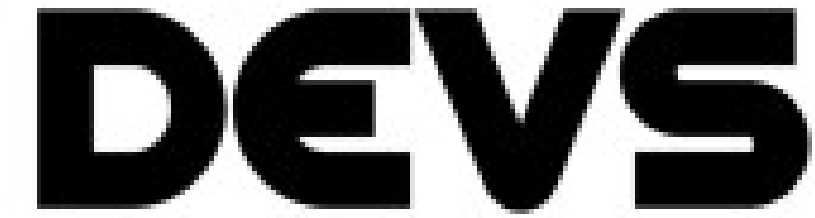

PRESENTS

Roadmap for ML becoming a

Successful ML engineer

ML Engineer Roadmap for Beginners

Following is the roadmap to learning **AI Engineer** (also known as **ML Engineer**) skills for a total beginner. It includes FREE learning resources for technical skills (or tool skills) and soft (or core) skills

Prerequisites: You must have skills or interests to build skills in Coding and Math. Without these two you cannot become an ML engineer.

Total Duration: **6 Months** (**2 hours** of study Every Day) Also,

ML Engineer = Data scientist + Software Engineer

Week 0: Do Proper Research and protect yourself from SCAMS.

Unfortunately, a lot of systematic scams are happening in ed tech, especially in the data field where aspirants are provided with false promises like a 100% job guarantee or trapped into "Masterclasses" which are nothing but sales pitches to upsell their low-grade courses at exorbitant prices. You need to do complete research about the market and mentors before starting your journey. Providing you the links to a few posts that we have made in this regard which will support your research.

Even though these posts are **NOT** sufficient, do your additional research.

- https://bit.ly/4at9Jaw
- https://bit.ly/477IOOs
- https://bit.ly/3GPD7dp

Week 1 and 2: Beginners Python

- **Topics**
  - Variables, Numbers, Strings
  - Lists, Dictionaries, Sets, Tuples
  - If condition, for loop
  - Functions, Lambda Functions
  - Modules (pip install) o Read, Write files
  - Exception handling
  - Classes, Objects
- **Learning Resources**

  Python Tutorials (Codebasics) on YouTube (first 16 videos)

- **Assignment**

  □ Track A: Finish all these exercises: https://bit.ly/3k1mof5
  □ Track B: Finish exercises and quizzes for relevant topics
  □ Create a professional-looking LinkedIn profile.

Week 3 and 4 : Data Structures and Algorithms in Python (Optional)

DEVS
CODE. COFFEE. REPEAT ...

Topics

- Data structures basics, Big O notation
- Data structures: Arrays, Linked List, Hash Table, Stack, Queue
- Data structures: Tree, Graph
- Algorithms: Binary search, Bubble sort, quick sort, merge sort
- Recursion

Learning Resources

- DSA YouTube Playlist: https://bit.ly/3uiW2Lf

- **Motivation**
  - How Kaggle helped this person become ML engineer**:** https://bit.ly/3RFVruy
- **Assignment**

□ Finish all these exercises in this same playlist: https://bit.ly/3uiW2Lf

Week 5, 6: Advance Python

- **Topics**
  - Inheritance, Generators, Iterators
  - List Comprehensions, Decorators
  - Multithreading, Multiprocessing
- **Learning Resources**
  - Python Tutorials (Codebasics) on YouTube (17th to 27th video) - https://bit.ly/3X6CCC7
- **Assignment**

□ Finish all these exercises in this same playlist: https://bit.ly/3X6CCC7

Week 7 Version Control (Git, Github)

- **Topics**
  - What is the versioncontrol system? What is Git and GitHub?
  - Basic commands: add, commit, push.
  - Branches, reverting change, HEAD, Diff and Merge
  - Pull requests.

- **Learning Resources**
  - YT playlist (codebasics): https://bit.ly/3SECQQ7
  - YT playlist (Corey): https://bit.ly/3T0Yrmb

Week 8, 9: SQL

- **Topics**
  - Basics of relational databases.
  - Basic Queries: SELECT, WHERE LIKE, DISTINCT, BETWEEN, GROUP BY, ORDER BY
  - Advanced Queries: CTE, Subqueries, Window Functions
  - Joins: Left, Right, Inner, Full
  - Database creation, indexes, stored procedures.

- **Learning Resources**
  - **Khan academy** SQL course: https://bit.ly/3WFku20
  - **sqlboltYT** video: https://youtu.be/Rm0xH2Vpfi0?si=6ZLK8A5LvGqN4NmT
  - **Ankit Bansal** video: https://youtu.be/Rm0xH2Vpfi0?si=6ZLK8A5LvGqN4NmT
  - **TechTFQ** video: https://youtu.be/Rm0xH2Vpfi0?si=6ZLK8A5LvGqN4NmT

- **Assignment**

  □ Participate in SQL resume project challenge on https://codebasics.io/

  - Link: https://codebasics.io/challenge/codebasics-resume-project-challenge/7
    These challenges help you improve technical skills, soft skills and business understanding.

Week 10: Numpy, Pandas, Data Visualization

- **Tech Skills**
  - **Numpy**
    - numpy YouTube playlist: https://bit.ly/3GTppa8
  - **Pandas, Matplotlib, Seaborn**
    - Go through chapter 3 in this course (entire chapter is free): https://codebasics.io/courses/math-and-statistics-for-data-science

Week 12,13,14,15: Math & Statistics for AI

- **Math and Statistics for AI**
  - Topics to Learn

- Basics: Descriptive vs inferential statistics, continuous vs discrete data, nominal vs ordinal data
- Linear Algebra: Vectors, Metrices, Eigenvalues and Eigenvectors
- Calculus: Basics of integral and differential calculus
- Probability basics Distributions: Normal distribution Correlation and covariance Central limit theorem Hypothesis
- testing: p value, confidence interval, type 1 vs type 2 error,
- Z test
- Measures of dispersion: variance, standard deviation
- Measures of central tendency: mean, median, mode
- Basic plots: Histograms, pie charts, bar charts, scatter plot etc.

- Learning Resources

  - statquest YouTube channel: https://www.youtube.com/@statquest
  - 3 blue 1 brown YT : https://www.youtube.com/@3blue1brown
  - Learn the above topics from this excellent Khan academy course on statistics and probability.
    Course link: https://www.khanacademy.org/math/statistics-probability

- **Assignment**
  - □ Finish all exercises in this playlist: https://bit.ly/3QrSXis
  - □ Finish all exercises in Khan academy course.
  - □ Track B: Finish exercises and quizzes for relevant topics.

Week 16: Exploratory Data Analysis (EDA)

- **Exploratory Data Analysis (EDA)**
  - https://www.kaggle.com/code?searchQuery=exploratory+data+analysis
  - Use the above link to search for exploratory data analysis notebooks.
  - Practice EDA using at least 3 datasets.
    - e.g. https://www.kaggle.com/datasets/rishabhkarn/ipl-auction-2023/data
- **Assignment**
  - □ Perform EDA (Exploratory data analysis on **at least 2 additional datasets** on Kaggle)

Week 17,18, 19, 20: Machine Learning

Machine Learning: Preprocessing

- Handling NA values, outlier treatment, data normalization
  One hot encoding, label encoding , Feature engineering,
  Train test split & Cross validation

Machine Learning: Model Building

- Types of ML: Supervised, Unsupervised
  Supervised: Regression vs Classification
- Linear models
  - Linear regression
  - logistic regression
  - Gradient descent
- Nonlinear models (tree-based models)
  - Decision tree
  - Random forest
  - XGBoost
- Model evaluation
  - Regression: Mean Squared Error, Mean Absolute Error, MAPE
  - Classification: Accuracy, Precision-Recall, F1 Score, ROC Curve, Confusion matrix
- Hyperparameter tunning: GridSearchCV, RandomSearchCV
- Unsupervised: K means, Hierarchical clustering, Dimensionality reduction (PCA)

- **Learning Resources**

  Complete ML Playlist : https://www.youtube.com/playlist?list=PLeo1K3hjS3uvCeTYTeyfe0-rN5r8zn9rw

- **Assignment**

  □ Complete all exercises in ML playlist: https://bit.ly/3io5qqX

  □ Work on **2 Kaggle ML notebooks**

  □ Write **2 LinkedIn posts** on whatever you have learnt in ML

Week 21: ML Ops

- **Topics**
  - o What is API? FastAPI for Python server development
  - o DevOps Fundamentals: CI/CD pipelines, containerization (Docker, Kubernetes)
  - o Familiarity with at least one cloud platform (AWS, Azure etc.)
- Learning Resources
  - Docker tutorial: https://bit.ly/3uCNpeE
  - FastAPI tutorial: https://bit.ly/497p6Ex

Week 22 - 26: Machine Learning Projects with Deployment

You need to finish **two** end to end ML projects. One on **Regression**, the other on **Classification**

Regression Project: Bangalore property price prediction

- o YouTube playlist link: https://bit.ly/3ivycWr
- o Project covers following
  - Data cleaning Feature engineering
  - Model building and hyper parameter tuning
  - Write flask server as a web backend
  - Building website for price prediction
  - Deployment to AWS

Classification Project: Sports celebrity image classification

- o YouTube playlist link: https://bit.ly/3ioaMSU
- o Project covers following
  - Data collection and data cleaning
  - Feature engineering and model training
  - Flask server as a web backend Building website and deployment