Course #
Netflix Pathways:
Advanced Data Science Boot Camp
Section 01

Course Information:	**Instructor Information:**	**Office Hours:**
Tuesdays 5:30pm - 8:30pm EDT	Rene Simon	Tuesdays 8:30-9:30pm EST
Thursdays 5:30pm - 8:30pm EDT	Monali Patel	Thursdays 8:30-9:30pm EST
3 Credits	Juilian Simmons	

COURSE DESCRIPTION AND PREREQUISITES

Apply advanced data engineering and machine learning tools and techniques to solve large-scale analytics problems. Design and manage data storage and pipeline solutions with ethical principles. Leverage big data tools and technologies to process data at any scale. Improve business processes and decision making by harnessing state-of-the-art machine learning and natural language processing algorithms.

Pre-requisites:

- Prior completion of Data Science Foundations Bootcamp
- CS Master's degree candidate (or alumni standing) OR current CS major with a data science concentration who demonstrates eligibility by completing course entry assessment.

COURSE RATIONALE

In this course, students will learn advanced data engineering, data science, and machine learning skills. Designed for computer science majors or learners currently enrolled in our Data Science Foundation course, this program will provide a crash course in modern data engineering, data science, and machine learning engineering practices commonly employed by modern day teams.

Over 16 weeks, this course covers essential technical and topics, including database engineering, SQL programming, ETL, big data analytics, and machine learning. This course is intended to broaden students' knowledge of advanced topics to prepare them to build well-crafted, production-ready code. By the end of this course, students will be well prepared to interview for internships and technical roles, and will have gained advanced skills in Python, SQL, ETL, data engineering, machine learning, neural networks and NLP.

By the end of this course, students will complete a series of projects and scenario-based assignments to supplement their professional portfolios. The course is recommended for students seeking future employment in technical roles such as data engineering, machine learning, or data science.

COURSE GOALS AND LEARNING OUTCOMES

Course Goals

By taking this course, students will be able to:

- Design and model data using relational databases.
- Perform ETL and create data pipelines for production.
- Analyze data at any scale with big data.
- Apply machine learning and artificial intelligence to automate data decisions and predictions.

- Create and share code online to showcase your visuals and results

Learning Outcomes

By the end of this course, students should be able to:

- Write SQL commands to perform Create, Read, Update, and Delete commands.
- Perform advanced queries and analytics using SQL.
- Use the ETL process (Extract, Transform, Load) to transform and consolidate data from
- multiple sources.
- Analyze data at any scale with big data tools such as MapReduce, Spark, and PySpark.
- Calculate and apply regression analysis, decision trees, and random forests to datasets.
- Apply feature engineering and selection to improve machine learning performance.
- Evaluate, tune, and optimize machine learning models.
- Design neural networks and deep learning architectures to solve complex problems.
- Apply natural language processing to interpret and utilize language and text for machine learning applications.
- Compare and contrast supervised learning and unsupervised learning
- Identify common problems and resolutions in data ethics
- Collaborate on code using open-source tools like Jupyter Notebooks and Git

Program Educational Outcomes

In partnership with Netflix and 2U, the University offers this course as part of its program mission to provide students with real-world skills to succeed in their future careers. Such skills include the ability to:

- Identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics;
- Plan and assess the impact of engineering solutions in global, economic, environmental, and societal contexts.

COURSE MATERIALS

Students will be provided with lesson plans, activity files, example code, and assignment materials. This course requires use of a laptop or desktop computer, headphones/speakers, video camera, and microphone. Students will be required to install specific software, including:

- Anaconda for Python 3.8
- Git and Git Bash
- Zoom, Slack

PRIMARY METHOD(S) OF INSTRUCTION

Lecture, discussion, demonstrations, and hands-on practice activities.

COURSE OUTLINE

This course runs for 16 weeks and is broken into four units, covering the following topics:

Unit	Description	Objectives

Unit I. Databases and Data Engineering	This module covers data engineering fundamentals including database structures, building and querying SQL databases. Use the Extract, Transform, Load (ETL) process to create data pipelines. The material will cover SQL database concepts using PostgreSQL.	• Use SQL programming to create, read, update, and delete data • Model and design tables and schemas for databases • Use SQL to analyze data. • Use Python and SQL to design data pipelines to extract, transform, and load data into databases
Unit II. Big Data and PySpark	In this unit, students will learn what constitutes big data and how it's handled. We'll start by reviewing Hadoop and its ecosystem. We'll cover MapReduce and how it has improved the process for handling big data. We'll then move on to PySpark, which has become the leading technology for handling big data.	• Define big data and describe the challenges associated with it. • Explain how tools like Hadoop and MapReduce can be used to process big data. • Query data at any scale with Spark and PySpark.
Unit III. Machine Learning	In this unit, students will use Python to build and evaluate several machine learning models to make predictions, provide recommendations, and determine patterns or group data with machine learning algorithms.	• Define the major components of machine learning and how they are applied in industry. • Prepare and train machine learning models for making data predictions and automating decisions. • Evaluate and tune machine learning models for production.
Unit IV. Neural Networks and Natural Language Processing	In this unit students will explore and implement neural networks using the TensorFlow platform in Python. They'll implement neural networks and deep neural networks across a number of different datasets, including image, natural language, and numerical datasets. Finally, they'll learn how to store and retrieve trained models for more robust uses.	• Describe a neural network model and its components. • Implement neural network models using TensorFlow and Keras. • Process text and language using natural language processing.

This outline is subject to change without notice at the discretion of the instructor or the progress of the class.

EVALUATION / ASSESSMENT METHODS
In-class activities, homework assignments, and projects will be used to determine course grades:

Factors in Final Grade	**Percentage of Final Grade**
Attendance and Activity Participation	10%
Homework: Weekly Assignments (x11)	30%
Unit Assessments (x4)	20%
Project 1: Mid Term	20%
Project 2: Capstone	20%

ATTENDANCE
A student is permitted one "unexcused" absence per semester hour of credit.

LATE SUBMISSION POLICY
Homework is due on the specified due date. Please notify the instructor in advance if your timely submission is problematic. Late submissions will be penalized. Products more than one week late will not be accepted.

MAKE-UP EXAM/ASSIGNMENT POLICY
Makeup assignments will not be given unless an approved medical or other absence is confirmed with the instructor. The approval should be sought prior to the due date. Failing assignment scores may be resubmitted at the instructor's discretion.

ACADEMIC HONESTY:

- **Norfolk State University** (Norfolk, Virginia)
- **Talladega College** (Talladega, Alabama):
- **Marymount University** (Arlington, Virginia)
- **Edward Waters College** (Jacksonville, Florida)
- **St. Edward's University** (Austin, Texas)
- **UC Irvine**

2U/TES: "see Enrollment Agreement"