SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
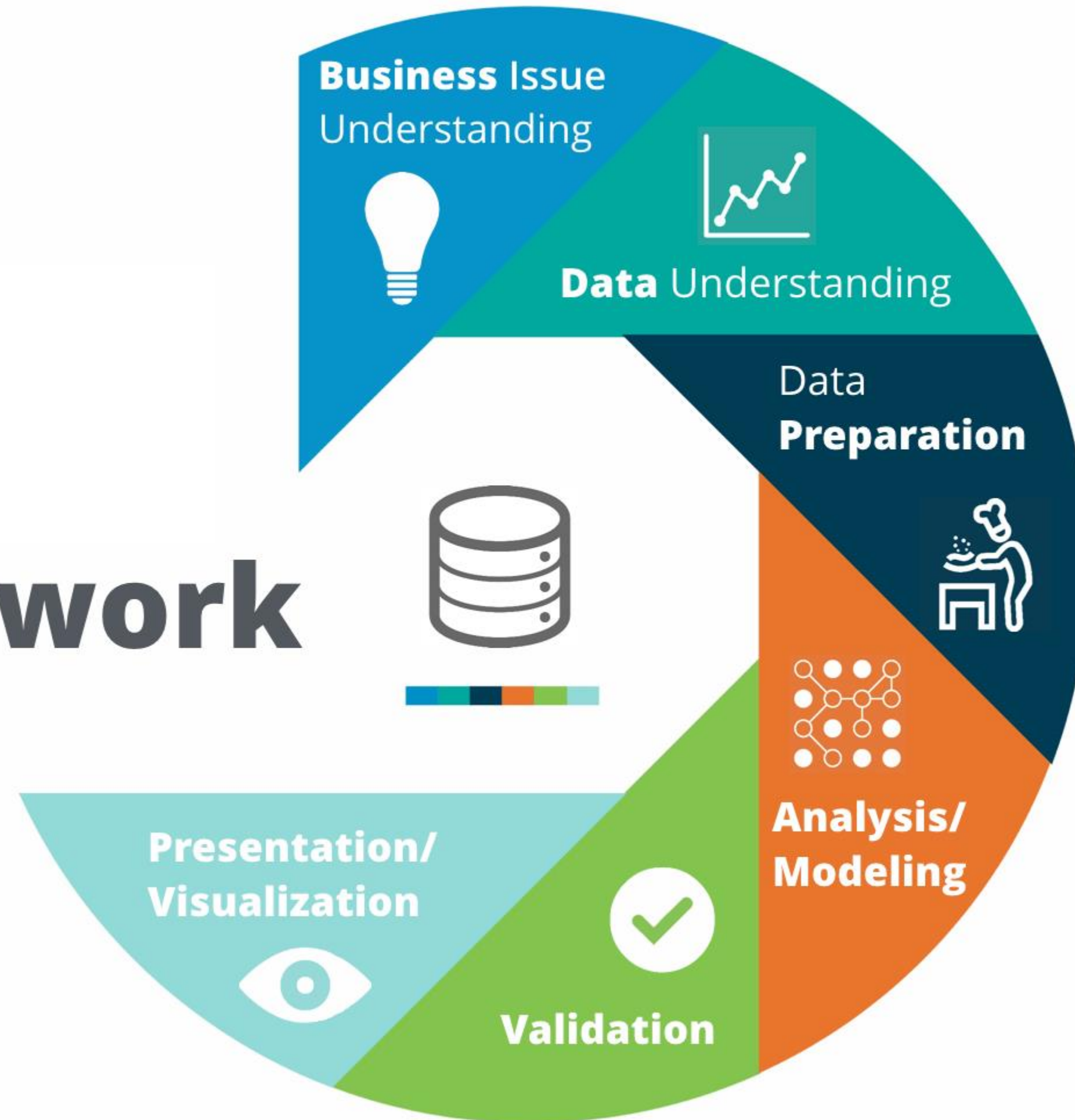
DEPARTMENT OF COMPUTER SCIENCE AND   ENGINEERING

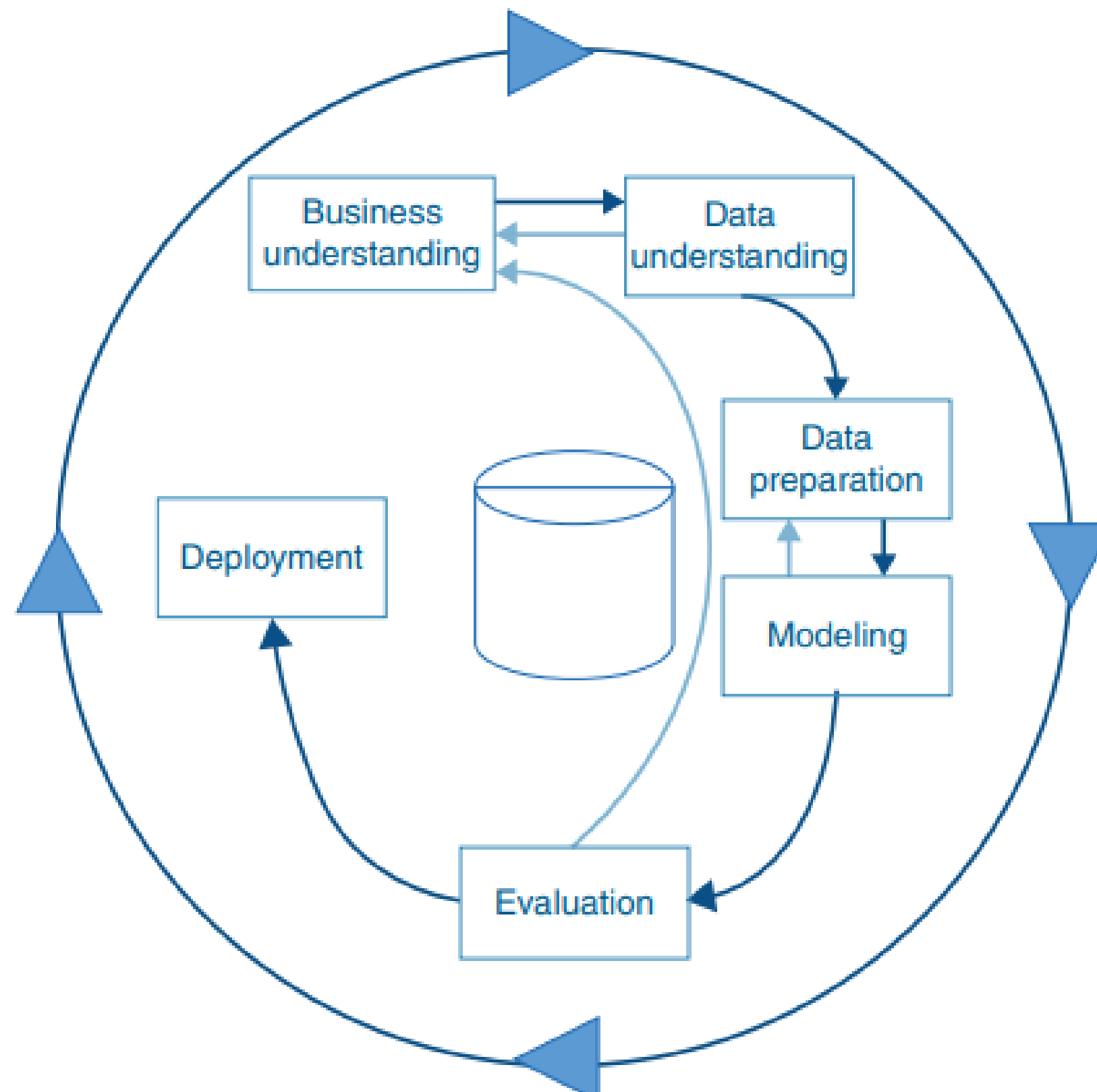**COURSE NAME :19CS407 DATA ANALYTICS WITH R**
II YEAR /IV SEMESTER

Unit 1- Introduction

Topic : CRISP-DM Methodology

DATA ANALYTICS/JEBAKUMAR IMMANUEL/CSE/SNSCE

The CRISP-DM Methodology

✓ CRoss-Industry Standard Process for Data Mining (CRISP-DM) is a six-step method, which, like the KDD process, uses a non-rigid sequential framework.

✓ Despite the six phases, CRISP-DM is seen as a perpetual process, used throughout the life of a company in successive iterations

Business understanding

✓ This involves understanding the business domain, being able to define the problem from the business domain perspective, and finally being able to translate such business problems into a data analytics problem

Data Understanding

✓ This involves collection of the necessary data and their initial visualization/summarization in order to obtain the first insights, particularly but not exclusively, about data quality problems such as missing data or outliers.

Data preparation

✓ This involves preparing the data set for the modeling tool, and includes data transformation, feature construction, outlier removal, missing data fulfillment and incomplete instances removal.

DATA ANALYTICS/JEBAKUMAR IMMANUEL/CSE/SNSCE

Modeling

- ✓ Typically there are several methods that can be used to solve the same problem in analytics, often with specific data requirements.

- ✓ This implies that there may be a need for additional data preparation tasks that are method specific. In such case it is necessary to go back to the previous step.

- ✓ The modeling phase also includes tuning the hyper-parameters for each of the chosen method(s)

Evaluation

✓ Solving the problem from the data analytics point of view is not the end of the process.

✓ It is now necessary to understand how its use is meaningful from the business perspective; in other words, that the obtained solution answers to the business requirements

DATA ANALYTICS/JEBAKUMAR IMMANUEL/CSE/SNSCE

Deployment

✓ The integration of the data analytics solution in the business process is the main purpose of this phase.

✓ Typically, it implies the integration of the obtained solution into a decision-support tool, website maintenance process, reporting process or elsewhere.

Assessment 1

To create your own CRISP DM Approach

References

1. João Moreira, Andre Carvalho, Tomás Horvath – "A General Introduction to Data Analytics" – Wiley -2018

Thank You