



Let's Reach For Excellence!

TAN DUC INFORMATION TECHNOLOGY SCHOOL JSC

Address: 103 Pasteur, Dist.1, HCMC

Tel: 08 38245819; 38239761

Email: traincert@tdt-tanduc.com

Website: www.tdt-tanduc.com; www.tanducits.com

Applied Unsupervised Learning with R

Duration: 02 days

Starting with the basics, Applied Unsupervised Learning with R explains clustering methods, distribution analysis, data encoders, and all features of R that enable you to understand your data better and get answers to all your business questions.

This course begins with the most important and commonly used method for unsupervised learning - clustering - and explains the three main clustering algorithms - k-means, divisive, and agglomerative. Following this, you'll study market basket analysis, kernel density estimation, principal component analysis, and anomaly detection. You'll be introduced to these methods using code written in R, with further instructions on how to work with, edit, and improve R code. To help you gain a practical understanding, the course also features useful tips on applying these methods to real business problems, including market segmentation and fraud detection. By working through interesting activities, you'll explore data encoders and latent variable models.

By the end of this course, you will have a better understanding of different anomaly detection methods, such as outlier detection, Mahalanobis distances, and contextual and collective anomaly detection.

AUDIENCE

Applied Unsupervised Learning with R is designed for business professionals who want to learn about methods to understand their data better, and developers who have an interest in unsupervised learning. Although the course is for beginners, it will be beneficial to have some basic, beginner-level familiarity with R. This includes an understanding of how to open the R console, how to read data, and how to create a loop. To easily understand the concepts of this course, you should also know basic mathematical concepts, including exponents, square roots, means, and medians.

AT COURSE COMPLETION

After completing this course, you will be able to:

- Implement clustering methods such as k-means, agglomerative, and divisive
- Write code in R to analyze market segmentation and consumer behavior
- Estimate distribution and probabilities of different outcomes
- Implement dimension reduction using principal component analysis
- Apply anomaly detection methods to identify fraud
- Design algorithms with R and learn how to edit or improve code

COURSE OUTLINE

1. Introduction to Clustering Methods

- Introduction
- Introduction to Clustering
- Introduction to the Iris Dataset
- Introduction to k-means Clustering
- Introduction to k-means Clustering with Built-In Functions
- Introduction to Market Segmentation
- Introduction to k-medoids Clustering

2. Advanced Clustering Methods

- Introduction
- Introduction to k-modes Clustering
- Introduction to Density-Based Clustering (DBSCAN)

3. Probability Distributions

- Introduction
- Basic Terminology of Probability Distributions
- Introduction to Kernel Density Estimation
- Introduction to the Kolmogorov-Smirnov Test

4. Dimension Reduction

- Introduction
- Market Basket Analysis

5. Data Comparison Methods

- Introduction
- Analytic Signatures

- Latent Variable Models – Factor Analysis

6. Anomaly Detection

- Introduction
- Univariate Outlier Detection
- Kernel Density