



LEARN DATASCIENCE IN 2023

COMPLETE ROADMAP

5 MONTHS

ZERO MONEY NEEDED

All Free

Who is the target audience and who can benefit the most from this roadmap?

- This roadmap is designed to benefit the following individuals:
 - Recent college graduates interested in learning Data Science and Machine Learning.
 - Individuals who have previously worked in a different field and want to transition to Data Science or Machine Learning.
 - Individuals from a Data Engineering or Data Analytics background who are interested in pursuing Data Science and Machine Learning

Are there any prerequisites?

- No hard prerequisites for this roadmap. However, a basic understanding of high school level mathematics is preferred. If you do not have this background, you can learn it along the way.
- It is also preferable to have some prior experience with any programming language. If you have no previous programming experience, we suggest looking for beginner-level knowledge on YouTube. Understanding basic concepts such as data types, loops, conditions, functions, compilation, and execution would be sufficient.

What you need to bring to the table?

- You should be able to dedicate 2-3 hours every day.
- Dedication, hard work, and perseverance.
- Will to learn and succeed.

- Introduction to Python for Data Science
 - Python Introduction
 - Python , Jupyter and Anaconda Installation
 - Wlkthrough of Jupyter and Anaconda
 - Python Syntax and Semantics
 - Python Data Structures
 - Python Functions and Lambdas
 - Python List comprehension and generators
 - Python regular expressions
 - Python File IO and Resource Operations
- Advanced Python for Data Science
 - Numpy deepdive
 - Pandas deepdive
 - Matplotlib & Seaborn – Visualization
 - Common Modules and Packages

Corey Shafer	https://www.youtube.com/watch?v=YYXdXT2l-Gg&list=PL-osiE80TeTt2d9bfVyTiXJA-UTHn6WwU *Pay attention to topics which are listed above. Corey has lot of videos which are more intended to full stack developers.
MLWithAP	https://www.youtube.com/@MLWithAP/playlists *_ My channel. Basics of Python, Pandas,Numpy and Matplotlib are covered here.

- Mathematics for Machine Learning (2 weeks)
 - Statistics – Descriptive and Inferential
 - Central Tendency - Mean, Median, Mode
 - Dispersion – Std Dev , Variance
 - Hypothesis Test
 - Types of Distributions
 - Central Limit Theorem
 - Correlation
 - Linear Algebra – Vectors and Matrices
 - Scalars, Vectors and n-dimensional matrices (Tensors)
 - Dot product, Transpose and Inverse of Matrices
 - Determinants, Eigen Value and Eigen Vectors
 - Calculus & Probability
 - Differential and Integral Calculus
 - Limit, Continuity and Partial derivatives
 - Step, Sigmoid, Logit, and ReLU Function
 - Maxima and Minima of a Function
 - Product and Chain Rule
 - Bayes Theorem
 - Conditional Probability

Statistics	Organic Chemistry Tutor	https://www.youtube.com/playlist?list=PL0o_zxa4K1BVszilRdfv4HI4UlqDZhXWV
Calculus	Khan Academy	https://www.khanacademy.org/math/calculus-1
All Maths	Imperial College London	https://www.youtube.com/playlist?list=PLiiljHvN6z193BBzS0Ln8NnqQmzimTW23
Probability	Khan Academy	https://www.khanacademy.org/math/statistics-probability
Both Khan Academy and Organic Chemistry Tutor are very good resources. If you get sometime, you can check out my playlist on Descriptive Statistics as well.		

Machine Learning – 1

5 Weeks

- Exploratory Data Analysis
 - Data Gathering and Extraction
 - SQL (MySQL). NoSQL (Mongo)
 - CSV and JSON files
 - Web Crawling /Scraping / APIs / Beautiful Soup
 - Cloud - S3
 - Unstructured data (Video, Speech , Text)

- Data Preprocessing & Analysis
 - Univariate, Bivariate and Multivariate Analysis
 - Outliers and Anomalies detection
 - Data Cleansing - Null value, Imputations, Duplicate treatment
 - Dispersion and Distribution
- Data Split
 - Train, Test, Validation set
 - Sampling technique and Stratification Strategy
 - Bias and Variance trade off
 - Handling Bias and Imbalanc

- Introduction to Machine Learning
 - Machine learning Landscape
 - Machine learning end to end project lifecycle
 - Regression Vs Classification
 - Supervised Vs Unsupervised
 - Batch Vs Online
 - Linear Regression
 - Multi Variable Linear Regression
 - Logistic Regression
 - Scikit learn and Statsmodel
- Feature Engineering and Feature Selection
 - Feature Engineering
 - New feature creation
 - Variable transformation
 - Feature Encoding
 - Handling Categorical and Numerical features
 - Binning

- Feature Selection
 - PCA
 - Dimensionality Reduction techniques
 - Multicollinearity
 - Forward/Backwar/Stepwise selection
 - Lasso
 - Filter/Wrapper/Embedded
- Feature Scaling
 - Standardization
 - Normalization

EDA	Edx	https://www.edx.org/course/analyzing-data-with-python
EDA	Coursera	https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning
ML	Josh Starmer	https://www.youtube.com/@statquest
ML	Andrew Ng	https://www.coursera.org/specializations/machine-learning-introduction

Side notes:

- Andrew Ng's courses are highly rated courses. Spend most of your time over here.
- Josh Starmer Youtube playlist is a quick and fun way to learn algorithms. It doesn't go very deep but gives you a good understanding of what algorithm is all about.
- EDA is super important. I have listed 2 resources. But skim through the internet(Medium, Towards Data Science and Youtube) if you are stuck or want to know more about a given topic.

- Advanced Supervised Learning
 - Naïve Bayes Classifier
 - k-NN Classifier
 - Support Vector Machines (Regressor and Classifier)
 - Ensemble Techniques
 - Decision Tress
 - Bagging
 - Random Forest
 - Boosting
- Model Selection and Tuning
 - Hyper Parameter Tuning
 - Model Performance measures
 - Bias and Variance
 - Overfitting vs Underfitting
 - Cross validation
 - GridSearchCV Vs RandomizedSearchCV
 - Regularization – L1 and L2
 - Pipelining
- Unsupervised Learning
 - K-means clustering
 - KNN
 - Hierarchical Clustering

- Anomaly detection
- Dimensionality Reduction Techniques / PCA
- SVD
- DBSCAN
- Production Deployment.
 - Deployment scenario and strategies
 - ML Pipeline
 - Flask and Heroku
 - Introduction to FastAPI
 - Deployment to AWS ECS
 - Monitoring and Continuous performance measure
 - MLOps

ML	Standford CS229	https://www.youtube.com/watch?v=jGwO_UgTS7I&list=PLoROMvody4rMiGQp3WXShMGgzqpfVfbU
ML	Kaggle	https://www.kaggle.com/learn
ML	Google	https://developers.google.com/machine-learning/crash-course

Data Engineering (Overview)

3 Weeks

- Data warehousing Concepts
- ETL (Extract, Transform, Load)

- Big data processing Hadoop, Spark
- Data modelling
- Fundamentals of Cloud computing: Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure
- Data quality
- Database management : One SQL and One NoSQL databases (MySQL and Mongo)
- Data streaming
- Data visualization - Tableau of PowerBI Model Selection and Tuning

There are various resources spread on internet. I have put some, but I would request you to take a look at Medium blog posts for individual topics .

You would only need foundational knowledge of this section.

Data Engineering	Coursera	https://www.coursera.org/learn/big-data-integration-processing?action=enroll&specialization=big-data
SQL	Khan Academy	https://www.youtube.com/watch?v=90NJg_XLY0Y&list=PLSQL0a2vh4HCtgJXRDRah4IIKPfehlhii
Data Engineering	Marc	https://www.youtube.com/@MarcLamberti/videos

- Neural Network and Deep Learning Fundamentals
 - Perceptron
 - Activation and loss function
 - Gradient Descent
 - Batch Normalization
 - Introduction to TensorFlow and Keras
 - Transfer learning and regularization
- Computer Vision (2 Weeks)
 - CNN
 - Convolution, Pooling and Padding
 - CNN architectures and ImageNet Challenge
 - Object Detection
 - Semantic Segmentation

- Natural Language Processing
 - RNNs
 - Tokenization, Stemming and Lemmatization
 - LSTMs and GRUs
 - Time Series analysis
 - Advance Language Models – BERT, GPT3
 - Attention is all you need

- Autoencoders and GANs
 - Generative Network and Adversarial Network
 - Variational Autoencoders
 - Convolution and DCGAN
 - Application of GANs

- Reinforcement Learning
 - RL framework
 - Markov Chain
 - Policy Gradient Methods
 - Type of RL systems
 - Q Learning

DL	Stanford CS230	https://www.youtube.com/watch?v=PySo_6S4ZAg&list=PLoROMvody4rOABXSygHTsbvUz4G_YQhOb
----	----------------	---

DL	Coursera	https://www.coursera.org/specializations/deep-learning
DL	MIT 6S.191	https://www.youtube.com/watch?v=7sB052Pz0sQ&list=PLtBw6njQRU-rwp5_7C0olVt26ZgjG9NI
DL	Stanford CS231n	https://www.youtube.com/watch?v=vT1JzLTH4G4&list=PLC1qU-LWwrF64f4QKQT-Vg5Wr4qEE1Zxk&index=2

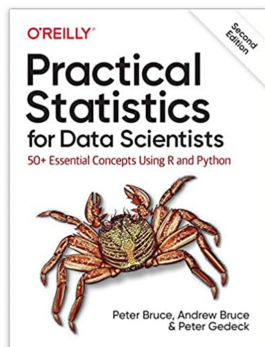
Side Note

- Deep learning can be exhausting and challenging. Whole ML/DL field is evolving and it can become daunting to cover everything. DO NOT let FOMO come in. Remember, you are in this for the long haul.
- Andrew's course on Coursera is very good. Devote most of the time there to that.

Books:

There are quite a few books in the market which you can refer. I am mentioning which I have personally used and gained from.

If you can spend some money, these books are completely worth it.



See all 2 images

Practical Statistics for Data Scientists, 2e: 50+ Essential Concepts Using R and Python Paperback – 29 June 2020

by Peter Bruce (Author), Andrew Bruce (Author), Peter Gedeck (Author)

★★★★☆ 763 ratings

See all formats and editions

Kindle Edition
₹2,634.93

Paperback
₹3,941.00 ✓ prime

Read with Our Free App

12 New from ₹3,941.00

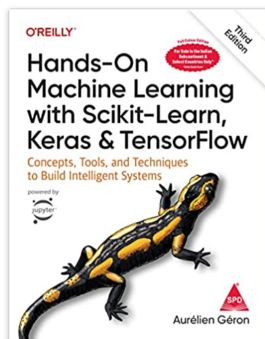
EMI starts at ₹188. No Cost EMI available EMI options

Save Extra with 4 offers

Cashback: ₹200 cashback & ₹1,800 welcome rewards on Amazon Pay ICICI Credit Card . 5% cashback every time on shopping. Apply now! T&C. Not applicable on Amazon Business Transactions |Details

No Cost EMI: Avail No Cost EMI on select cards for orders above ₹3000 |Details

See 2 more



Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Third Edition (Full Colour Print) Paperback – 10 October 2022

by Aurélien Géron (Author)

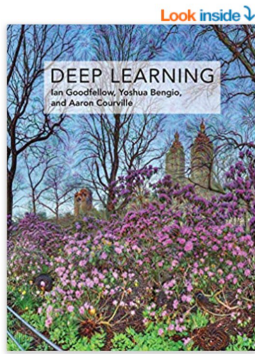
★★★★☆ 4 ratings

See all formats and editions

Paperback
₹3,000.00 ✓ prime

2 New from ₹3,000.00

EMI starts at ₹143. No Cost EMI available EMI options



See this image

Deep Learning (Adaptive Computation and Machine Learning series) Hardcover – 18 November 2016

by Aaron Courville (Author), Ian Goodfellow (Author), Yoshua Bengio (Author)

★★★★★ 1,917 ratings

Part of: Adaptive Computation and Machine Learning series (13 books)

See all formats and editions

Kindle Edition
₹4,346.25

Read with Our **Free App**

Hardcover
₹5,500.00

8 New from ₹5,490.00

Paperback
from ₹3,000.00

2 Used from ₹3,000.00

1 New from ₹4,000.00

EMI starts at ₹263. No Cost EMI available EMI options



Save Extra with 4 offers

Cashback: ₹200 cashback & ₹1,800 welcome rewards on Amazon Pay ICICI Credit Card . 5% cashback every time on shopping. Apply now! T&C. Not applicable on Amazon Business Transactions | [Details](#)

No Cost EMI: Avail No Cost EMI on select cards for orders above ₹3000 | [Details](#)

Podcasts:

Lex Friedman

<https://www.youtube.com/c/lexfridman>

Adrej Karpathy – Not a Podcast – but a Youtube Channel

<https://www.youtube.com/@AndrejKarpathy>

Other things which one should do in parallel :

1. Build few (3+) end to end project portfolio and showcase it on LinkedIn
 - a. Don't build Titanic / Iris data set. Do something new and novel to stand out
 - b. Do it end to end with deployment on cloud,if possible.

2. Build an Online presence on LinkedIn by posting and engaging with other ML community members.
 - a. LinkedIn is new resume. Spend time and polish it.
3. Practice, Practice and Practice.
4. Learn from Others – Especially Kaggle notebooks and how they approach any problem.

Thank you ! If you want to connect with me :

Connect me on LinkedIn : <https://www.linkedin.com/in/anas-p-296b38258/>

Connect me on Youtube: <https://www.youtube.com/@MLWithAP>



