Arabian Institute For Training

Certificate in Data Science

Why Attend

- Certificate in Data Science will expose participants to Data Science best practices, introduce them to the essentials of the Big Data ecosystem and opportunities for Artificial Intelligence. It doesn't limit itself to analytics, but to all disciplines to which modern data relates to as well. By the end of this course, participants will become specialists in techniques and technologies that will allow them to get meaningful knowledge from their data, and deal professionally with experts in all advanced data management fields.

Course Methodology

- All analytical methods and solutions are elaborated with step-by-step case studies with practical, hands on experiences. An exhaustive documentation will cover analytical topics with an exclusive face-to-face comparison between SAS, SPSS, STATISTICA, Excel, R and Python.

Course Objectives

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

- Understand and design data for efficient analysis
- Compare solutions related to Data Analysis vs. Machine Learning
- Differentiate between predictive models and pattern finding ones
- Decide between "proprietary" and "open source" technologies
- Outline the modern data flow from sources to reports
- Manage Data Science projects with project management best practices

Target Audience

- This course is for specialists who aspire to become accustomed with data science components, and how they can be applied coordinately to solve data and business problems, as well as research issues. The course is specifically suited for managers and persons involved in marketing, CRM, research, manufacturing, quality control, app developers and IT analysts from almost any sector, such as banks, insurance companies, retail, governments, manufacturers, healthcare, telecom, transport and distributors.

Target Competencies

- Business data analysis
- Data analytic validity
- Judging AI algorithms
- Evaluating IoT platforms
- Comparing big data results

Data Analysis and Visualization

- Types of data and data visualization
- Evaluating the representative quality of data
- Using descriptive statistics to summarize data
- Profiling two or more groups with statistical tests
- Visualizing multiple analytics with powerful smart charts
- Simple Linear Regression
- Simple Logistic Regression
- Managing and removing outliers

Machine Learning – Supervised

- Multiple linear regressions
- Multiple logistic regressions
- Discriminant analysis: Functions and probabilistic models
- Decision trees: CART – CHAID and Random Forests
- Support vector machines
- K-nearest neighbors
- Naïve Bayes
- Neural networks, deep learning and AI possibilities

Business Intelligence Forecasting – R vs. Python

- Business Intelligence
- Databases: collection and sources
- ETL
- Storage: Data warehouses, data marts and data lakes
- Analytics: BI Tools, OLAP, Dashboards, etc.
- Forecasting
- Trends
- Exponential smoothing: Additive and multiplicative methods
- Time Series: Additive and multiplicative methods
- ARIMA models
- R vs. Python
- Statistical Tests
- Machine Learning algorithms

Machine Learning: Unsupervised

- Principle Component Analysis
- Clustering: Hierarchical and K Means
- Simple correspondence analysis
- Multi-dimensional scaling
- Quadrant analysis

PMP for Data Scientists

- PMP
- Integration, Cost, Scope
- Time, Cost, Quality, Communication
- Risk, Procurement and Stakeholders

IoT and Big Data Ecosystem

- IoT essentials - M2M and Embedded Systems
- Basic IoT protocols
- Big Data: "where" and "when"
- Big Data distributed files with HDFS
- MapReduce vs. Spark Data Sharing
- Big Data Ecosystem bird's eye view: Spark, Mongo DB, Cassandra, Flume, Cloudera, Oozie, Mahout

Arabian Institute For Training