

# MASTERING DATA SCIENCE WITH DEEP LEARNING



University of  
Massachusetts  
Global A nonprofit  
affiliate

*Machine Learning patents grew at a 34% Compound Annual Growth Rate (CAGR) between 2013 and 2017, the third-fastest growing category of all patents granted. Additionally, International Data Corporation (IDC) forecasts that spending on artificial intelligence (AI) and machine learning (ML) will grow from \$12B in 2017 to \$57.6B by 2021.\**

## Course Overview

Mastering Applied Data Science + Deep Learning is a project-driven course that will teach students the practical aspects of Data Science, such as collecting data by web scraping, validation of information in data by data analysis, comparing models created by ML and deep learning (DL) algorithms by interpreted metrics, and more. Additionally, we dive into the more enhanced aspects of data science by introducing topics, such as recommender systems, natural language processing (NLP), DL, and Computer Vision, which are all in everyday applications of AI.

This course is offered through UMass Global's School of Extended Education in partnership with theDevMasters.



theDev  
Masters

theDevMasters is a privately owned, Artificial Intelligence and Data Science Company headquartered in Irvine California. The company is nationally ranked and recognized as a Top 10 Data Science Bootcamp by switchup and Course Report.

## Imagine Your Future In Data Science\*\*

NATIONWIDE

**34,338**

JOB POSTINGS

**↑ 19%**

PROJECTED GROWTH  
(10 YEARS)

**TOP 10  
ORGANIZATIONS HIRING**

- Amazon
- Facebook
- IBM
- Capital One
- Bayer Corporation
- Deloitte
- KPMG
- Anthem Blue Cross
- Accenture
- Booz Allen Hamilton Inc.

**\$116,198** AVERAGE SALARY

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## PROGRAM STRUCTURE

### PRE-WORK

Students will work through a collection of tutorials to be completed prior to the start of Data Science Course. These instructor-led mentoring sessions are meant to cover the basics of Python, Stats, SQL, and Web. This will ensure that all students start from a similar base of knowledge to build upon.

### APPLIED LABS

Our classes provide background and insights while solidifying the ideals through hands-on in-class projects and in depth real life business projects. Students are expected to present their findings and documentations on their journey throughout their project and explain further on what would be their next step to solve the business problem.

### PORTFOLIO DEVELOPMENT

Student Portfolio development (LinkedIn, blog, Github, Masterminds, data storytelling, partner corporations introductions, and networking)

## PROGRAM OVERVIEW

Pre-Work (Python, Stats, SQL, Web)

Student Portfolio development (LinkedIn, blog, Github, Masterminds, data storytelling, partner corporations introductions, and networking)

Machine Learning (Includes Advanced ML with 3 PBL Projects, and 2 Kaggle Challenges)

Deep Learning (Major 3 Libraries: Theano, TensorFlow and Keras, NLP using DL, NLP Times series)

Big Data & AWS Platform (ML Modeling Spark, Splunk, AWS AMI Deep Learning, NLP)

Artificial Intelligence (Recommended Systems, NLP, ChatBot, Deep Learning Fundamentals, Computer Vision)

Project Based Learning PBL (Total 10 projects including Business Focus Deep Learning)

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## TEXTBOOK & COURSE MATERIALS

Grus, J. (2015). Data Science From Scratch: First Principles with Python (First ed.). Sebastopol, CA: O'Reilly Media.

Matthes, E. (2016). Python Crash Course. San Francisco, CA: No Starch Press, Incorporated.

## DOWNLOADS

Anaconda Python 3 (<https://www.anaconda.com>)

## WHO SHOULD ENROLL:

- Managers, Directors, BIs, CTOs, CIOs and Data Analysts
- Graduates looking to capitalize on the growing amount of data
- College students and recent graduates looking for hands-on experience and to fill the gap between school and real world applications.
- This course is open for students of all backgrounds. Pre-Work is required for all students enrolling in the course to assure students are set up for success.

---

## PREWORK

### Python 101

Whether you are familiar with programming or not, our Python PreWork sessions introduce the fundamentals of Python, such as variables, string fundamentals, if-else statements, try & except statements, for loops, while loops, break & continue statements, & lambda functions, as well as certain data types relevant to data science, like lists, tuples, dictionaries, & sets for beginning exposure. The activities done in these sessions will guide students to moving forward.

### Stats 101

The hands-on portion of statistics in PreWork is to establish an understanding of concepts such as mathematical variables, numerical vs categorical, nominal vs ordinal, interval vs discrete; measurements of statistics, when to use mean, when to consider median, & when to revised to mode; relationship between variables, like correlation & independence; ending with hypothesis testing & p-value, but only to the degree of applying the mindset towards data science. These concepts will be reviewed in Day 2 of the program to ensure student's clarifications are addressed.

### Web 101

An introduction to HTML & CSS is key to future project building & publication of a blog students' write tracking their progress throughout the program. The access of relevant open source data is available throughout the web for people to utilize. Using the most open source methods, like HTML & CSS, to grab that information within our Python environments will be introduced in Day 3. Once students are in Project Based Learning, GitHub portfolios are best displayed in themes that students choose & customize with HTML & CSS.

### SQL 101

While some of the tools used in Python will take the place of SQL functions & methods, it is still beneficial to understand the origins of these tools as well as be able to replicate them when applied in future work's expectations. A solid portion of demand in data science jobs ask for big-query experience with SQL, such as: Microsoft SQL & PostgreSQL vs NoSQL, MongoDB & DynamoDB, which students will learn scenarios to further solidify their application to the business problems.

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## APPLIED LABS

### SESSION 1: Introduction to Data Science with Python

In our first class, we will go over some intermediate functions in Python as review & move onto introducing what is the expected mindset of a data scientist versus the traditional viewpoint & how to take full advantage of the program by using the Applied Labs environment. We will encourage students to introduce themselves to each other & gather each other's strengths, along with the instructor's experience to not only grasp the skills & tools a data scientist is expected to know, but know exactly when to use which tools & why through peer & real-life learning. There will also be an introduction to the CRISP-DM data science methodology & chosen framework with the distinctions between the two mindsets of machine learning: supervised learning & unsupervised learning. The session has two miniature projects, Temperature & Christmas, to wrap up Python essentials.

**Project 1:** How Much Longer Until Christmas?

---

### SESSION 2: Exploratory Data Analysis

We start by asking the questions that data science can help answer for students to identify the difference between a data analytical question vs a data science question. We further breakdown what are the key checklist items in form of questions that CRISP-DM individual stages require before moving further in the cycle. We again showcase the distinctiveness between supervised learning & unsupervised learning & explain why sometimes supervised learning is the method that most of us will encounter, but unsupervised learning will elaborate more patterns in data than we can ever imagine. We introduce the self-checking mindset of what is considered good data for data science projects: what is good data & how can we detect bad data from good data, & we let the students ponder how we can tackle dirty data. We then give the attributes to help students identify big data

from small data through the Four V's. A small review on what are the differences between mathematical variables, numerical vs categorical along with a short case of where statistics are required the most in data science: the data analysis phase. The hands-on portion of the class familiarizes students with NumPy and Pandas and showcasing how to clean, manipulate, and analyze data by applying those concepts. Students will be given the data set for Titanic, a Kaggle competition known for introductory data science methods & cleaning, practicing data analysis skills on the Titanic dataset with Pandas to get students in the data science mindset of result-oriented, instead of process-oriented.

**Project 2:** Exploration of Titanic

---

### SESSION 3: Data Visualizations & Information Analysis

We start off by asking what's the purpose of visualization in data science, broadening on student's experiences with charting & decision making with charts. A review of NumPy functions for generating different types of data is done before a brief introduction to Matplotlib's figure attributes & properties. Instructors will continue with explaining what are the most common analysis-based visuals, such as histograms & scatterplots. An intermediate approach to Titanic is used for exercises with graphing in Matplotlib & analyzing whether the graph is deemed useful or not. We continue with creating a Python-based method for web-scraping & introduction to JSON. There are further functions & helpful tips to consider analyzing data with Pandas, such as common Excel functions implemented to insights. The day ends with a project on what happened during the 2012 election & whether the data of polls can give us clues into who was more likely to win. A GitHub repository is expected to be created by the end of this session & students will learn how to create their own blog & begin to publish content.

**Project 3:** Election Day Results

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## SESSION 4: Machine Learning

We will review by explaining the difference between supervised learning and unsupervised learning, asking students why certain scenarios will not be effective for supervised learning. Furthermore, an explanation on the two result-oriented methods of supervised learning, regression & classification are introduced. The day is dedicated to determining a regression problem, immediate analysis to modeling using regression methods, assessing the models, then optimizing for the best results by different metrics. Afterwards, students will work on building one of the regression models introduced, such as linear, polynomial, ridge, lasso, gradient, robust, & an introduction to logistic regression for classification. The day end with a Kaggle based project using regression.

**Project 4:** Optimizing House Price Predictions

---

## SESSION 5: Advance Machine Learning

Revisiting the results students derived for their House Pricing project we will give more hints & clues to how to approach the project further. We will then dive into the second supervised learning

need: classification algorithms, such as Naive Bayes, Decision Trees, Random Forest, and other methods based on regression. Students are expected to be able to identify when a certain algorithm should be used based on the data, which methods to use to optimize classification algorithms, what is appropriate for insights & decision making. Students will also learn metrics such as R-squared, MSE & RMSE, & scoring using precision, recall, sensitivity, specificity, accuracy score, AUC, ROC, along with gains & lift charts. The session ends with a Spam Classifier project, which eludes to the processes of Natural Language Processing.

**Project 5:** Classification of Spam Emails

---

## SESSION 6: Hack Day

Students will be separated into groups to practice their skills, emphasizing on visualization & modeling with machine learning, with a live Kaggle competition. During this time working with others, students will also be encouraged to identify the gaps in their skills, especially in analysis & modeling, in the project & review as much as possible moving forward to other projects in the continual sessions.

**Project 6:** Baseline Kaggle Competition

## PROJECT BASED LEARNING – LEVEL 1

### Project 1: Skill Assessment

The first project is designed to solidify the skills that are crucial to any data science project. This project will let students understand where they need more review in.

### Project 2: End-to-End Development

In this project, students will complete a data science project to the deployment stage; their results will be displayed in a report from an acceptable medium, an app or a webpage.

### Project 3: Domain-Specific Project

Students are given the option of choosing a project in a domain from below.

- Real Estate
- Ecommerce
- Telecommunications
- Sports
- Online Gambling
- Finance
- Insurance Healthcare

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## SESSION 7: Recommender Systems

Students will review machine learning algorithms and be introduced to types of recommender systems, such as collaborative filtering with k-nearest, using either items or users, similar to Amazon's. Then students will start by building their own recommender system with the MovieLens dataset, elaborating on what to consider as the best method for selection & integrating with what viewers of recommender results will use best; understanding dimension reduction with PCA, principle component analysis; explore SVM, support vector machines; and learn A/B Testing with T-Tests and P-Value methods.

**Project 7:** MovieLens Through Recommendations

---

## SESSION 8: Natural Language Processing and Sentiment Analysis

Students will explore the Natural Language Toolkit to process and extract text data: learning about tokenization of words & sentences, part-of-speech tagging & stemming with lemmatization for the best analysis of textual data. Students will then start a Natural Language Processing project with Yelp data before we move onto Sentimental Analysis to predict positive versus negative Yelp reviews.

**Project 8:** Yelp Reviews & the Truth from Customers

---

## SESSION 9: Big Data with Spark & Splunk

Students will be introduced to Big Data and data engineering with the Hadoop ecosystem, the MapReduce paradigm, Apache Spark, and the up-and-coming Splunk, where real-time data is represented in a dashboard format for easier assessment. An existing project, such as MovieLens, will be transferred to AWS to expose students to the difference.

**Project 9:** MovieLens Through Big Data & Splunk

---

## SESSION 10: Deep Learning and Time Series

Instructors will make sure that student's understanding of unsupervised learning & supervised learning is re-clarified & where does deep learning come in. We will be introducing deep learning through Tensorflow, training neural network, and visualizing what a neural network has learned using TensorFlow Playground. Students will also learn time series, what makes them special, loading and handling time series in Pandas. Students will understand how seasonality affects trends. Projects for this session include handwriting recognition & digital face recognition.

**Project 10:** Hand-writing Recognition

---

## SESSION 11: Computer Vision with OpenCV and Hack Project

After initial installation, we will expand on the notion why letting computers understand images is harder said than done when compared to the way humans & eyes process images. Then, students will be introduced to computer vision fundamentals using OpenCV to detect faces, people, cars, and other objects, even when images are manipulated in rotations or scaling situations. Projects will use sensors such as student's webcam to create a real-time facial recognition program & object recognition program.

**Project 11:** Facial Recognition

---

## SESSION 12: Hack Day

In the last session, we will host a private Kaggle competition amongst the students. Students will be grouped into teams and will showcase their group project at the end of class. Students will be assessed based on their presentation skills, as well as their ability to solve the business problem while utilizing the skills they've developed thus far.

**Project 12:** Private Kaggle Competition

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## PROJECT BASED LEARNING – LEVEL 2

### Project 1: Skill Assessment

The first project is designed to solidify the skills that are crucial to any data science project. This project will let students understand where they need more review in.

### Project 2: End-to-End Development

In this project, students will complete a data science project to the deployment stage; their results will be displayed in a report from an acceptable medium, an app or a webpage.

### Project 3: Domain-Specific Project

Students are given the option of choosing a project in a domain from below.

- Real Estate
- Ecommerce
- Telecommunications
- Sports
- Online Gambling
- Finance
- Insurance Healthcare

## ADVANCE EXPERTISE USING DEEP LEARNING

### Deep Learning mastering Theano, TensorFlow and Keras libraries

Develop large models on GPUs cheaply in the cloud

Crash course in Multilayer Perceptrons

Develop your first Neural Network with Keras

Evaluate the performance of Deep Learning Models

Use Keras Models with Scikit-Learn for General Machine Learning

### Advanced Multilayer Perceptrons and Keras

Save your models for later with Serialization

Keep the best models during training with Check Pointing

Understand model behavior during training by Plotting History

Reduce overfitting with Dropout Regularization

Lift performance with learning Rate Schedules

### Convolutional Neural Networks

Course in Convolutional Neural Networks

Project: Object Recognition in Photographs

### Recurrent Neural Networks

Course in Recurrent Neural Networks

---

# MASTERING DATA SCIENCE WITH DEEP LEARNING

## DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING

### Develop Deep Learning Models with Keras

Develop a Neural Bag-of-Words Model for Sentiment Analysis  
Develop Word Embeddings with Gensim

### Neural Language Modeling

Develop a Word-Based Neural Language Model  
Neural Image Caption Generation  
Neural Network Models for Caption Generation  
Load and Use a Pre-Trained Object Recognition Model

### Machine Translation

Neural Machine Translation  
Encoder-Decoder Models for Neural Machine Translation  
Configure Encoder-Decoder Models for Machine Translation

## DEEP LEARNING FOR TIME SERIES FORECASTING

### Taxonomy of Time Series Forecasting Problems

Develop a Skillful Forecasting Model  
Transform time series to a Supervised Learning Problem

### Simple and Classical Forecasting Methods

Prepare time series data for CNNs and LSTMs

### Develop LSTMs for Time Series Forecasting

Grid search Deep Learning Models for Univariate Forecasting

## PROJECT BASED LEARNING – DEEP LEARNING

### Business Focus Deep Learning Level 1 Deployment

Using Theano, TensorFlow, and Keras libraries  
Project 1: Object Recognition in Photographs  
Project 2: Sentiment from Business Reviews Box

### Business Focus Deep Learning Level 2 Deployment

Deep Learning with Natural Language Processing  
Project 1: Develop a Neural Language Model for Text Generation  
Project 2: Prepare a Photo Caption Dataset for Modeling  
Project 3: Develop a Neural Image Caption Generation Model

### Business Focus Deep Learning Level 3 Deployment

Deep Learning with Time Series  
Project 1: Explore Household Energy Usage Data  
Project 2: Text Generation with Alice in Wonderland