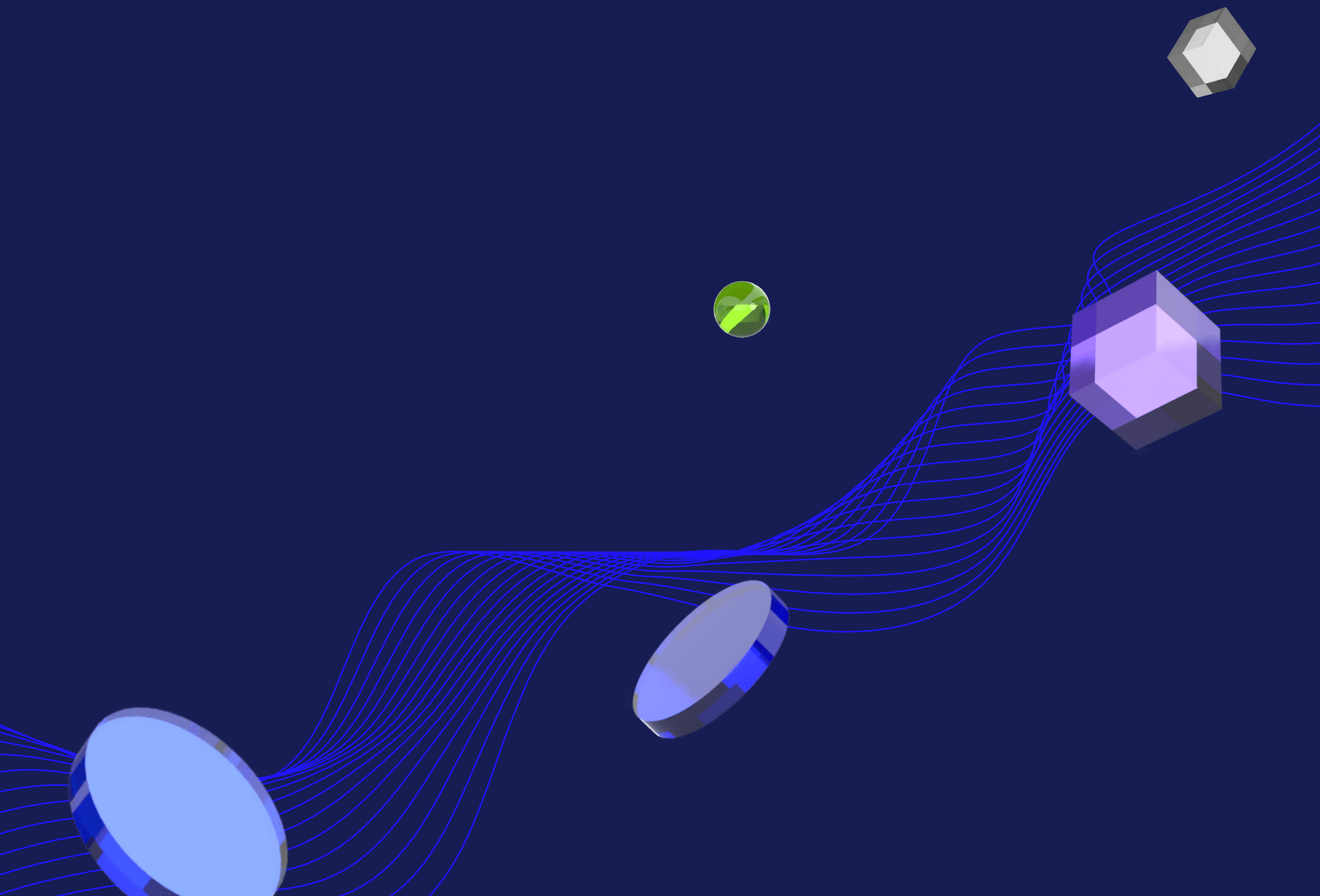UDACITY

SCHOOL OF DATA SCIENCE

Data Scientist

Nanodegree Program Syllabus

Overview

Build effective machine learning models, run data pipelines, build recommendation systems, and deploy solutions to the cloud with industry-aligned projects.

Learning Objectives

A graduate of this program will be able to:

- Use Python and SQL to access and analyze data from several different data sources.

- Use principles of statistics and probability to design and execute A/B tests and recommendation engines to assist businesses in making data-automated decisions.

- Deploy a data science solution to a basic flask app.

- Manipulate and analyze distributed datasets using Apache Spark.

- Communicate results effectively to stakeholders.

Built in collaboration with:

BERTELSMANN   figure eight   IBM Watson®   Insight   kaggle   STARBUCKS

Program information

Estimated Time

4 months at 10hrs/week*

Skill Level

Advanced

Prerequisites

The Data Scientist Nanodegree program is an advanced program and requires previous competence in the following areas:

- Programming
- Probability and statistics
- Mathematics
- Data wrangling
- Data visualization with matplotlib
- Machine learning

Required Hardware/Software

Learners need access to a computer running OS X or Windows.

*The length of this program is an estimation of total hours the average student may take to complete all required coursework, including lecture and project time. If you spend about 5-10 hours per week working through the program, you should finish within the time provided. Actual hours may vary.

Solving Data Science Problems

Learn the data science process, including how to build effective data visualizations, and how to communicate with various stakeholders.

Course Project

Write a Data Science Blog Post

In this project, learners will choose a dataset, identify three questions, and analyze the data to find answers to these questions. They will create a GitHub repository with their project, and write a blog post to communicate their findings to the appropriate audience. This project will help learners reinforce and extend their knowledge of machine learning, data visualization, and communication.

Lesson 1

The Data Science Process

- Apply the CRISP-DM process to business applications.
- Wrangle, explore, and analyze a dataset.
- Apply machine learning for prediction.
- Apply statistics for descriptive and inferential understanding.
- Draw conclusions that motivate others to act on your results.

Lesson 2

Communicating with Stakeholders

- Implement best practices in sharing your code and written summaries.
- Learn what makes a great data science blog.
- Learn how to create your ideas with the data science community.

Software Engineering for Data Scientists

Develop software engineering skills that are essential for data scientists, such as creating unit tests and building classes.

Lesson 1

Software Engineering Practices

- Write clean, modular, and well-documented code.
- Refactor code for efficiency.
- Create unit tests to test programs.
- Write useful programs in multiple scripts.
- Track actions and results of processes with logging.
- Conduct and receive code reviews.

Lesson 2

Object Oriented Programming

- Understand when to use object oriented programming.
- Build and use classes.
- Understand magic methods.
- Write programs that include multiple classes, and follow good code structure.
- Learn how large, modular Python packages, such as pandas and scikit-learn, use object oriented programming.
- Portfolio Exercise: Build your own Python package.

Lesson 3

Wen Development

- Learn about the components of a web app.
- Build a web application that uses Flask, Plotly, and the Bootstrap framework.
- Portfolio Exercise: Build a data dashboard using a dataset of your choice and deploy it to a web application.

Data Engineering for Data Scientists

Learn to work with data through the entire data science process, from running pipelines, transforming data, building models, and deploying solutions to the cloud.

Course Project

Build Disaster Response Pipelines with Figure Eight

Figure Eight (formerly Crowdflower) crowdsourced the tagging and translation of messages to apply artificial intelligence to disaster response relief. In this project, learners will build a data pipeline to prepare the message data from major natural disasters around the world. They'll build a machine learning pipeline to categorize emergency text messages based on the need communicated by the sender.

Lesson 1

ETL Pipelines

- Understand what ETL pipelines are.
- Access and combine data from CSV, JSON, logs, APIs, and databases.
- Standardize encodings and columns.
- Normalize data and create dummy variables.
- Handle outliers, missing values, and duplicated data.
- Engineer new features by running calculations.
- Build a SQLite database to store cleaned data.

Lesson 2

Natural Language Processing

- Prepare text data for analysis with tokenization, lemmatization, and removing stop words.

- Use scikit-learn to transform and vectorize text data.

- Build features with bag of words and tf-idf.

- Extract features with tools such as named entity recognition and part of speech tagging.

- Build an NLP model to perform sentiment analysis.

Lesson 3

Machine Learning Pipelines

- Understand the advantages of using machine learning pipelines to streamline the data preparation and modeling process.

- Chain data transformations and an estimator with scikitlearn's pipeline.

- Use feature unions to perform steps in parallel and create more complex workflows.

- Grid search over pipeline to optimize parameters for entire workflow.

- Complete a case study to build a full machine learning pipeline that prepares data and creates a model for a dataset.

Course 4

Experiment Design and Recommendations

Learn to design experiments and analyze A/B test results. Explore approaches for building recommendation systems.

Design a Recommendation Engine with IBM

IBM has an online data science community where members can post tutorials, notebooks, articles, and datasets. In this project, learners will build a recommendation engine, based on user behavior and social network in IBM Watson Studio's data platform, to surface content most likely to be relevant to a user.

Lesson 1

Experiment Design

- Understand how to set up an experiment, and the ideas associated with experiments vs. observational studies.
- Defining control and test conditions.
- Choosing control and testing groups.

Lesson 2

Statistical Concerns of Experimentation

- Applications of statistics in the real world.
- Establishing key metrics.
- SMART experiments: Specific, Measurable, Actionable, Realistic, Timely.

Lesson 3

A/B Testing

- How it works and its limitations.
- Sources of bias: novelty and recency effects.
- Multiple comparison techniques (FDR, Bonferroni, Tukey).
- Portfolio Exercise: Use a technical screener from Starbucks to analyze the results of an experiment and write up your findings.

Lesson 4

Introduction to Recommendation Engines

- Distinguish between common techniques for creating recommendation engines including knowledge-based, content-based, and collaborative filtering-based methods.
- Implement each of these techniques in Python.
- List business goals associated with recommendation engines and be able to recognize which of these goals are most easily met with existing recommendation techniques.

Lesson 5

Matrix Factorization for Recommendations

- Understand the pitfalls of traditional methods and pitfalls of measuring the influence of recommendation engines under traditional regression and classification techniques.
- Create recommendation engines using matrix factorization and FunkSVD.
- Interpret the results of matrix factorization to better understand latent features of customer data.
- Determine common pitfalls of recommendation engines like the cold start problem and difficulties associated with usual tactics for assessing the effectiveness of recommendation engines using usual techniques, and potential solutions.

Course 5

Data Science Projects

Leverage learnings from this program to build an open-ended data science project. This project will demonstrate valuable, job-ready skills.

Data Science Capstone Project

In this capstone project, learners will leverage what they've learned throughout the program to build a data science project of their choosing. They will define the problem they want to solve, identify and explore the data, then perform their analyses and develop a set of conclusions. Learners will present the analysis and their conclusions in a blog post and GitHub repository. This project will serve as a demonstration of their ability as a data scientist, and will be an important component of their job-ready portfolio.

Lesson 1

Elective 1: Dog Breed Classification

- Use convolutional neural networks to classify different dogs according to their breeds.
- Deploy your model to allow others to upload images of their dogs and send them back the corresponding breeds.
- Complete one of the most popular projects in Udacity history, and show the world how you can use your deep learning skills to entertain an audience.

Lesson 2

Elective 2: Starbucks

- Use purchasing habits to arrive at discount measures to obtain and retain customers.
- Identify groups of individuals that are most likely to be responsive to rebates.

Lesson 3

Elective 3: Arvato Financial Services

- Work through a real-world dataset and challenge provided by Arvato Financial Services, a Bertelsmann company.
- Top performers have a chance at an interview with Arvato or another Bertelsmann company.

Lesson 4

Elective 4: Spark for Big Data

- Take a course on Apache Spark and complete a project using a massive, distributed dataset to predict customer churn.

- Learn to deploy your Spark cluster on either AWS or IBM Cloud.

Lesson 5

Elective 5: Your Choice

- Use your skills to tackle any other project of your choice.

Meet your instructors.

Josh Bernhard

Data Scientist at Nerd Wallet

Josh has been sharing his passion for data for nearly a decade at all levels of university, and as lead data science instructor at Galvanize. He's used data science for work ranging from cancer research to process automation.

Juno Lee

Curriculum Lead at Udacity

Juno is the curriculum lead for the School of Data Science. She has been sharing her passion for data and teaching, building several courses at Udacity. As a data scientist, she built recommendation engines, computer vision and NLP models, and tools to analyze user behavior.

Luis Serrano

Machine Learning Engineer

Luis was formerly a machine learning engineer at Google. He holds a PhD in mathematics from the University of Michigan, and a postdoctoral fellowship at the University of Quebec at Montreal.

Andrew Paster

Instructor

Andrew has an engineering degree from Yale, and has used his data science skills to build a jewelry business from the ground up. He has additionally created courses for Udacity's Self-Driving Car Engineer Nanodegree program.

Mike Yi

Data Analyst Instructor

Mike is a content developer with a multidisciplinary academic background, including math, statistics, physics, and psychology. Previously, he worked on Udacity's Data Analyst Nanodegree program as a support lead.

David Drummond

VP of Engineering at Insight

David is VP of Engineering at Insight where he enjoys breaking down difficult concepts and helping others learn data engineering. David has a PhD in Physics from UC Riverside.
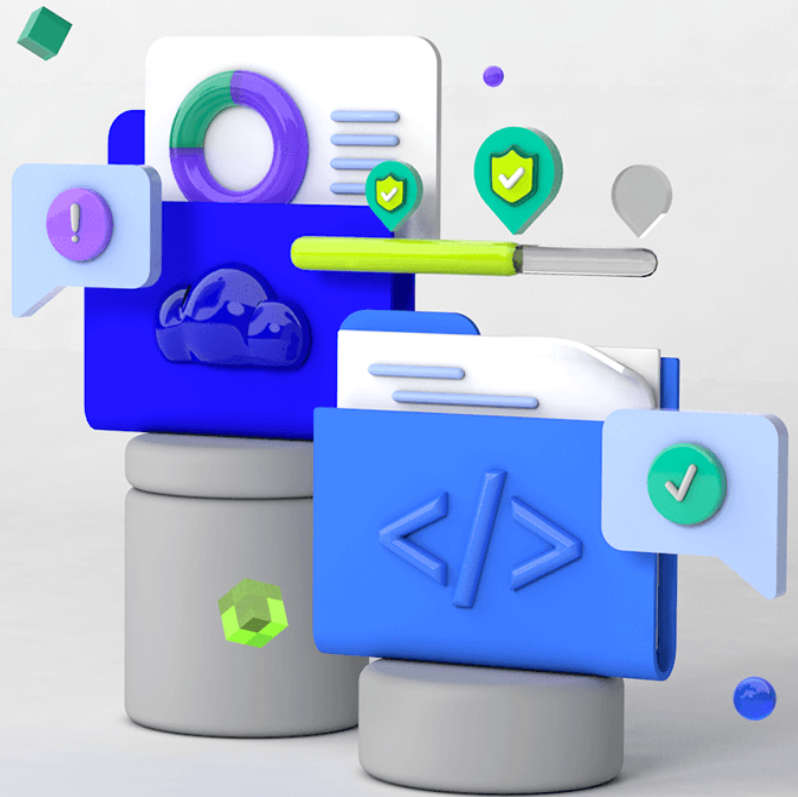
Judit Lantos

Senior Data Engineer at Netflix

Judit is a senior data engineer at Netflix. A former data engineer at Split, she worked on the statistical engine of their full-stack experimentation platform. She has also instructed at Insight Data Science, helping software engineers and academic coders transition to DE roles.

Udacity's learning experience

Hands-on Projects

Open-ended, experiential projects are designed to reflect actual workplace challenges. They aren't just multiple choice questions or step-by-step guides, but instead require critical thinking.

Quizzes

Auto-graded quizzes strengthen comprehension. Learners can return to lessons at any time during the course to refresh concepts.

Knowledge

Find answers to your questions with Knowledge, our proprietary wiki. Search questions asked by other students, connect with technical mentors, and discover how to solve the challenges that you encounter.

Custom Study Plans

Create a personalized study plan that fits your individual needs. Utilize this plan to keep track of movement toward your overall goal.

Workspaces

See your code in action. Check the output and quality of your code by running it on interactive workspaces that are integrated into the platform.

Progress Tracker

Take advantage of milestone reminders to stay on schedule and complete your program.

Our proven approach for building job-ready digital skills.
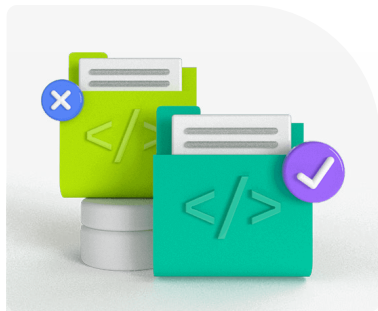
Experienced Project Reviewers

Verify skills mastery.

- Personalized project feedback and critique includes line-by-line code review from skilled practitioners with an average turnaround time of 1.1 hours.
- Project review cycle creates a feedback loop with multiple opportunities for improvement—until the concept is mastered.
- Project reviewers leverage industry best practices and provide pro tips.

Technical Mentor Support

24/7 support unblocks learning.

- Learning accelerates as skilled mentors identify areas of achievement and potential for growth.
- Unlimited access to mentors means help arrives when it's needed most.
- 2 hr or less average question response time assures that skills development stays on track.
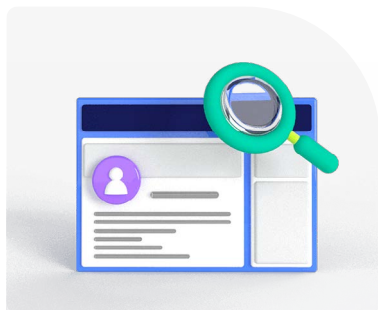
Personal Career Services

Empower job-readiness.

- Access to a Github portfolio review that can give you an edge by highlighting your strengths, and demonstrating your value to employers.*
- Get help optimizing your LinkedIn and establishing your personal brand so your profile ranks higher in searches by recruiters and hiring managers.
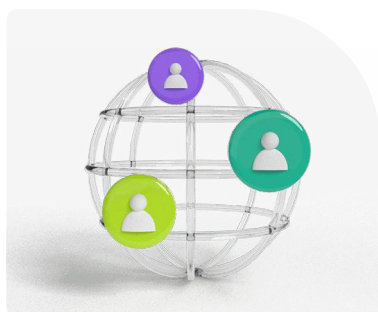
Mentor Network

Highly vetted for effectiveness.

- Mentors must complete a 5-step hiring process to join Udacity's selective network.
- After passing an objective and situational assessment, mentors must demonstrate communication and behavioral fit for a mentorship role.
- Mentors work across more than 30 different industries and often complete a Nanodegree program themselves.

*Applies to select Nanodegree programs only.

UDACITY

Learn more at

www.udacity.com/online-learning-for-individuals →