Final (last) project requires both report and e-presentation. Others only require the report.
    - Two Quizzes: 12.5% + 17.5%

- Coding: R or Python
    - No prerequisite on the skill level. Individual assignments will be challenging but manageable for a rookie coder.
    - **Because I only use R, if you choose Python, I might not be able to help you effectively. For example, the coding examples for projects are written in R.**
    - **Python does not handle categorical variables well. This means that you need to write more codes to preprocess data.**
    - 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. 11th	Introduction to Data Science & Tips for Coding	-
2	Aug. 18th	Linear Regression	Assignment 1 Due
3	Aug. 25th	Linear Regression 2 & Bias-Variance Tradeoff	Assignment 2 Due
4	Sep. 1st	Regularization: Lasso, Ridge, PCR	Assignment 3 Due
5	Sep. 8th	Project 1	Assignment 4 Due
6	Sep. 15th	Trees and Enhancement & Quiz 1 Review	Project 1 Due
-	Sep. 22nd	Recess Week	-
7	Sep. 29th	Classification 1 & AdaBoost	In-class Quiz 1
8	Oct. 6th	Gradient Boosting & Logistic Regression	Assignment 5 Due
9	Oct. 13th	Project 2	Assignment 6 Due
10	Oct. 20st	SVM & Overfitting in Validation	Project 2 Due
11	Oct. 27th	Smart Choice of Objective & Time-series & Quiz 2 Review	Final Project Proposal & Assignment 7 Due
12	Nov. 3rd	Unsupervised Learning	In-class Quiz 2
13	Nov. 10th	Final Project e-Presentation	Final Project Due

Resources

- Datacamp.com may be the best way to start learning data science coding.

    - You should be able to use DataCamp for free with your **NUS email** (domain: @u.nus.edu).
    - 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

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

Here the XP percentage is calculated as your **new** XP divided by the XP target which is **95%** of the total XP. This 5% gap buffers for evaluation mistakes, DataCamp gliches, and too technical problems.

For example, if you complete all assignments and achieve 80% as the XP percentage, then your grade will be $70\% \times 100\% + 30\% \times 80\% = 94\%$.

For students who have completed some DataCamp courses before, you need to earn **950 XP** for each completed course. However, there is no requirement on the courses/projects you complete. I hope you could share the good ones with me.

Three 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 use the same programming language.* 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.

Presentation(Pre-recorded Video): the final project requires both report and e-presentation. Each group will have 15-min e-presentation. Please video record the presentation and upload to LumiNUS. Explain the problem, managerial decisions you are making, data source, approaches, evaluation and insights. The content of the presentation should highlight the key findings.

Two Quizzes

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

If you could not attend the class synchronized, please let me know ASAP.

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".

FAQ

1. I have completed some assigned DataCamp courses before but I am not satisfied with my XPs. What should I do?

   - You might earn more XPs by doing DataCamp projects.

2. If I choose Python track, could I also access R courses?

   - You could access any DataCamp course and project for free and learn XP by doing it. However, your completion score is only determined by the assigned ones.

3. My DataCamp code is identical to the solution, but DataCamp considers it wrong. What should I do?

- It also happens to Long. Thus, the XP target is 95% of the total XP to compensate the existence of such thing.

4. I am also interested in other DataCamp courses, could I use hints freely for other courses?

   - Yes, you could use hints without hurting your grades of individual assignments. However, there is a risk that the courses you do will be assigned as homework in the future. Thus, try not use hints too aggressively.

5. Is my DataCamp permanent?

   - Yes, it is permanent but the premium only lasts 6 months. The starting date is Aug. 2nd, 2020.