



UDACITY
FOR ENTERPRISE

THE SCHOOL OF ARTIFICIAL INTELLIGENCE

Machine Learning DevOps Engineer



NANODEGREE SYLLABUS

Overview

Machine Learning DevOps Engineer Nanodegree Program

The Machine Learning DevOps Engineer Nanodegree program focuses on the software engineering fundamentals needed to successfully streamline the deployment of data and machine-learning models in a production-level environment. Students will build the DevOps skills required to automate the various aspects and stages of machine learning model building and monitoring over time.

A graduate of this program will be able to:

- Implement production-ready Python code/processes for deploying ML models outside of cloud-based environments facilitated by tools such as AWS SageMaker, Azure ML, etc.
- Engineer automated data workflows that perform continuous training (CT) and model validation within a CI/CD pipeline based on updated data versioning
- Create multi-step pipelines that automatically retrain and deploy models after data updates
- Track model summary statistics and monitor model online performance over time to prevent model-degradation

Program Information



TIME

4 months
Study 10 hours/week



LEVEL

Specialist



PREREQUISITES

Prior experience with Python and Machine Learning



HARDWARE/SOFTWARE REQUIRED

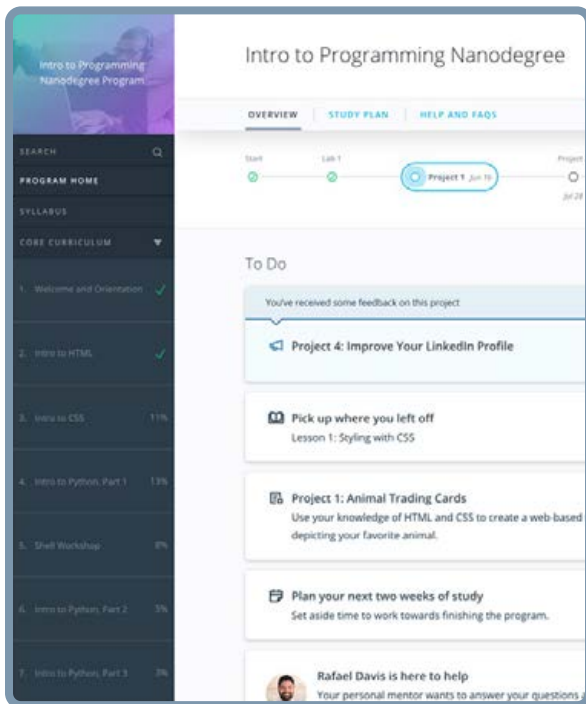
A 64-bit computer, at least 8GB of RAM and administrator account permissions sufficient to install programs including Anaconda with Python 3.x and supporting packages.



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

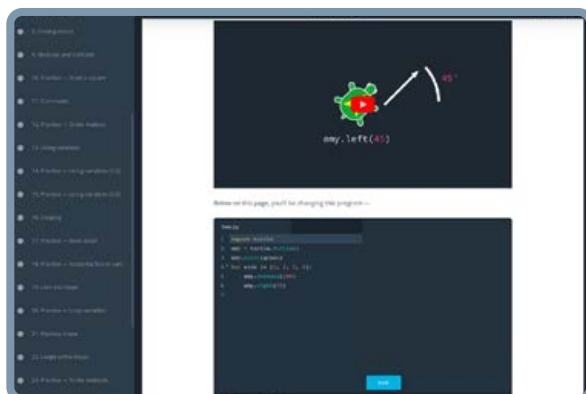
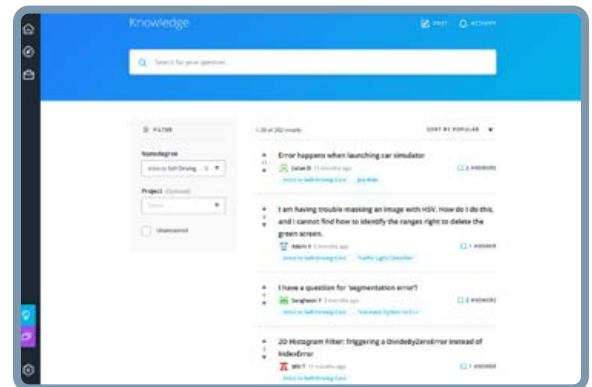
Learners leverage the power of community through a simple, yet powerful chat interface built within the classroom. Learner Hub connects learners with their technical mentor and fellow learners.

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



Joshua Bernhard

DATA SCIENTIST AT THUMBTRACK

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



Giacomo Vianello

LEAD DATA SCIENTIST

Giacomo is an end-to-end data scientist with a passion for state-of-the-art but practical technical solutions. He is Lead Data Scientist at Cape Analytics, where he develops AI systems to extract intelligence from geospatial imagery bringing cutting-edge AI solutions to the insurance and real estate industries.



Justin Clifford Smith, Ph.D.

SENIOR DATA SCIENTIST

Justin is a Senior Data Scientist at Optum where he works to make healthcare more efficient with natural language processing and machine learning. Previously he was a Data Scientist at the US Census Bureau. His doctorate is from the University of California, Irvine where he studied theoretical physics.



Bradford Tuckfield

DATA SCIENTIST AND WRITER

Bradford is a data scientist and writer. He has worked on applications of data science in a variety of industries. He's the author of Dive Into Algorithms, forthcoming with No Starch Press.



Ulrika Jägare

HEAD OF AI/ML STRATEGY
EXECUTION IN ERICSSON

Ulrika has been with Ericsson for 21 years in various leadership roles, out of which 11 years in the Data and AI space. Ulrika holds a Master of Science degree from University of Lund in Sweden and is also author of seven published books in Data Science.

Nanodegree Program Overview

Course 1: Clean Code Principles

Develop skills that are essential for deploying production machine learning models. First, you will put your coding best practices on autopilot by learning how to use PyLint and AutoPEP8. Then you will further expand your Git and Github skills to work with teams. Finally, you will learn best practices associated with testing and logging used in production settings to ensure your models can stand the test of time.

Project

Predict Customer Churn with Clean Code

In this project, you will implement your learnings to identify credit card customers most likely to churn. The completed project will include a Python package for a machine learning project that follows coding (PEP8) and engineering best practices for implementing software (modular, documented and tested). The package will also have the flexibility to run interactively or from the command-line interface (CLI). This project will give you practice using your skills for testing, logging and coding best practices from the lessons. It will also introduce you to a problem data scientists across companies face all the time: How do we identify (and later intervene with) customers who are likely to churn?

LESSON TITLE

LEARNING OUTCOMES

CODING BEST PRACTICES

- Write clean, modular and well-documented code
- Refactor code for efficiency
- Follow PEP8 Standards
- Automate use of PEP8 standards using PyLint and AutoPEP8

WORKING WITH OTHERS USING VERSION CONTROL

- Work independently using git and Github
- Work with teams using git and Github
- Create branches for isolating changes in git and Github
- Open pull requests for making changes to production code
- Conduct and receive code reviews using best practices

Nanodegree Program Overview



LESSON TITLE	LEARNING OUTCOMES
PRODUCTION READY CODE	<ul style="list-style-type: none">• Correctly use try-except blocks to identify errors• Create unit tests to test programs• Track actions and results of processes with logging• Identify model drift and when automated or non-automated retraining should be used to make model updates

Nanodegree Program Overview

Course 2: Building a Reproducible Model Workflow

This course empowers the students to be more efficient, effective and productive in modern, real-world ML projects by adopting best practices around reproducible workflows. In particular, it teaches the fundamentals of MLOps and how to: a) create a clean, organized, reproducible, end-to-end machine learning pipeline from scratch using MLflow b) clean and validate the data using pytest c) track experiments, code and results using GitHub and Weights & Biases d) select the best-performing model for production and e) deploy a model using MLflow. Along the way, it also touches on other technologies like Kubernetes, Kubeflow, and Great Expectations and how they relate to the content of the class.

Project

Build an ML Pipeline for Short-term Rental Prices in NYC

Students will write a Machine Learning Pipeline to solve the following problem: a property management company is renting rooms and properties in New York for short periods on various rental platforms. They need to estimate the typical price for a given property based on the price of similar properties. The company receives new data in bulk every week, so the model needs to be retrained with the same cadence, necessitating a reusable pipeline. The students will write an end-to-end pipeline covering data fetching, validation, segregation, train and validation, test, and release. They will run it on an initial data sample, then re-run it on a new data sample simulating a new data delivery.

LESSON TITLE

LEARNING OUTCOMES

MACHINE LEARNING PIPELINES

- MLOps fundamentals
- Version data and artifacts
- Write a ML pipeline component
- Link together ML components

DATA EXPLORATION AND PREPARATION

- Execute and track the Exploratory Data Analysis (EDA)
- Clean and pre-process the data
- Segregate (split) datasets

Nanodegree Program Overview



LESSON TITLE	LEARNING OUTCOMES
DATA VALIDATION	<ul style="list-style-type: none">• Use pytest with parameters for reproducible and automatic data tests• Perform deterministic and non-deterministic data tests
TRAINING, VALIDATION AND EXPERIMENT TRACKING	<ul style="list-style-type: none">• Tame the chaos with experiment, code and data tracking• Track experiments with W&B• Validate and choose best-performing model• Export model as an inference artifact• Test final inference artifact
RELEASE AND DEPLOY	<ul style="list-style-type: none">• Release pipeline code• Options for deployment and how to deploy a model

Nanodegree Program Overview

Course 3: Deploying a Scalable ML Pipeline in Production

This course teaches students how to deploy a machine learning model into production. En route to that goal, students will learn how to put the finishing touches on a model by taking a fine-grained approach to model performance, checking bias and ultimately writing a model card. Students will also learn how to version control their data and models using Data Version Control (DVC). In the last piece of preparation for deployment, students will learn Continuous Integration and Continuous Deployment accomplished using GitHub Actions and Heroku. Finally, students will learn how to write a fast, type-checked and auto-documented API using FastAPI.

Project

Deploying a Machine Learning Model on Heroku with FastAPI

In this project, students will deploy a machine learning model on Heroku. The students will use Git and DVC to track their code, data and model while developing a simple classification model on the Census Income Data Set. After creating the model, the students will finalize the model for production by checking its performance on slices and writing a model card encapsulating key knowledge about the model. Students will put together a Continuous Integration and Continuous Deployment framework and ensure their pipeline passes a series of unit tests before deployment. Lastly, an API will be written using FastAPI and tested locally. After successful deployment, the API will be tested live using the requests module.

After completion, you will have a working API that is live in production, a set of tests, a model card and a full CI/CD framework. On its own, this project can be a portfolio piece but can also be applied to other projects, e.g., continuous integration, to flesh them further out.

LESSON TITLE

LEARNING OUTCOMES

PERFORMANCE TESTING AND PREPARING A MODEL FOR PRODUCTION

- Analyze slices of data when training and testing models
- Probe a model for bias using common frameworks such as Aequitas
- Write model cards that explain the purpose, provenance and pitfalls of a model

Nanodegree Program Overview



LESSON TITLE	LEARNING OUTCOMES
DATA AND MODEL VERSIONING	<ul style="list-style-type: none">• Version control data/models/etc locally using DVC• Set up remote storage for use with DVC• Create pipelines and track experiments with DVC
CI/CD	<ul style="list-style-type: none">• Follow software engineering principles by automating, testing and versioning code• Set up Continuous Integration using GitHub Actions• Set up Continuous Deployment using Heroku
API DEPLOYMENT WITH FASTAPI	<ul style="list-style-type: none">• Write an API for machine learning inference using FastAPI• Deploy a machine learning inference API to Heroku• Write unit tests for APIs using the requests module

Nanodegree Program Overview

Course 4: Automated model scoring and monitoring

This course will help students automate the DevOps processes required to score and re-deploy ML models. After model deployment, you will set up regular scoring processes, learn to reason carefully about model drift, and whether models need to be retrained and re-deployed. Students will learn to diagnose operational issues with models, including data integrity and stability problems, timing problems and dependency issues. Finally, students will learn to set up automated reporting with APIs.

Project

A Dynamic Risk Assessment System

In this project, you will make predictions about attrition risk in a fabricated dataset. Begin by setting up processes to ingest data and score, retrain and re-deploy ML models that predict attrition risk while writing scripts that automatically check for new data and model drift. You'll also set up APIs that allow users to access model results, metrics and diagnostics. After completing this project, students will have an end-to-end, automated ML project that performs risk assessments. This project can be a valuable addition to students' portfolios, and the concepts they apply in the project can be applied to business problems across a variety of industries.

LESSON TITLE

LEARNING OUTCOMES

MODEL TRAINING AND DEPLOYMENT

- Ingest data
- Automatically train models
- Deploy models to production
- Keep records about processes
- Automate processes using cron jobs

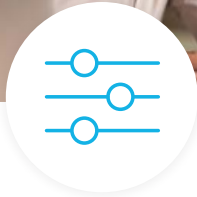
Nanodegree Program Overview



LESSON TITLE	LEARNING OUTCOMES
MODEL SCORING AND MODEL DRIFT	<ul style="list-style-type: none">• Automatically score ML models• Keep records of model scores• Check for model drift using several different model drift tests• Determine whether models need to be retrained and re-deployed
DIAGNOSING AND FIXING OPERATIONAL PROBLEMS	<ul style="list-style-type: none">• Check data integrity and stability• Check for dependency issues• Check for timing issues• Resolve operational issues
MODEL REPORTING AND MONITORING WITH APIS	<ul style="list-style-type: none">• Create API endpoints that enable users to access model results, metrics and diagnostics• Set up APIs with multiple, complex endpoints• Call APIs and work with their results



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

Udacity © 2021

2440 W El Camino Real, #101
Mountain View, CA 94040, USA - HQ

For more information visit: www.udacity.com/enterprise