PDS Assignment 01_Group A

Page :
Date :

Q. Write in details, What is Data science? Description of various phases in data science projects in details also mention various tools used in every phase. Explaine one application of data science thoroughly.

What is Data Science?

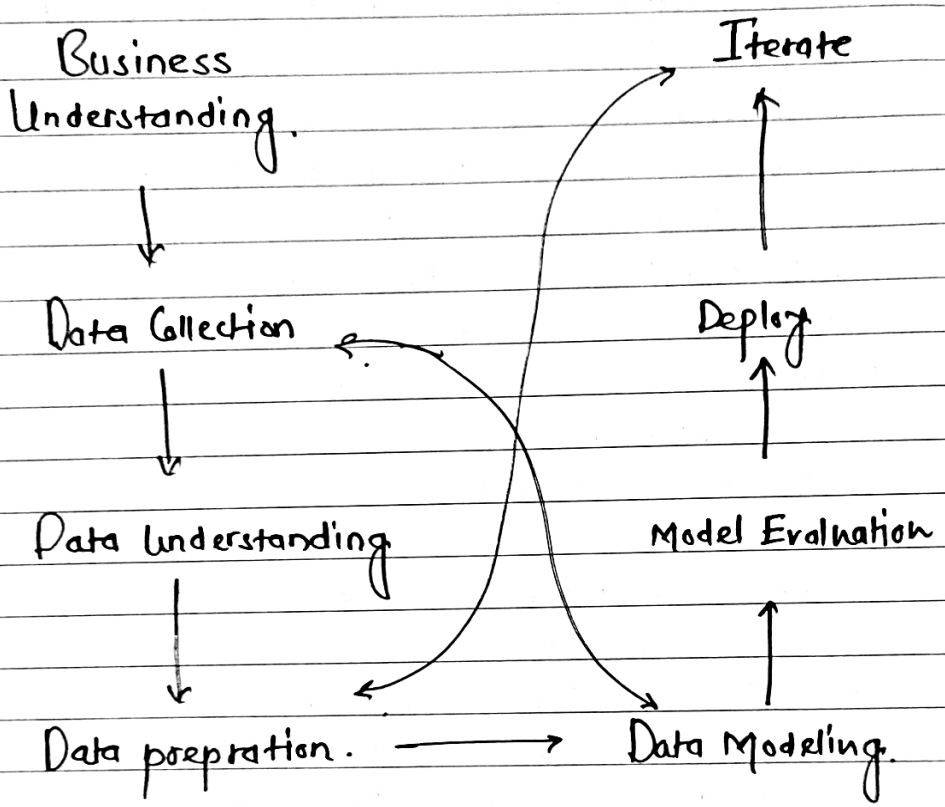
Data science combines multiple fields, including statistics, scientific methods, artificial intelligence (AI), and data analysis, to extract value from data. Those who practise data science are called as scientists and they combine a range of skills to analyze data collected from the web, smartphones, customers, sensors and other sources to derive actionable insights.

Data Science encompass preparing data for analysis, including cleansing, aggregating and manipulating the data to perform advanced data analysis. Analytic applications and data scientists can then review the results to uncover pattern and enable business leaders to draw informed insights.

rudraksh.karpe.cs@ghrcem.raisoni.net

Page :
Date :

Various phases in a Data Science project:



(1) Business Understanding (Asking the right question).

Access to data and the computing power have been increased tremendously in the last ~~year~~ decade the success of an Organisation still largely depends upon the quality of questions ask of their data sets. The amount of data collected and the Compute-power allocated is less of a differentiator.

rudraksh.karpe.cs@ghrcem.raisoni.net

Business Understanding is the term coined by IBM Data scientist John B. Rollins. eg. Google in its beginning would have asked "What constitutes relevant search result". As Google as a search engine it started showing ads. then the right question to ask would've been "which ads are relevant to users".
A few right questions that successful businesses have asked in the past of their data science teams.

- AMAZON → How much compute (EC2) and storage (S3) space could they lease out during lean periods?
- Uber → What percentage of time do drivers drive? How steady is their income?

(2) Data Collection

The data scientist needs to know which ingrediants are required, how to source and collect them, and how to prepare data to meet the desired outcome.

eg. For AMAZON the type of data required would be

- Number of Computational servers lying free during lean period.
- Number of storage servers lying unused during lean period.
- The amount of money being spent to maintain these machines.

Methods used for Data Collection →
Case studies, Checklists, Interviews, Observation sometimes and Surveys or Questionnaries are tools to collect data

Page :
Date :

(3) Data prepration →

Data prepration is the process of data gathering, combining, structuring and organizing so it can be used in business intelligence (BI), analytics and data visualization applications. The components of Data pre processing include data pre-processing, profiling, cleansing, validation and transformation it often also involves pulling together data from different internal systems and external sources.

Tools used for Data prepration →

tye, Data ladder, Microsoft power Bi, Tableau prep, Infogix Data360, Tamr Unify, Talend, Altreyx Analysis, Altair Monarch, paxata, Trifacta.

(4) Data Modeling →

Data Modeling is a process of creating visual representation of either a whole information on system or parts of it to communicate connections between data points and structures.
The goal is to illustrate types of data used and stored within system, the relationships among these datatypes, the way data can be organised and grouped and its format attributes.

Tools used for Data Modelling →

Draw.io, Lucidchart, SQWirreL SQL Client, MySQL Workbench, Amundsen, erwin Data Modeler, ER/studio, Postico, Navicat, Datagrip.

(5) Deploy and Iterate

Eventually we deploy all data science projects into the real world. The deployment could be through an Android or just an ios app just like cred or it could be through a webApp like monkeycontrol.com or it could be as an enterprise software like IBM Watson.

The concept of deployment in Data science refers to the application of model for prediction using new data. Building a model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organised and presented in a way that the consumer can use it.

Depending upon the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repetable data science processes. In many cases it will be consumer not the data analyst who will carry out the deployment steps.

Tools used for deployment →

1. Docker
2. Gradio
3. Kubernetes
4. SageMaker
5. MLFlow
6.

rudraksh.karpe.cs@ghrcem.raisoni.net

Page :
Date :

Application of Data science in 'Augmented Reality'

- Augmented Reality is one of the most exciting uses of technology.
- As a VR headset incorporates computer expertise, algorithms, and data to provide you with the greatest viewing experience, Data science and Virtual Reality have a connection. The popular game pokemon Go is a modest step in the right direction.

The ability to wander about and gaze at pokemon on walls, streets, and other non-existent objects. To determine the location of pokemon and gyms, the game designers used data from Ingress, the Company's previous software.

Data science will make more sense if VR economy becomes more affordable and Consumers begins to utilize it in the same way they do other applications.

rudraksh.karpe.cs@ghrcem.raisoni.net