Data Science

Machine Learning

GenerativeAI | MLOps

Full-Stack Roadmap

Build Your Strong Data Science ML Gen AI Portfolio/Personal Brand & Get Tons of Job Offers in 12 weeks →

Modules→

1 → **Python for Machine Learning** 🚀

2 → **Data Structures & Algorithms**

3 → **Git & GitHub Make Recruiters reach You, Build your stunning profile**

4 → **Data Science X NumPy, Pandas, Matplotlib, Seaborn**

5 → **Mathematics in Machine Learning**

6 → **Machine Learning   Algorithms X Data Processing**

7 → **Natural Language Processing X Deep Learning**

8 → **Generative AI - GANs, VAEs, LLMs**

9 → **Computer Vision X Deep Learning**

10 → **MLOps | Machine Learning Operations**

11 → **Machine Learning System Design**

12 → **SQL - Structured Query Language**

13 → **Major Capstone Projects**

14 → **Data Science, ML, GenAI Interview**

15 → **Personal Branding & Portfolio**

16 → **Others**

Technology Stack

- Python
- Data Structures
- NumPy ● Pandas
- Matplotlib
- Seaborn
- Scikit-Learn
- Statsmodels
- Natural Language Toolkit (NLTK)
- PyTorch
- OpenCV
- Tableau
- Structure Query Language (SQL)
- PySpark
- Azure Fundamentals
- Azure Data Factory
- Databricks
- 5 Major Projects
- Git and GitHub
- AWS
- GCP
- Azure

1 | Python Programming and Logic Building

I will prefer Python Programming Language. Python is the best for starting your programming journey. Here is the roadmap of Python for logic building.

1 | Introduction and Basics

- Installation
- Python Org, Python 3
- Variables
- Print function
- Input from user
- Data Types
- Type Conversion
- First Program

2 | Operators

- Arithmetic Operators
- Relational Operators
- Bitwise Operators
- Logical Operators
- Assignment Operators
- Compound Operators
- Membership Operators
- Identity Operators

3 | Conditional Statements

- If Else
- If
- Else
- El If (else if)
- If Else Ternary Expression

4 | While Loop

- While loop logic building
- Series based Questions
- Break
- Continue
- Nested While Loops
- Pattern-Based Questions
- pass
- Loop else

5 | Lists

- List Basics
- List Operations
- List Comprehensions / Slicing
- List Methods

6 | Strings

- String Basics
- String Literals
- String Operations
- String Comprehensions / Slicing
- String Methods

7 | For Loops

- Range function
- For loop
- Nested For Loops
- Pattern-Based Questions
- Break
- Continue
- Pass
- Loop else

8 | Functions

- Definition
- Call
- Function Arguments
- Default Arguments
- Docstrings
- Scope
- Special functions Lambda, Map, and Filter
- Recursion
- Functional Programming and Reference Functions

9 | Dictionary

- Dictionaries Basics
- Operations
- Comprehensions
- Dictionaries Methods

10 | Tuple

- Tuples Basics
- Tuples Comprehensions / Slicing
- Tuple Functions
- Tuple Methods

11 | Set

- Sets Basics
- Sets Operations
- Union
- Intersection
- Difference and Symmetric Difference

12 | Object-Oriented Programming

- Classes
- Objects
- Method Calls

- Inheritance and Its Types
- Overloading
- Overriding
- Data Hiding
- Operator Overloading

13 | File Handling

- File Basics
- Opening Files
- Reading Files
- Writing Files
- Editing Files
- Working with different extensions of file
- With Statements

14 | Exception Handling

- Common Exceptions
- Exception Handling
- Try
- Except
- Try except else
- Finally
- Raising exceptions
- Assertion

15 | Regular Expression

- Basic RE functions
- Patterns
- Meta Characters
- Character Classes

16 | Modules & Packages

- Different types of modules
- Inbuilt modules
- OS

- Sys
- Statistics
- Math
- String
- Random
- Create your own module
- Building Packages
- Build your own python module and deploy it on pip

17 | Data Structures

- Stack
- Queue
- Linked Lists
- Sorting
- Searching
- Linear Search
- Binary Search

18 Higher-Order Functions

- Function as a parameter
- Function as a return value
- Closures
- Decorators
- Map, Filter, Reduce Functions

19 | Python Web Scrapping

- Understanding BeautifulSoup
- Extracting Data from websites
- Extracting Tables
- Data in JSON format

20 | Virtual Environment

- Virtual Environment Setup

21 | Web Application Project

- Flask
- Project Structure
- Routes
- Templates
- Navigations

22 | Git and GitHub

- Git - Version Control System
- GitHub Profile building
- Manage your work on GitHub

23 Deployment

- Heroku Deployment
- Flask Integration

24 | Python Package Manager

- What is PIP?
- Installation
- PIP Freeze
- Creating Your Own Package
- Upload it on PIP

25 | Python with MongoDB Database

- SQL and NoSQL
- Connecting to MongoDB URI
- Flask application and MongoDB integration
- CRUD Operations
- Find
- Delete
- Drop

26 | Building API

- API (Application Programming Interface)
- Building API
- Structure of an API
- PUT
- POST
- DELETE
- Using Postman

27 Statistics with NumPy

- Statistics
- NumPy basics
- Working with Matrix
- Linear Algebra operations
- Descriptive Statistics

28 | Data Analysis with Pandas

- Data Analysis basics
- Dataframe operations
- Working with 2-dimensional data
- Data Cleaning
- Data Grouping

29 | Data Visualization with Matplotlib

- Matplotlib Basics
- Working with plots
- Plot
- Pie Chart
- Histogram

30 | What to do Now?

- Discussions on how to process further with this knowledge.

2 | Data Structure & Algorithms

Data Structure is the most important thing to learn not only for data scientists but for all the people working in computer science. With data structure, you get an internal understanding of the working of everything in software.

0 | Data Structures & Algorithms Starting Point

- Getting Started
- Variables
- Data Types
- Data Structures
- Algorithms
- Analysis of Algorithm
- Time Complexity
- Space Complexity
- Types of Analysis
- Worst
- Best
- Average
- Asymptotic Notations
- Big-O
- Omega
- Theta

Data Structures - Phase 1

1 | Stack

2 | Queue

3 | Linked List

4 | Tree

5 | Graph

Algorithms - Phase 2

Interviews Questions & Solutions

3 | Pandas Numpy Matplotlib

Python supports n-dimensional arrays with NumPy. For data in 2 dimensions, Pandas is the best library for analysis. You can use other tools but tools have drag-and-drop features and limitations. Pandas can be customized as per the need as we can code depending upon the real-life problem.

Numpy

- Vectors, Matrix
- Operations on Matrix
- Mean, Variance, and Standard Deviation
- Reshaping Arrays
- Transpose and Determinant of Matrix
- Diagonal Operations, Trace
- Add, Subtract, Multiply, Dot, and Cross Product.

Pandas

- Series and DataFrames
- Slicing, Rows, and Columns
- Operations on DataFrame
- Different ways to create DataFrame

- Read, Write Operations with CSV files
- Handling Missing values, replacing values, and Regular Expression
- GroupBy and Concatenation

Matplotlib

- Graph Basics
- Format Strings in Plots
- Label Parameters, Legend
- Bar Chart, Pie Chart, Histogram, Scatter Plot

4 | Statistics

Descriptive Statistics

- Measure of Frequency and Central Tendency
- Measure of Dispersion
- Probability Distribution
- Gaussian Normal Distribution
- Skewness and Kurtosis
- Regression Analysis
- Continuous and Discrete Functions
- Goodness of Fit
- Normality Test
- ANOVA
- Homoscedasticity
- Linear and Non-Linear Relationship with Regression

Inferential Statistics

- t-Test
- z-Test
- Hypothesis Testing

- Type I and Type II errors
- t-Test and its types
- One way ANOVA
- Two way ANOVA
- Chi-Square Test
- Implementation of continuous and categorical data

5 | Machine Learning

The best way to master machine learning algorithms is to work with the Scikit-Learn framework. Scikit-Learn contains predefined algorithms and you can work with them just by generating the object of the class. These are the algorithm you must know including the types of Supervised and Unsupervised Machine Learning:

- Linear Regression
- Logistic Regression
- Decision Tree
- Gradient Descent
- Random Forest
- Ridge and Lasso Regression
- Naive Bayes
- Support Vector Machine
- KMeans Clustering

Other Concepts and Topics for ML

- Measuring Accuracy
- Bias-Variance Trade-off
- Applying Regularization
- Elastic Net Regression
- Predictive Analytics
- Exploratory Data Analysis

6 | MLOps

You can master any one of the cloud services providers from AWS, GCP, and Azure. You can switch easily once you understand one of them.

We will focus on AWS—Amazon Web Services first

- Deploy ML models using Flask

- Amazon Lex—Natural Language Understanding

- AWS Polly—Voice Analysis

- Amazon Transcribe—Speech to Text

- Amazon Textract—Extract Text

- Amazon Rekognition—Image Applications
  Amazon SageMaker—Building and deploying models

- Working with Deep Learning on AWS

7 | Natural Language Processing

If you are interested in working with Text, you should do some of the work an NLP Engineer do and understand the working of Language models.

- Sentiment analysis
- POS Tagging, Parsing,
- Text preprocessing
- Stemming and Lemmatization
- Sentiment classification using Naive Bayes
- TF-IDF, N-gram,
- Machine Translation, BLEU Score
- Text Generation, Summarization, ROUGE Score
- Language Modeling, Perplexity
- Building a text classifier
- Identifying the gender

8 | Computer Vision

To work on image and video analytics we can master computer vision. To work on computer vision we have to understand images.

- PyTorch Tensors
- Understanding Pretrained models like AlexNet, ImageNet, and ResNet.
- Neural Networks
- Building a perceptron
  Building a single-layer neural network
- Building a deep neural network
- Recurrent neural network for sequential data analysis

Convolutional Neural Networks

- Understanding the ConvNet topology
- Convolution layers
- Pooling layers
- Image Content Analysis
- Operating on images using OpenCV-Python
- Detecting edges
- Histogram equalization
- Detecting corners
- Detecting SIFT feature points

9 | Data Visualization with Tableau

How to use it Visual Perception

- What is it, How it works, Why Tableau
- Connecting to Data
- Building charts
- Calculations
- Dashboards
- Sharing our work
- Advanced Charts, Calculated Fields, Calculated Aggregations
  Conditional Calculation, Parameterized Calculation

10 | Structured Query Language (SQL)

- Fundamental to SQL syntax and Installation
- Creating Tables, Modifiers
- Inserting and Retrieving Data, SELECT INSERT UPDATE DELETE
- Aggregating Data using Functions, Filtering, and RegEX
- Subqueries, retrieve data based on conditions, grouping of Data.
- Practice Questions
- JOINs
- Advanced SQL concepts such as transactions, views, stored procedures, and functions.
- Database Design principles, normalization, and ER diagrams.
- Practice, Practice, Practice: Practice writing SQL queries on real-world datasets, and work on projects to apply your knowledge.

11 | Data Engineering

BigData

- What is BigData?
- How is BigData applied within Business?

PySpark

- Resilient Distributed Datasets
- Schema
- Lambda Expressions
- Transformations
- Actions

Data Modeling

- Duplicate Data
- Descriptive Analysis of Data

- Visualizations
- ML lib
- ML Packages
- Pipelines

Streaming

- Packaging Spark Applications

12 | Data System Design

What is system design?

- IP and OSI Model
- Domain Name System (DNS)
- Load Balancing
- Clustering
- Caching
- Availability, Scalability, Storage

Databases and DBMS

- SQL databases
- NoSQL databases
- SQL vs NoSQL databases
- Database Replication
- Indexes
- Normalization and Denormalization
- CAP theorem

System Design Interview

- URL Shortener
- Whatsapp, Twitter, Netflix, Uber

13 | Five Major Projects and Git

We follow project-based learning and we will work on all the projects in parallel.

14 | Interview Preparation

15 | Git & GitHub

- Understanding Git
- Commands and How to commit your first code?
- How to use GitHub?
- How to make your first open-source contribution? ● How to work with a team?—Part 1 ● How to create your stunning GitHub profile?
- How to build your own viral repository?
- Building a personal landing page for your Portfolio for FREE ● How to grow followers on GitHub?
- How to work with a team? Part 2—issues, Milestones, and projects