# Big Data & Data Science Program

# Diploma Courses

Date: March 2019– v 3.0

## Diploma Structure

The Big Data & Data Science Diploma requires the attendance of 4 courses and 1 hands-on group project according to the following structure:

**Semester #1 (2 Courses)**

1- Introduction to Big Data, Developing with Spark and Hadoop (42 Hours, 14 Lectures)

2- Introduction to Machine Learning and Statistical Analysis (42 Hours, 14 Lectures)

**Semester #2 (2 Courses)**

3- Advanced Big Data Analytics Technologies and Applications (42 Hours, 14 Lectures)

4- Only 1 of the 3 following courses:

- Practical Data Mining (42 Hours, 14 Lecture)

*OR*

- Practical Data Science Using Machine Learning Technique (42 Hours, 14 Lectures)

*OR*

- Selected topics in Deep Learning (42 Hours, 14 Lectures)

**Semester #3 (Final Project)**

- Hands-on group project based on real life use case (14 Weeks of Mentoring)

Please refer to ***Appendix A*** for the description of each of those courses.

## Important Notes

- All enrollments are subject to the admission rules and acceptance criteria of Nile University and the Big Data and Data Science Program.
- The default training location in Nile University premises and any change will be decided upon case by case by the program management team.
- Timing, lecture distribution, assigned instructors and schedules will be assigned and announced to students upon registration completion subject to Nile University and the program administrative decisions.
- The courses details and outlines might get changed due to continuous development and enhancements to cope with trending theories, technologies, methods and applications in this domain.

For more details and pricing, please contact us: bigdata@nu.edu.eg

# Appendix A: Course Descriptions

## CIT-651: Introduction to Machine Learning and Statistical Analysis (42 Hours, 14 Lectures)

### Description

This course provides an introduction to machine learning and statistical data analysis. The course provides an introduction to the basic probability theory, statistics, and statistical data analysis. Topics such as parameter estimation, hypothesis testing and regression analysis will be covered in the course. In addition, the course will focus on machine learning topics including: Bayes classifiers, K-nn, decision trees, SVM, K-means, principal component analysis, independent component analysis and Neural Nets.

During this course, students will learn how to solve real-life problems using state-of-the-art technologies developed for cloud-based machine learning computing and data analyses.

### Reference Textbook

Main textbook: “Pattern Recognition and Machine Learning," by Christopher M. Bishop; Springer, 2006

Other References:

“Applied Statistics and Probability for Engineers,” by Douglas Montgomery, George Runger; John Wiley, 2003

“Machine Learning: A Probabilistic Perspective”, by Kevin P. Murphy; MIT Press 2012

“Machine Learning,” by Tom Mitchell; McGraw Hill, 1997

### Course Outlines

Part 1: Introduction and Background

1. Introduction
    1.1. Applications
    1.2. Relation between Statistics and Learning
    1.3. Supervised, Unsupervised and Reinforcement Learning
2. Linear Algebra Review
    2.1. Vector and Matrix Operations
    2.2. Matrix Inverse and Decomposition
    2.3. The Eigenvalue Problem
3. Analysis Tools
    3.1. R-Programming
    3.2. Waikato Environment for Knowledge Analysis (WEKA)
    3.3. Azure Platform

Part 2: Statistical Analysis

4. Probability Theory Review
    4.1. Marginal and joint Probabilities
    4.2. Conditional Probabilities
    4.3. Bayes' Rule
    4.4. Prior and Posterior Probabilities
    4.5. Probability Distributions
    4.6. Expected Value, Variance and Covariance
5. Statistical Parameter Estimation:
    5.1. Types of Estimators
    5.2. Random Sampling of a Population
    5.3. Estimation of the Mean and Variance
    5.4. Detection of Outliers
    5.5. Data representation and Visualization
6. Hypothesis Testing
    6.1. Confidence Interval and p-value
    6.2. Alternative Hypotheses
    6.3. Z-test and T-test
7. Regression Analysis
    7.1. Assumptions of Linear Regression
    7.2. Simple Linear Regression
    7.3. Error Analysis

Part 3: Machine Learning

8. Linear Classification:
    8.1. Discriminant Functions:
        8.1.1. Discriminant Functions Properties
        8.1.2. Least Squares Classifier
        8.1.3. Fisher's Linear Discriminant
        8.1.4. Perceptron
    8.2. Probabilistic Generative Models:
        8.2.1. Maximum Likelihood Estimation of Gaussian Generative Model
        8.2.2. Naïve Bayes Classifier
    8.3. Probabilistic Discriminative Models:
        8.3.1. Logistic Regression
9. Non-linear Classification:
    9.1. Instance-based Learning:
        9.1.1. K-nearest Neighbor Classifier
        9.1.2. Cross-validation
        9.1.3. Weighted K-nearest Neighbor Classifier
    9.2. Support Vector Machines
    9.3. Decision Tree Learning
    9.4. Artificial Neural Networks:
        9.4.1. Network Architecture
        9.4.2. Back-propagation Learning
10. Introduction to Reinforcement Learning:
    10.1. Markov Decision Process
    10.2. Q-learning
    10.3. Non-deterministic Rewards and Actions