ML-Roadmap-for-2022

- **A curated list of Machine learning videos, links, projects and datasets to help you conquer the ML landscape in 6 months**

Levels of Learning:

- Testing the waters
- Gaining Conceptual depth
- Learning Practical Concepts
- Diving into different domains
- Pushing it with Projects

1. Testing the waters (Est. time 6-8 Weeks)

- The goal of this level is to get you familiar with the ML universe. You will learn a bit of everything.

Learn Python (Est. time - 2 weeks)

1. Basics of Python

https://www.youtube.com/playlist?list=PLKnIA16_Rmvb1RYR-iTA_hzckhdONtSW4

2. OOP in Python

- **Lecture 1** - https://www.youtube.com/watch?v=1s869EfxoDo
- **Lecture 2** - https://www.youtube.com/watch?v=8To-A6VPL90

3. Advance Topics

- **File Handling** - https://www.youtube.com/watch?v=ixEeeNjjOJ0
- **Exception Handling** - https://www.youtube.com/watch?v=NIWwJbo-9_8
- **Regular Expressions** - https://www.youtube.com/watch?v=K8L6KVGG-7o
- **Functional Programming** - https://www.youtube.com/watch?v=SvK_GErE2nM
- **Basics of Flask** - https://www.youtube.com/watch?v=swHI1H7DVsQ

4. Practice Problems

https://docs.google.com/document/d/1E_xCNijOWZ4Bm7r7DVj-1OA-oUopEFmv4tRm0YNuFWQ/edit?usp=sharing

Learn Numpy (Est. time 3 Days)

1. **Numpy Playlist**

   https://www.youtube.com/watch?v=CpPLLp3snK4&list=PLKnIA16_Rmvb-ToL3RQ_bwxG4_ND-0-DT

2. **Numpy Practice Problems**

   https://github.com/rougier/numpy-100

Learn Pandas (Est. time 4 Days)

1. **Pandas Playlist**

   https://www.youtube.com/watch?v=kq9Vmg5d7Sk&list=PLKnIA16_RmvbR85fgbfVRKOiMokUKVupy

2. **Pandas Problems**

   https://github.com/ajcr/100-pandas-puzzles

Learn Data Visualization (Est. time 1 Week)

1. **Matplotlib**

   https://www.youtube.com/playlist?list=PL-osiE80TeTvipOqomVEeZ1HRrcEvtZB_

2. **Seaborn**

   https://www.youtube.com/playlist?list=PLKnIA16_RmvbB1bFGjvS6a8T0mnqawejo

Descriptive Statistics (Est. time 4 Days)

1. **Statistics Playlist**

   https://www.youtube.com/watch?v=tPhzDKjQBpo&list=PLKnIA16_RmvbVrE0eZO2bCaFln6jaNq-1

Learn Data Analysis Process (Est. time 1 week)

1. Playlist

https://www.youtube.com/watch?v=ZhacwtUR0SU&list=PLKnIA16_RmvZAqJzKstVHywcRNMn6pcGD

Learn Exploratory Data Analysis (EDA) (Est. time 1 Week)

1. Understanding your data - https://www.youtube.com/watch?v=mJlRTUuVr04

2. Univariate Analysis - https://www.youtube.com/watch?v=4HyTlbHUKSw

3. Bivariate and Multivariate Analysis - https://www.youtube.com/watch?v=6D3VtEfCw7w

4. Pandas Profiling - https://www.youtube.com/watch?v=E69Lg2ZgOxg

5. EDA on House Prices Dataset

https://www.kaggle.com/pmarcelino/comprehensive-data-exploration-with-python

6. EDA on Titanic Dataset - https://www.kaggle.com/startupsci/titanic-data-science-solutions

7. EDA on Haberman's Survival Dataset -

https://www.kaggle.com/gokulkarthik/haberman-s-survival-exploratory-data-analysis

8. EDA on Heart Disease Dataset

https://www.kaggle.com/kralmachine/analyzing-the-heart-disease

9. EDA on IPL Dataset - https://www.kaggle.com/ash316/let-s-play-cricket

10. EDA on Wine Review Dataset -

https://www.kaggle.com/kabure/wine-review-s-eda-recommend-systems

11. EDA on PIMA Diabetes Dataset

https://www.kaggle.com/shrutimechlearn/step-by-step-diabetes-classification-knn-detailed

12. EDA on Breast Cancer Dataset -

https://www.kaggle.com/kanncaa1/statistical-learning-tutorial-for-beginners

13. EDA on Olympics Dataset - https://www.youtube.com/watch?v=5nQXhusiu7s

14. EDA on Covid Data - https://www.youtube.com/watch?v=ll0aZVNnOP8

15. WhatsApp Chat Analysis Project - https://www.youtube.com/watch?v=Q0QwvZKG_6Q

Learn Machine Learning Basics (Est. time 1 Week)

1. What is Machine Learning? https://www.youtube.com/watch?v=ZftI2fEz0Fw

2. AI vs ML vs DL https://www.youtube.com/watch?v=1v3_AQ26jZ0

3. Types of Machine Learning - https://www.youtube.com/watch?v=81ymPYEtFOw

4. Batch Machine Learning - https://www.youtube.com/watch?v=nPrhFxEuTYU

5. Online Machine Learning - https://www.youtube.com/watch?v=3oOipgCbLIk

6. Instance based vs Model based learning https://www.youtube.com/watch?v=ntAOq1ioTKo

7. Challenges in Machine Learning - https://www.youtube.com/watch?v=WGUNAJki2S4

8. Applications of Machine Learning - https://www.youtube.com/watch?v=UZio8TcTMrI

9. Machine Learning Development Lifecycle

https://www.youtube.com/watch?v=iDbhQGz_rEo

10. Data Engineer V Data Analyst V Data Scientist V ML Engineer -

https://www.youtube.com/watch?v=93rKZs0MkgU

11. How to frame a Machine Learning problem? -

https://www.youtube.com/watch?v=A9SezQlvakw

12. Installing and using software for data science -

https://www.youtube.com/watch?v=82P5N2m41jE

13. How to work with CSV files? - https://www.youtube.com/watch?v=a_XrmKlaGTs

14. Working with JSON and SQL data - https://www.youtube.com/watch?v=fFwRC-fapIU

15. **Building an End to End Machine Learning Project** -

https://www.youtube.com/watch?v=dr7z7a_8lQw

2. Gaining Conceptual depth (Est. time 6-8 Weeks)

- The goal of this level is to learn the core machine learning concepts and algorithms

Learn about tensors (Est. time - 1 Day)

1. **What are Tensors?** - https://www.youtube.com/watch?v=vVhD2EyS41Y

Advance Statistics

1. **Covariance**
2. **Pearson Correlation Coefficient**
3. **QQ Plot**
4. **Confidence Interval**
5. **Hypothesis Testing**
6. **Chi square Test, Anova Test**
7. **Playlist link** -

https://www.youtube.com/watch?v=qtaqvPAeEJY&list=PLKnIA16_Rmvbe9wDJGXc28KKr6lp5Jn2g

Probability Basics

1. **Conditional Probability**
2. **Independent Events**
3. **Bayes Theorem**
4. **Uniform Distribution**
5. **Binomial Distribution**
6. **Bernoulli Distribution**
7. **Poisson Distribution**
8. **Playlist Link** -

https://www.youtube.com/watch?v=Ty7knppVo9E&list=PLKnIA16_RmvYNbPMB6ofVLRCcTPUAftdY

Linear Algebra Basics

1. **Representing Tabular Data**
2. **Vectors**
3. **Matrices**
4. **Matrix Multiplication**
5. **Dot Product**
6. **Equation of line in N-dim**
7. **EigenVector and Eigenvalues**
8. **Playlist Link** -

https://www.youtube.com/watch?v=e9h-ZZ_ahRg&list=PLKnIA16_RmvYu0fS_RuIB2eTbJcTFdrAA

Basics of Calculus

1. **Big Picture of Derivatives**
2. **Maxima and Minima**
3. **Playlist link - (first 4 videos only)**

https://www.youtube.com/playlist?list=PLBE9407EA64E2C318

Machine Learning Algorithms

1. **Linear Regression** -

https://www.youtube.com/watch?v=UZPfbG0jNec&list=PLKnIA16_Rmva-wY_HBh1gTH32ocu2SoTr

2. **Gradient Descent** -

https://www.youtube.com/watch?v=ORyfPJypKuU&list=PLKnIA16_RmvZvBbJex7T84XYRmor3IPK1

3. **Logistic Regression** -

https://www.youtube.com/watch?v=XNXzVfltWGY&list=PLKnIA16_Rmvb-ZTsM1QS-tlwmlkeGSnru

4. Support Vector Machines -

https://www.youtube.com/watch?v=ugTxMLjLS8M&list=PLKnIA16_RmvbOIFee-ra7U6jR2oIbCZBL

5. Naive Bayes -

https://www.youtube.com/watch?v=Ty7knppVo9E&list=PLKnIA16_RmvZ67wQaHoBuzXaDAfPz-a6l

6. K Nearest Neighbors -

https://www.youtube.com/watch?v=BYaoDZM1IcU&list=PLKnIA16_RmvZiE-lEdN5RDi18-u-T43zd

7. Decision Trees -

https://www.youtube.com/watch?v=gwgmSSTdiXs&list=PLKnIA16_RmvYGY_n9PP8zN-0LG9MoZRjU

8. Random Forest -

https://www.youtube.com/watch?v=bHK1fE_BUms&list=PLKnIA16_RmvZyqP3WGUo7iVzilIea_1bp

9. Bagging -

https://www.youtube.com/watch?v=LUiBOAy7x6Y&list=PLKnIA16_RmvZ7iKIcJrLjUoFDEeSejRpn

10. Adaboost -

https://www.youtube.com/watch?v=sFKnP0iP0K0&list=PLKnIA16_RmvZxriy68dPZhorB8LXP1PY6

11. Gradient Boosting -

https://www.youtube.com/watch?v=fbKz7N92mhQ&list=PLKnIA16_RmvaMPgWfHnN4MXl3qQ1597Jw

12. Xgboost -

https://www.youtube.com/watch?v=BTLB-ppqBZc&list=PLKnIA16_RmvbXJbBW4zCy4Xbr81GRyaC4

13. Principle Component Analysis (PCA) -

https://www.youtube.com/watch?v=ToGuhynu-No&list=PLKnIA16_RmvYHW62E_lGQa0EFsph2NquD

14. KMeans Clustering -

https://www.youtube.com/watch?v=5shTLzwAdEc&list=PLKnIA16_RmvbA_hYXlRgdCg9bn8ZQK2z9

15. Hierarchical Clustering - https://www.youtube.com/watch?v=Ka5i9TVUT-E

16. DBSCAN - https://www.youtube.com/watch?v=RDZUdRSDOok

17. T-sne - https://www.youtube.com/watch?v=NEaUSP4YerM and https://distill.pub/2016/misread-tsne/

Machine Learning Metrics -

https://www.youtube.com/watch?v=Ti7c-Hz7GSM&list=PLKnIA16_RmvZJGOqRjqhOhTEmQW3vDdbQ

Bias Variance Tradeoff - https://www.youtube.com/watch?v=74DU02Fyrhk

Regularization -

https://www.youtube.com/watch?v=aEow1QoTLo0&list=PLKnIA16_RmvZuSEZ24Wlm13QpsfLlJBM4

Cross-Validation - https://www.youtube.com/watch?v=S5NkE-xgx98

3. Learn Practical Concepts (Est. time 6-8 Weeks)

- The goal of this level is to get you introduced to the practical side of machine learning. What you learn at this level would really help you out there in the wild.

Data Acquisition (Est. time - 2 Days)

1. **Web Scraping** - https://www.youtube.com/watch?v=8NOdgjC1988

    **** Project*** *- Create a Pandas dataframe of Indian cuisines from some website using web scraping.*

2. **Fetch data from API** - https://www.youtube.com/watch?v=roTZJaxjnJc

    **** Project*** *- Create a Pandas dataframe of movies from TMDB API.*

Working with missing values (Est. time - 3 Days)

1. **Complete Case Analysis** - https://www.youtube.com/watch?v=aUnNWZorGmk
2. **Handling missing numerical data** - https://www.youtube.com/watch?v=mCL2xLBDw8M
3. **Handling missing categorical data** - https://www.youtube.com/watch?v=l_Wip8bEDFQ
4. **Missing indicator** - https://www.youtube.com/watch?v=Ratcir3p03w
5. **KNN Imputer** - https://www.youtube.com/watch?v=-fK-xEev2I8
6. **MICE** - https://www.youtube.com/watch?v=a38ehxv3kyk
7. **Kaggle Notebooks and Practice Datasets** -

https://docs.google.com/document/d/1_9Y6kxNc6QTym2Y2JGEBbnCUbE1qZWLVzVXlT2eX_FQ/edit?usp=sharing

Feature Scaling/Normalization (Est. time - 2 Days)

1. **Standardization** - https://www.youtube.com/watch?v=1Yw9sC0PNwY
2. **Normalization** - https://www.youtube.com/watch?v=eBrGyuA2MIg

Feature Encoding Techniques (Est. time - 2 Days)

1. **Ordinal Encoding and Label Encoding** -

https://www.youtube.com/watch?v=w2GglmYHfmM

2. **One Hot Encoding** - https://www.youtube.com/watch?v=U5oCv3JKWKA

3. **Encoding high cardinality categorical features** - https://www.kaggle.com/general/16927
4. **Feature hashing** -

https://datasciencestunt.com/dealing-with-categorical-features-with-high-cardinality-feature-hashing/

Feature Transformation(Est. time - 2 Days)

1. **Log Transform** - https://www.youtube.com/watch?v=cTjj3LE8E90

2. **Box Cox Transform** - https://www.youtube.com/watch?v=lV_Z4HbNAx0

3. **Yeo Johnson Transform** - https://www.youtube.com/watch?v=lV_Z4HbNAx0

4. **Discretization** - https://www.youtube.com/watch?v=kKWsJGKcMvo

Working with Pipelines(Est. time - 2 Days)

1. **Column Transformer** - https://www.youtube.com/watch?v=5TVj6iEBR4I

2. **Sklearn Pipelines** - https://www.youtube.com/watch?v=xOccYkgRV4Q

Handing Time and Date data(Est. time - 1 Day)

1. **Working with time and date data** - https://www.youtube.com/watch?v=J73mvgG9fFs

Working with Outliers (Est. time - 3 Days)

1. **What are Outliers?** - https://www.youtube.com/watch?v=Lln1PKgGr_M

2. **Outlier detection and removal using Z-score method** -

https://www.youtube.com/watch?v=OnPE-Z8jtqM

3. **Outlier detection and removal using IQR method** -

https://www.youtube.com/watch?v=Ccv1-W5ilak

4. **Percentile method** - https://www.youtube.com/watch?v=bcXA4CqRXvM

Feature Construction (Est. time - 1 Day)

1. **Feature Construction** - https://www.youtube.com/watch?v=ma-h30PoFms

Feature Selection (Est. time - 3 Days)

1. **Feature selection using SelectKBest and Recursive Feature Elimination** - https://www.youtube.com/watch?v=xlHk4okO8Ls&t=1s
2. **Chi-squared Feature Selection** - https://www.youtube.com/watch?v=fMIwIKLGke0
3. **Backward Feature Elimination** - https://www.youtube.com/watch?v=zW1SvA0Z-l4&t=2s
4. **Dropping features using Pearson correlation coefficient** - https://www.youtube.com/watch?v=FndwYNcVe0U
5. **Feature Importance using Random Forest** - https://www.youtube.com/watch?v=R47JAob1xBY
6. **Feature Selection Advise** - https://www.youtube.com/watch?v=YaKMeAlHgqQ

Cross Validation (Est. time - 2 Days)

1. **What is cross-validation?** - https://www.youtube.com/watch?v=fSytzGwwBVw
2. **Holdout Method** - https://www.youtube.com/watch?v=4NnI3SBuww4
3. **K-Fold Cross Validation** - https://www.youtube.com/watch?v=gJo0uNL-5Qw
4. **Leave 1 Out Cross Validation** - https://www.youtube.com/watch?v=yxqcHWQKkdA
5. **Time series cross validation** - https://www.youtube.com/watch?v=g9iO2AwTXyI

Modeling - Stacking and Blending (Est. time - 1 Week)

1. **Stacking** - https://www.youtube.com/watch?v=O-aDHBGMqXA
2. **Blending** - https://www.youtube.com/watch?v=TuIgtitqJho
3. **LightGBM** - https://www.youtube.com/watch?v=n_ZMQj09S6w

4. **CatBoost** - https://www.youtube.com/watch?v=8o0e-r0B5xQ

Model Tuning (Est. time - 4 Days)

1. **GridSearchCV** - https://www.youtube.com/watch?v=4Im0CT43QxY
2. **RandomSearchCV** - https://www.youtube.com/watch?v=Q5dH5mOQ_ik
3. **Hyperparameter Tuning** - https://www.youtube.com/watch?v=355u2bDqB7c

Working with imbalanced data (Est. time - 3 Days)

1. **How to handle imbalanced data** - https://www.youtube.com/watch?v=JnlM4yLFNuo
2. **Kaggle Notebook** - https://www.kaggle.com/kabure/credit-card-fraud-prediction-rf-smote
3. **SMOTE on Quora Dataset** -

https://www.kaggle.com/theoviel/dealing-with-class-imbalance-with-smote

Handling Multicollinearity(Est. time - 2 Days)

1. **What is multicollinearity?** - https://www.youtube.com/watch?v=ekuD8JUdL6M
2. **Practical Example** - https://www.youtube.com/watch?v=ATH4urDitI8
3. **VIF in Multicollinearity** - https://www.youtube.com/watch?v=GMAp_tP1ZQ0

Data Leakage - (Est. time - 2 Days)

1. **What is Data Leakage?** -

https://machinelearningmastery.com/data-leakage-machine-learning/

2. **Practical - Data Leakage on Quora Question Pair Dataset** -

https://www.kaggle.com/sudalairajkumar/simple-leaky-exploration-notebook-quora

3. **Practical - Data Leakage on Credit Card data** -

https://www.kaggle.com/dansbecker/data-leakage

Serving your model(Est. time - 1 Week)

1. **Pickling your model** - https://www.youtube.com/watch?v=yY1FXX_GSco
2. **Flask Tutorial** - https://www.youtube.com/watch?v=swHI1H7DVsQ
3. **Streamlit Tutorial** - https://www.youtube.com/watch?v=Klqn--Mu2pE
4. **Deploy model on Heroku** - https://www.youtube.com/watch?v=YncZ0WwxyzU
5. **Deploy model on AWS** - https://www.youtube.com/watch?v=_rwNTY5Mn40
6. **Deploy model to GCP** - https://www.youtube.com/watch?v=fw6NMQrYc6w
7. **Deploy model to Azure** - https://www.youtube.com/watch?v=qnbJcbjh-3s
8. **ML model to Android App** - https://www.youtube.com/watch?v=ax3WyB-_LJY

Working with Large Datasets

1. **What is Out of core ML?** - https://www.youtube.com/watch?v=9e4nUuq2Hmg
2. **Practical implementation of Out of core ML** -

https://www.youtube.com/watch?v=sRCuvcdvuzk

3. **NYC Cab Dataset Project** -

https://vaex.io/blog/ml-impossible-train-a-1-billion-sample-model-in-20-minutes-with-vaex-and-scikit-learn-on-your

4. Diving into different domains (Est. time 6-8 Weeks)

- This is the level where you would dive into different domains of Machine Learning. Mastering these will make you a true Data Scientist.

SQL (Est. time - 2 Days)

1. **Complete SQL Roadmap** - https://www.youtube.com/watch?v=FGBme8dWR_M

2. **SQL learning resources** -

https://docs.google.com/document/d/1wCALgWubTOvuvlXJ3Eweh7AgJj4sPq2pW92y3viPZbs/edit?usp=sharing

3. **The only video you need to see** - https://www.youtube.com/watch?v=nopIGY1zJE0

Recommendation Systems

1. **Movie Recommendation System** - https://www.youtube.com/watch?v=1xtrIEwY_zY

2. **Book Recommender System** - https://www.youtube.com/watch?v=sf93xpq8vaA

3. **Fashion Recommender System** - https://www.youtube.com/watch?v=xanJe6e8Xuw

Association Rule Learning

1. **Association Rule Mining(Apriori Algorithm)** -

https://www.youtube.com/watch?v=guVvtZ7ZClw

2. **Eclat Algorithm** - https://www.youtube.com/watch?v=oBiq8cMkTCU

3. **Market Basket Analysis** - https://www.youtube.com/watch?v=Y7Xkqqfz1UU

Anomaly Detection

1. **Anomaly Detection Lecture from Microsoft Research** -

https://www.youtube.com/watch?v=12Xq9OLdQwQ

2. **Novelty Detection Lecture** - https://www.youtube.com/watch?v=vIDcjbpwY3k

NLP

1. **Complete NLP Roadmap** - https://www.youtube.com/watch?v=PKv_okm1H-k

2. **Complete NLP Playlist** -

https://www.youtube.com/watch?v=zlUpTlaxAKI&list=PLKnIA16_RmvZo7fp5kkIth6nRTeQQsjfX

3. **NLP Project Ideas** - https://www.youtube.com/watch?v=oWJe2T29kAo

4. **Email Spam Classifier Project** - https://www.youtube.com/watch?v=YncZ0WwxyzU

5. **Building a Chatbot** - https://www.youtube.com/watch?v=Nb21OhaW8GY

Time Series(Coming Soon)

Computer Vision(Coming Soon)

Fundamentals of Neural Network -

https://www.youtube.com/playlist?list=PLKnIA16_RmvYuZauWaPlRTC54KxSNLtNn

5. Pushing it with Projects (Est. time 6-8 Weeks)

- The objective of this level is to sharpen your knowledge that you have accumulated in the previous 4 levels

8 types of Projects for your portfolio - https://www.youtube.com/watch?v=SQHfry4xmdM

How to select a project - https://www.youtube.com/watch?v=kH--k1VKFt4

Car Price Predictor - https://www.youtube.com/watch?v=iRCaMnR_bpA

Banglore House Price Predictor - https://www.youtube.com/watch?v=DVxkI1VmpCk

Posture Detection using ML5.js - https://www.youtube.com/watch?v=kRvIcdLhDtU

Laptop Price Predictor - https://www.youtube.com/watch?v=BgpM2IiCH6k

Which Bollywood celebrity are you? - https://www.youtube.com/watch?v=X67rclJcIL0

Finding similar GOT characters - https://www.youtube.com/watch?v=ygGknomFEWY

IPL win probability predictor - https://www.youtube.com/watch?v=ygGknomFEWY

T20 score predictor - https://www.youtube.com/watch?v=ygGknomFEWY

Titanic Survivor Prediction - https://www.youtube.com/watch?v=Bnp94fpxZjY

Diabetes Prediction using ML - https://www.youtube.com/watch?v=xUE7SjVx9bQ

Fake news prediction - https://www.youtube.com/watch?v=nacLBdyG6jE

Loan Status Prediction - https://www.youtube.com/watch?v=XckM1pFgZmg

Gold Price Prediction - https://www.youtube.com/watch?v=9ffkBvh8PTQ

Handwriting Classifier - https://www.youtube.com/watch?v=1B3YIkyPNk0

Flight Fare Prediction - https://www.youtube.com/watch?v=y4EMEpEnElQ

Link for 500+ ML+DL projects -

https://github.com/ashishpatel26/500-AI-Machine-learning-Deep-learning-Computer-vision-NLP-Projects-with-code