IMIx
IMI EXECUTIVE PROGRAMS

Institute for Management & Innovation
UNIVERSITY OF TORONTO
MISSISSAUGA

IMIx Course Overview

Title: Supervised & Unsupervised Machine Learning

Length: 1 day

Rationale/purpose of the course:

- What does the course/module seek to achieve?
    - We will discuss step-by-step hands-on solutions using Python with a free software machine learning library (Scikit-learn) for real-world business problems using widely available data mining techniques in the area of supervised and unsupervised learning applied to real-world data sets.

- Why is the course important?
    - Corporations and institutions worldwide are learning to apply data mining and predictive analytics, in order to increase profits. A significant constraint on realizing value from big data is a shortage of talent, particularly of people with deep expertise in statistics and machine learning, and the managers and analysts who know how to operate companies by using insights from big data.

- Who should attend?
    - Managers, data analysts, and others who need to keep abreast of the latest methods for enhancing return on investment.

Instructor: Gerhard Trippen, Associate Professor, Operations Management

Issues/Topics to be covered:

- Perhaps the most common data mining task is that of *classification*. In classification, there is a target categorical variable, (e.g., income bracket), which is partitioned into predetermined classes or categories, such as high income, middle income, and low income. The data mining model examines a large set of records, each record containing information on the target variable as well as a set of input or predictor variables. We would like to be able to classify the income bracket of persons not currently in the database, based on the other characteristics associated with that person, such as age, gender, and occupation.

- In *estimation*, we approximate the value of a numeric target variable using a set of numeric and/or categorical predictor variables. Models are built using “complete” records, which provide the value of the target variable, as well as the predictors. Then, for new observations, estimates of the value of the target variable are made, based on the values of the predictors.

- *Clustering* refers to the grouping of records, observations, or cases into classes of similar objects. A cluster is a collection of records that are similar to one another, and dissimilar to records in other clusters. Clustering differs from classification and estimation in that there is no target variable for clustering. The clustering task does not try to classify or estimate the value of a target variable. Instead, clustering algorithms seek to segment the whole data set into relatively homogeneous subgroups or clusters, where the similarity of the records within the cluster is maximized, and the similarity to records outside of this cluster is minimized.

Modes of instruction:

- Discussion
- Interactive Sessions including Coding

*This course is a requirement for the Certificate in Data Analytics