Module 1: Data Science Fundamentals

Quick Reference Guide

By the end of this module, you will be able to:

- Discuss the ways data/signals are generated.
- Correlate business examples with the correct model type.
- Sort elements according to the five modeling layers.
- Explain potential errors in simplified representations of data.

The Dimensions of Readiness

The framework of analytics, machine learning and artificial intelligence rely on three major concepts:

- **Data**: are collected from business environments/organizations and, through models, inform the design and improvement of intelligent decision processes and systems
- **Models**: serve data-enabled languages that allow organizations to leverage data into the design of intelligent and improved decision processes and systems
- **Processes**: comprise the sets of activities and resources organizations use to convert input to output

The decision process includes:

- **Input**: data, predictions, and decisions
- **Resources**: databases, tools, and human experts
- **Activities**: meetings, discussions, analyses, and algorithms
- **Output**: actions

Defining Models

A model:

- Is a concept that captures relationships between different quantities in the real world.
- Captures the relationship between the model's input that is usually observed and the output or outcome that can be either observed or unobserved
- Has parameters that describe the relationship between the input and the output and they are estimated from data

In order for a business to have a holistic view of the market and how an organization competes efficiently within that market, it requires a robust analytic environment which includes:

- Descriptive analytics: uses data aggregation and data mining to provide insight into the past and answer: “What has happened?”

- Predictive analytics: uses statistical models and forecasting techniques to understand the future and answer: “What could happen?”
- Prescriptive analytics: uses optimization and simulation algorithms to advise on possible outcomes and answer: “What should we do?”

Factors affecting the choice of models:

- The accuracy of different models
- The computational complexity
- The amount of data needed to use each model
- The interpretability of each model

Qualitative and Quantitative models

The qualitative model

- It is constituted by the concepts of clinical and logistical readiness
- It does not provide exact equations that describe precise connections or relationships
- It is called a descriptive model, as it describes the system and processes involved in your patient’s journey
- It is an important first step but does not lend itself to the use of large-scale data

The data-enabled quantitative model

- It takes the concept of readiness, both clinical and logistical readiness, and translate them to a model that is data-enabled
- It is also known as a representation model

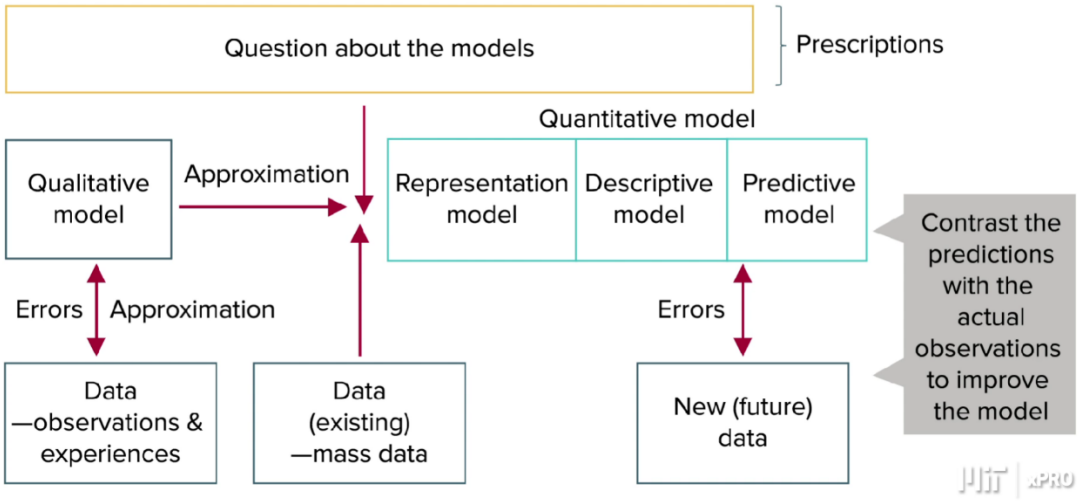
Modeling Layers

The way you choose to interpret the data will become more nuanced over time. To understand these changes, you can use the following layers of modeling as a scale to guide organizations down the right path to get the most from their data.

1. Data/signals
2. Representation models
3. Descriptive models
4. Predictive models
5. Prescriptive/optimization models

Data and Models

The figure below shows the interaction between data and models. Based on the data, you can convert a qualitative model into a quantitative data-enabled model and finally into a predictive model. The predictions of the predictive models can be contrasted with the actual observations which will enable better decision making.



Data obtained through observations and experience is approximated to create a qualitative model, with some error introduced. The qualitative model along with prescriptions and existing data is approximated to a quantitative model which is build up as the representation model, descriptive model and predictive model. The predictive model provides new or future data. Predictions are contrasted with actual data to improve the model.