National University of Singapore
School of Continuing & Lifelong Education (SCALE)

TBA2104 Predictive Analytics
Group Project

To study and research an interesting business analytics problem. In your group, you are to discuss and decide on an interesting problem that requires some form of predictive analytics. A list of possible topics can be found at the end of the document, but you are encouraged to propose your own topic.

Tasks

- Form groups of 2-4.
- Choose an interesting business analytics problem.
- Submit a project proposal so that the teaching team can better advise you.
- Apply the CRISP-DM process, focusing on the first 5 steps:
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modeling
  - Evaluation
- Prepare the presentation slides for the project presentation.

Deliverables

1. Project Proposal
2. Presentation

Note: The group presentation will happen during the last instructional week.

Project Proposal

- Decide on a topic. (Refer to the bottom for a list of possible topics or suggest your own topics)
- Choose a group name and indicate the members of the group on the project proposal.
- Write a **1-page** project proposal describing the topic you have chosen.
- **One** person in the group is supposed to submit the proposal in pdf format to workbin. (Deliverables Submission > Project Proposal folder)
- Upload the softcopy of the proposal by **Fri, 18 Feb 2022 2359hrs (Week 6)**.
- The proposal is worth 5% of the course grade.

Presentation

- Each group is supposed to give a 15 minutes presentation + 5 minutes Q&A on week 13.
- Please prepare the presentation slides and upload the slides to workbin after your presentation.
- **All members of the group should present.**
- Your presentation should include the following (where applicable):
  1. Introduction
     - Introduce what this project is about.
  2. Problem/Data
     - Describe the business domain and the data used.
  3. Analysis Process
     - Describe the steps to undertake this project.
     - Discuss any challenges and the steps taken to address these challenges (if applicable).
  4. Data Exploration
     - Describe the various data exploration tasks performed (e.g. data visualization, descriptive analysis, etc).
  5. Data Preparation

- Describe the data preparation process such as data cleaning (if applicable).

6. Experiments
    - Describe the modeling techniques used (e.g. regression, classification, clustering).
    - Compare and contrast different modeling algorithms.
    - Explain and discuss the experimental results.
    - Discuss any limitations of the approach.
7. Discussion and Conclusion
    - Discuss any learning points.
    - Suggest possible future improvements.

Evaluation Criteria

- Weightage: 30% of the course grade
    - 5% project proposal
    - 10% presentation
    - 15% analysis

Consultation

- Throughout the semester, you are encouraged to see the lecturer regularly to ensure that you are on the right track and to feedback on any problems you are facing.
- We will guide you along if you encounter any problems and will provide as much help as possible.

Data

- You will be required to upload the data used in the experiment (or provide the download link to the data if the data is too large for uploading)

Topics

Since we will be covering mainly regression, classification, and clustering in this course, the following suggested topics will be along these three main areas. **You are free to propose any interesting topics and to use any techniques not covered in the course. Try to propose a project that has practical applications.** When deciding the project topic, you might also want to take a look at websites such as kaggle.com for datasets you can use for the project.

Email Spam Classification

Build a system/propose an approach to classify emails to be either spam or not spam.

HDB Pricing Prediction System

Build a system/propose an approach to predict house pricing.

Movie Rating Prediction System

Build a system/propose an approach to predict the rating of a movie.

Election Prediction System

Build a system/propose an approach to predict which candidate will win an election.

HR Analytics

Predict whether someone is likely to resign.

Forecasting/Prediction Systems

Build a system/propose an approach that will perform forecasting/prediction.

Classification Systems

Build a system/propose an approach that will perform classification.