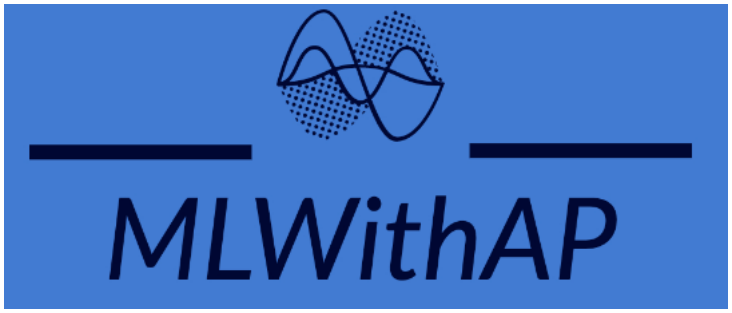


LEARN DATASCIENCE IN 2023

COMPLETE ROADMAP

5 MONTHS

ZERO MONEY NEEDED

All Free

Python 4 Weeks

- 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
MLWithAP	https://www.youtube.com/watch?v=Kk-ob0bQKhg&list=PL47eid5T-F5S1U_xiTfFR3f8myu7YGF2l

Maths for Data Science — 2 Weeks

- Mathematics for Machine Learning (2 weeks)
  - Statistics – Descriptive and Inferential
  - Linear Algebra
  - Calculus & Probability

  Khan Academy - https://www.khanacademy.org/math/statistics-probability
  Khan Academy - https://www.khanacademy.org/math/calculus-1
  Organic Chemistry Tutor - https://www.youtube.com/@TheOrganicChemistryTutor/playlists

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 Imbalance

Coursera / IBM - https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning

- 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

Coursera Andrew Ng's -
https://www.coursera.org/specializations/machine-learning-introduction

Josh Starmer
https://www.youtube.com/@statquest

Machine Learning - 2 — 5 Weeks

- 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

Stanford CS229 –

https://www.youtube.com/watch?v=jGwO_UgTS7I&list=PLoROMvodv4rMiGQp3WXShtMGgzqpfVfbU

Kaggle :
https://www.kaggle.com/learn

Deep Learning	5 Weeks

- 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

Andrew Ng's CS230 Stanford –

https://www.youtube.com/watch?v=PySo_6S4ZAg&list=PLoROMvodv4rOABXSygHTsbvUz4G_YQhOb

Coursera / Deeplearning.AI

https://www.coursera.org/specializations/deep-learning

MIT 6S.191

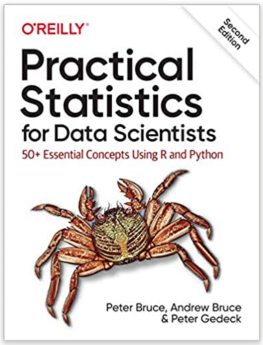
https://www.youtube.com/watch?v=7sB052Pz0sQ&list=PLtBw6njQRU-rwp5__7C0oIVt26ZgjG9NI

Stanford CS231n
https://www.youtube.com/watch?v=vT1JzLTH4G4&list=PLC1qU-LWwrF64f4QKQT-Vg5Wr4qEE1Zxk&index=2

Google's Machine Learning Course –

https://developers.google.com/machine-learning/crash-course

Books :



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)

4.5 out of 5 stars 763 ratings

See all formats and editions

Kindle Edition	Paperback
₹2,634.93	**₹3,941.00** prime
Read with Our **Free App**	12 New from ₹3,941.00

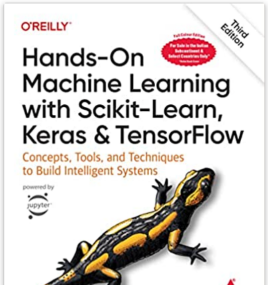
EMI starts at ₹188. No Cost EMI available EMI options

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

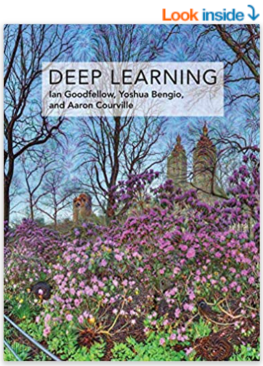
by Aurélien Géron (Author)

4.5 out of 5 stars 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)

4.5 out of 5 stars 1,917 ratings

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

See all formats and editions

Kindle Edition	Hardcover	Paperback
₹4,346.25	**₹5,500.00**	from ₹3,000.00
Read with Our **Free App**	8 New from ₹5,490.00	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 :

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.
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.