UDACITY
FOR ENTERPRISE

Data Scientist

NANODEGREE SYLLABUS

Overview

In Collaboration With

BERTELSMANN   IBM Watson®   figure eight   Insight   kaggle   STARBUCKS

The Data Scientist Nanodegree program prepares learners to perform well as data scientists. Graduates of this program emerge equipped 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.

Program Information

TIME
4 months
Study 10 hours/week

LEVEL
Specialist

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
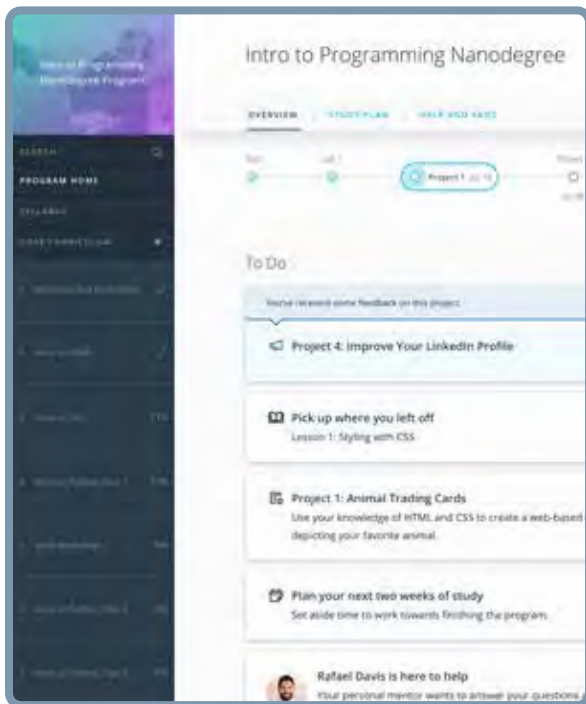- Machine Learning

HARDWARE/SOFTWARE REQUIRED
Computer running OS X or Windows

LEARN MORE ABOUT THIS NANODEGREE
Contact us at enterpriseNDs@udacity.com.

Our Classroom Experience

REAL-WORLD PROJECTS
Learners build new skills through industry-relevant projects and receive personalized feedback from our network of 900+ project reviewers. Our simple user interface makes it easy to submit projects as often as needed and receive unlimited feedback.

KNOWLEDGE
Answers to most questions can be found with Knowledge, our proprietary wiki. Learners can search questions asked by others and discover in real-time how to solve challenges.

LEARNER HUB
Learners leverage the power of community through a simple, yet powerful chat interface built within the classroom. Learner Hub connects learners with their technical mentor and fellow learners.

WORKSPACES
Learners can check the output and quality of their code by testing it on interactive workspaces that are integrated into the classroom.

QUIZZES
Understanding concepts learned during lessons is made simple with auto-graded quizzes. Learners can easily go back and brush up on concepts at anytime during the course.
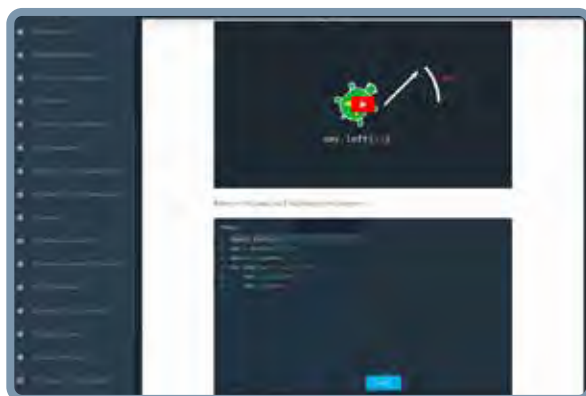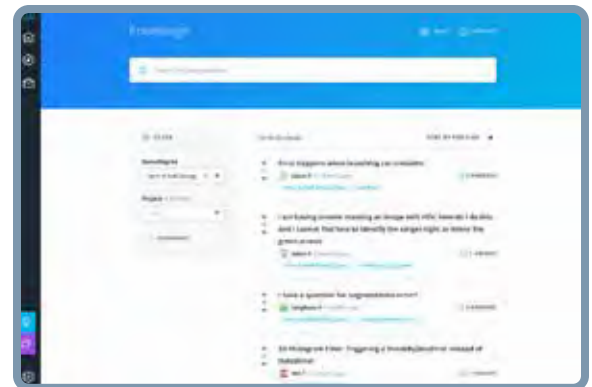
CUSTOM STUDY PLANS
Mentors create a custom study plan tailored to learners' needs. This plan keeps track of progress toward learner goals.

PROGRESS TRACKER
Personalized milestone reminders help learners stay on track and focused as they work to complete their Nanodegree program.

Learn with the Best

Josh Bernard
DATA SCIENTIST AT NERD WALLET

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

Juno Lee
DATA SCIENCE INSTRUCTOR

As a data scientist, Juno built a recommendation engine to personalize online shopping experiences, computer vision and natural language processing models to analyze product data, and tools to generate insight into user behavior.

Luis Serrano
INSTRUCTOR

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.

Course 1: Solving Data Science Problems

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

Project	Write a Data Science Blog Post

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

LESSON TITLE	LEARNING OUTCOME
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.
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.

Course 2: Software Engineering for Data Scientists

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

LESSON TITLE	LEARNING OUTCOME
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.
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.
WEB 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.

Course 3: 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.

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, you'll build a data pipeline to prepare the message data from major natural disasters around the world. You'll build a machine learning pipeline to categorize emergency text messages based on the need communicated by the sender.

LESSON TITLE	LEARNING OUTCOME
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.
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.
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 & Recommendations

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

Project	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, you 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 TITLE	LEARNING OUTCOME
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.
STATISTICAL CONCERNS OF EXPERIMENTATION	• Applications of statistics in the real world • Establishing key metrics • SMART experiments: Specific, Measurable, Actionable, Realistic, Timely.
A/B TESTING	• How it works and its limitations. • Sources of Bias: Novelty and Recency Effects. • Multiple Comparison Techniques (FDR, Bonferroni, Tukey). • Portfolio Exercise: Using a technical screener from Starbucks to analyze the results of an experiment and write up your findings.

Course 4: Experiment Design & Recommendations, cont.

LESSON TITLE	LEARNING OUTCOME
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.
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 what you've learned throughout the program to build your own open-ended Data Science project. This project willserve as a demonstration of your valuable abilities as a Data Scientist.

Project	Data Science Capstone Project

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

| LESSON TITLE	LEARNING OUTCOME
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.
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.
ELECTIVE 3: ARVATO FINANCIAL SERVICES	· Work through a real-world dataset and challenge provided by Arvato Financial Services, a Bertelsmann company.
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.
ELECTIVE 5: YOUR CHOICE	· Use your skills to tackle any other project of your choice.

Our Nanodegree Programs Include
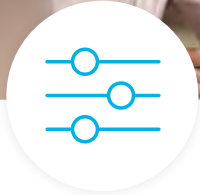
Pre-Assessments

Our in-depth workforce assessments identify your team's current level of knowledge in key areas. Results are used to generate custom learning paths designed to equip your workforce with the most applicable skill sets.

Dashboard & Progress Reports

Our interactive dashboard (enterprise management console) allows administrators to manage employee onboarding, track course progress, perform bulk enrollments and more.

Industry Validation & Reviews

Learners' progress and subject knowledge is tested and validated by industry experts and leaders from our advisory board. These in-depth reviews ensure your teams have achieved competency.

Real World Hands-on Projects

Through a series of rigorous, real-world projects, your employees learn and apply new techniques, analyze results, and produce actionable insights. Project portfolios demonstrate learners' growing proficiency and subject mastery.

Our Review Process

Real-life Reviewers for Real-life Projects

Real-world projects are at the core of our Nanodegree programs because hands-on learning is the best way to master a new skill. Receiving relevant feedback from an industry expert is a critical part of that learning process, and infinitely more useful than that from peers or automated grading systems. Udacity has a network of over 900 experienced project reviewers who provide personalized and timely feedback to help all learners succeed.

Vaibhav
UDACITY LEARNER

"I never felt overwhelmed while pursuing the Nanodegree program due to the valuable support of the reviewers, and now I am more confident in converting my ideas to reality."

now at
CODING VISIONS INFOTECH

All Learners Benefit From:

Line-by-line feedback for coding projects

Industry tips and best practices

Advice on additional resources to research

Unlimited submissions and feedback loops

How it Works
Real-world projects are integrated within the classroom experience, making for a seamless review process flow.

- Go through the lessons and work on the projects that follow
- Get help from your technical mentor, if needed
- Submit your project work
- Receive personalized feedback from the reviewer
- If the submission is not satisfactory, resubmit your project
- Continue submitting and receiving feedback from the reviewer until you successfully complete your project

About our Project Reviewers

Our expert project reviewers are evaluated against the highest standards and graded based on learners' progress. **Here's how they measure up to ensure your success.**

900+
Expert Project Reviewers
Are hand-picked to provide detailed feedback on your project submissions.

1.8M
Projects Reviewed
Our reviewers have extensive experience in guiding learners through their course projects.

3
Hours Average Turnaround
You can resubmit your project on the same day for additional feedback.

4.85 /5
Average Reviewer Rating
Our learners love the quality of the feedback they receive from our experienced reviewers.

UDACITY
FOR ENTERPRISE