Data Science and Big Data Analytics v2 (Exam: DEA-7TT2)

Course Outline

Big Data, Analytics, and the Data Scientist Role

- Define and describe the characteristics of Big Data

- Describe the business drivers for Big Data analytics and data science

- Describe the Data Scientist role and related skills

Data Analytics Lifecycle

- Describe the data analytics lifecycle purpose and sequence of phases

- Discovery - Describe details of this phase, including activities and associated roles

- Data preparation - Describe details of this phase, including activities and associated roles

- Model planning - Describe details of this phase, including activities and associated roles

- Model building - Describe details of this phase, including activities and associated roles

Initial Analysis of the Data

- Explain how basic R commands are used to initially explore and analyze the data

- Describe the most important statistical measures and effective visualizations of data

- Describe the theory, process, and analysis of results for hypothesis testing and its use in evaluating a model

Advanced Analytics - Theory, Application, and Interpretation of Results for Eight Methods

Describe theory, application, and interpretation of results for the following methods:

- K-means clustering
- Association rules
- Linear regression
- Logistic Regression
- Naïve Bayesian classifiers
- Decision trees
- Time Series Analysis
- Text Analytics

Advanced Analytics for Big Data - Technology and Tools

- Describe the technological challenges posed by Big Data

- Describe the nature and use of MapReduce and Apache Hadoop

- Describe the Hadoop ecosystem and related product use cases

- Describe in-database analytics and SQL essentials

- Describe advanced SQL methods: window functions, ordered aggregates, and MADlib

Operationalizing an Analytics Project and Data Visualization Techniques

- Describe best practices for communicating findings and operationalizing an analytics project

- Describe best practices for building project presentations for specific audiences

- Describe best practices for planning and creating effective data visualizations