

Introduction to AI, Data Science & Machine Learning with Python

Course LT1264

5 Days

Instructor-led, Hands on

Course Information

Data science is a field that has exploded in popularity in recent years, and for good reason. Companies across industries are increasingly relying on data to inform their decision-making, and skilled data scientists are in high demand. In this comprehensive course, you'll learn the foundational skills and techniques you need to succeed in this exciting field.

You'll start by exploring the role of a data scientist and the lifecycle of data science efforts within an organization. Then, you'll dive into the technical skills you need, such as using Python and its relevant libraries for data analysis and visualization, preprocessing unstructured data, and building AI/ML models.

You'll also explore key machine learning algorithms, including linear regression, decision tree classifiers, and clustering algorithms. And, you'll learn how to apply these techniques to real-world problems, such as predicting customer churn and building recommendation engines.

Throughout the data science training, you'll have the opportunity to work on hands-on exercises and projects, allowing you to practice your skills and build your portfolio. By the end of the course, you'll have a deep understanding of the data science process, the tools and techniques used by data scientists, and the ability to apply these skills to real-world problems.

At Course Completion

Upon successful completion of this course, students will be able to:

- Consider how traditional Predictive Machine Learning techniques combine Generative AI, Agentic AI, Multimodal AI and Deepseek-R1 approaches in the current Data Science landscape.
- Translate everyday business questions and problems into Machine Learning tasks to make data-driven decisions.
- Use Python Pandas, Matplotlib & Seaborn libraries to explore, analyze, and visualize data from various sources, including the web, word documents, email, NoSQL stores, databases, and data warehouses.
- Explore the concepts behind Foundation Models, Generative Pre-trained Transformers (GPTs), and Retrieval Augmented Generation (RAGs).

Contact ISInc for more information at 916.920.1700 or by visiting our website at <http://www.isinc.com>

- Train a Machine Learning Classifier using different algorithmic techniques from the Scikit-Learn library, such as Decision Trees, Logistic Regression, and Neural Networks.
- Re-segment your customer market using K-Means and Hierarchical algorithms to better align products and 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.
- Test your knowledge with the included end-of-course exam.
- Leverage continued support with after-course one-on-one instructor coaching and computing sandbox

Prerequisites

Prior to attending this course, students should know the fundamentals of using Python.

Course Outline

Module 1: The Role of a Data Scientist: Combining Technical and Non-Technical Skills

- What is the required skillset of a Data Scientist?
- Combining the technical and non-technical roles of a Data Scientist
- The difference between a Data Scientist and a Data Engineer
- Exploring the entire lifecycle of Data Science efforts within the organization
- Turning business questions into Machine Learning (ML) and Artificial Intelligence (AI) models
- Exploring diverse and wide-ranging data sources that you can use to answer business questions
- Explore the concepts behind Foundation Models, Generative Pre-trained Transformers (GPTs), and Retrieval Augmented Generation (RAGs)

Module 2: Data Manipulation and Visualization using Python's Pandas and Matplotlib Libraries

- Introducing the features of Python that are relevant to Data Scientists and Data Engineers
- Viewing Data Sets using Python's Pandas library
- Importing, exporting, and working with all forms of data, from Relational Databases to Google Images

Contact ISInc for more information at 916.920.1700 or by visiting our website at <http://www.isinc.com>

- Using Python Selecting, Filtering, Combining, Grouping, and Applying Functions from Python's Pandas library
- Dealing with Duplicates, Missing Values, Rescaling, Standardizing, and Normalizing Data
- Visualizing data for both exploration and communication with the Pandas, Matplotlib, and Seaborn Python libraries

Module 3: Preprocessing and Analyzing Unstructured Data with Natural Language Processing

- Preprocessing Unstructured Data such as web adverts, emails, and blog posts for AI/ML models
- Exploring the most popular approaches to Natural Language Processing (NLP), such as stemming and "stop" words
- Preparing a term-document matrix (TDM) of unstructured documents for analysis
- Review the architectures of Foundation Models, Generative Pre-trained Transformers (GPTs), and Retrieval Augmented Generation (RAGs)
- Look at how Data Scientists can integrate Large Language Models (LLMs) in their work

Module 4: Linear Regression and Feature Engineering for Business Problem Solving

- Expressing a business problem, such as customer revenue prediction, as a linear regression task
- Assessing variables as potential Predictors of the required Target (e.g., Education as a predictor of Salary Build)
- Interpreting and Evaluating a Linear Regression model in Python using measures such as RMSE
- Exploring the Feature Engineering possibilities to improve the Linear Regression model

Module 5: Classification Models and Evaluation for Predictive Analysis

- Learning how AI/ML Classifiers are built and used to make predictions such as Customer Churn
- Exploring how AI/ML Classification models are built using Training, Test, and Validation
- Evaluating the strength of a Decision Tree Classifier

Module 6: Alternative Approaches to Classification and Model Evaluation

- Examining alternative approaches to classification

- Considering how Activation Functions are integral to Logistic Regression Classifiers
- Delve into the architecture of Neural Networks and investigate the explosive growth of Deep Learning approaches in AI
- Exploring the probability foundations of Naive Bayes classifiers
- Reviewing different approaches to measuring the performance of AI/ML Classification Models
- Reviewing ROC curves, AUC measures, Precision, Recall, and Confusion Matrices

Module 7: Clustering Techniques for Customer and Product Segmentation

- Uncovering new ways of segmenting your customers, products, or services using clustering algorithms
- Exploring what the concept of similarity means to humans and how you can implement it programmatically through distance measures on descriptive variables
- Performing top-down clustering with Python's Scikit-Learn K-Means algorithm
- Performing bottom-up clustering with Scikit-Learn's hierarchical clustering algorithm
- Examining clustering techniques on unstructured data (e.g., Tweets, Emails, Documents, etc.)

Module 8: Association Rules and Recommender Systems for Business Applications

- Building models of customer behaviors or business events from logged data using Association Rules
- Evaluating the strength of these models through probability measures of support, confidence, and lift
- Employing feature engineering approaches to improve the models
- Building a recommender for your customers that is unique to your product/service offering

Module 9: Network Analysis for Organizational Insights

- Analyzing your organization, its people, and its environment as a network of inter-relationships
- Visualizing these relationships to uncover previously unseen business insights
- Exploring ego-centric and socio-centric methods of analyzing connections critical to your organization

Module 10: Big Data Analytics, Communication, and Ethics

- Examining Cloud (Microsoft, Amazon, Google) approaches to handling Big Data analytics

Contact ISInc for more information at 916.920.1700 or by visiting our website at <http://www.isinc.com>



- Exploring the communications and ethics aspects of being a Data Scientist
- Discuss the ethical implications of recent developments in AI
- Surveying the paths of continual learning for a Data Scientist