zenius × Kampus Merdeka INDONESIA JAYA

Introduction to Data Science

Put the Date Here

Accelerated Machine Learning Program

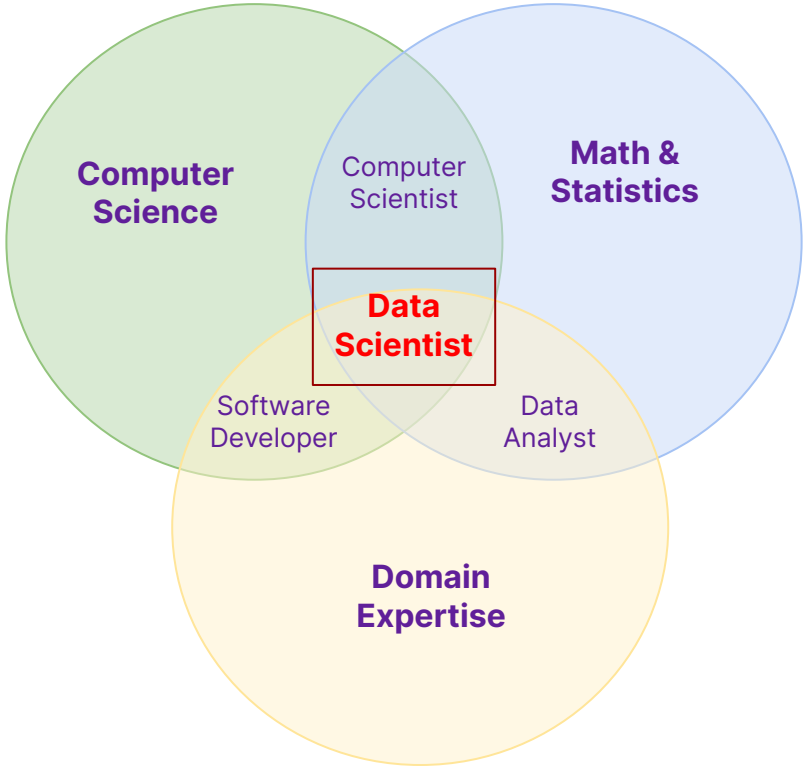
Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka

1. What is Data Science?
2. Data Science Use Cases in Real Life
3. Data Science Methodology & Life Cycle
4. Tools & Tech-Stacks for Data Scientist

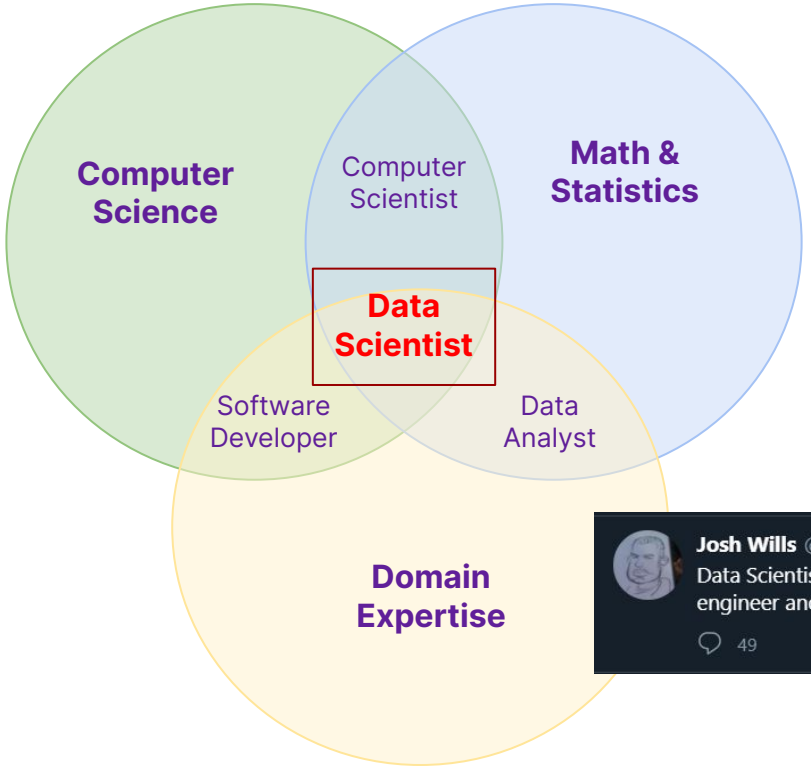
What is Data Science?

Introducing Data Science and Machine Learning

What is Data Science?

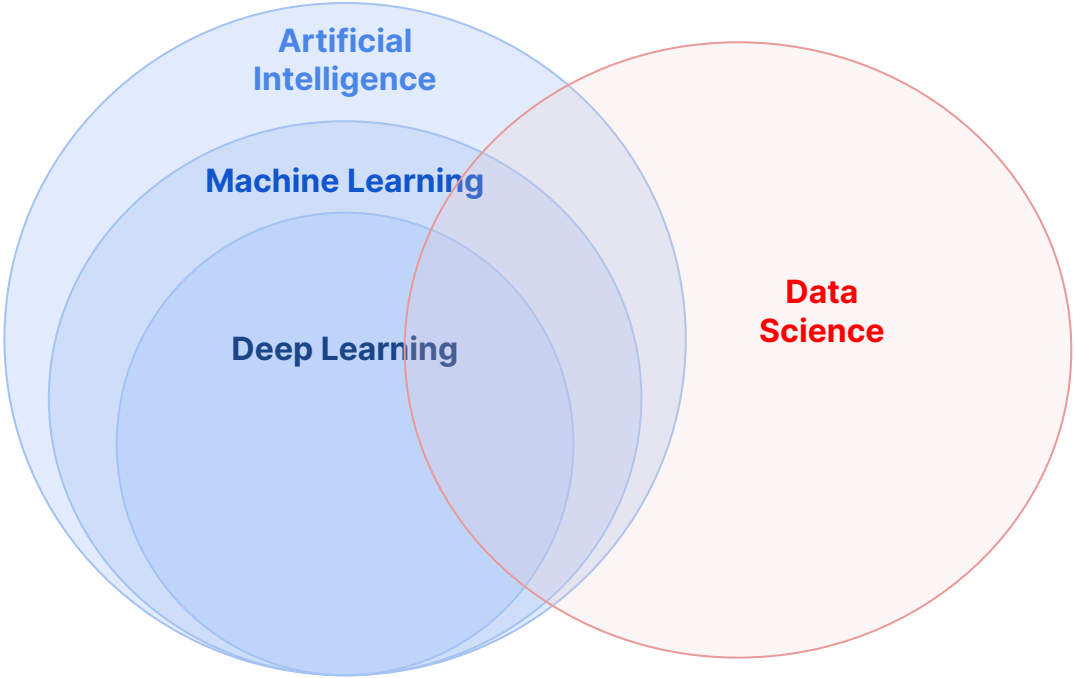


What is Data Science?

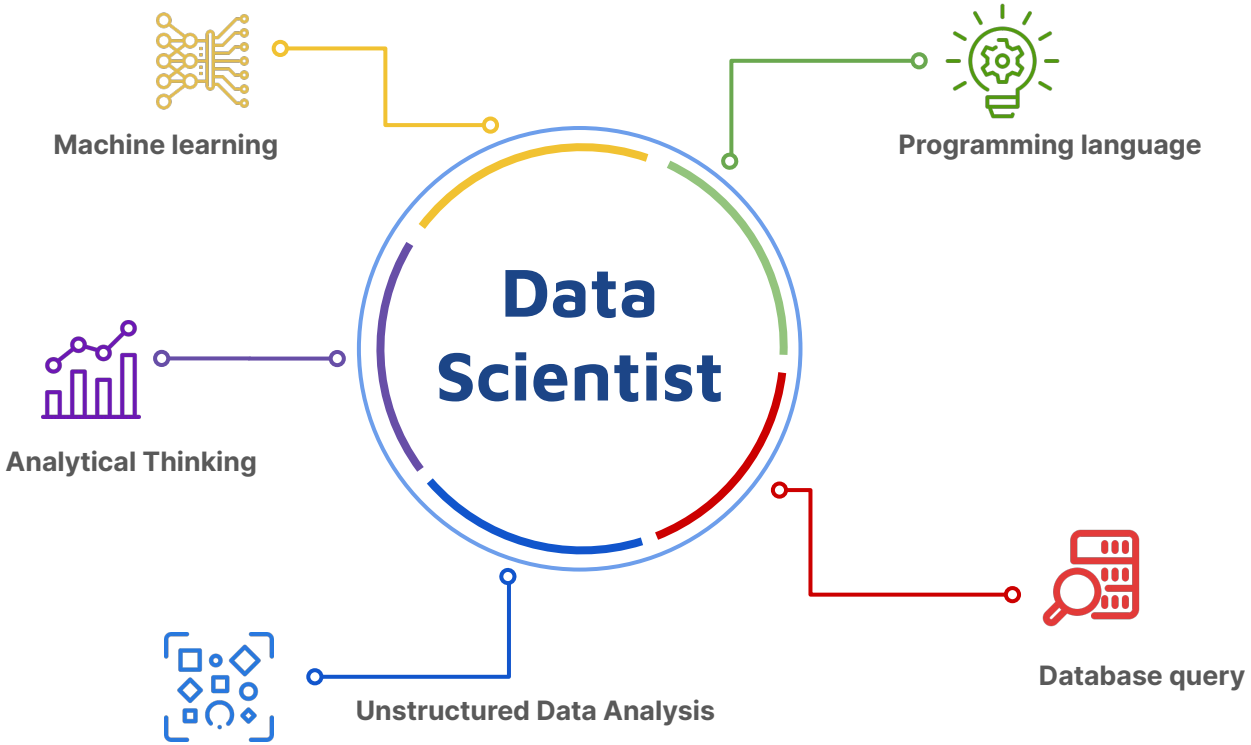


Data Science is an art of data: from extracting, cleaning, analyzing, and turning data into insights, predictions, and decisions

Data Science is not only about the Jargons

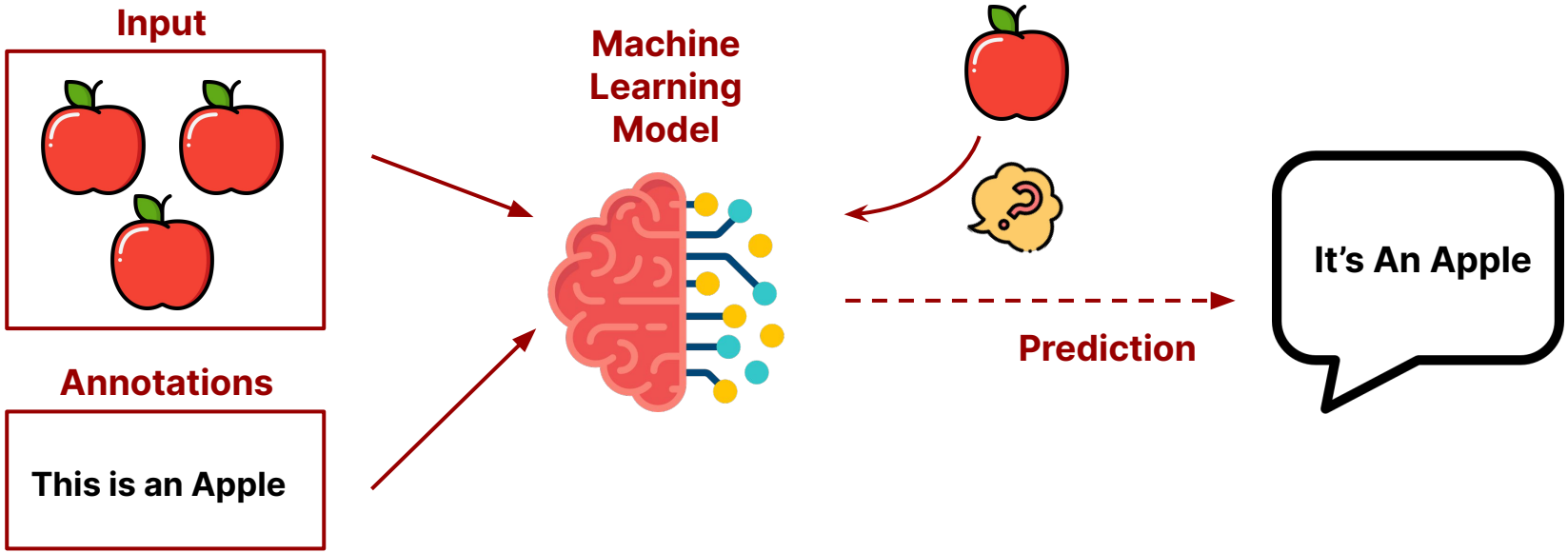


Skills make a Data Scientist

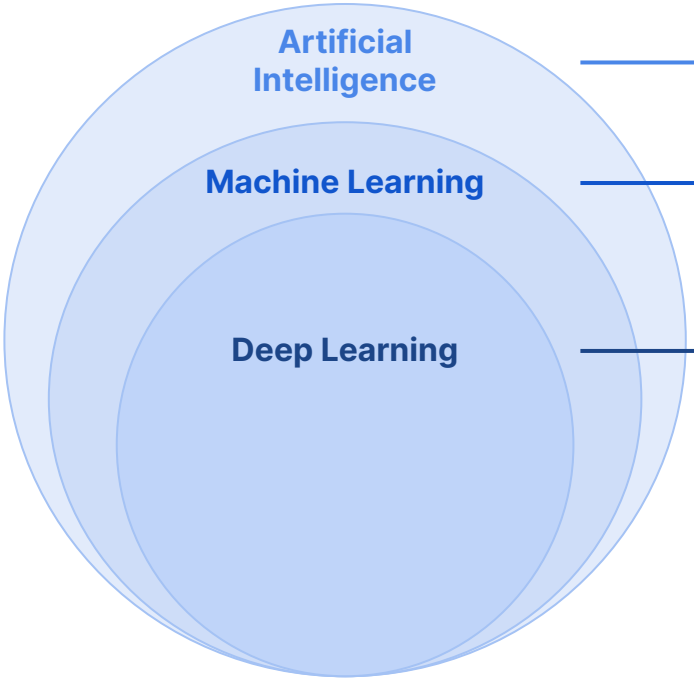


What is Machine Learning?

Simply, machine learning finds patterns in data and uses them to make predictions.



AI vs Machine Learning vs Deep Learning



Programs with the ability to learn like humans

Algorithms with the ability to learn without being explicitly programmed

Subset of machine learning in which artificial neural networks adapt & learn from data

Pop Quiz!

What are rule-based systems (like “chess playing program”) included to ??

A. AI with Machine Learning

B. AI without Machine Learning

Pop Quiz!

Predicting house prices with linear regression is included as deep learning.

A. True

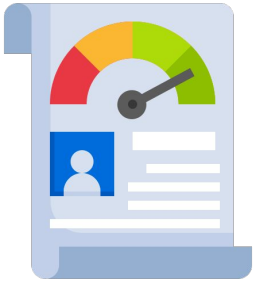
B. False

Data Science Use Cases in Real Life

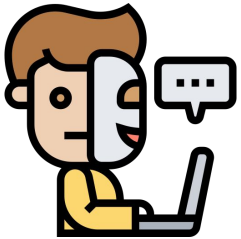
Real applications and use cases

1. Financial & Risk Management

Credit Scoring



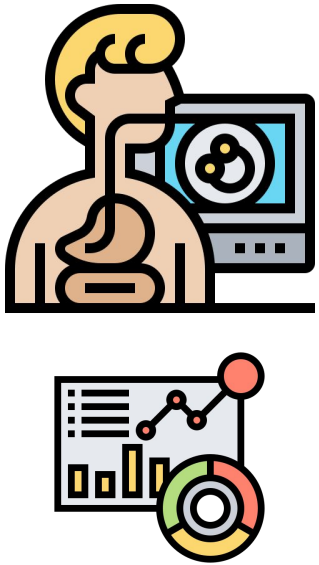
Fraud Detection



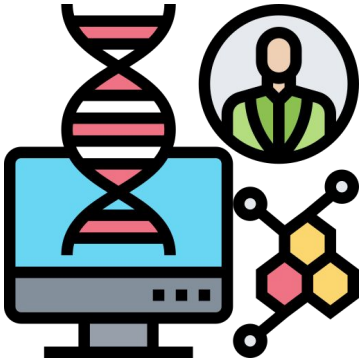
Stock Market Price Prediction

2. Healthcare

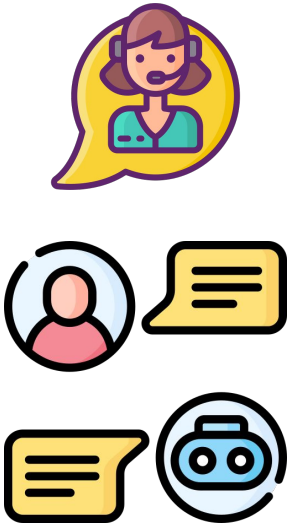
Medical Image Analysis



Genetics & Genomics

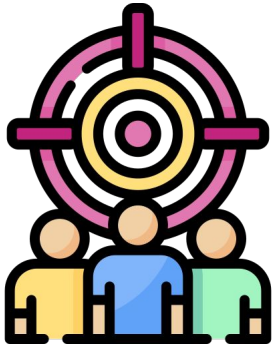


Virtual Assistance

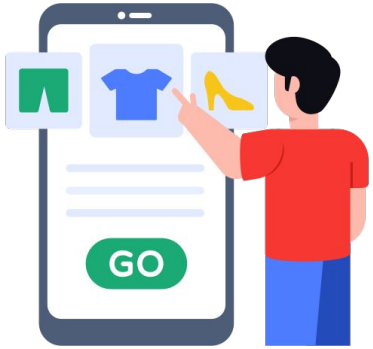


3. Marketing

Targeted Ads/Campaigns



Product Recommendation

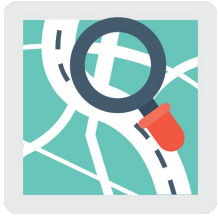


Customer Segmentation

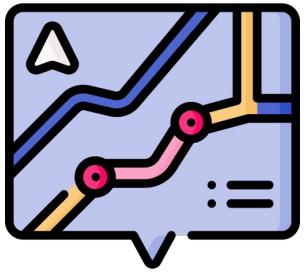
4. Transport

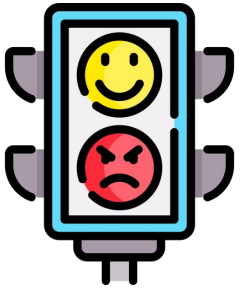
Self Driving Car

Routes Optimization



Traffic Management

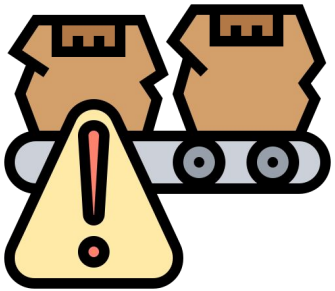




5. Manufacturing

Monitoring Systems

Anomaly Detection



Scheduling

Pop Quiz!

Netflix utilizes data science.

A. True

B. False

Pop Quiz!

Recommendation engines provide random recommendations.

A. True

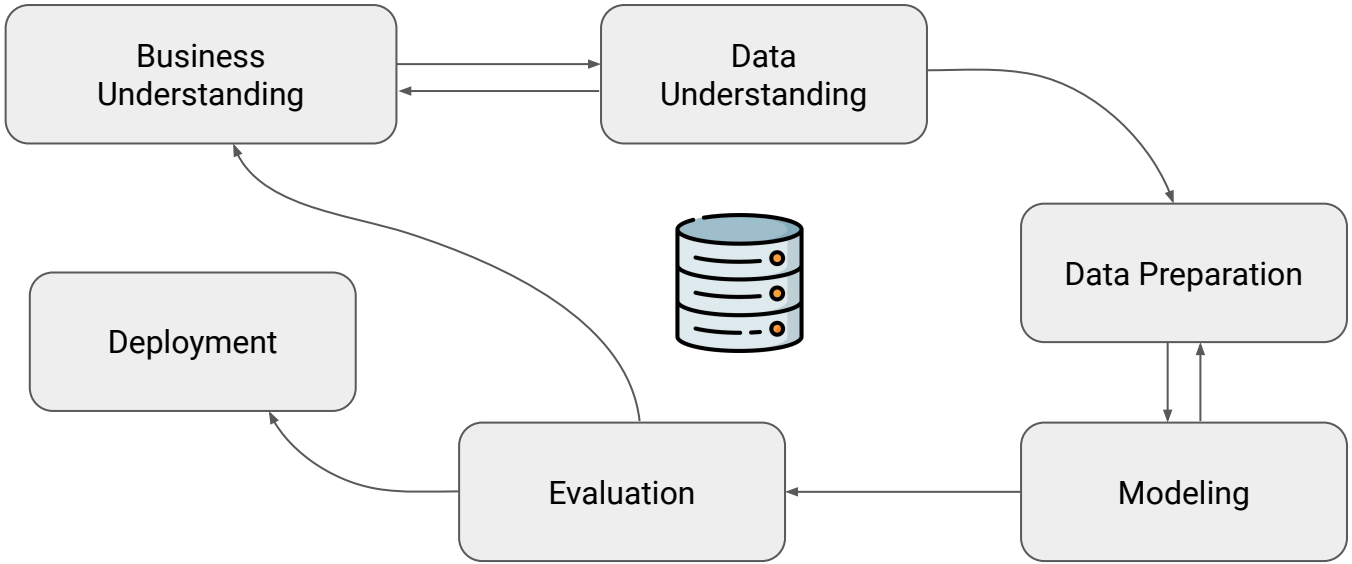
B. False

Data Science Methodology & Life Cycle

Methodology, Workflow, Cycle

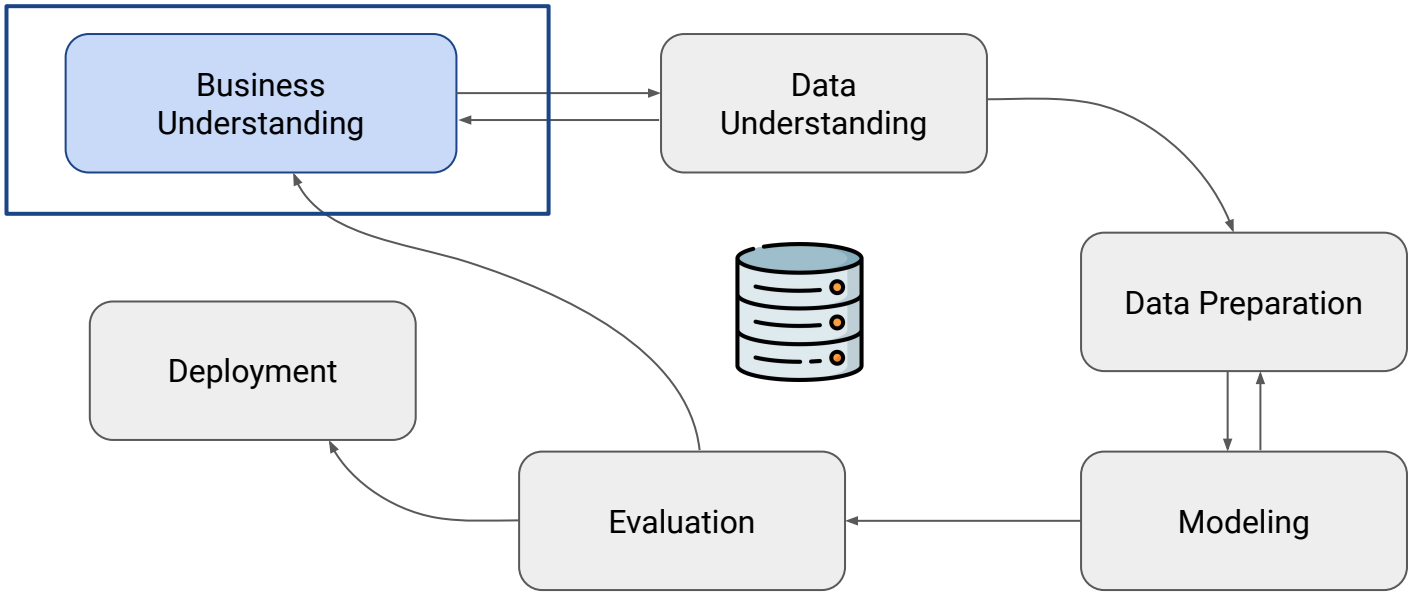
CRISP - DM

Cross Industry Standard Process for Data Mining



CRISP - DM

Cross Industry Standard Process for Data Mining



CRISP - DM

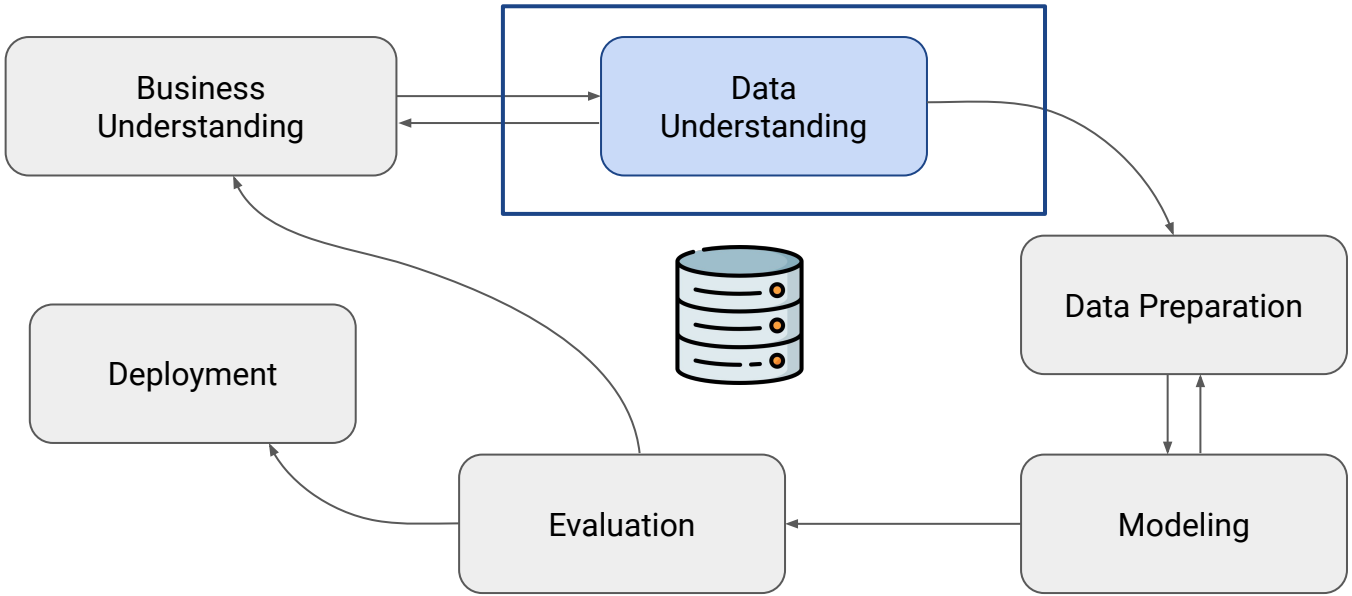
Cross Industry Standard Process for Data Mining

Business Understanding

This entails the understanding of a project's objectives and requirements from the business viewpoint. Such business perspectives are used to figure out what business problems to solve via the use of data mining.

CRISP - DM

Cross Industry Standard Process for Data Mining



CRISP - DM

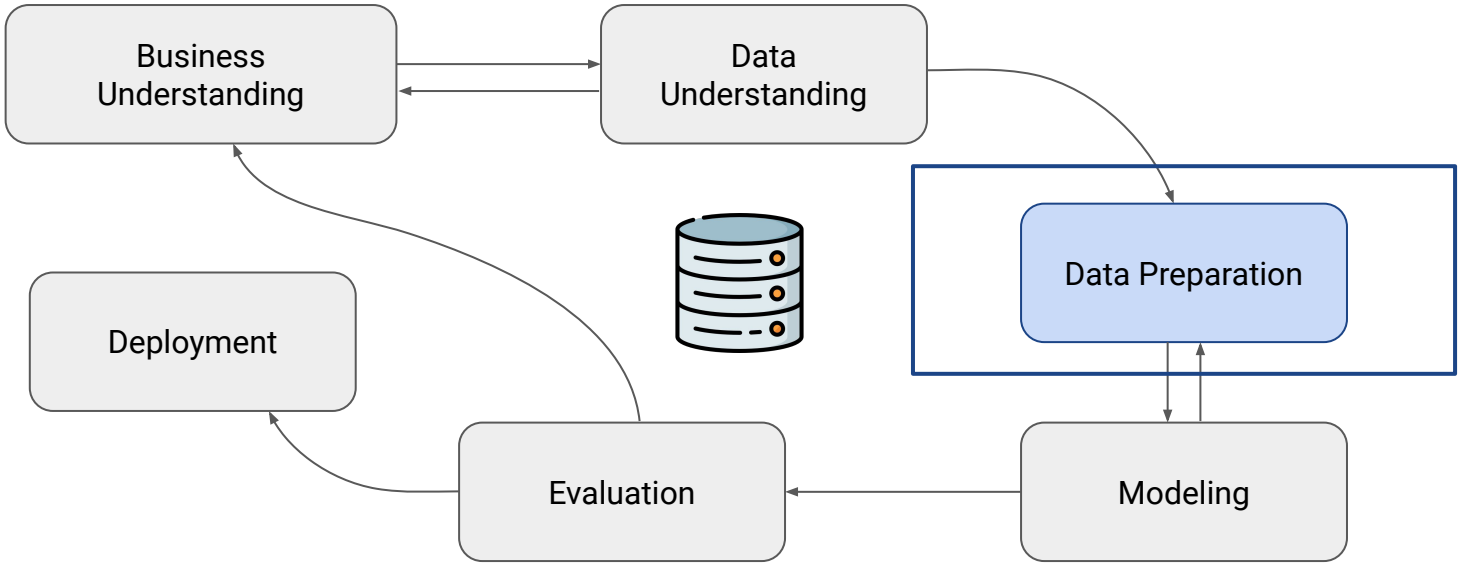
Cross Industry Standard Process for Data Mining

Data Understanding

This phase allows us to become familiarize with the data and this involves performing exploratory data analysis. Such initial data exploration may allow us to figure out which subsets of data to use for further modeling as well as aid in the generation of hypothesis to explore.

CRISP - DM

Cross Industry Standard Process for Data Mining



CRISP - DM

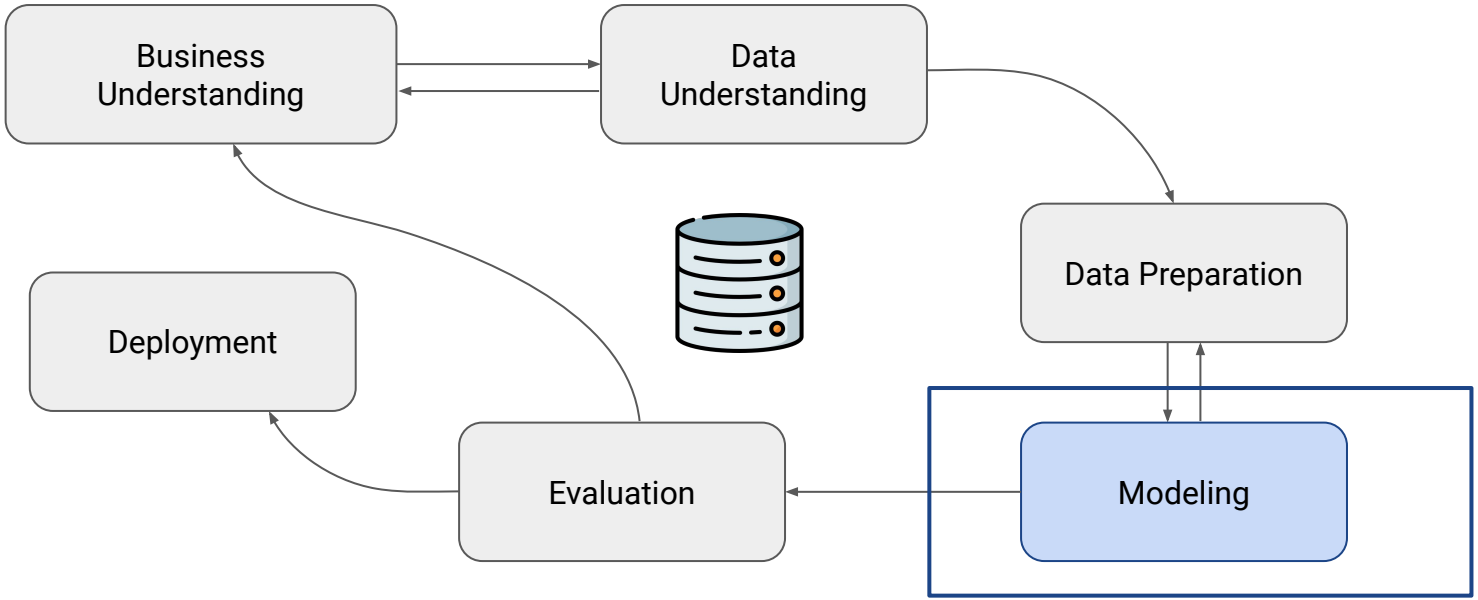
Cross Industry Standard Process for Data Mining

Data Preparation

This can be considered to be the most time-consuming phase of the data mining process as it involves rigorous data cleaning and pre-processing as well as the handling of missing data.

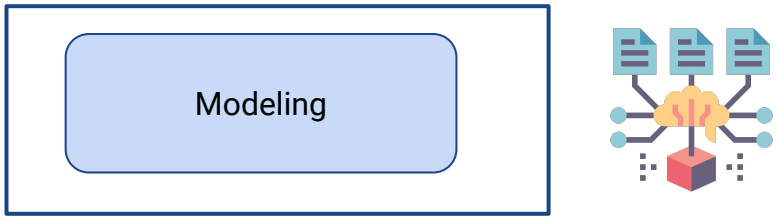
CRISP - DM

Cross Industry Standard Process for Data Mining



CRISP - DM

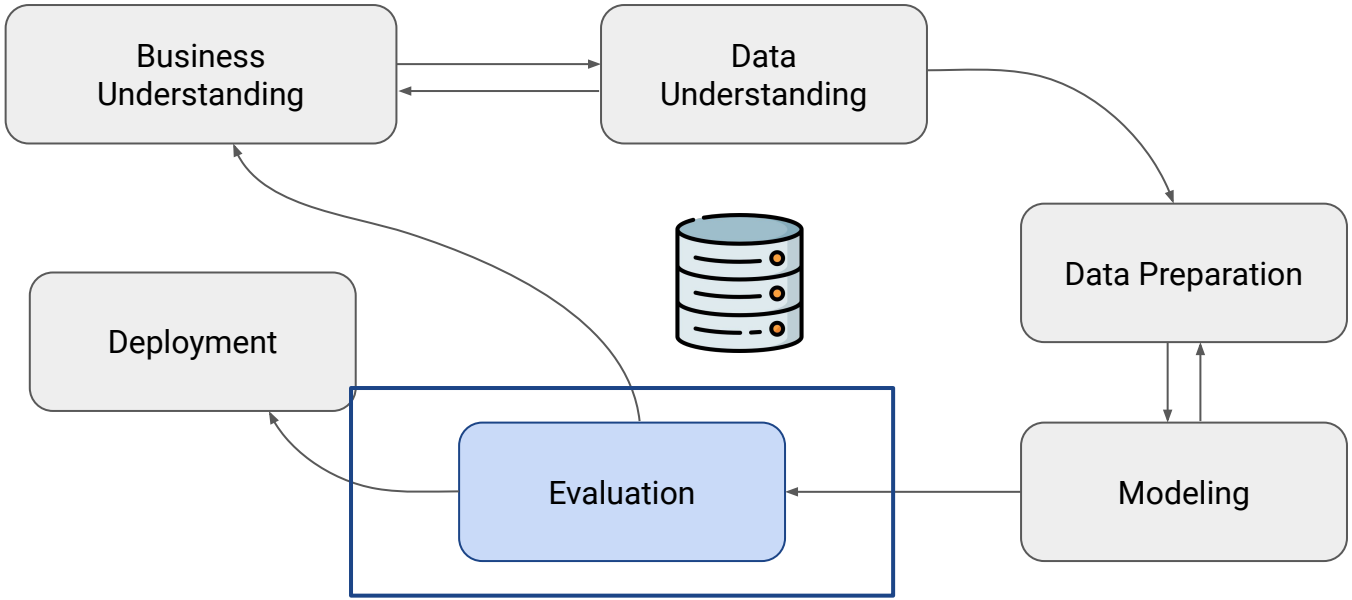
Cross Industry Standard Process for Data Mining



The pre-processed data are used for model building in which learning algorithms are used to perform multivariate analysis.
Iterate model building and assessment until you strongly believe that you have found the best model(s).

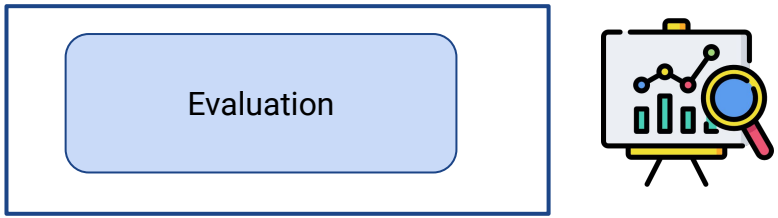
CRISP - DM

Cross Industry Standard Process for Data Mining



CRISP - DM

Cross Industry Standard Process for Data Mining

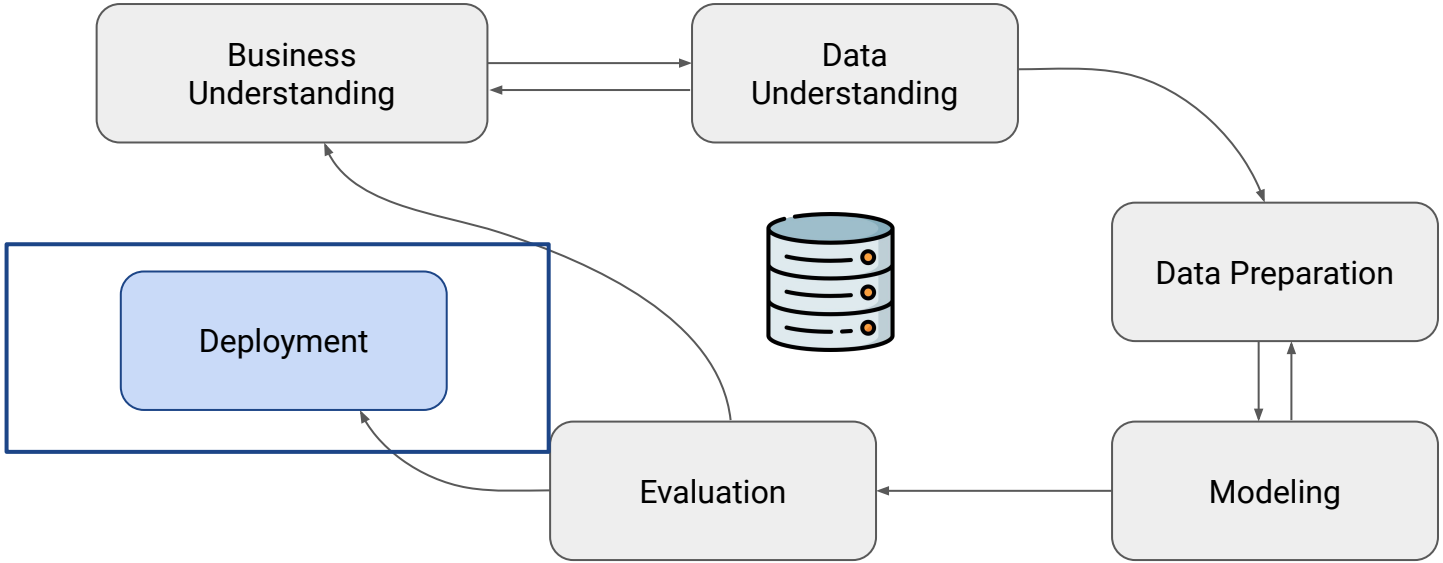


It is important to evaluate the model results and review the process performed to determine whether the originally set business objectives are met or not.

If deemed appropriate, some steps may need to be performed again. Rinse and repeat. Once it is deemed that the results and process are satisfactory then we are ready to move to deployment. Additionally, in this evaluation phase, some findings may ignite new project ideas for which to explore

CRISP - DM

Cross Industry Standard Process for Data Mining

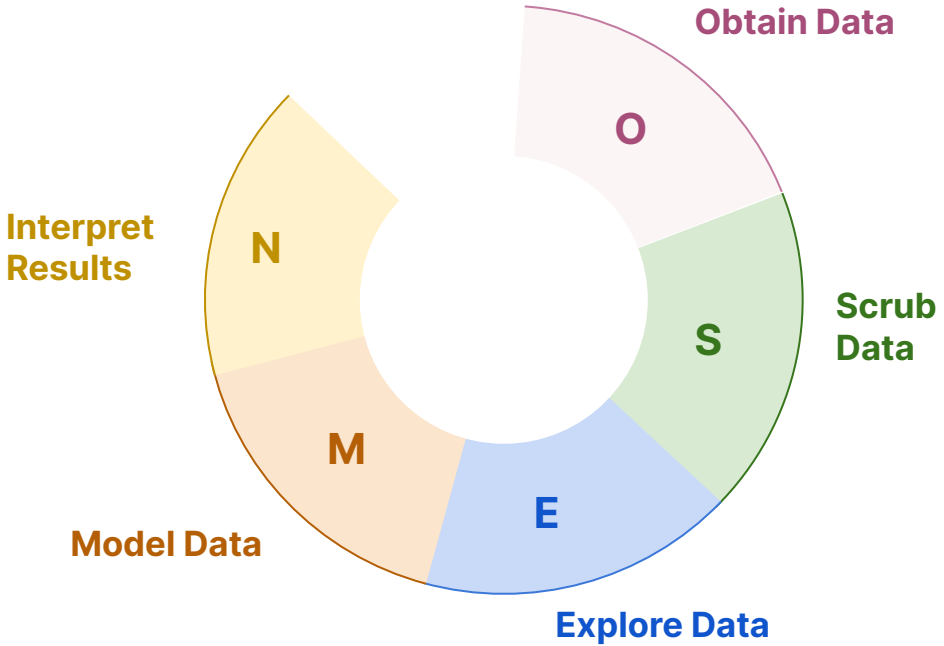


CRISP - DM

Cross Industry Standard Process for Data Mining

Once the model is of satisfactory quality, the model is then deployed, which may range from being a simple report, an API that can be accessed via programmatic calls, a web application, etc.

OSEMN



OSEMN

Obtain Data

Data forms the requisite of the data science process and data can come from pre-existing ones or from newly acquired data (from surveys), from newly queried data (from databases or APIs), downloaded from the internet (e.g. from repositories available on the cloud such as GitHub) or extracted.

OSEMN

Scrub Data

Scrubbing the data is essentially data cleaning and this phase is considered to be the most time-consuming as it involves handling missing data as well as pre-processing it to be as error-free and uniform as possible.

OSEMN

Explore Data

This is essentially exploratory data analysis and this phase allows us to gain an understanding of the data such that we can figure out the course of actions and areas that we can to explore in the modeling phase. This entails the use of descriptive statistics and data visualizations.

OSEMN

Model Data

Here, we make use of machine learning algorithms in efforts to make sense of data and gain useful insights that are essential for data-driven decision-making.

OSEMN

Interpret Results

This is perhaps one of the most important phase and yet the least technical as it pertains to actually making sense of the data by figuring out how to simplify and summarize results from all the models built.

This is including draws meaningful conclusion and rationalizing actionable insights that would essentially allow us to figure out what the next course of actions are. For example, what are the most important features that influences the class labels.

Pop Quiz!

Which of these is NOT part of the CRISP DM Data Understanding phase?

A. Defining the problems that we want to solve.

B. Finding and identifying any problems within the data sets.

C. Cleaning and addressing any problems with the data sets.

Pop Quiz!

The CRISP DM phase of Evaluation is similar to which step at OSEMN?

A. O

B. S

C. E

D. M

E. N

Tools & Tech-Stacks for Data Scientist

Terima kasih!

Ada pertanyaan?