SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data science and Analytics
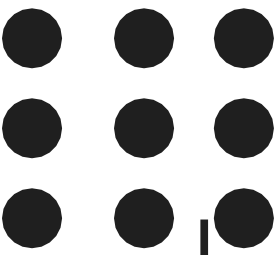
II Year / IV Semester

Unit 1 – Introduction to Data science

Facets of Data

Introduction to Data

• Data is crucial for data science and comes in different forms.

•Understanding data's characteristics is key for effective analysis.

•Data can be categorized into structured, unstructured, and semi-structured.

•These types impact analysis methods and tools used.

Types of Data

Structured Data: Organized in fixed formats (tables, databases).
Unstructured Data: No predefined format (social media, images).
Semi-Structured Data: Combination of structured and unstructured (JSON, XML).
Each type presents unique challenges and opportunities for analysis.

Structured Data

Structured data is highly organized and easy to store in databases.
Examples include sales records, customer details, and transactions.
It is easy to analyze using conventional tools.
This type of data is highly efficient for traditional data analysis.

Unstructured Data

Unstructured data lacks a predefined format, making it complex to process.
Examples include social media text, images, and sensor data.
It requires advanced techniques like NLP and image recognition.
Processing unstructured data demands sophisticated tools and technologies.

Semi-Structured Data

Unstructured data lacks a predefined format, making it complex to process.
Examples include social media text, images, and sensor data.
It requires advanced techniques like NLP and image recognition.
Processing unstructured data demands sophisticated tools and technologies.

The Five Vs of Data

Volume: The sheer amount of data generated.

Velocity: Speed at which data is generated and processed.

Variety: The diverse formats and sources of data.

Veracity:Data quality and reliability.

Value: The actionable insights derived from data.

Challenges in Data

Large volumes of data require scalable storage solutions.
High-velocity data demands real-time processing capabilities.
Ensuring data quality (veracity) is critical for accurate analysis.
Data diversity calls for robust integration and preprocessing techniques.
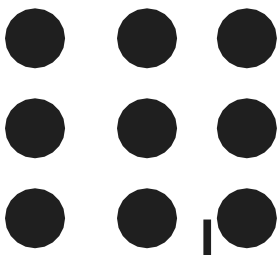
Opportunities in Data

Data can drive innovation and strategic decision-making.
Effective data analysis can lead to valuable business insights.
Understanding data facets helps organizations address challenges.
Proper tools and techniques unlock the potential of data for growth.

Conclusion

Grasping the facets of data is essential for successful analysis.

Data types and the "Five Vs" influence analysis methods.

Overcoming data challenges enhances decision-making and innovation.

Emphasizing data quality and diversity leads to actionable insights.

THANK YOU