

Arya Nilesh Kumbhar

Roll no. - A47 TY [CSE]

PRN - 2122000375

Course Name - Data Science for Everyone.

No. : 1.

Date :

Assignment : 1 -

1. What is Data Science ?

- i. Data science is the study of data to extract meaningful insights for business.
- ii. It is a multidisciplinary approach that combines principles & practice from the fields of mathematics, statistics, artificial intelligence and computer engineering to analyze large amounts of data.
- iii. This analysis helps data scientists to ask & answer questions like what happened, why it happened, what will happen & what can be done with the results.
- iv. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio & more to produce artificial intelligence (AI) system to perform tasks that ordinarily require human intelligence.

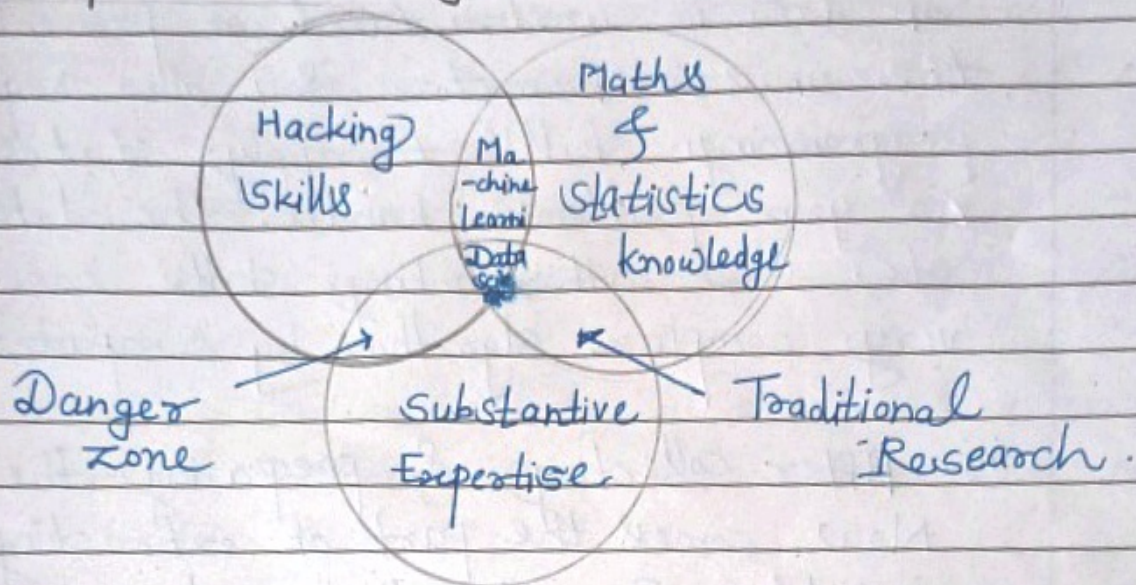
KIT College Of Engineering, Kolhapur.
(An Autonomous Institute)

2. Explain the importance of data science.

- i. Data science is important because it combines tools, methods & technology to generate meaning from data.
- ii. Modern Organizations are inundated with data; there is a proliferation of devices that can automatically collect & store information.
- iii. Online systems & payment portals capture more data in the fields of e-commerce, medicine, finance & every other aspect of human life.
- iv. We have text, audio, video, image data available in vast quantities.
- v. They are important for,
 - Unlocking Insights
 - Efficient data handling
 - Predictive Power &
 - Competitive Edge.

KIT College Of Engineering, Kolhapur.
(An Autonomous Institute)

3. Explain Venn Diagram.



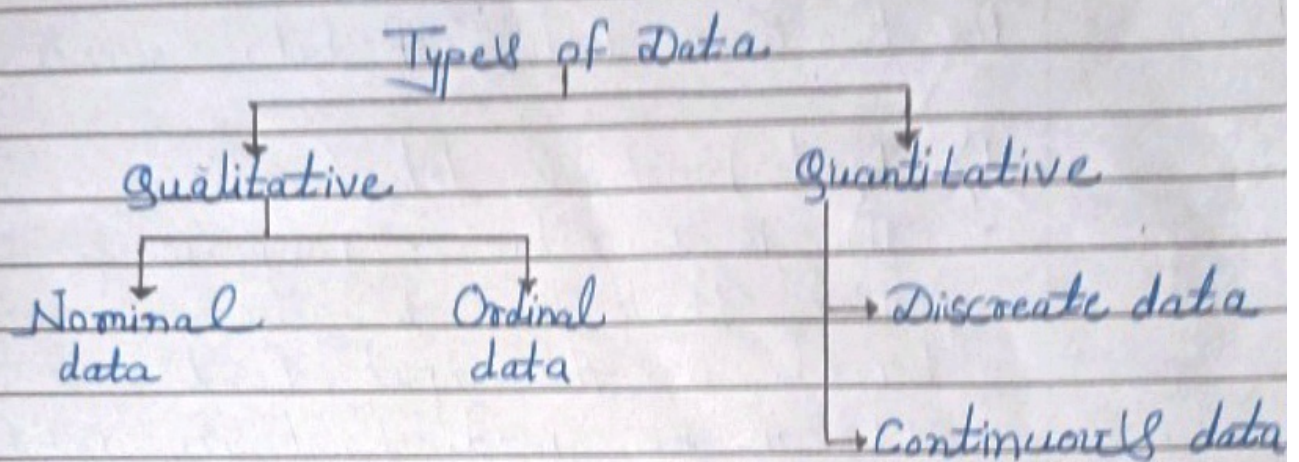
- i. According to data science Venn diagram, Machine Learning involves the knowledge of computer programming & maths & but without any domain expertise. This just means that you just need to throw your data into the model without necessarily knowing about the details of the data such as what data it is, what it means, etc.
- ii. Hacking requires, great coding skills. Coding is important because it helps you

to gather & prepare the data. A lot of data is unstructured or present in the unusual formats. You also require programming skills to apply statistics to your problems, handle the database, etc. One with hacking skills can apply very complex algorithm by computer.

iii. After collecting & preparing the data Now comes the part of extracting the insights from it. Mathematics is important for analyzing the data. For Any Analyzing the data, you will require several tools from maths such as probability, algebra, etc.

It helps in the diagnosis of the problem by applying various mathematical & statistical approaches to data.

4. Explain different data types -



1. Nominal data -

- Nominal data is used to label variables without any order or quantitative value. The colour of hair can be considered as a Nominal data, as one colour can't be compared with other.

- With the help of Nominal data, we can't do any numerical tasks or can't give any order to sort the data. These data don't have any meaningful order, their values are distributed to distinct categories.

2. Ordinal Data -

- Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale.

These data are used for observation, but we can't do any arithmetical tasks on them.

- The Ordinal data is qualitative data for which the values have some kind of relative position.

These kind of data can be considered as "in-between" the qualitative data & the quantitative data.

- The Ordinal data only shows the sequences & cannot use for statistical analysis.

3. Discrete Data -

- The term discrete means distinct or separate. The discrete data contain the values that fall under integers or whole numbers.

- The total number of student in a class is an example of discrete data.

- These data can't be broken into decimal or fraction values.
- The discrete data are countable & have finite values.
- These data are presented mainly by a bar graph, number line or frequency table.
- E.g. - Cost of a cell phone.

4. Continuous Data-

- Continuous data are in the form of fraction, or fractional numbers. They represent information on that can be divided into smaller levels.
- They can vary in any range.
- The key difference between discrete & continuous data is that discrete data contains whole number.
- They store fractional numbers to record different data like temperature, height, width, time, speed, etc.
- E.g. → Height of Person.

5. Mention any two case studies.

The two case studies of data science are -

1) Text Emotions detection :

- This is a natural language processing problem that aims to generate emojis based on the input text.
- It can be used to train artificial intelligence chatbots that can express emotions.

2) Hotel Recommendation System :

- This is a collaborative filtering problem that aims to predict which hotel a user is most likely to choose from among all hotels.
- It can be done using customer ratings & reviews.