



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

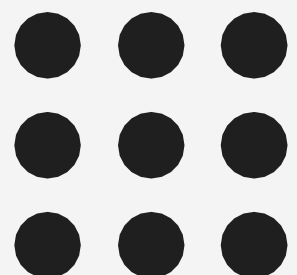
Department of Information Technology

19IT601 - Data Science and Analytics

III Year / VI Semester

Unit 1 – Introduction

Topic 1- Data Science Introduction





Data Science

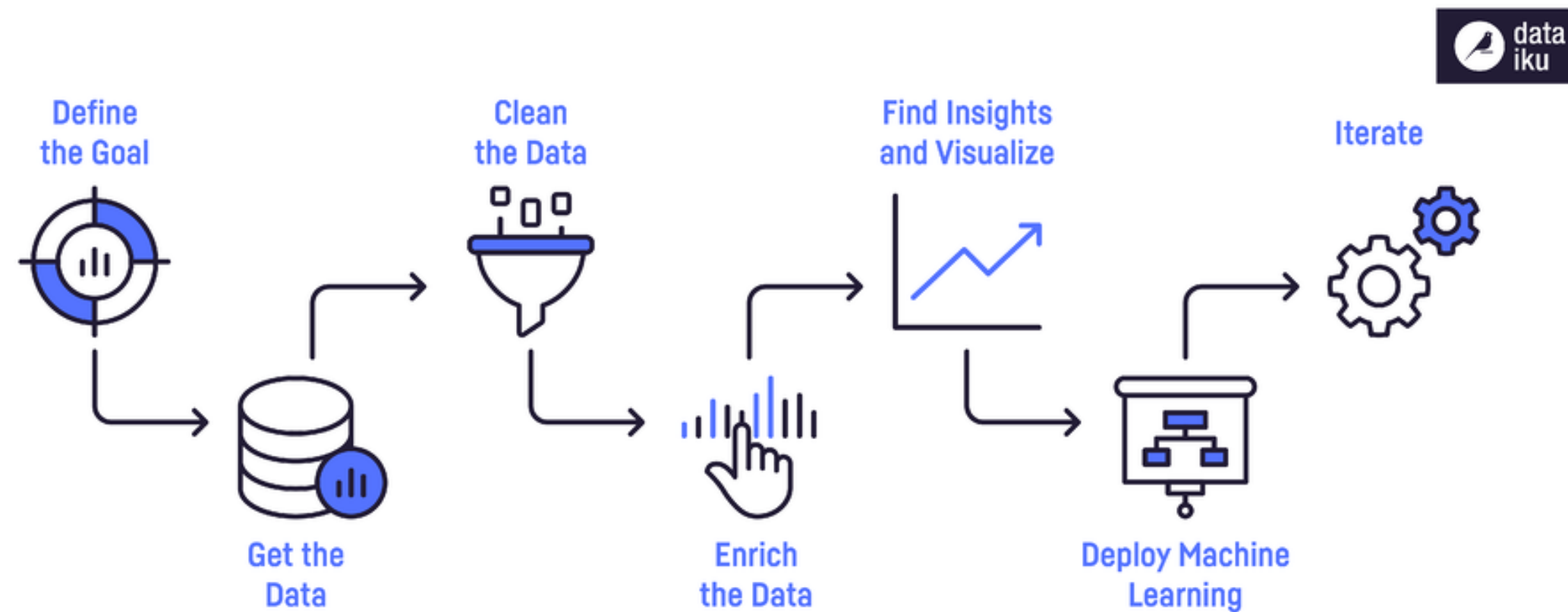


- Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data.
- Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful 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.

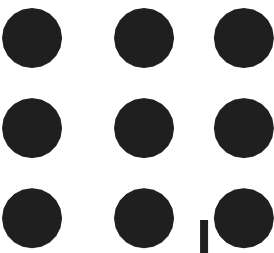


Data Science Lifecycle

- 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.

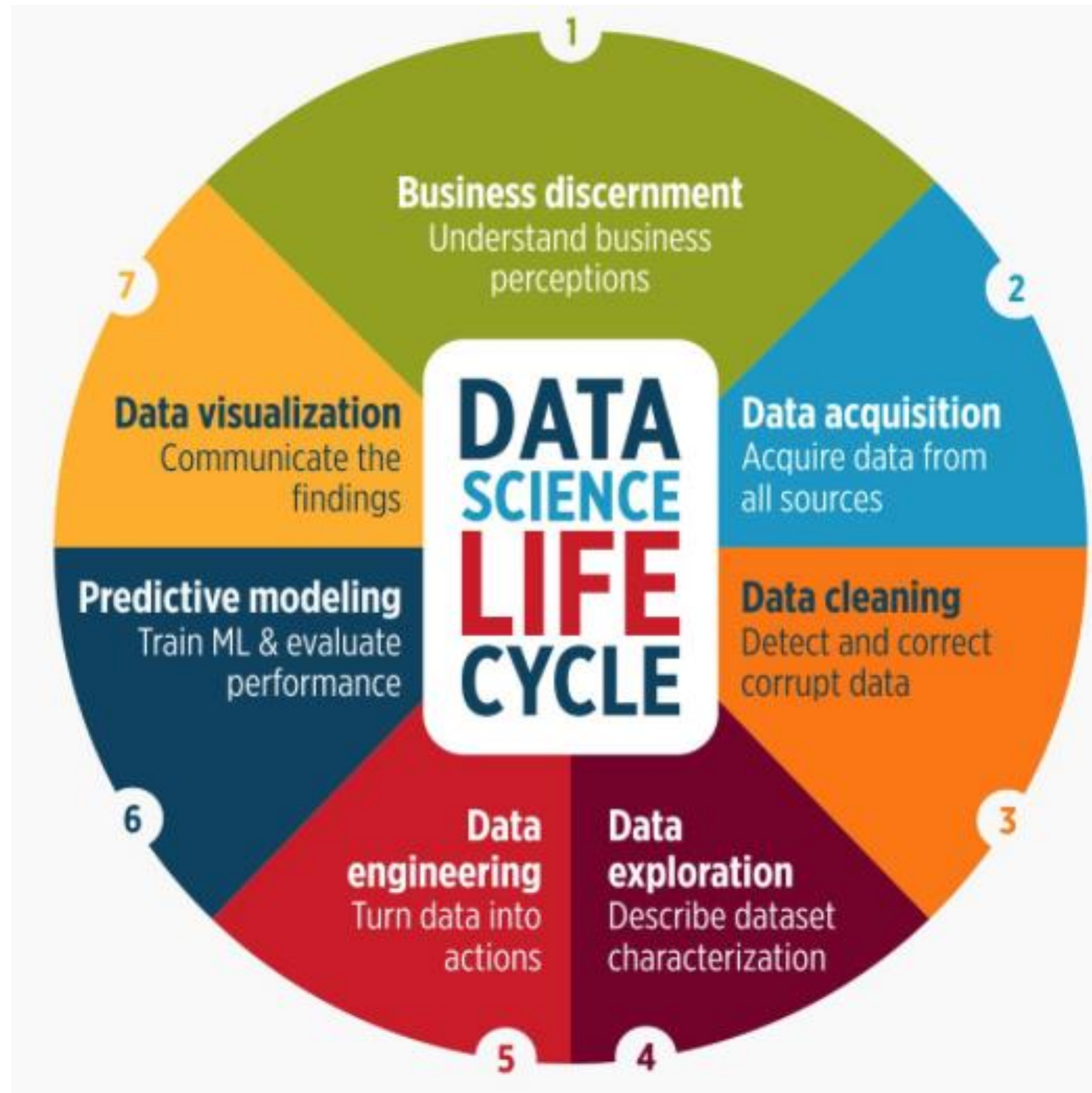
Prepare and maintain: This involves putting the raw data into a consistent format for analytics or machine learning or deep learning models. This can include everything from cleansing, deduplicating, and reformatting the data, to using ETL (extract, transform, load) or other data integration technologies 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





Data Scientist

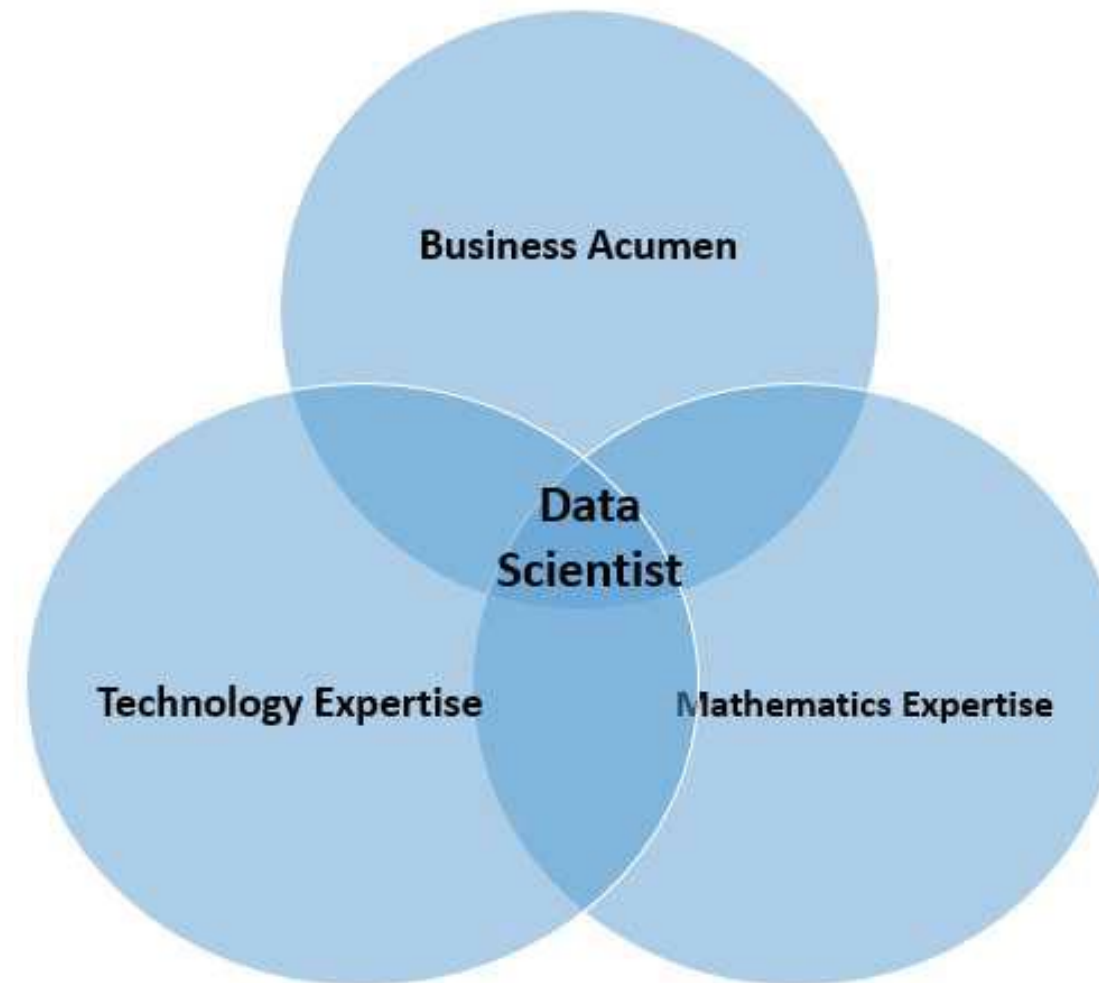
What Does a Data Scientist Do?

- A data scientist analyzes business data to extract meaningful insights. In other words, a data scientist solves business problems through a series of steps, including:
- Before tackling the data collection and analysis, the data scientist determines the problem by asking the right questions and gaining understanding.
- The data scientist then determines the correct set of variables and data sets.
- The data scientist gathers structured and unstructured data from many disparate sources—enterprise data, public data, etc.
- Once the data is collected, the data scientist processes the raw data and converts it into a format suitable for analysis. This involves cleaning and validating the data to guarantee uniformity, completeness, and accuracy.
- After the data has been rendered into a usable form, it's fed into the analytic system—ML algorithm or a statistical model.

Data Scientist

- This is where the data scientists analyze and identify patterns and trends. When the data has been completely rendered, the data scientist interprets the data to find opportunities and solutions.
- The data scientists finish the task by preparing the results and insights to share with the appropriate stakeholders and communicating the results.

Skills Needed By Data Scientist





Data Scientist



Business Acumen Skills

A data scientist should have business acumen skills to counter the pressure of business:

- Understanding of domain
- Business strategy
- Problem solving
- Communication
- Presentation
- Inquisitiveness

Technology Expertise Skills

A data scientist should be technology expert to convert the business into business logic:

- Good database knowledge such as RDBMS.
- Good NoSQL database knowledge such as MongoDB, Cassandra, HBase, etc.
- Programming languages such as Java, Python, etc.
- Open-source tools such as Hadoop, R.
- Datawarehousing, Datamining.
- Visualization such as Tableau, Flare, Google visualization APIs, etc.



Data Scientist



Mathematics Expertise Skills

A data scientist should be mathematics expert to formulize and analyze data:

- Mathematics.
- Statistics

Areas of Technical Proficiency

- Artificial Intelligence (AI).
- Machine learning.
- Pattern recognition.
- Natural Language Processing



Data Scientist



Data Science Tools

The data science profession is challenging, but fortunately, there are plenty of tools available to help the data scientist succeed at their job.

Data Analysis: SAS, Jupyter, R Studio, MATLAB, Excel, RapidMiner

Data Warehousing: Informatica/ Talend, AWS Redshift

Data Visualization: Jupyter, Tableau, Cognos, RAW, Matlibpro, Power BI, Zoho Analytics

Machine Learning: Spark MLib, Mahout, Azure ML studio, Scikilearn, Pytorch, Tensorflow



THANK YOU