

Top 20 Data Science MOOCs

On August 16, 2015 September 5, 2015 By devendradesale In Data Science

By Devendra Desale.

Data science is big landscape and self-learning is the necessary skill if anyone wants to become a good data scientist. MOOCs had been Major source of treasure for the data scientist. Though there are many sites offering MOOCs, but Coursera (<https://www.coursera.org/>), Edx (<https://www.edx.org/>) and Udacity (<https://www.udacity.com/>) have been leaders. Whether, your language is R, python, Java or C/C++ we have captured all of them. If, you are a beginner and understanding what data science is exactly or you are an expert looking for your next frontiers. You can search through this exhaustive list as per needed.

In this post we have kept only single courses, will write a separate post for the specializations or degrees related to data science. There are some upcoming promising courses in this list too.

Some general guidelines about the source details:

-
- The level of the course is decided by considering the prerequisites, the efforts required and duration of the course.
-
- All courses assume basic background in the statistics.
-
- The courses are arranged w.r.t. level of expertise, i.e. beginners courses are listed ahead of expert level courses.
-
- The tools are considered as a programming language, or software tools used in the course.
-

The Analytics Edge (<https://www.edx.org/course/analytics-edge-mitx-15-071x-0>) (MIT)

Level: Beginners-Expert

Effort: 10-15 hrs/week

Status: Achieved

Duration: 12 weeks

Prerequisite: None

Tools: R

One of the best course to learn data science and analytics using R. The course provides in-depth lectures on multiple business cases, along with extensive exercises. Keep in mind, it is a very demanding course in terms of time commitment, but it is worth. The examples include Moneyball, eHarmony, the Framingham Heart Study, Twitter, IBM Watson, and Netflix. Through these examples and many more, we will teach you the following analytic methods: linear regression, logistic regression, trees, text analytics, clustering, visualization, and optimization.



Machine Learning (<https://www.coursera.org/learn/machine-learning/home/info>) (Stanford University)

Level: Beginners-Expert

Effort: 7-12 hrs/week

Status: On-demand

Duration: 11 weeks

Prerequisite: Programming

Tools: Octave

Whenever you will listen about the machine learning MOOCs, this course has to be there. Excellent course taught by one of the best professors in machine learning domain, Andrew Ng. The way complete course is well-organized and covers all core concepts of machine learning. Topics include: (i) Supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks). (ii) Unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning). (iii) Best practices in machine learning (bias/variance theory; innovation process in machine learning and AI).

Data Science and Machine Learning Essentials (<https://www.edx.org/course/data-science-machine-learning-essentials-microsoft-dat203x>) (Microsoft) (24 Sep 2015 onwards)

Level: Beginners-Intermediate
Status: Upcoming
Prerequisite: None

Effort: 3-4 hrs/week
Duration: 5 weeks
Tools: R

Learn data science essentials with experts from M.I.T and the industry, partnering with Microsoft to help develop your career as a data scientist. By the end of this course, you will know how to build and derive insights from data science and machine learning models. You will learn key concepts in data acquisition, preparation, exploration and visualization along with examples on how to build a cloud data science solution using the Azure Machine Learning, R & Python. This course is organized into 5 weekly modules, each concluding with a quiz.

Databases (<https://lagunita.stanford.edu/courses/Home/Databases/Engineering/about>) (Stanford University)

Level: Beginners
Status: Self-paced
Prerequisite: None

Effort: 8-10 hrs/week
Duration: 10 weeks
Tools: SQL, XML query

If you are dealing with data, databases are inevitable. This course covers database design and the use of database management systems for applications. It includes extensive coverage of the relational model, relational algebra, and SQL. It also covers XML data, including DTDs and XML Schema for validation, and the query and transformation languages XPath, XQuery, and XSLT. The course includes database design in UML, and relational design principles based on dependencies and normal forms.

Coding the Matrix: Linear Algebra through Computer Science Applications
(<https://www.coursera.org/course/matrix>) (Brown University)

Level: Beginner-Intermediate
Status: Achieved
Prerequisite: None

Effort: 10-14 hrs/week
Duration: 10 weeks
Tools: Python

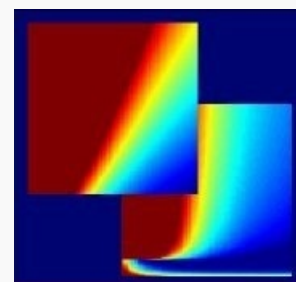
Linear algebra is one the important building block of not only computer science, but also machine learning, graphics and statistics. This is a brilliant course guides you through the real examples and excellent python assignments. You will write programs to implement basic matrix and vector functionality and algorithms, and use these to process real-world data to achieve such tasks as: two-dimensional graphics transformations, face morphing, face detection, image transformations such as blurring and edge detection, image perspective removal, classification of tumors as malignant or benign, integer factorization, error-correcting codes, and secret-sharing. Another, more basic course is [LAF](http://www.ulaff.net/) (<http://www.ulaff.net/>) by The University of The Texas Austin.

Learning From Data (<http://work.caltech.edu/telecourse.html>) (California Institute of Technology)

Level: Intermediate-Expert
Status: Archived
Prerequisite: probability, matrices, calculus

Effort: 10-14 hrs/week
Duration: 10 weeks
Tools: No restriction

One of the best MOOC ever for machine learning enthusiasts. This is an introductory course in machine learning (ML) that covers the basic **theory, algorithms, and applications**. But requires one to have good linear algebra, calculus and probability background, along with coding skills. The course is taught by [Yaser S. Abu-Mostafa](http://work.caltech.edu/) (<http://work.caltech.edu/>), who is a Professor of Electrical Engineering and Computer Science at the California Institute of Technology. He is the co-author of Amazon's machine learning bestseller Learning From Data and great professor who simplifies the learning.



CSCI E-109 Data Science (<http://cs109.github.io/2015/>) (Harvard Extension School)

Level: Beginners-Expert
Status: Achieved
Prerequisite: None

Effort: 7-12 hrs/week
Duration: 16 weeks
Tools: Python, d3

Excellent course, recommended to all the data science aspirants. This course introduces methods for five key facets of an investigation: data wrangling, cleaning, and sampling to get a suitable data set; data management to be able to access big data quickly and reliably; exploratory data analysis to generate hypotheses and intuition; prediction based on statistical methods such as regression and classification; and communication of results through visualization, stories, and interpretable summaries.

Introduction to Data Science (<https://www.coursera.org/course/datasci>) (University of Washington)

Level: Beginner-Intermediate
Status: Achieved
Prerequisite: Programming

Effort: 10-14 hrs/week
Duration: 10 weeks
Tools: Python, R, SQL

Introduce yourself to the basics of data science and leave armed with practical experience extracting value from big data. This course teaches the basic techniques of data science, including both SQL and NoSQL solutions for massive data management (e.g., MapReduce and contemporaries), algorithms for data mining (e.g., clustering and association rule mining), and basic statistical modelling (e.g., linear and non-linear regression).

Networks, Crowds and Markets (<https://www.edx.org/course/networks-crowds-markets-cornellx-info2040x#!>) (Cornell University)

Level: Beginners-Expert
Status: Achieved
Prerequisite: None

Effort: 4-8 hrs/week
Duration: 10 weeks
Tools: None

The course examines the interconnectedness of modern life through an exploration of fundamental questions about how our social, economic, and technological worlds are connected. Students will explore game theory, the structure of the Internet, social contagion, the spread of social power and popularity, and information cascades. Another important source of knowledge for link analysis is [SNAP \(https://snap.stanford.edu/data/index.html\)](https://snap.stanford.edu/data/index.html).

Data Analysis: Take It to the MAX() (<https://www.edx.org/course/data-analysis-take-it-max-delftx-ex101x>) (DelftX) (1 Sep 2015 onwards)

Level: Intermediate
Status: Upcoming
Prerequisite: Basic Spreadsheet exp.

Effort: 4-6 hrs/week
Duration: 8 weeks
Tools: MS-Excel, python

Even in the era of the big data, there is a huge number of data analyst who rely heavily on the spreadsheets to gather the insights and its still relevant. This is an excellent course for those who want to enhance analytical skills using excel. You will take a deep dive into data analysis with spreadsheets: PivotTables, VLOOKUPS, Named ranges, what-if analyses, making great graphs - all those will be covered in the first weeks of the course. After that, you will investigate the quality of the spreadsheet model, and especially how to make sure your spreadsheet remains error-free and robust. Finally, you will also look into how Python, a programming language, can help us with analyzing and manipulating data in spreadsheets.

Text Mining and Analytics (<https://www.coursera.org/course/textanalytics>) (University of Illinois at Urbana-Champaign)

Level: Intermediate-Expert
Status: Achieved
Prerequisite: Programming

Effort: 5-10 hrs/week
Duration: 5 weeks
Tools: C++

This course will cover the **major techniques for mining and analyzing text data** to discover interesting patterns, extract useful knowledge, and support decision making, with an emphasis on statistical approaches that can be generally applied to arbitrary text data in any natural language with no or minimum human effort. You will learn the basic concepts, principles, and major algorithms in text mining and their potential applications.



Mining Massive Datasets (<https://www.coursera.org/course/mmds>) (Stanford University) (Sep 12 - Oct 31, 2015)

Level: Expert
Status: Upcoming
Prerequisite: Calculus, data structure

Effort: 8-10 hrs/week
Duration: 7 weeks
Tools: C++ or Java

If you are planning to develop your next distributed system which is suppose to turn big data into insights, this is the course you should be taking. The course covers MapReduce, PageRank, locality-sensitive hashing, and many more advance algorithms. It is an extensive course involves a great amount of dedication and coding. Though coding assignments are optional it is highly recommended to complete them.



Neural Networks for Machine Learning (<https://www.coursera.org/course/neuralnets>) (University of Toronto)

Level: Intermediate-Expert
Status: Achieved
Prerequisite: None

Effort: 7-9 hrs/week
Duration: 8 weeks
Tools: Octave

If you want to explore the current "hot topic" deep learning, you should explore this course. Taught by the Prof. Goffrey Hilton, whose research has been revolutionizing the field. The course covers all parts right from the perceptron till the auto-encoders. The course will explain the new learning procedures that are responsible for current advances in the field of neural network, including effective new procedures for learning multiple layers of non-linear features, and give you the skills and understanding



required to apply these procedures in many other domains. If you want to learn deep learning in-depth, consider following these courses: [CS224d: Deep Learning for Natural Language Processing](http://cs224d.stanford.edu/index.html) (<http://cs224d.stanford.edu/index.html>) and Nvidia's [Deep Learning Courses](https://developer.nvidia.com/deep-learning-courses) (<https://developer.nvidia.com/deep-learning-courses>).

Convex Optimization

(<https://lagunita.stanford.edu/courses/Engineering/CVX101/Winter2014/about>) (Stanford University)

Level: Intermediate-Expert

Status: Achieved

Prerequisite: Probability, Optimization

Effort: 8-10 hrs/week

Duration: 10 weeks

Tools: Matlab

Advance course, if you are interested in optimization problems. This course concentrates on recognizing and solving convex optimization problems that arise in applications. The syllabus includes: convex sets, functions, and optimization problems; basics of convex analysis; least-squares, linear and quadratic programs, semidefinite programming, minimax, extremal volume, and other problems; optimality conditions, duality theory, theorems of alternative, and applications; interior-point methods; applications to signal processing, statistics and machine learning, control and mechanical engineering, digital and analog circuit design, and finance.

Process Mining: Data science in Action (<https://www.coursera.org/course/procmin>) (Eindhoven University of Technology) (7 Oct- 2 Dec 2015)

Level: Intermediate-Expert

Status: Upcoming

Prerequisite: None

Effort: 4-6 hrs/week

Duration: 8 weeks

Tools: Prom, Disco

Process mining is the missing link between model-based process analysis and data-oriented analysis techniques. Through concrete data sets and easy to use software the course provides data science knowledge that can be applied directly to analyze and improve processes in a variety of domains. The course explains the key analysis techniques in process mining. Participants will learn various process discovery algorithms. These can be used to automatically learn process models from raw event data.

Bioinformatic Methods I & II (<https://www.coursera.org/learn/bioinformatics-methods-1>) (University of Toronto)

Level: Intermediate-Expert

Status: On-demand

Prerequisite: None

Effort: 12-18 hrs/week

Duration: 10 weeks

Tools: No restriction

Good course to cover all your basics in Bioinformatics. This is a two part course deals with databases, Blast, multiple sequence alignments, phylogenetic, selection analysis and metagenomics. Later, in part II, it covers motif searching, protein-protein interactions, structural Bioinformatics, gene expression data analysis, and cis-element predictions.

Model Building and Validation (<https://www.udacity.com/course/model-building-and-validation--ud919>) (AT & T)

Level: Intermediate-Expert

Status: Open

Prerequisite: ML, modelling, python

Effort: 6 hrs/week

Duration: 8 weeks

Tools: Python, SQL

In this course you will take a more general approach, walking through the questioning, modelling and validation steps of the model building process. The goal is to get you to practice thinking in depth about a problem and coming up with your own solutions. Many examples we will attempt may not have one correct answer but, will require you to work through the problems applying the methods we hope to illustrate throughout this class.

Intro to Hadoop and MapReduce (<https://www.class-central.com/mooc/1470/udacity-intro-to-hadoop-and-mapreduce>)(Cloudera)

Level: Beginners-Intermediate

Status: Open

Prerequisite: None

Effort: 6-10 hrs/week

Duration: 4 weeks

Tools: Java, Apache Hadoop

Learn the fundamental principles behind it, and how you can use its power to make sense of your Big Data. How Hadoop fits into the world (recognize the problems it solves). Understand the concepts of HDFS and MapReduce (find out how it solves the problems). Write MapReduce programs using java and apache Hadoop.

Real-Time Analytics with Apache Storm (<https://www.udacity.com/course/real-time-analytics-with-apache-storm--ud381>) (Tweeter)

Level: Intermediate-Expert
Status: Open
Prerequisite: Data Structures

Effort: 6-10 hrs/week
Duration: 4 weeks
Tools: Java, python, d3

Realtime analytics picking up its pace, and this is the course which specifically explains the details, challenges and solutions. Starting from basic distributed concepts presented during our first Udacity-Twitter Storm Hackathon, link Storm concepts to Storm syntax to scalably drive Word Cloud visualizations with Vagrant, Ubuntu, Maven, Flask, Redis, and d3. Learn how to link to the public Twitter gardenhose stream to process live tweets, parse embedded URLs, and calculate Top worldwide hashtags. Extend beyond Storm basics by exploring multi-language capabilities in Python, integrate open source components, and implement real-time streaming joins.

Introduction to Recommender Systems (<https://www.coursera.org/learn/recommender-systems>) (University of Minnesota)

Level: Beginners-Intermediate
Status: Self-paced
Prerequisite: None

Effort: 8-10 hrs/week
Duration: 8 weeks
Tools: No restriction

Retrieval systems are widespread in current softwares, whether it is web search, movie recommendation, document searching. The algorithms you will study include content-based filtering, user-user collaborative filtering, item-item collaborative filtering, dimensionality reduction, and interactive critique-based recommenders. The approach will be hands-on, with six week projects, each of which will involve implementation and evaluation of some type of recommender.

[About these ads \(<https://wordpress.com/about-these-ads/>\)](https://wordpress.com/about-these-ads/)

Data ScienceMOOC

One thought on “Top 20 Data Science MOOCs”

- 1.
2. Pingback: [Distilled News | Data Analytics & R](#)
- 3.

[Blog at WordPress.com.](#) & [The lovecraft Theme.](#)

%d bloggers like this: