

SNS COLLEGE OF TECHNOLOGY

**An Autonomous Institution
Coimbatore-35**



DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

II YEAR IV SEM

UNIT I – INTRODUCTION TO DATA SCIENCE

SETTING THE RESEARCH GOAL

Defining Research Goals

EMPATHY:

To understand the project, three concept must understand: what, why and how.

a) What is expectation of company or organization?

b) Why does a company's higher authority define such research value?

c) How is it part of a bigger strategic picture?

Defining Research Goals in Data Science



Identify Business Problem



Set Objectives



Define Key Metrics



Understand Data Requirements



Align with Business Goals & Outcomes

1. Learning the business domain :

DEFINE:

- Understanding the domain area of the problem is essential. In many cases, data scientists will have deep computational and quantitative knowledge that can be broadly applied across many disciplines.
- Data scientists have deep knowledge of the methods, techniques and ways for applying heuristics to a variety of business and conceptual problems.

2. Resources :

- As part of the discovery phase, the team needs to assess the resources available to support the project. In this context, resources include technology, tools, systems, data and people.

3. Frame the problem :

IDEATE:

- Framing is the process of stating the analytics problem to be solved. At this point, it is a best practice to write down the problem statement and share it with the key stakeholders.
- Each team member may hear slightly different things related to the needs and the problem and have somewhat different ideas of possible solutions.

4. Identifying key stakeholders:

- The team can identify the success criteria, key risks and stakeholders, which should include anyone who will benefit from the project or will be significantly impacted by the project.
- When interviewing stakeholders, learn about the domain area and any relevant history from similar analytics projects.

TESTING:

5. Interviewing the analytics sponsor:

- The team should plan to collaborate with the stakeholders to clarify and frame the analytics problem.
- At the outset, project sponsors may have a predetermined solution that may not necessarily realize the desired outcome.
- In these cases, the team must use its knowledge and expertise to identify the true underlying problem and the appropriate solution.
- When interviewing the main stakeholders, the team needs to take time to thoroughly interview the project sponsor, who tends to be the one funding the project or providing the high-level requirements.
- This person understands the problem and usually has an idea of a potential working solution.

PROTOTYPING:

6. Developing initial hypotheses:

- This step involves forming ideas that the team can test with data. Generally, it is best to come up with a few primary hypotheses to test and then be creative about developing several more.
- These Initial Hypotheses form the basis of the analytical tests the team will use in later phases and serve as the foundation for the findings in phase.

7. Identifying potential data sources:

- Consider the volume, type and time span of the data needed to test the hypotheses. Ensure that the team can access more than simply aggregated data.
- In most cases, the team will need the raw data to avoid introducing bias for the downstream analysis.