hemansnation / **Data-Science-ML-Full-Stack** Public

<> Code   ⊙ Issues 2   ⣼ Pull requests   ▷ Actions   ▦ Projects   📖 Wiki   ⊘ Security   ⬚ Insights   ⚙ Settings

⎇ master ▾   ⋯

Data-Science-ML-Full-Stack / **README.md**

hemansnation Update README.md   ⟲

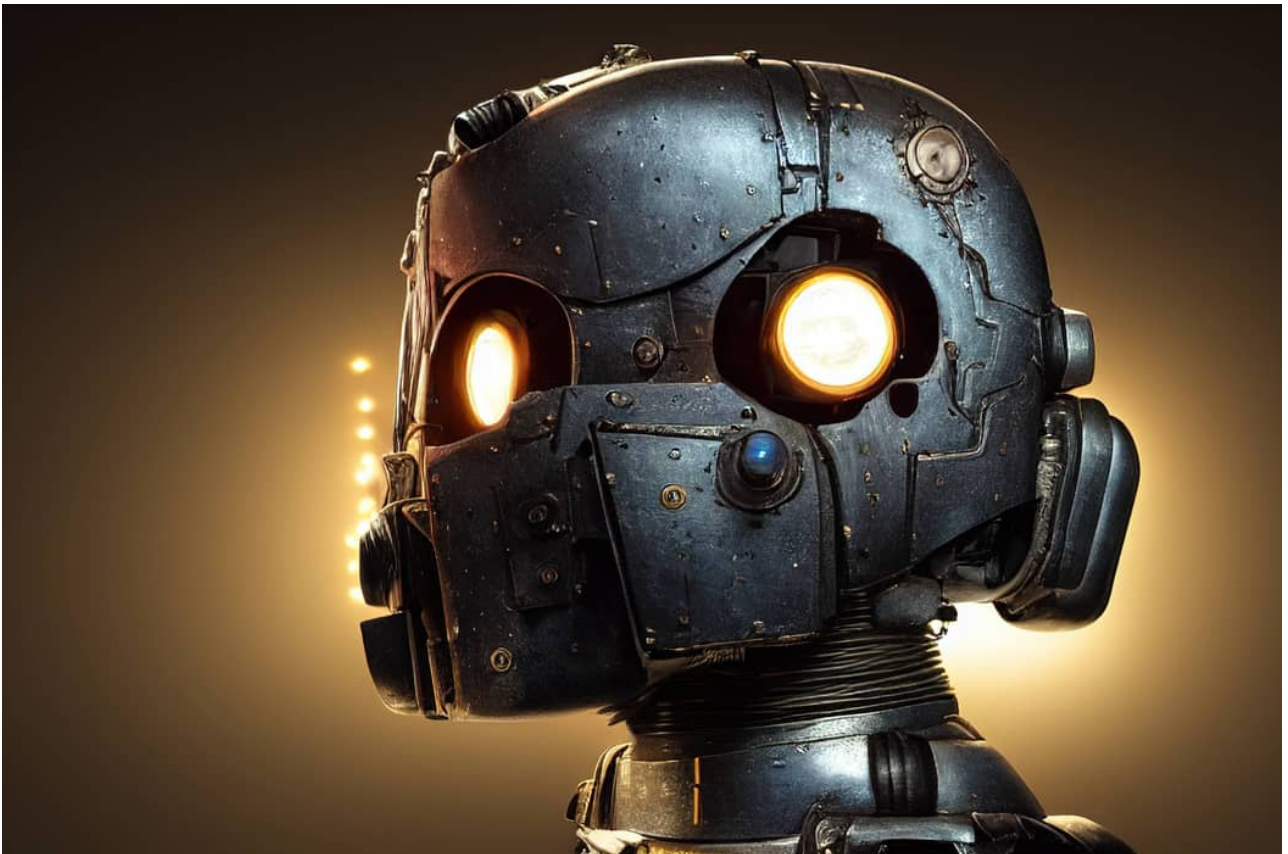⣿ **2 contributors**   👤 🏛

≔   362 lines (268 sloc) | 9.59 KB   ⋯

🔗 Data Science ML Full Stack

What we will do and gain?

- Build an in-depth understanding of all the data concepts.
- Create your strong social media profile on LinkedIn and GitHub.
- Build 15+ projects including 5+ Major Projects.
- Showcase your skills with a portfolio of real projects.
- Work on Live projects in parallel to understand how companies create end-to-end software solutions and apply ML models to real-life problems.

The Roadmap is divided into 12 Sections

Duration: 100 Hours (6 Months)

1. Python Programming and Logic Building
2. Data Structure & Algorithms
3. Pandas Numpy Matplotlib
4. Statistics
5. Machine Learning
6. ML Operations
7. Natural Language Processing
8. Computer Vision
9. Data Visualization with Tableau
10. Structure Query Language (SQL)
11. Data Engineering
12. Data System Design
13. Five Major Projects
14. Git & GitHub
15. Interview Preparations
16. Personal Branding and portfolio

Resources

- Dataset Collection

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

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

- Python basics, Variables, Operators, Conditional Statements
- List and Strings
- While Loop, Nested Loops, Loop Else
- For Loop, Break, and Continue statements
- Functions, Return Statement, Recursion
- Dictionary, Tuple, Set
- File Handling, Exception Handling
- Object-Oriented Programming
- Modules and Packages

In-Depth Roadmap of Python

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.

Understand these topics

- Types of Algorithm Analysis
- Asymptotic Notation, Big-O, Omega, Theta
- Stacks
- Queues
- Linked List
- Trees
- Graphs
- Sorting
- Searching
- Hashing

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 have 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, replace 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

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, 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 | 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

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 on Data
- Visualizations
- ML lib
- ML Packages
- Pipelines

Streaming

- Packaging Spark Applications

12 | Data System Design

- 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

13 | Five Major Projects and Git

Git - Version Control System

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

14 | Git & GitHub

15 | Interview Preperation

16 | Personal Profile & Portfolio

Resources

Datasets

1 Awesome Public Datasets This list of a topic-centric public data sources in high quality.

2 NLP Datasets Alphabetical list of free/public domain datasets with text data for use in NLP.

3 Awesome Dataset Tools A curated list of awesome dataset tools.

4 Awesome time series database A curated list of time series databases.

5 Awesome-Cybersecurity-Datasets A curated list of amazingly awesome Cybersecurity datasets.

6 Awesome Robotics Datasets Robotics Dataset Collections.

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?

For One-on-One sessions for Python, Data Science, Machine Learning, and Data Engineering.
Email your requirements Here: connect@himanshuramchandani.co

Give feedback