**Course Outline for BC2406 Analytics I: Visual & Predictive Techniques**

Pre-requisite  :  AB1202 Statistical & Quantitative Methods, or equivalent.

Important  :  Students to bring their laptop with R Studio installed in every class.

## Course Description

Most organizations are data rich and information poor. The large volumes of data in an organization are "oilfields" rich in information content that are pending extraction with the right tools and models. Analytics involves the art of data exploration, visualization, communication and the science of analyzing large quantities of data in order to discover meaningful patterns and useful insights to support decision-making. The primary objective of this course is to introduce students to various techniques available to extract useful insights from the large volumes of data. At the end of the course, students will not only see the substantial opportunities that exist in real world, but also learn techniques that allow them to exploit these opportunities. This course focus on the use of open source R software, which is one of the key analytics software used in various industries and a critical skillset required in the job market for analytics and data science professionals.

## Course Learning Objectives

At the end of the module, students will:

1.  Identify aspects of business problems that could be fruitfully solved by predictive techniques.

2.  Apply selected predictive techniques to solve the business problem.

3.  Explain the results of the selected predictive techniques in the context of the business problem.

4.  Evaluate the performance of the predictive techniques.

5.  Propose business solutions based on the results of the predictive techniques.

## Learning & Teaching Methods

Duration: Approx 2.5 hrs of readings and exercises at home and 1.5 – 2 hrs max of face-to-face class session (Physical or online) per week.

Refer to weekly checklists for the pre-class readings and exercises. These must be completed before attending face to face class sessions. Please bring your laptop with RStudio installed in every class.

**Main Textbook**

Chew C. H. (2020). Artificial Intelligence, Analytics and Data Science, Vol 1: Core Concepts and Models. Cengage.

**Supplementary Textbooks**

Sanchez (2018). Handling Strings with R. eBook: https://www.gastonsanchez.com/r4strings/

Siegel and Robinson (2018). Text Mining with R. O'Reilly. eBook: https://www.tidytextmining.com/

It is important to complete the weekly pre-class readings and complete as much as you can, the pre-class exercises. Use the online discussion forum to ask questions and unblock your learning. The 90 mins face to face class session (physical or online) will focus only on selected advanced concepts, extension of exercises and discussion of exercise solutions.

*"I hear and I forget.*
*I see and I remember.*
*I do and I understand."*
*--- Confucius (孔子)*

## Course Schedule

| # | Topic & Deadlines | Content | Remarks |
|---|---|---|---|
| 1 | Course Overview and Introduction to Analytics | **Course Introduction**<br>• Course Policy and Assessment.<br>• Analytics Opportunities & Applications.<br>**Software Introduction**<br>• R & R Studio installation. | View online interactive video Analytics Opportunities & Applications, at home.<br><br>No class. |
| 2 | Fundamental Analytics Concepts and Industry Practice | **Concepts**<br>• Data-Analytic Thinking<br>• Fundamental Concepts<br>**R Exercise**<br>• Importing & Exporting Data<br>• Errors Culture & Corrections | Read Main Textbook Chap 2.<br><br>Physical class in NTU. |
| 3 | Data Exploration and Summaries | **Concepts**<br>• Exploring Data<br>• Useful Statistics<br>**R Exercise**<br>• Statistical Functions<br>• Slice, Dice & Transform Data | Read Main Textbook Chap 3.<br><br>Online Zoom Class. |

| | | | |
|---|---|---|---|
| 4 | Data Structures and Visualization | **Concepts**<br>• Data Types & Structure<br>• Useful Visualization<br>**R Exercise**<br>• Data Visualization<br>• Variable Data Structures | Read Main Textbook Chap 4.<br><br>Physical class in NTU. |
| 5 | Data Cleaning and Preparation<br><br>*Assignment submission deadline on Sunday 11pm.* | **Concepts**<br>• Data Cleaning<br>• Data Imputation<br>**R Exercise**<br>• Best Practices<br>• Verifying data changes. | Read Main Textbook Chap 5.<br><br>Online Zoom Class. |
| 6 | Linear Regression | **Concepts**<br>• When to use Linear Reg.<br>• Model Assumptions<br>**R Exercise**<br>• Model Coefficients<br>• Regression Diagnostics<br>• Train-Test Split | Read Main Textbook Chap 6.<br><br>Physical class in NTU. |
| 7 | Logistic Regression | **Concepts**<br>• When to use Logistic Reg.<br>• Logistic Function<br>• Relative Risks & Odds Ratios<br>**R Exercise**<br>• Model Coefficients<br>• Odds Ratios & Odds Ratio CI | Read Main Textbook Chap 7.<br><br>Read Freitas et. al. (2012) Factors Influencing Hospital High Length of Stay Outliers.<br>Online Zoom Class. |
| *Break Week* | | |
| 8 | Decision Tree (CART Part 1) | **Concepts**<br>• Node Impurity<br>• Tree Pruning<br>**R Exercise**<br>• 10 fold Cross Validation<br>• Automated Decision Rules | Read Main Textbook Chap 8.<br><br>Physical class in NTU. |
| 9 | Decision Tree (CART Part 2) | **Concepts**<br>• Regression Tree<br>• Surrogates<br>**R Exercise**<br>• rpart(…, method = "anova").<br>• Surrogates. | Read Main Textbook Chap 8.<br><br>Physical class in NTU. |
| 10 | Clustering for Analytics<br><br>*E-Learning week. To be Confirmed.* | **Concepts**<br>• Clusters<br>• Similarity, Neighbourhood.<br>**R Exercise**<br>• Clustering<br>• Improving Model Accuracy | Refer to eLearning instructions & resources provided.<br><br>No Class. |

| 11 | Text-mining<br><br>*Project submission deadline on Sunday 11pm.* | **Concepts**<br>• Strings<br>• Bag of Words<br>• Tokens<br>**R Exercise**<br>• string functions<br>• stringr package<br>• Text Mining package<br>• Sentiment Analysis | Read:<br>- Handling Strings with R Chap 6 & 7.<br>- Main Textbook Chap 10.<br><br>Physical class in NTU. |
|---|---|---|---|
| 12 | Project Presentation (part 1).<br><br>*CBA question paper release on Monday 12 noon.* | Some Teams present project in class if physical class not cancelled.<br><br>CBA question paper release after Project submission. | According to Team presentation schedule randomly assigned.<br><br>Physical class in NTU. |
| 13 | Project Presentation (part 2).<br><br>*CBA submission due on Monday 12 noon.* | Other Teams (if applicable) presents project in class if physical class not cancelled. | According to Team presentation schedule randomly assigned.<br><br>Physical class in NTU. |

## Instructors

| Class | Instructor | Office | Email |
|---|---|---|---|
| 1, 4, 6. | Neumann Chew C.H. (Course Coordinator) | S3 B2C 104 | neumann.chew@ntu.edu.sg |
| 2, 3, 5. | Hyeokkoo Eric Kwon | S3 B2B 71 | eric.kwon@ntu.edu.sg |
| 7 | Pinar Darendeli | TBC | sp.darendeli@ntu.edu.sg |

*For consultation with your instructor, please email to arrange. Due to Covid-19, this might be online.*

## Course Assessment

| # | Components | Weightage | Type |
|---|---|---|---|
| 1 | Participation & Individual Presentation | 30 | Individual |
| 2 | Assignment | 10 | Team |
| 3 | Project | 30 | Team |
| 4 | Computer Based Assessment | 30 | Individual |
| | **Total** | **100** | |

**Assessment Measures**

Class Participation
- Critical Thinking

Individual Presentation
- Critical Thinking

Assignment
- Problem Solving & Decision Making

Project
- Written & Oral Communication

Computer Based Assessment
- Problem Solving & Decision Making

**Assessment Rubric**

Critical Thinking:
- Identifies and summarize the issue at hand.
- Identifies and considers key assumptions and the influence of the context on the issue.
- Identifies and assesses conclusions, implications and consequences.

Problem Solving & Decision Making:
- Define the Problem.
- Devise Strategies to Solve the Problem.
- Evaluate Outcomes.

Written & Oral Communication:
- Presents relevant information.
- Substantiates claims with strong evidence.
- Has a clear message for audience.
- Maximizes likelihood of audience accepting the message.

*Note: This is just a summary list for assessment. Refer to Assessment Components and Rubrics PDF for detailed scale of measurements.*