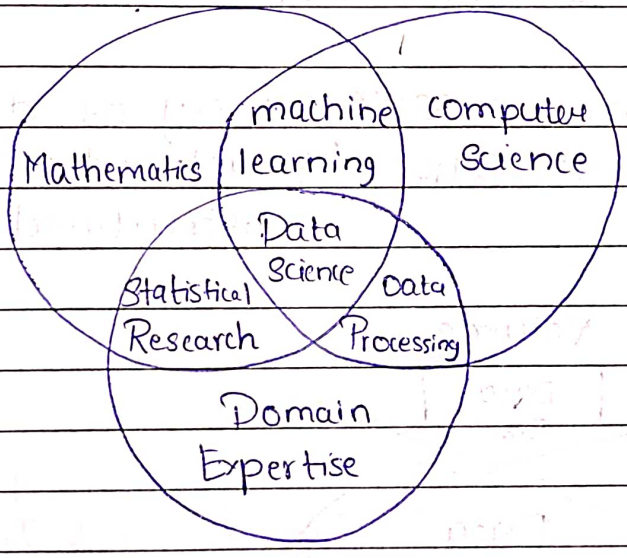
Introduction to Data Science

* • Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data.
• Data science improves the quality of data.
• Data science improves quality of services and products.
• Real-life examples
(i) Internet search
(ii) digital advertisements
(iii) Gaming world
(iv) Healthcare
(v) Transport.



Data Science and Associated fields.

Page No.
Date

★ Big Data

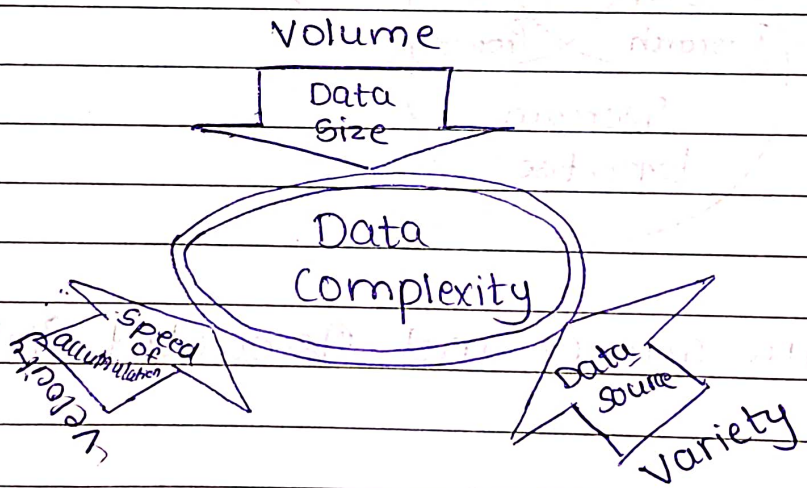
- Large volume of data.
- collection of large datasets that cannot be processed using traditional computing techniques
- Big data is any data that is expensive to manage and hard to extract value from.

★ The 3 V's

(I) Velocity :- The speed at which data is accumulated.

(II) Volume :- The size and scope of data.

(III) Variety and Complexity :- The massive array of data and types (structured and unstructured)



3V's in data Science.

Page No.
Date

★ Types of Data.

(i) Structured Data.

⇒ • Data or record that stored in fixed field within file
• Data is stored in predefine and searchable format of DBMS/RDms or warehouse or Excel files.
• Generally it contains numbers, text and generated by machine or manually.
• Example:- ~~ATM machine~~./ Excel

(ii) Semi-Structured Data.

⇒ • Data or record that have self-describing structure.
• Such data does not fit into stored in predefine and searchable format of DBMS/RDms or warehouse or Excel files.
• Generally it contains numbers, text and generated by machine or manually.
• Eg-HTML files.

(iii) Unstructured Data.

• Data or record that is not in predefined structured format or not having any kind of data model.
• Data is not easy to fit into a data model because the content is context-specific or varying.
• Example: Email.