Text as Data: Text Mining Techniques using R

At the end of the course, participants will be able to:

- How to tidy unstructured text data into a structured dataset
- Execute and cross-validate the results from text classification using supervised learning
- Perform topic modelling using Latent Dirichlet Allocation in Natural Language Processing
- Internalise integration of supervised and unsupervised learning algorithms and techniques within the workflow of analysis of text data

Course objectives

From social media to product reviews, textual data is replacing other forms of unstructured data due to how inexpensive and current it is. To draw meaningful insights, text mining deals with helping computers understand the "meaning" of the text and some of the common text mining applications include sentiment analysis and text classification.

Course outline

This course provides guidance to wrangle text data using the principles of Tidyverse and execute both supervised and unsupervised machine learning techniques, as learnt from previous modules, and analyse text data and cross-validate analyses.

Course details
1 week

Certificated by Singapore Management University (SMU)

Who should attend
- Anyone whose work interfaces with data analysis who wants to learn key concepts, formulations, algorithms, and practical examples of what is possible in machine learning and artificial intelligence
- Managers who need the vision and understanding of the many opportunities, costs, and likely performance hurdles in predictive modelling, especially as they pertain to large amounts of textual (or similar) data
- Professionals looking for a deeper understanding and hands-on experience with SMU adjunct faculty and industry expert

Pre-requisites
No prior experience or background required

Tools
Latent Dirichlet Allocation

Model of training
Classroom,
Field trip