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 Artificial Intelligence and Data Science

Course Name – Foundation of Data Science
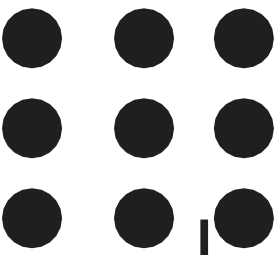
III Year / V Semester

Unit 1 – Introduction

Topic 1- Data Science Introduction

Data Science

Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data.

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions.
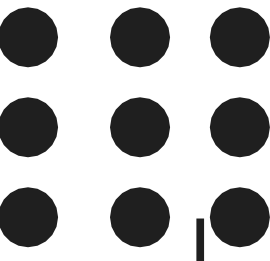
Data science uses complex machine learning algorithms to build predictive models.
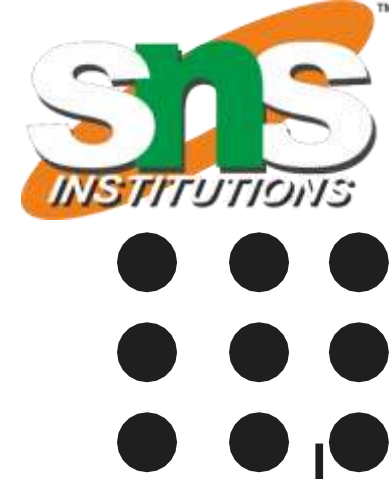
Data Science

Data science encompasses
- preparing data for analysis and processing,
- performing advanced data analysis, and
- presenting the results to reveal patterns and
- enable stakeholders to draw informed conclusions.
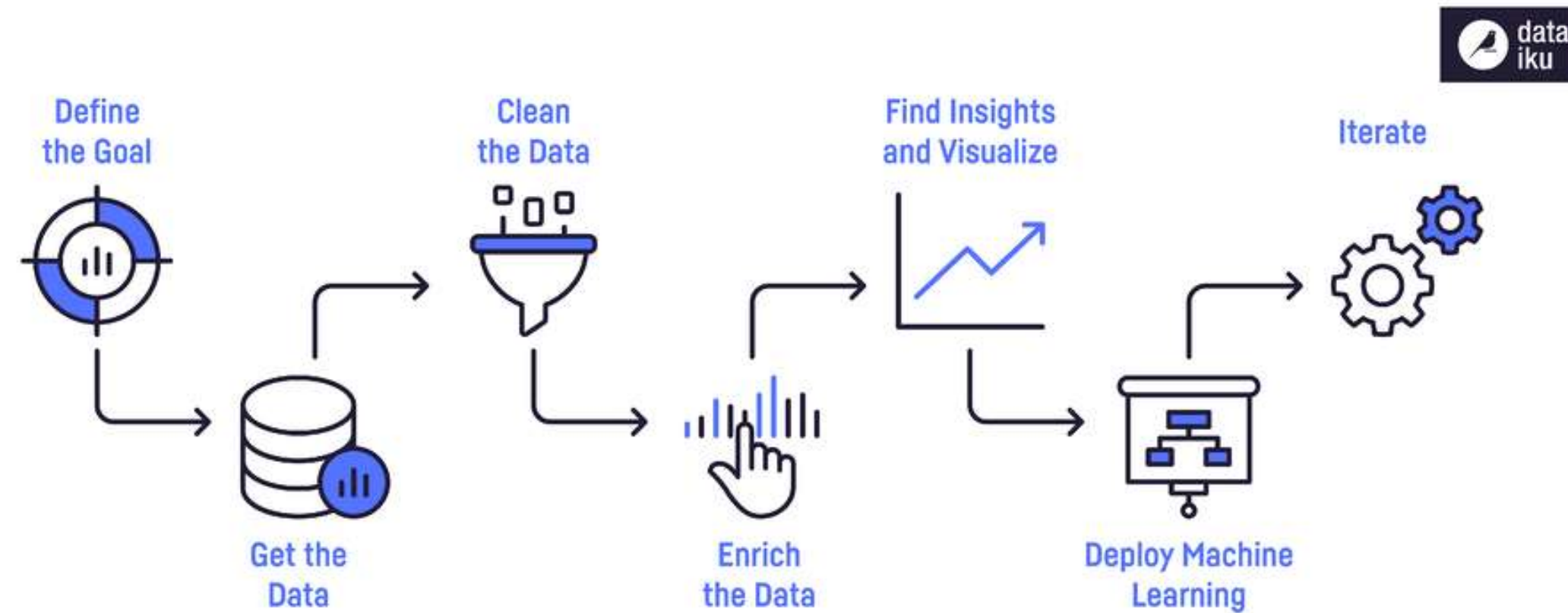
Data Science

- Data preparation can involve cleansing, aggregating, and manipulating it to be ready for specific types of processing.

- Analysis requires the development and use of algorithms, analytics and AI models.

- It's driven by software that combs through data to find patterns within to transform these patterns into predictions that support business decision-making.

- And the results should be shared through the skillful use of data visualization tools that make it possible for anyone to see the patterns and understand trends.
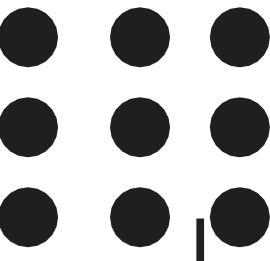
Data Science Lifecycle

- Capture
- Prepare and Maintain
- Preprocess or Process
- Analyze
- Communicate

Capture: This is the gathering of raw structured and unstructured data from all relevant sources via just about any method—from manual entry and web scraping to capturing data from systems and devices in real time.
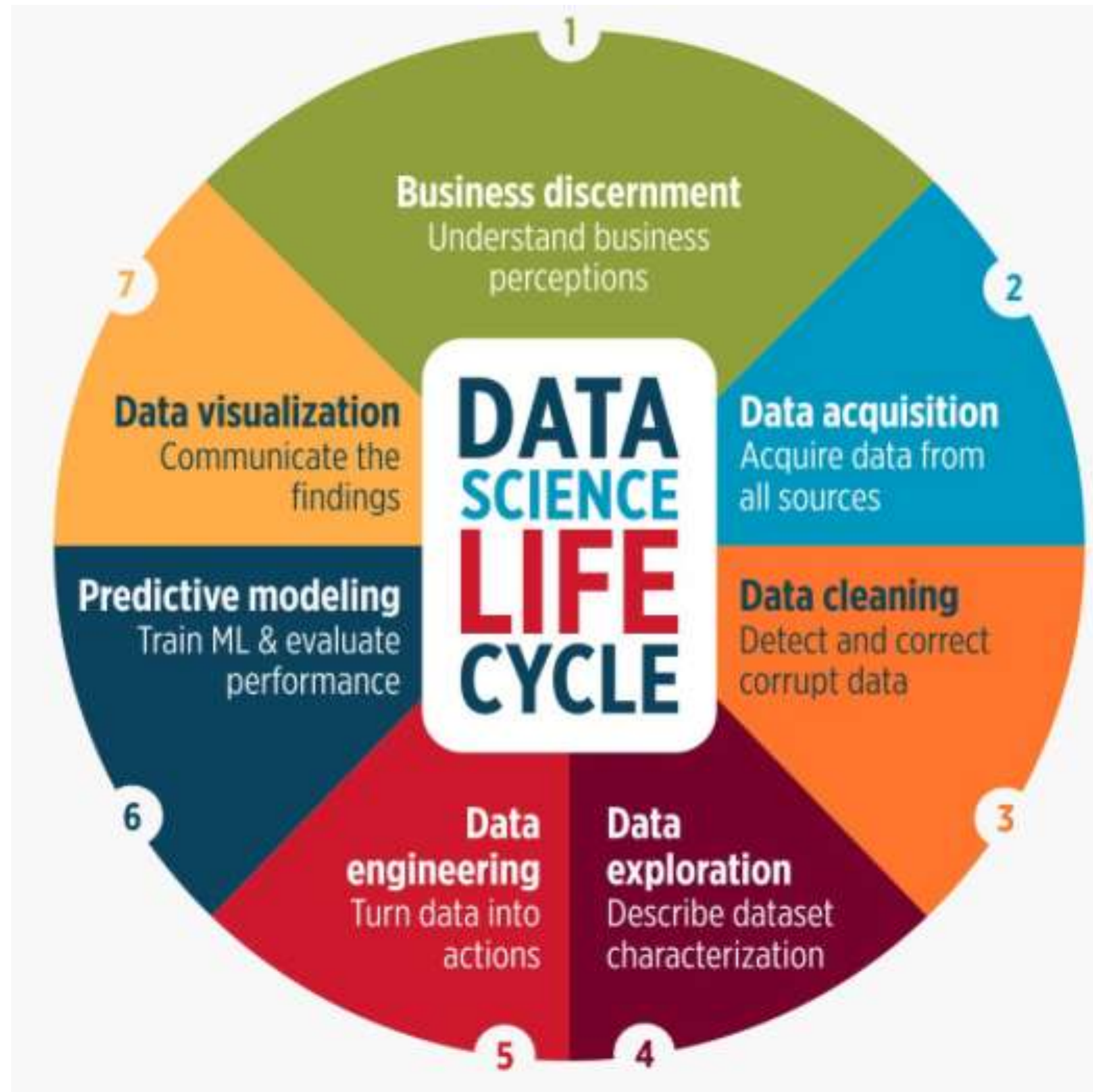
Prepare and maintain: This involves putting the raw data into a consistent format for analytics or machine learning or deep learning models. This can include everything from cleansing, deduplicating, and reformatting the data, to using ETL (extract, transform, load) or other data integration technologies to combine the data into a data warehouse, data lake, or other unified store for analysis.

Data Science Lifecycle

- Preprocess or process: Here, data scientists examine biases, patterns, ranges, and distributions of values within the data to determine the data's suitability for use with predictive analytics, machine learning, and/or deep learning algorithms (or other analytical methods).

- Analyze: This is where the discovery happens—where data scientists perform statistical analysis, predictive analytics, regression, machine learning and deep learning algorithms, and more to extract insights from the prepared data

- Communicate: Finally, the insights are presented as reports, charts, and other data visualizations that make the insights—and their impact on the business—easier for decision-makers to understand.

Data Science Lifecycle

THANK YOU