Data Science

Machine Learning

Full Stack Roadmap

Himanshu Ramchandani
M.Tech | Data Science

The Roadmap is divided into 12 Sections

Duration: 100 Hours (4 to 5 Months)

1. Python Programming and Logic Building

2. Data Structure & Algorithms

3. Pandas Numpy Matplotlib

4. Statistics

5. Machine Learning

6. Natural Language Processing

7. Computer Vision

8. Data Visualization with Tableau

9. Structure Query Language (SQL)

10. Big Data and PySpark

11. Development Operations with Azure

12. Five Major Projects and Git

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

1 | Python Programming and Logic Building

Basics

01. Variables
02. Print function
03. Input from user
04. Data Types
    a. Numbers
    b. Strings
    c. Lists
    d. Dictionaries
    e. Tuples
    f. Sets
    g. Other Types
05. Operators
    a. Arithmetic Operators
    b. Relational Operators
    c. Bitwise Operators
    d. Logical Operators
06. Type conversion

Control Statements

1. If Else
    a. If
    b. Else
    c. Else If
    d. If Else Ternary Expression
2. While Loops

    a. Nested While Loops
    b. Break
    c. Continue
    d. pass
    e. Loop else

Lists

1. List Basics
2. List Operations
3. List Comprehensions
4. List Methods

Strings

1. String Basics
2. String Literals
3. String Operations
4. String Comprehensions
5. String Methods

For Loops

1. Functions
2. Nested For Loops
3. Break
4. Continue
5. Pass
6. Loop else

Functions

1. Definition
2. Call
3. Function Arguments
4. Default Arguments
5. Docstrings
6. Scope
7. Special functions Lambda, Map, and Filter
8. Recursion
9. Functional Programming and Reference Functions

Dictionaries

1. Dictionaries Basics
2. Operations
3. Comprehensions
4. Dictionaries Methods

Tuples

1. Tuples Basics
2. Tuples Comprehensions
3. Tuple Methods

Sets

1. Sets Basics
2. Sets Operations
3. Union

4. Intersection
5. Difference and Symmetric Difference

File Handling

1. File Basics
2. Opening Files
3. Reading Files
4. Writing Files
5. Editing Files
6. Working with different extensions of file
7. With Statements

Exception Handling

1. Common Exceptions
2. Exception Handling
    a. Try
    b. Except
    c. Try except else
    d. Finally
    e. Raising exceptions
    f. Assertion

Object-Oriented Programming

1. Classes
2. Objects
3. Method Calls
4. Inheritance and Its Types
5. Overloading
6. Overriding
7. Data Hiding
8. Operator Overloading

Regular Expression

1. Basic RE functions
2. Patterns
3. Meta Characters
4. Character Classes

Modules & Packages

1. Different types of modules
2. Create your own module
3. Building Packages
4. Build your own python module and deploy it on pip

Magic Methods

1. Dunders
2. Operator Methods

2 | Data Structure & Algorithms

Analysis of Algorithms

Types of analysis

Asymptotic Notations

- Big O
- Omega
- Theta

Recursion and Backtracking

Stack

Queue

- Circular Queue

Trees

Linked Lists

- Insertion with Stack
- Insertion with Queue
- Deletion

Sorting

- Bubble Sort | Selection Sort | Insertion Sort | Quick Sort
- Merge Sort

Searching

- Linear Search | Binary Search

3 | Pandas Numpy Matplotlib

Numpy

1. Understanding Numpy
2. Basic working
3. Working with dimensions and matrix
4. Statistics basics Mainly descriptive
5. Linear algebra operations

Pandas

1. Dataframe basics
2. Different ways of creating a data frame
3. Read-write to excel
4. Handling missing values
5. Grouping data
6. Merging and Concat data frames

Matplotlib

1. Introduction
2. Formatting strings
3. Legend, grid, axis, labels
4. Bar chart
5. Histogram
6. Pie chart

4 | Statistics

Descriptive Statistics

Measure of Frequency and Central Tendency
Measure of Dispersion

Probability Distribution

Gaussian Normal Distribution
Skewness and Kurtosis

Hypothesis Testing

Type I and Type II errors
t-Test and its types

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
One way ANOVA
Two way ANOVA
Chi-Square Test
Implementation of continuous and categorical data

5 | Machine Learning

Linear Regression

1. Simple Linear Regression
    a. Evaluating the fitness of the model with a cost function
    b. Solving OLS for simple linear regression
    c. Evaluating the model
2. Multiple Linear Regression Polynomial regression
3. Applying linear regression
4. Exploring the data
5. Fitting and evaluating the model
6. Gradient descent
7. Working with Different datasets.
8. How to approach data science problems
9. Datasets
    a. House Price Prediction
    b. Salary prediction based on GMAT score
    c. Predicting the sold price of players in IPL
10. Summary

Logistic Regression

1. Logistic Regression
2. Binary Classification
3. Performance Matrix
4. Accuracy

5. Precision and Recall
6. F1 measure
7. ROC AUC
8. How to approach Classification problems
9. Datasets
    a. Predicting Insurance
    b. Spam filtering
    c. Digit Classification
    d. Titanic Dataset
10. Summary

Decision Tree

1. Decision Tree
2. Nonlinear Classification and Regression
3. Training decision trees
4. Selecting the questions
5. Information gain
6. Gini impurity
7. Implementation with Scikit-learn
8. Working with datasets
    a. Salary Prediction
9. Summary

Random Forest

1. Ensemble
2. Bagging
3. Bosting
4. Stacking
5. Fast parameter optimization with randomized search
6. Datasets
7. Summary

Naive Bayes

1. Naive Bayes mathematical concept
2. Bayes' theorem
3. Generative and discriminative models
4. Naive Bayes
5. Assumptions of Naive Bayes
6. Solving dataset with problems
7. Summary

Understanding Interview questions

Data Science and Machine Learning interview questions with answers.

Support Vector Machines

1. Support Vector Machines
2. Linear SVM Classification
3. Nonlinear SVM Classification
    a. Polynomial Kernel
    b. Adding Similarity Features
4. SVM Regression
    a. Under the Hood
5. Hyperparameter optimization
6. Summary

Machine Learning Advanced Concepts

1. Gradient Descent
2. GD for Linear Regression
3. Steps for Building Machine Learning Models
4. Measuring Accuracy
5. Bias-Variance Trade-off
6. Applying Regularization
7. Ridge Regression
8. LASSO Regression
9. Elastic Net Regression
10. Predictive Analytics
11. Exploratory Data Analysis.

Clustering

1. How clustering works
2. Euclidean Distance
3. K-means clustering
4. Feature normalization
5. Working with datasets
6. Cluster interpretation
7. Summary

Recommendation Systems

1. Association rules
2. Collaborative filtering
3. Similarities
4. Surprise library
5. Building Recommendation Engine
6. Euclidean distance score
7. Pearson correlation score
8. Generating movie recommendations
9. Summary

6 | Natural Language Processing

Text Analytics

1. Sentiment analysis
2. Working with dataset
3. Text preprocessing
4. Stemming and Lemmatization
5. Sentiment classification using Naive Bayes
6. TF-IDF
7. N-gram
8. Building a text classifier
9. Identifying the gender
10. Summary

Speech Recognition

1. Understanding Audio Signals
2. Transforming audio signals into the frequency domain
3. Generating audio signals with custom parameters
4. Synthesizing music
5. Extracting frequency domain features
6. Building Hidden Markov Models
7. Building a speech recognizer
8. Summary

7 | Computer Vision with PyTorch

Neural Networks

1. Introduction
2. Building a perceptron
3. Building a single layer neural network
4. Building a deep neural network
5. Building a recurrent neural network for sequential data analysis
6. Visualizing the characters in an optical character recognition database
7. Building an optical character recognizer using neural networks
8. Summary

Convolutional Neural Networks

1. Introducing the CNN
2. Understanding the ConvNet topology
3. Understanding convolution layers
4. Understanding pooling layers
5. Training a ConvNet
6. Putting it all together
7. Applying a CNN
8. Summary

Image Content Analysis

1. Introduction
2. Operating on images using OpenCV-Python
3. Detecting edges
4. Histogram equalization
5. Detecting corners
6. Detecting SIFT feature points
7. Building a Star feature detector
8. Building an object recognizer
9. Summary

Biometric Face Recognition

1. Introduction
2. Capturing and processing video from a webcam
3. Building a face detector using Haar cascades
4. Building eye and nose detectors
5. Performing Principal Components Analysis
6. Performing Kernel Principal Components Analysis
7. Performing blind source separation
8. Building a face recognizer
9. Summary

Integration with Web Apps

1. Understanding Flask
2. Recalling HTML CSS JavaScript.
3. Integrate Flask and Machine Learning

Deployment

1. Flask
2. Heroku

Extra Projects

1. Breast Cancer Classification using Scikit Learn
2. Fashion Class classification using TensorFlow and PyTorch
3. Directing Customers to Subscription Through App Behavior Analysis
4. Minimizing churn rate through analysis of financial habits.
5. Credit Card fraud detection.
6. Live Sketch with Webcam using OpenCV
7. Building Chatbot with Deep Learning.

8 | Data Visualization with Tableau

How to use it
Visual Perception

Tableau

What is it

How it works

Why Tableau

Installing Tableau

Connecting to Data

Building charts

Calculations

Dashboards

Sharing our work

Advanced Charts

Calculated Fields

Calculated Aggregations

Conditional Calculation

Parameterized Calculation

9 | Structure Query Language (SQL)

Setup SQL server

Basics of SQL

Writing queries

Data Types

Select

Creating and deleting tables

Filtering data

Order

Aggregations

Truncate

Primary Key

Foreign Key

Union

MySQL

Complex Questions

Solving Interview Questions

10 | Big Data and PySpark

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 on Data
Visualizations

ML lib

ML Packages
Pipelines

Streaming

Packaging Spark Applications

11 | Development Operations with Azure, GCP or AWS

Foundation of Data Systems

Data Models

Storage

Encoding

Distributed Data

Replication

Partitioning

Derived Data

Batch Processing

Stream Processing

Microsoft Azure

Azure Data Workloads

Azure Data Factory

Azure HDInsights

Azure Databricks

Azure Synapse Analytics

Relational Database in Azure

Non-relational Database in Azure

12 | Five Major Projects and Git

Git - Version Control System

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

Data Science ML Full Stack Roadmap
https://github.com/hemansnation/Data-Science-ML-Full-Stack-2022

Join Telegram for Data Science ML AI Resources:
https://t.me/+sREuRiFssMo4YWJl

Connect with me on these platforms:
LinkedIn: https://www.linkedin.com/in/hemansnation/

Twitter: https://twitter.com/hemansnation

GitHub: https://github.com/hemansnation

Instagram: https://www.instagram.com/masterdexter.ai/

Are you a professional?
DM for One-on-One sessions for Python, Data Science, Machine Learning, and Data Engineering.
Here: https://bit.ly/3U6zQvQ

Contact for any Query: +91 9074919189

End of Document