



Data Science and AI Principles

Syllabus

Data Science Principles makes the fundamental topics in data science approachable and relevant by using real-world examples and prompts learners to think critically about applying these new understandings to their own workplace. Get an overview of data science and AI systems with a nearly code- and math-free introduction to prediction, causality, visualization, data wrangling, privacy, ethics. See how these concepts translate from theory to practice with examples relevant to your own work environment.

Modules		Case Studies	Takeaways	Key Exercises
Module 1	Data 101	Flu Detection	<ul style="list-style-type: none">Explain why data collection is importantIdentify factors that may affect data qualityRecognize that not all data is numericalExplain how the organization of data can affect the information you are able to extract from it	<ul style="list-style-type: none">List sources of dataDiscuss what can be done with dataCategorize data by various factorsDetermine whether data is high-quality or not
Module 2	Predictions and Recommendations	Predicting Sepsis	<ul style="list-style-type: none">Understand the basic structure of a predictive algorithmIdentify where human decisions shape predictive systemsEvaluate the success of a predictive system	<ul style="list-style-type: none">Examine how weather forecasts workUse data to create a predictionSort types of training dataSimulate a predictive system
Module 3	Cause and Effect	The Google Tax	<ul style="list-style-type: none">Explain why it is important to establish causal relationshipsIdentify barriers to establishing causal relationships in a variety of settingsIdentify why randomization can help establish a causal relationship but also create other problems	<ul style="list-style-type: none">Classify relationships based on correlation or causationExamine the relationship between variablesIdentify potential common causes for correlated events
Module 4	Data Governance and Privacy	Privacy and Facial Recognition	<ul style="list-style-type: none">Explain why data privacy is importantDescribe what can constitute a violation of privacyCritique existing privacy policiesCreate a set of ethical tenets to guide data work at their own organizations	<ul style="list-style-type: none">Formulate data privacy guidelinesDiscuss the risks of data reidentificationEvaluate existing data privacy policies for ethics
Module 5	Beyond the Spreadsheet	Burning Glass and Text Data	<ul style="list-style-type: none">Identify sources of non-numerical dataExplain why it would be useful to use non-numerical dataDescribe the differences in approach for supervised and unsupervised learningIdentify use cases for neural networks	<ul style="list-style-type: none">Perform a sentiment analysisDetermine what types of data an algorithm cannot readExamine how computers intake visual and audio dataExperiment with facial recognition

Modules		Case Studies	Takeaways	Key Exercises
Module 6	Introduction to Algorithms	Reducing food waste with Shelf Engine	<ul style="list-style-type: none">Describe some algorithms commonly used in data scienceUnderstand basic workhorse algorithms in data science such as regressionExplain why and how such tools are made substantially more complexExplain the crucial role humans have in overseeing and maintaining algorithmsExplain some of the trade-offs between more sophisticated algorithms, including the costs of running and evaluating their success	<ul style="list-style-type: none">Examine how to evaluate the performance of an algorithmIdentify variables that can be used to predict consumer demandSelect appropriate algorithms for different purposes
Module 7	Data Science Ecosystems	Harvard Link	<ul style="list-style-type: none">Explain the importance of data transformation and wranglingList the common technologies used within data science ecosystemsDescribe the connection between data science tasks, software tools, and hardware toolsIdentify potential sources of bottlenecks in the data science process	<ul style="list-style-type: none">Identify and order the lifecycle of dataDefine what “the cloud” isEstimate the size of various data streams
Module 8	Data Science and AI	AXA Insurance	<ul style="list-style-type: none">Understand how AI can move beyond rules-based automation to perform more complex analysisDescribe why AI needs quality data to perform tasks and how you can gather necessary dataExplain how AI can help organizations move from reactive to proactive approaches to problem-solvingRecognize AI adoption as a leadership challenge	<ul style="list-style-type: none">Map AI capabilities to real-world business tasksAssess data readiness for complex AI systemsExplore AI techniques for different data modalitiesPropose leadership strategies that balance innovation, safety, and ethical constraints
Module 9	The Road Ahead	Healthcare Prioritization	<ul style="list-style-type: none">Recognize a problem that an algorithm might be able to solveRecognize the challenges created by using data science tools in ways outside their intended useIdentify steps within the data science process that need auditing	<ul style="list-style-type: none">Choose types of data to ingest into an algorithmEvaluate the risks of solely using an algorithm to make decisionsDiscuss how algorithms can reinforce biasesCreate a set of guidelines to evaluate projects

Learning requirements: In order to earn a Certificate of Completion from Harvard Online and Harvard Business School Online, participants must thoughtfully complete all 8 modules, including associated quizzes, by stated deadlines.