# DBA3803: PREDICTIVE ANALYTICS IN BUSINESS

NUS Business School
Department of Analytics & Operations (DAO)

## ADMINISTRATIVE INFORMATION

| | | | |
|---|---|---|---|
| **Instructor:** | QUEK Ser Aik | **Office:** | BIZ1 7-44 |
| **Consultation Hours:** | As per link from Canvas | **Email:** | seraik@nus.edu  ser@u.nus.edu  seraik@trinity.cantab.net |
| **Co-Instructor:** | | **Email:** | |

**Prerequisite**: DAO2702 Programming for Business Analytics.  An innate love of programming is necessary.  It will not be possible to contrive a liking for attention to details—such focus is needed to enjoy and master the software coding involved.  You need to be habitually and reflexively asking, "I wonder what will happen if I change this part…"  You must be logical and love to tinker or expect poor module result.

**Evaluation**: Assignment 30%, Quiz 30 %, Project 30%, Class Participation 10%

**Text**: Class notes will stand in for a textbook.

**Optional References**:

A Whirlwind Tour of Python, by Jake VanderPlas.
https://s3-us-west-2.amazonaws.com/python-notes/a-whirlwind-tour-of-python-2.pdf
Classification and Regression: In a Weekend, by Ajit Jaokar & Dan Howarth.
https://www.darwinpricing.com/training/Classification_and_Regression.pdf
The Elements of Statistical Learning, by Hastie, Tibshirani and Friedman (more advanced).
https://web.stanford.edu/~hastie/ElemStatLearn

## COURSE OUTLINE

The objective of this course is to build on prior Python knowledge to develop an understanding of data science with an emphasis on forecasting as a common tool for analyzing complex issues and solving business problems. We will make productive use of analytics tools available in Python. While the class focuses on simplified models, it aims to bridge classroom knowledge and business applications.  Topics might be dynamically adjusted or accelerated, depending on students' rate of learning.

References are to the optional Hastie et al, unless otherwise indicated.

| Week | Topic | Reference | Remark |
|---|---|---|---|
| 1 | Introduction to Machine Learning | Chapter 1 | |
| 2 | Python Programming Revision | Whirlwind Tour | |
| 3 | Predicting Numerical Values: Linear Methods for Regression 1 | Section 2.3, 3.1-3.3 | ASGN1 Due |
| 4 | Predicting Numerical Values: Linear Methods for Regression 2 | Section 3.4-3.5 | |
| 5 | Finding Temporal Patterns: Time Series Forecasting 1 | Notes | Quiz 1 |
| 6 | Finding Temporal Patterns: Time Series Forecasting 2 | Notes | |
| 7 | Predicting Categories: Neighbors and Clusters | Section 4.1-4.3 | ASGN2 Due |
| 8 | Predicting Categories: Classification Tree | Section 9.3 | |
| 9 | Understanding Model Complexity and Over-Fitting | Section 7.1-7.6 | ASGN3 Due |

| 10 | Resampling Methods 1 | Section 7.10-7.11 | |
|---|---|---|---|
| 11 | Resampling Methods 2 | As above | Quiz2 |
| 12 | Other Machine Learning Methods & Project Discussion | | |
| 13 | Project Presentations & Peer Assessments | | Project Due |

## ASSIGNMENT (30%)

Assignment will generally be due on Saturday before 2359. Everyone should turn in **individual soft copies** via Canvas (unless otherwise stated).  Late assignment submission will result in a lower grade.

## QUIZ (30%)

There will be 2 quizzes. The quiz will be cumulative and comprehensive. They will be Open book. The quiz duration will be 40 minutes in class.  Dates will be on LumiNUS' Module Details — Assessment.

Our School mandated that absence from any quiz without valid excuse may result in the student receiving a zero mark for that quiz, which may affect the final grade.

## TEAM PROJECT (30%)

The project for this class is to do a careful data analysis of sizeable real-world data.  A dataset will be assigned.  Data preparation, analytic modelling, and an executive summary will be assessed.

The project is a group effort, to be done in groups of 4-5 students, depending on class enrollment. Each group member will evaluate the participation of the other members of the group, and this will be considered in grading.

Assessment of the group project will include the following aspects:
   a. Innovation: are the identified issues interesting?
   b. Rigor: are the methods used appropriate?
   c. Reporting: will executives be attracted to read the report *Introduction* and then, further, its *Conclusion*?

Everyone will also evaluate other groups' project presentations, if any.

**Team Project Part I – Presentation (10%):**
Each group will have 5 to 10 minutes of technical-cum-managerial presentation that might be replaced by a recorded video if the learning environment is constrained.

**Team Project Part II – Final report (20%):**
The final report should contain the main results and details of the project, with 2 out of 3 parts targeted at non-technical and out-of-context executives.

## ATTENDANCE

If there is more than one weekly session, students may always attend either session without notice, except when there is a quiz or project presentation.  Attendance will count towards Class Participation.

## QUESTIONS

For questions regarding course materials, please post on Canvas. Posting and answering on Canvas will be counted towards Class Participation.  Questions during consultations may be re-posted and answered on Canvas.

## LEARNING OUTCOME

It is hoped that students will habitually dream along the data dimension for future workplace and personal decisions.  It is intended that graduates will continue to prepare to lead in expected digital conversations.

## INSTRUCTOR PROFILE



https://linkedin.com/in/seraik
https://alum.mit.edu/www/aik
https://cal.berkeley.edu/aik

QUEK Ser Aik was trained in Statistics that transitioned naturally into machine learning and artificial intelligence.  He always hopes to make data evangelists out of students, who sometimes refer to him jokingly as the "short old man" (which is correct on all 3 counts, but not as succinct as just "aik" 😊).

He loves to attract students to be sharing members of https://fb.com/groups/OpsBA to continue their analytics journeys together.