Data Science for Beginners: 2023 - 2024 Complete Roadmap.

Data handling is a field that has grown ten folds in the recent few years since the adoption of the internet age. The amount of raw data received is so enormous that it is estimated that Google alone handles around 20 Petabytes of data that is around 20,000,000 Gigabytes of data. That data in its raw form alone is useless unless it is used to give meaningful insights and give conclusions toward bettering processes it is not valuable. The process of cleaning, analyzing presenting and visualizing data that’s where Data Science comes in.

Data Science is afield that deals with mathematics, statistics and programming to clean and analyze raw data such that you get meaningful insights from it. It entails complex mathematical and scientific calculations as well as programming concepts to train machine models to handle big data and provide better prediction models. A data science project goes through data extraction, preparation, cleaning, modelling and evaluation. To be efficient in this a Data Scientist needs to be armed wit a group skills that we will discuss.

# **Programming.**

Programming is one of the key pillars in data science as this is what you will use to create algorithm and use in machine learning and data processing. The major programming languages are Python ,R,SAS,SQL. This enables you to create good programming logics to model your data. It is eminent to learn any of this languages and any frameworks related to them as far as data science is concerned

# **Statistics and Probability**

To create high quality machine learning models and algorithms statistics and probability are crucial it is very important to understand statistics concepts like mean, mode, variance and standard deviation. Also its important to know Probability distributions

Over and under sampling, Bayesian and frequentist statistics,Dimension reduction

# **Data Wrangling and Database Management**

Data wrangling is the process of cleaning and organizing huge datasets to make them easy to analyze and work with . The data can then be classified into tabular models and stored in a database so as to make the access and analysis of data simpler. Data wrangling tools include: Altair ,Talend, Alteryx. Popular database systems include MySQL, MongoDB and Oracle. This tools are imperative to you Data Science journey

# **Machine Learning and Deep Learning**

Machine learning involve focusing on data and algorithm in away that human interacts with the data to help come up with a prediction model close to human behavior. Deep learning on the other hand involves teaching computer systems human behavior. This helps improve the predictive results of a model. Its important to know Linear Regression , Logical regression, Naïve Bayes, Decision Tree and other relevant algorithms

# **Data Visualization**

This involves the use of charts, graphs and dashboard to present data With strong visualization skills, you can present your work to stakeholders so that the data tells a compelling story of the business insights. Data Visualization tools include Tableau, Microsoft Excel, Power Bi

# **Cloud Computing**

This helps data scientists access clouds based databases and frameworks that help in advancing technology. Major cloud computing tools include Amazon Web Service , Google Cloud and Microsoft Azure

As a data scientist it is crucial to be competent in the above skills to help grow your career