Data Science Roadmap for Beginners

Agenda of this Notebook:

- What is Data Science?
- What are the tools and techniques, and aplications of Data Science?
- Programming Language for Data Science
- Practice Projects
- Version Control System (Git and Github)
- Libraries for Data Science
- Mathematics for Data Science
  - Statistics
  - Probability
  - Linear Algebra
  - Calculus
  - Optimization
- Data Acquisition Skills for Data Science
  - Web Scraping
  - SQL and NOSQL
- Exploratory Data Analysis
- Introduction of Machine Learning
- Introduction of Deep Learning
- Introduction of Big Data
- Real World Projects

In [ ]:

What is Data Science?

- **A data scientist is a professional responsible for collecting, analyzing and interpreting extremely large amounts of structured and unstructured data in order to gain useful insights to grow the business.**
- **Applications of Data Science**:
  - There are many applications like **social media**(Recommendation Systems, Sentiment Analysis, Ad placement), **E-Commerce**(Upselling Cross selling, Discount Price Optimization, Business Forecasting), **Banking** (Credit Scoring, Fraud Detection, Anti-Money Laundering, Price Optimization), **Traveling** (Dynamic Pricing,Predict Flight Delay, Best route selection), **Automation** (Self-driving cars, Robots), **HealthCare** (Disease Prediction, Medical Imaging) and so on.
- **Skills set of a Data Scientist**:
  - Statistics, Programming Languages, Data extraction & processing, Data wrangling & exploration, Big Data processing, Machine Learning framework , Data Visualization.

Programming Language for Data Science:

There are lots of programming languages for Data Science but the famous are `Python` and `R` . Python is a base for any Data Scientist. There are many reasons to select this powerful programming language, so it's up to you which reason will be main. You should definitely consider Python due to its possibilities and ongoing improvement, which will help you to build amazing products and help businesses.

- You can learn python from this Youtube Channel for Python
- For Practice you can use this website Python practice exercises
- Reading Book for Python is available at this link Starting out with Python

In [ ]:

Projects for Beginners

After completing your python, I will recommend you to work on some projects. Here is list of some

- Hotel Management System
- School Management System
- Library Management System
- Tic-Tac-Toe Game
- Personal Contact Book

In [ ]:

Version Control System(Git and Github)

- Git and GitHub are essential tools for any data science professional who wants to code.
- Learn git and github in One Video
- Reading material for git and github is available at this link New to Git and GitHub?

In [ ]:

Python Libraries for Data Science Tasks

- Tools and libraries are softwares that are used to apply DS Data Scientist techniques to perform a task. There are several important libraries of Python for Data Science:
  - NumPy
  - Pandas
  - Matplotlib & Seaborn
  - SciPy
  - Tensorflow
  - PyTorch
  - ScikitLearn
  - Beautifulsoup etc.
- There are lot of resources available on internet to learn these libraries with practice questions.
  - Learn NumPy from this NumPy Youtube Channel
  - Practice Questions related to NumPy are available at this link NumPy Exercises with Solution
  - Learn Pandas from this Pandas Youtube Channel
  - Practice Questions related to Pandas are available at this link Pandas Exercies with solutions
  - Learn Matplotlib from this Matplotlib Youtube Channel
  - Practice Questions related to Matplotlib are available at this link Matplotlib Exercies with solutions
  - Learn Seaborn from this Seaborn Youtube Channel

In [ ]:

Mathematics for Data Science

- Mathematics is very important in the field of data science as concepts within mathematics aid in identifying patterns and assist in creating algorithms. The understanding of various notions of Statistics and Probability Theory are key for the implementation of such algorithms in data science. Notions include: Regression, Maximum Likelihood Estimation, the understanding of distributions (Binomial, Bernoulli, Gaussian (Normal)) and Bayes' Theorem.
  - Learn Statistics and Probability for Data Science from this youtube channel statistics 101 course
  - Leanr linear algebra for Data Science from this youtube channel Linear Algebra for Data Science
  - Learn calculus for Data Science from this youtube channel Calculus for Data Science
  - Learn what is optimization in one video from this youtube channel Introduction to Optimization

In [ ]:

Data Acquisition skills for Data Science

- Data acquisition is the processes for bringing data that has been created by a source outside the organization, into the organization, for production use.
  - Learn web scraping skills from this youtube channel Python Web Scraping
  - Learn SQL skills from this youtube channel SQL for Data Science
  - Learn NoSQL Skills from this youtube channel NoSQL for Data Science

In [ ]:

Exploratory Data Analysis

- Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.
  - Learn Feature engineering from this youtube channel Feature engineering in machine learning
  - Learn EDA from this youtube channel EDA for Data Science

In [ ]:

Exploratory Data Analysis Projects

- Exploratory data analysis on various datasets from FiveThirtyEight and Udacity coursework.
- Here is link of all projects Exploratory Data Analysis Projects

In [ ]:

Introduction to Machine Learning

- Machine Learning is defined as the study of computer programs that leverage algorithms and statistical models to learn through inference and patterns without being explicitly programed. Machine Learning field has undergone significant developments in the last decade.
  - Learn Machine Learning from this link Introduction to Machine Learning
  - Practice questions for ML is available at this link Machine learning - Scikit-learn Exercises with Solutions

In [ ]:

Introduction to Deep Learning

- Deep learning is a branch of machine learning. Unlike traditional machine learning algorithms, many of which have a finite capacity to learn no matter how much data they acquire, deep learning systems can improve their performance with access to more data: the machine version of more experience.
  - Learn deep learning from this link Introduction to Deep Learning

In [ ]:

Overview of Big Data

- Big Data is a term that is used for denoting a collection of datasets that is large and complex, making it very difficult to process using legacy data processing applications
  - Learn Big data skills from this linkIntroduction to Big Data

In [ ]:

Projects list

- Projects are very important to make a strong grip on any field. Here is list of some important projects that accelerate your profile
  - BigMart Sales Prediction ML Project.
  - Music Recommendation System ML Project.
  - Stock Prices Predictor using TimeSeries.
  - MNIST Handwritten Digit Classification.
  - Social Media Sentiment Analysis Using Twitter Dataset
  - Coupon Purchase Prediction
  - Data Mining for Sales Prediction in Tourism Industry
  - Web Data Mining To Detect Online Spread Of Terrorism

In [ ]:

Other Points to Remember

- Keep your eye on Kaggle
- Keep your eye on LinkedIn Profile for job interviews
- Keep your eye on Github trending repositories
- Domain specific knowledge helps in getting a job
- Good communication skills and spoken english helps!
- Staying ahead of the competition helps
- I wish you all the best in learning data science. Hope all this information helps

In [ ]:

Note: If you have any query or ambiguity, then you can contact with me at LinkedIn and Instagram

Ehtisham Sadiq

In [ ]: