Introduction Data Science

Ronny Fahrudin

TARGET

- Understanding What is Data Science
- Understanding Need for Data Science
- Understanding Data Science Field
- Understanding Prerequisites Data Science
- Different Bussines Inteligent and Data Science
- Understanding CRISP-DM for Data Science Framework

What is Data Science?

- Data: Whatever we sense by our five senses is data.
- Science: knowledge about everything which was discovered to solve the problem.

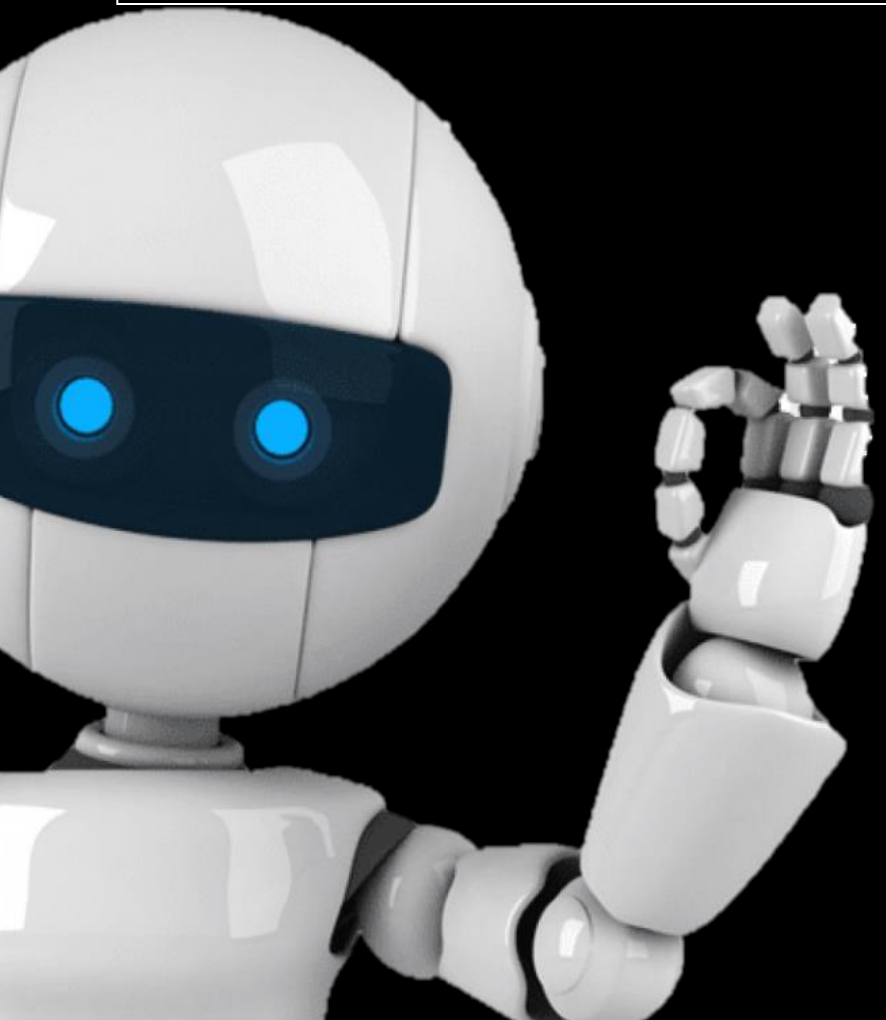
Data science is the process of understanding problem and make solution with data.

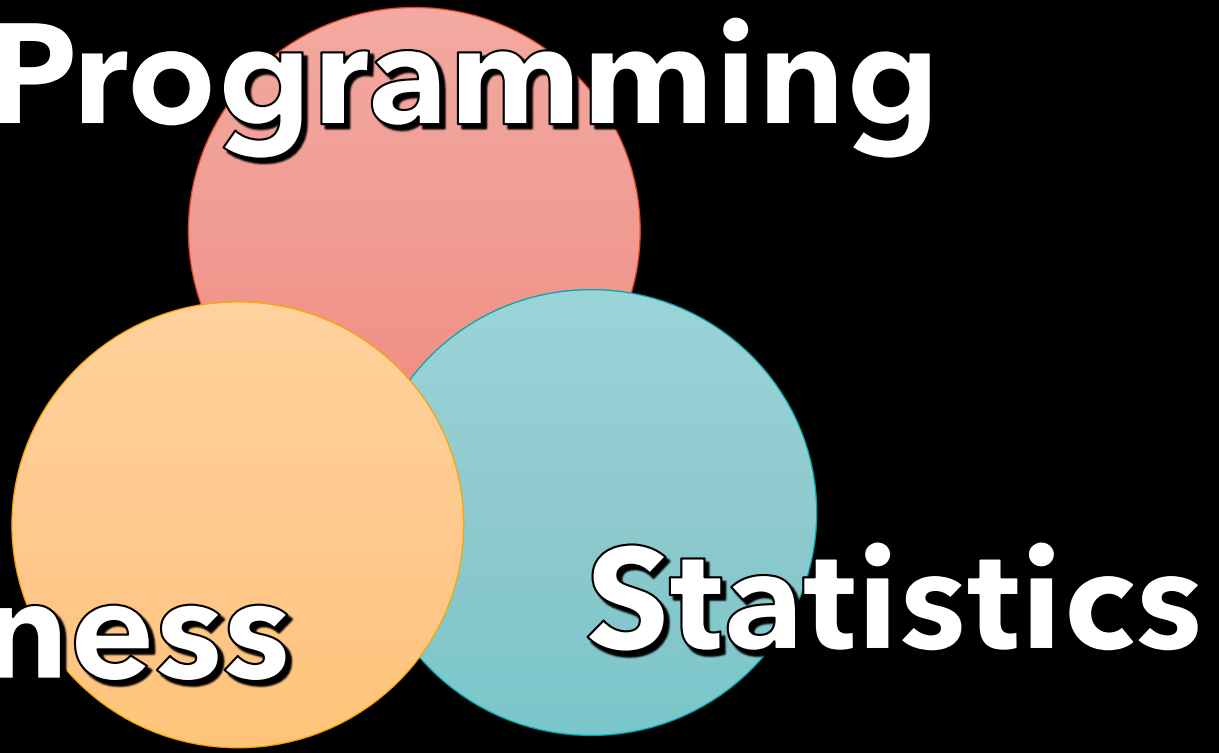
Why does data science exist?

Business Intelligence vs Data Science

Criterion	BI	DS
Data Source	Structured data. Like data warehouse	Unstructured data, like web logs
Method	Analytical	Scientific
Skill	Statistics, Visualization	Statistics, Visualization, Machine Learning
Focus	Past and Present Data	Present data & Future Prediction

The Essential of Data Science





Data Science Field

Educations

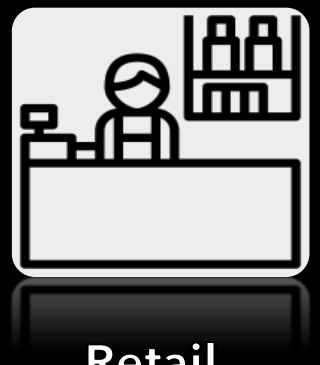
1. **Social-emotional skills**
2. **Monitoring student requirement**
3. **Innovating curriculum**
4. **Measuring instructor performance**

Transportations

1. **Automation driving car**
2. **Automatic traffic setting**

Governments

Making smart city



Retail

1. **Customer segmentation**
2. **Recommender system product**

Banking

1. Fraud detection credit
2. Risk analysis
3. Operating optimization

Prerequisites for Data Science

Following 3 essentials

Curiosity

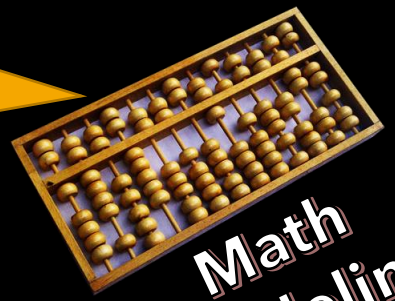
Communication Skill

Prerequisites for Data Science

Visualization & Analysis

Statistics

Hight calculating



Math modeling

Programming

Technical operating

predict

Databases

Technical operating in DB

Skills and Tools

Skill: R, Python, Statistics

Tools: SAS, Jupyter, R Studio, MATLAB, Excel, Rapidminer

Data Analysis

Skills : ETL, SQL, Hadoop, Apache Spark

Tools:Informatica/ Talend, AWS Redshift

Data Warehouse

Skills: Algebra, ML Algoritms, Statistics

Tools: Spark Mlib, Mahout, Azure ML Studio

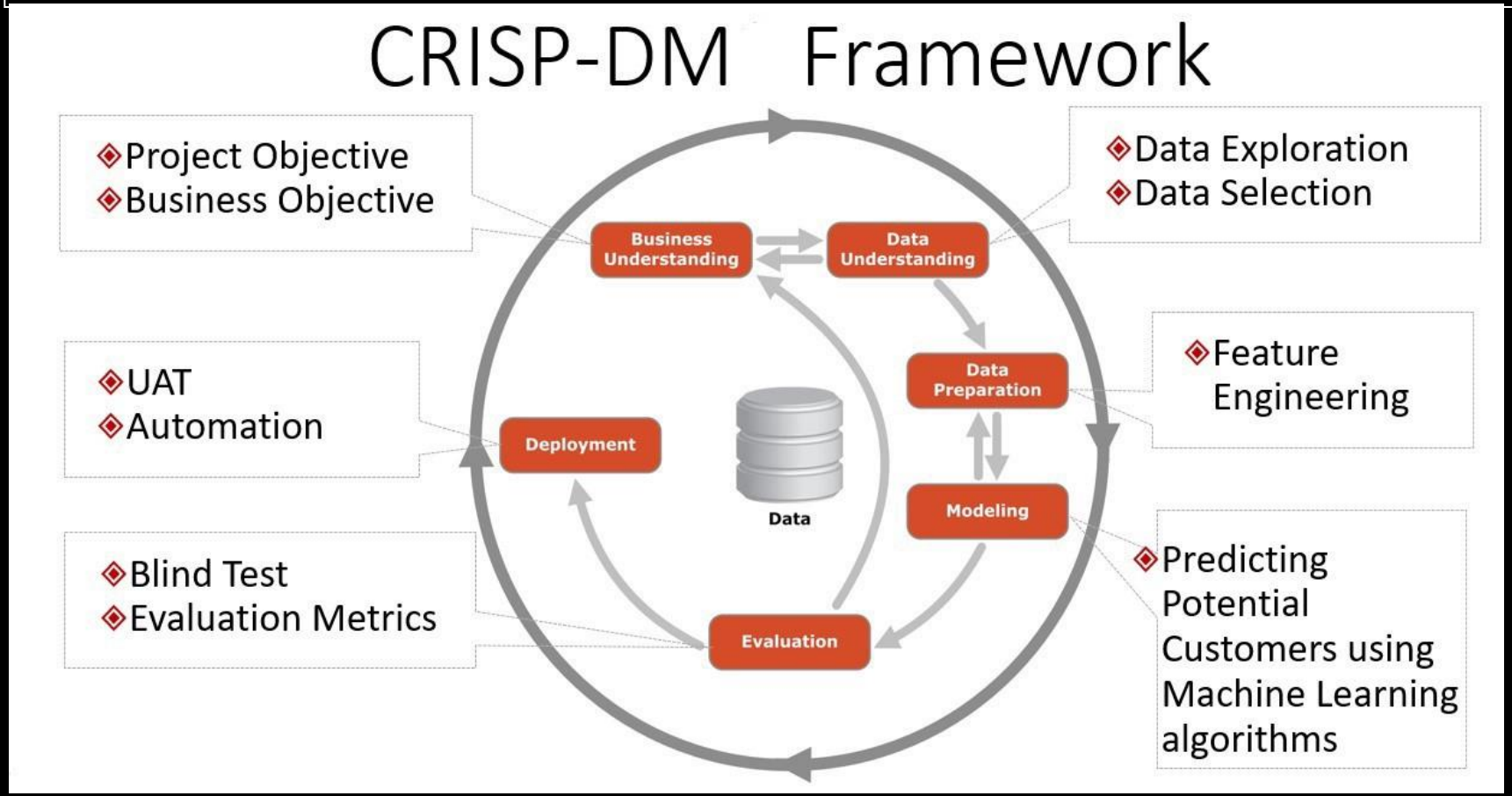
Machine Learning

Skills: R, Python libraries

Tools: Jupyter, Tableau, Cognos, RAW

Data Visualization

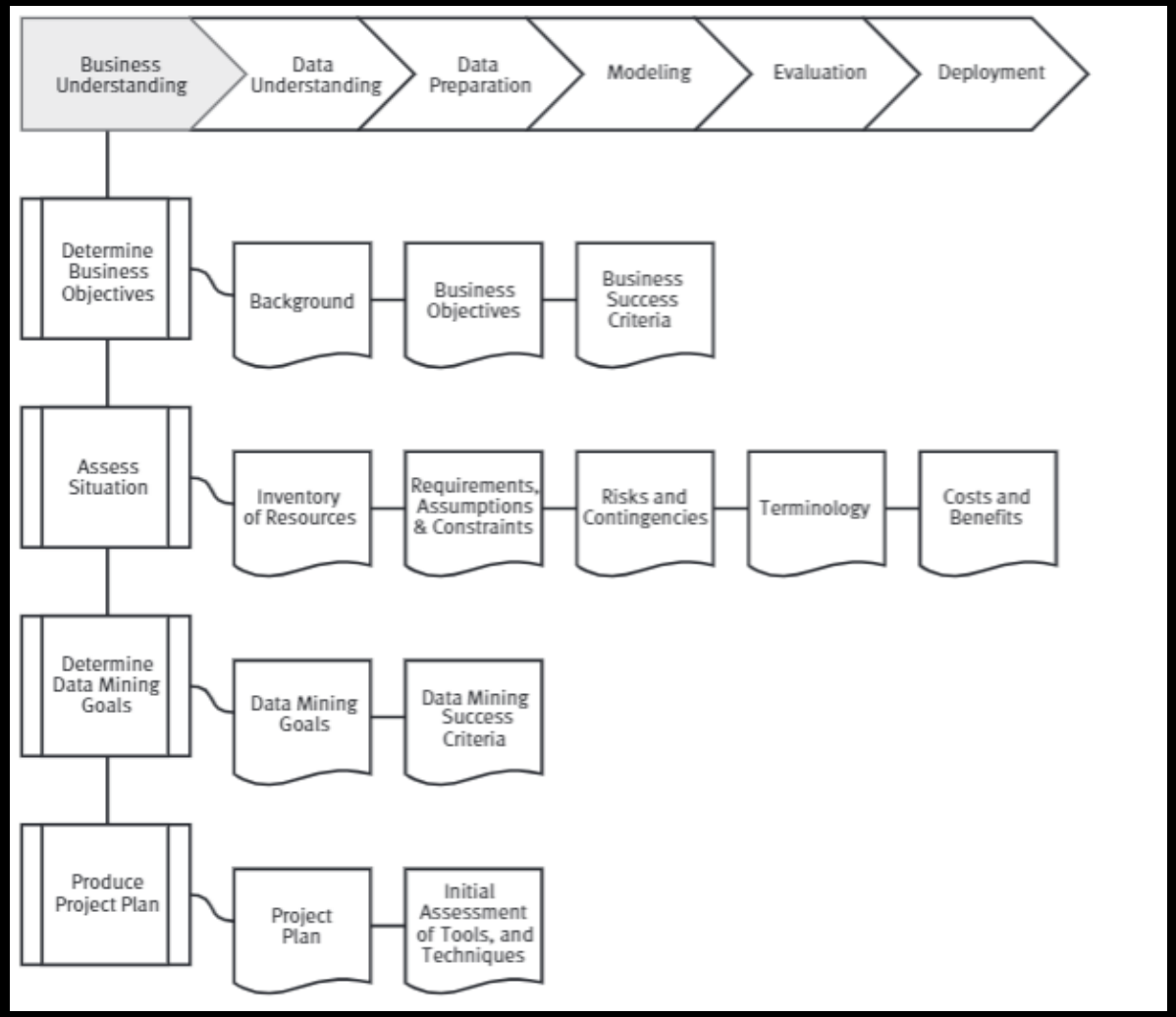
Workflow Data Science With



Cross-Industry Standard Process for Data mining

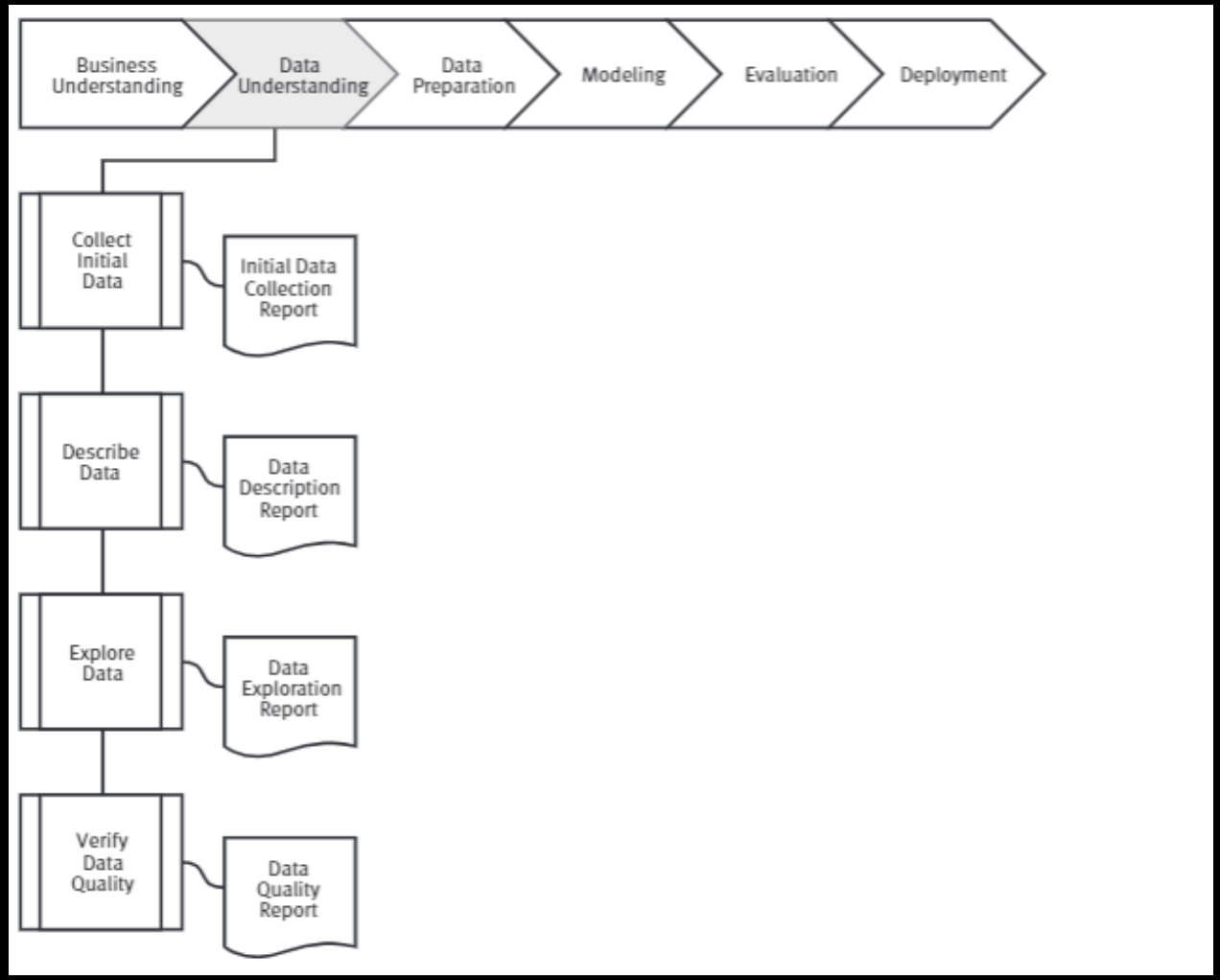
Business Understanding

1. Looking for problem base on business perspective.
2. Define decision base on problem that you find
3. Looking for information that needed to inform those decision
4. Determine type analysis can provide the information needed to inform those decisions



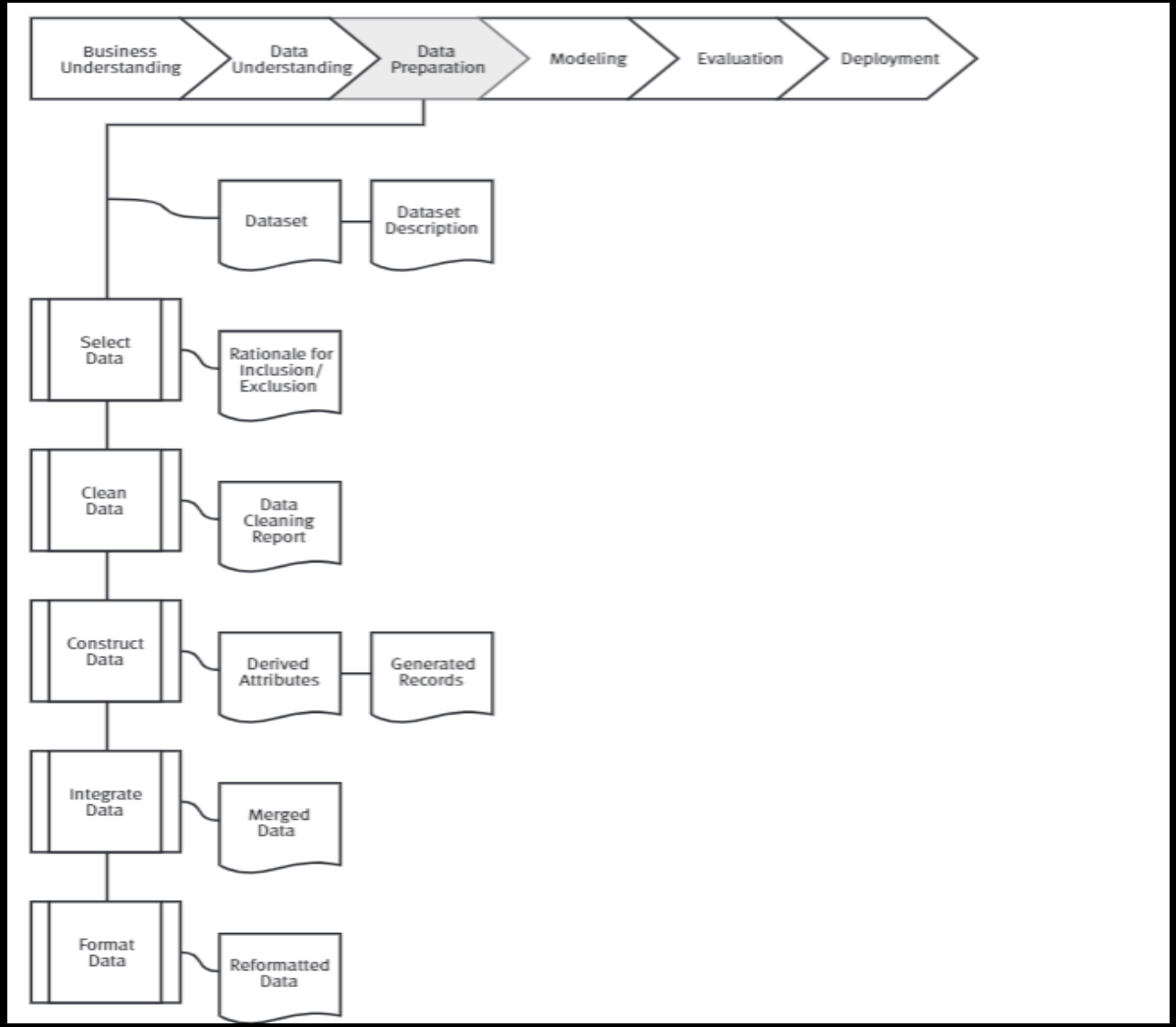
Data Understanding

1. Initial data collection and proceeds with activities in order to get familiar with data,
2. To identify data quality problems and that's characteristics
3. Discovered first insights into the data or to detect interesting subset to form hypotheses for hidden information



Data Preparation

1. In this step we build dataset to make modeling from raw data.
2. We can iterate step until data is clean and great to make modeling.



Modelling

1. Determine what methodology to use to solve the problem
2. Determine the important factors or variables that will help solve the problem
3. Build a model to solve the problem
4. Run the model and move to the evaluation phase

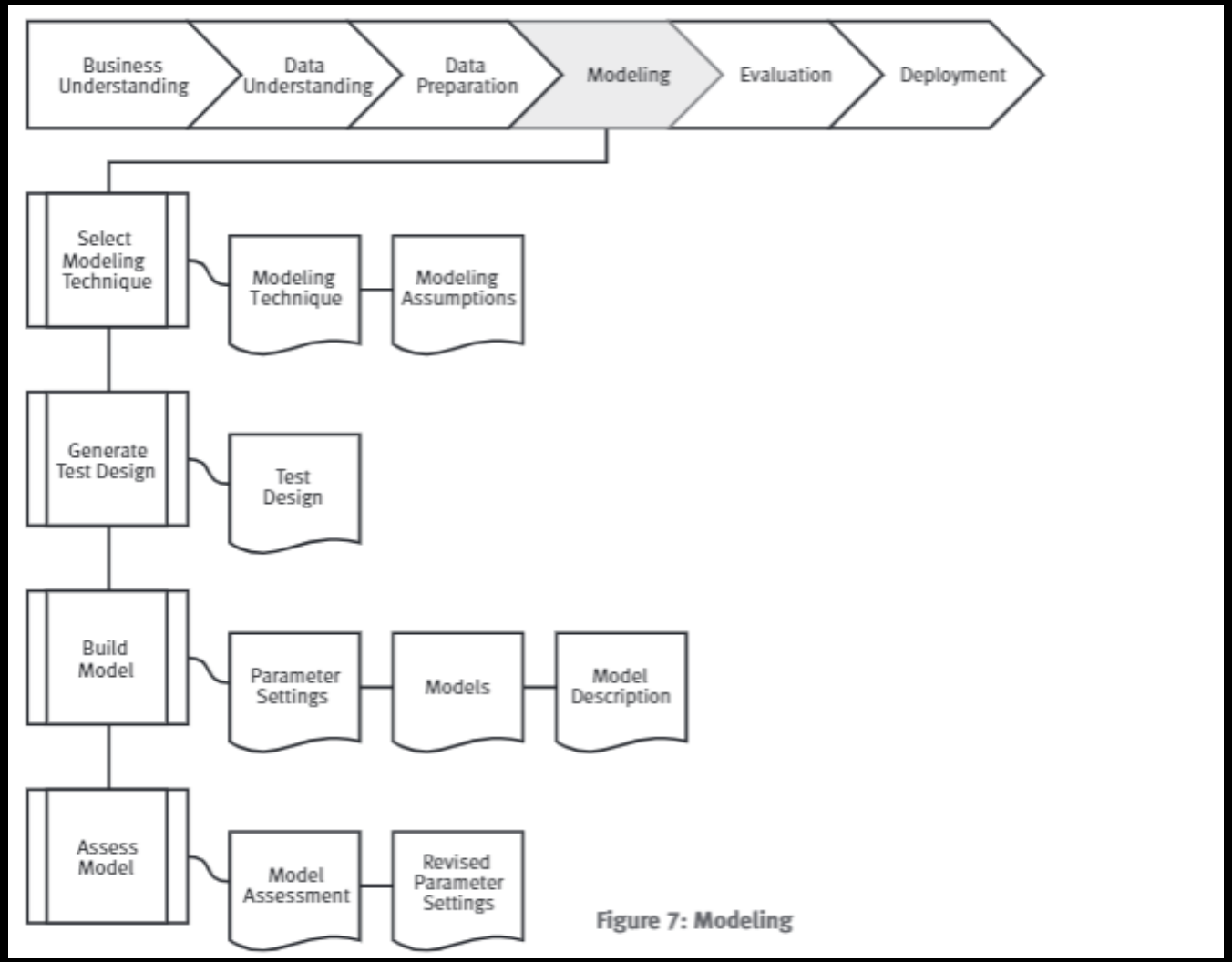
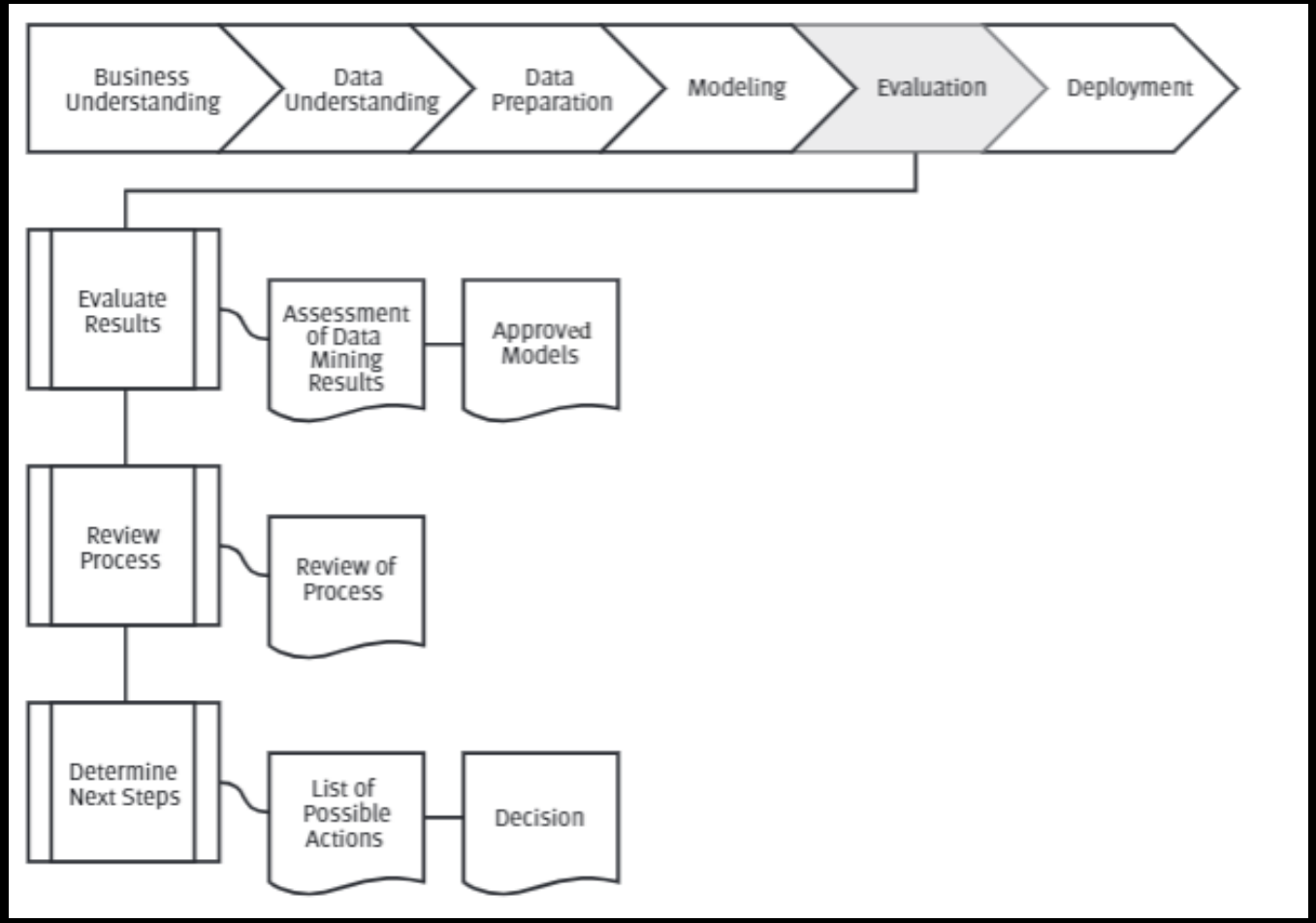


Figure 7: Modeling

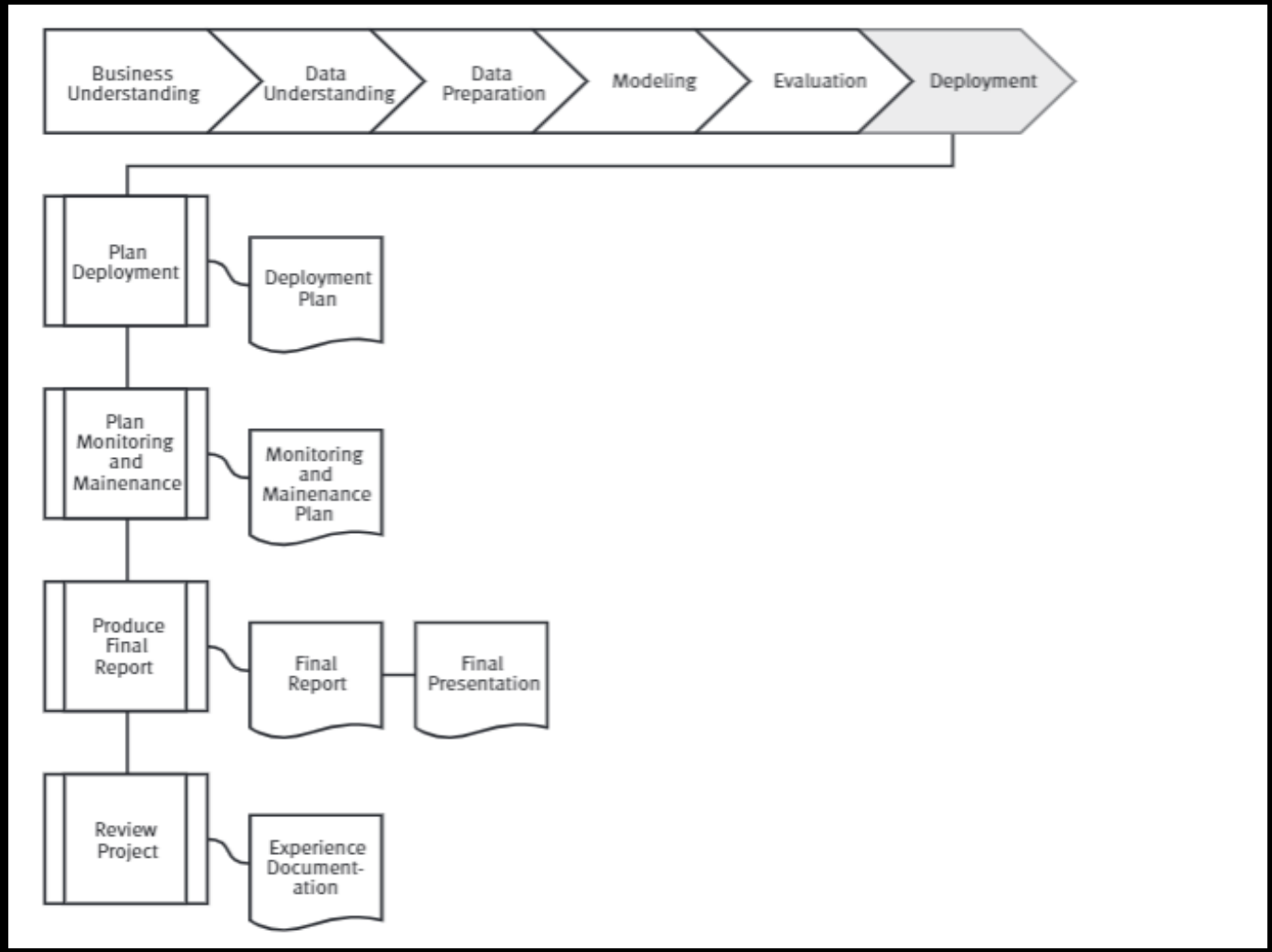
Evaluation

1. To checking the quality model objectively and how effective model to solve the problem
2. Observe the key results on the model
3. Ensure the results make sense within the content of the business problem
4. Determine whether to proceed to the next step or return to a previous phase
5. Repeat as many times necessary



Deployment/Presentation

Knowledge gained will need to be **organized** and **presented** in a way that the **customer can use** it. However, depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise.



Role in Data World

Job	Role
Data Scientist	1. Defining business problem, 2. Understanding the data, 3. Data exploration, modeling until evaluation
Data Analyst	Focus on Analyzing the data to be able convert the data into information that valuable for business
Data Engineer	Prepare the insfrastructure needed in a DS project suc as a dabases, server, create data processing pipeline and deploy a model
Business Intelligence	Defining business problem/goal, transform data into insight that drive business value, and deliver to user.
Machine Learning Engineer	Focus on implementing the model made by Data Scientist. Feeding the data into model, and optimize it. Scale the model into production-ready model.

Thank You