

Course duration

- 5 days

Course Benefits

- Translate everyday business questions as well as more complex problems into Machine Learning tasks in order to make truly data-driven decisions
- Use Python Pandas, Matplotlib & Seaborn libraries to Explore, Analyze & Visualize data from varied sources (the Web, Word documents, Email, Twitter, NoSQL stores, Databases, Data Warehouses & more) for patterns and trends relevant to your business
- Train a Machine Learning Classifier using different algorithmic techniques from the Scikit-Learn library (e.g., Decision Trees,
- Logistic Regression, Neural Networks)
- Re-segment your customer market using K-Means & Hierarchical algorithms for better alignment of products & services to customer needs
- Discover hidden customer behaviors from Association Rules and build a Recommendation Engine based on behavioral patterns
- Investigate relationships & flows between people and business relevant entities using Social Network Analysis
- Build predictive models of revenue and other numeric variables using Linear Regression

Course Outline

1. Lesson 1

1. What is the required Skill-set of a Data Scientist
2. Combining the Technical and Non-technical roles of a Data Scientist
3. The difference between a Data Scientist and a Data Engineer
4. Explore the full lifecycle of Data Science efforts within the organization
5. Discuss how to turn business questions into Machine Learning (ML) and Artificial Intelligence (AI) models
6. Explore diverse and wide-ranging data sources, internal and external to the organization that can be used to answer business questions

2. Lesson 2

1. Introduce the features of Python that make it an ideal tool for Data Scientists and Data Engineers alike
2. Viewing Data Sets using Python's Pandas library
3. Importing, Exporting and working with all forms of Data, from Relational Databases to Google Images using the Python
4. Selecting, Filtering, Combining, Grouping and Applying Functions using Python's Pandas library
5. Dealing with Duplicates, Missing Values, Rescaling, Standardizing and

Normalizing Data

6. Visualizing Data for both Exploration and Communication with the Pandas, Matplotlib and Seaborn Python libraries

3. Lesson 3

1. Preprocess Unstructured Data such as web adverts, emails, blog posts, in order to use it our AI/ML models
2. Explore the most popular approaches to Natural Language Processing (NLP) such as stemming, and “stop” words
3. Prepare a term-document matrix (TDM) of unstructured documents in preparation for analysis

4. Lesson 4

1. Express a business problem such as customer revenue prediction as a linear regression task
2. Assess variables as potential Predictors of the required Target eg. Education as a predictor of Salary
3. Build, Interpret and Evaluate a Linear Regression model in Python using measures such as RMSE
4. Explore the Feature Engineering possibilities to improve the Linear Regression model

5. Lesson 5

1. Learn how AI/ML Classifiers are built and used to make predictions such as Customer Churn
2. Explore how AI/ML Classification models are built using Training, Test and Validation Datasets
3. Build, Apply and Evaluate the strength of a Decision Tree Classifier

6. Lesson 6

1. Examine some alternative approaches to classification
2. Consider how Activation Functions are integral to Logistic Regression Classifiers
3. Investigate how Neural Networks and Deep Learning are used to build self-driving cars
4. Explore the probability foundations of Naive Bayes classifiers
5. Review different approaches to measuring the performance of AI/ML Classification Models
6. ROC curves, AUC measures, Precision, Recall, Confusion Matrix

7. Lesson 7

1. Uncover new ways of segmenting your customers, products or services through the use of clustering algorithms
2. Explore what the concept of similarity means to humans and how it can be implemented programmatically through
3. distance measures on descriptive variables
4. Perform top-down clustering with Python's Scikit-Learn K-Means algorithm
5. Perform bottom-up clustering with Scikit-Learn's hierarchical clustering algorithm
6. Examine clustering techniques on unstructured data (eg. Tweets, Emails, Documents, etc)

8. Lesson 8

1. Build models of customer behaviors or business events from logged data using

Association Rules

2. Evaluate the strength of these models through probability measures of support, confidence, and lift
3. Employ feature engineering approaches to improve the models
4. Build a recommender for your customers that is unique to your product/service offering

9. Lesson 9

1. Analyze your organization, its people and environment as a network of inter-relationships
2. Visualize these relationships to uncover previously unseen business insights
3. Explore ego-centric and socio-centric methods of analyzing connections important to your organization

10. Lesson 10

1. Examine Cloud (Microsoft, Amazon, Google) approaches to handling Big Data analytics
2. Explore the communications and ethics aspects of being a Data Scientist
3. Survey the paths of continual learning for a Data Scientist

Class Materials

Each student will receive a comprehensive set of materials, including course notes and all the class examples.

Class Prerequisites

Experience in the following *is required* for this Python class:

- An interest in gaining foundational knowledge of data science. This data scientist training course is designed for technical and non-technical beginners.