1.1 Introduction to Data Science

Jawad
DSAI

8, Jan 2025

Module Overview

Agenda

- Introduction to data science
- Introduction to database
- Introduction to data science toolbox
- Demo and code-along

Terminology

Artificial Intelligence

Artificial Intelligence (AI) is an *umbrella term* for any technology where a computer program performs a task that typically requires human intelligence
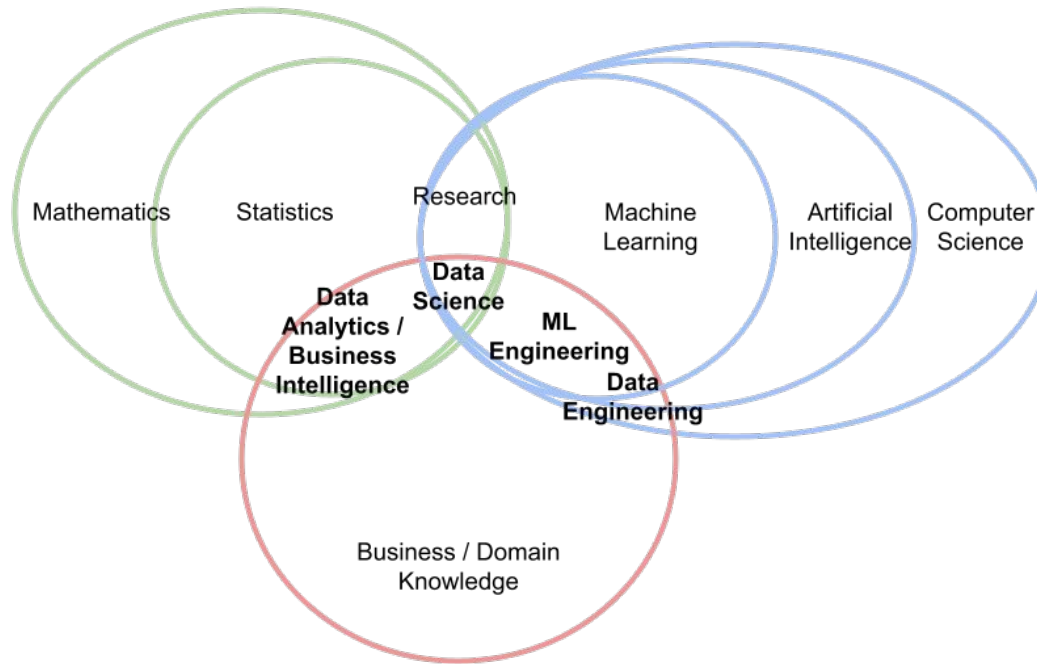
Machine Learning

Machine Learning (ML) is a *subset of AI* that involves training algorithms to make predictions or decisions based on data (without explicitly being programmed to do so).

Data Science

Data Science is an *interdisciplinary field* that uses statistics & ML methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.

Data Functions and Roles

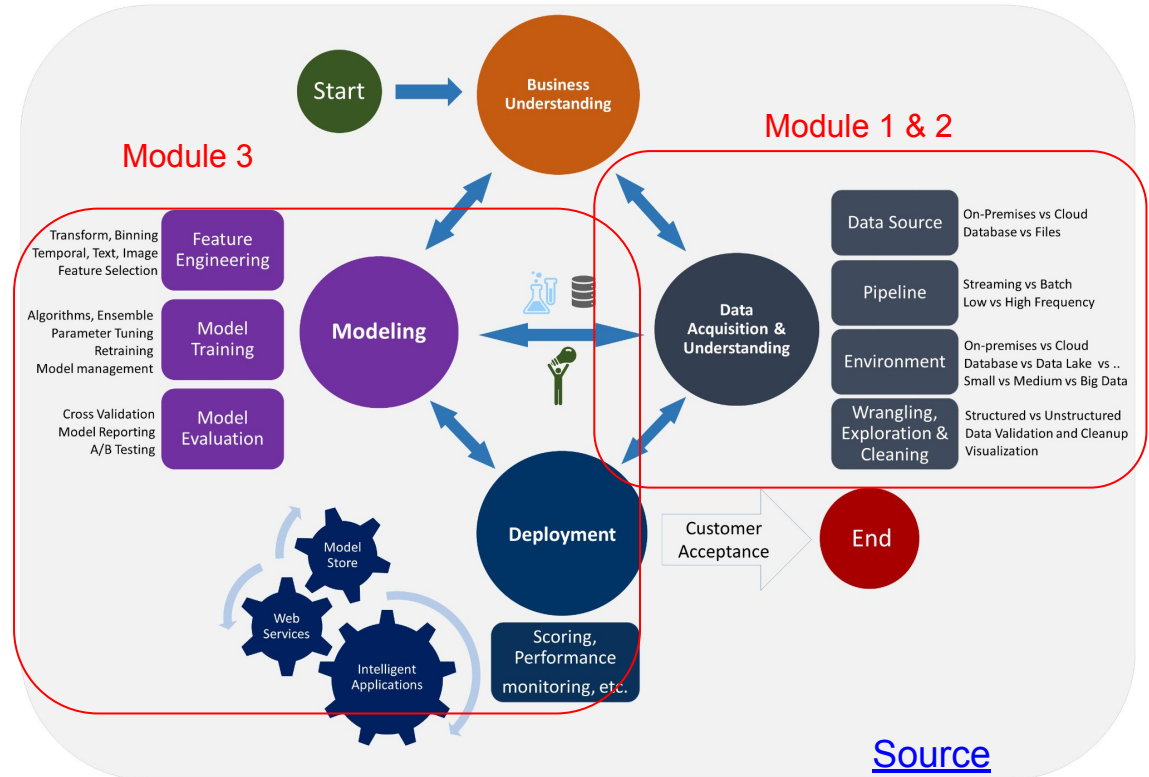Data Analytics / Business Intelligence -> Data Analyst

Data Science -> Data Scientist

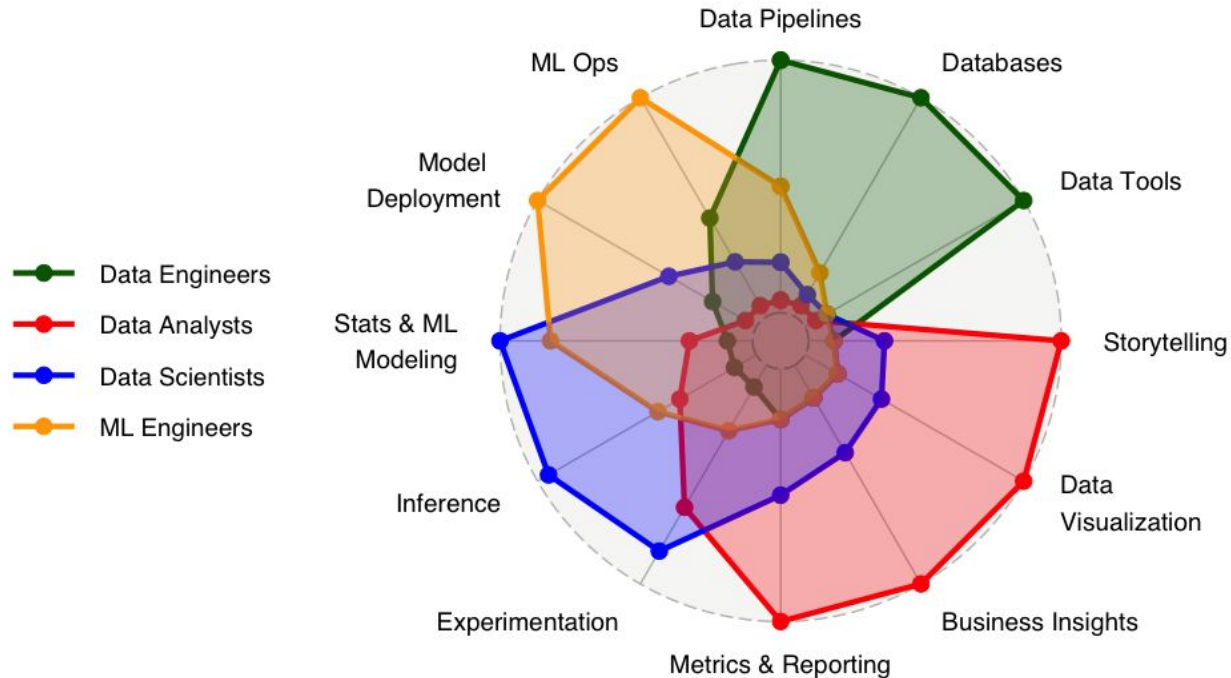ML Engineering -> ML Engineer

Data Engineering -> Data Engineer

Hidden technical debt in machine learning

Data Science Lifecycle

1. Business understanding
2. Data acquisition and understanding
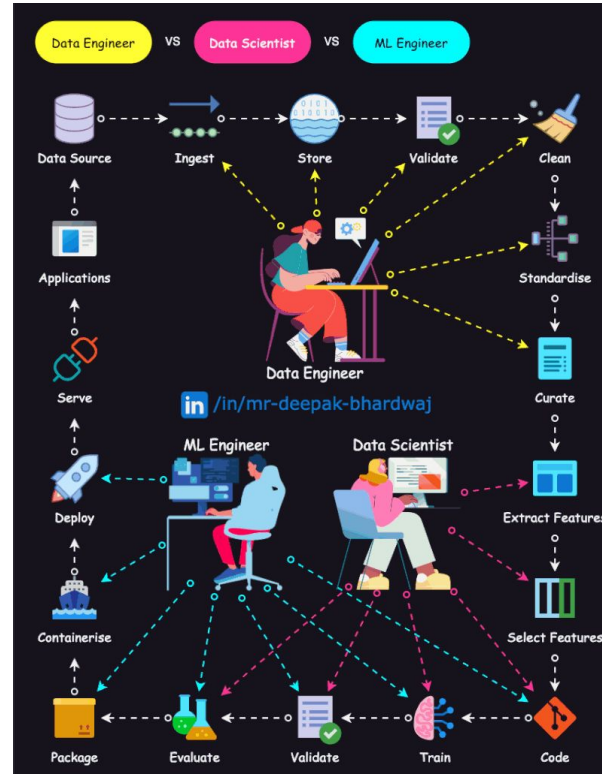3. Modeling
4. Deployment
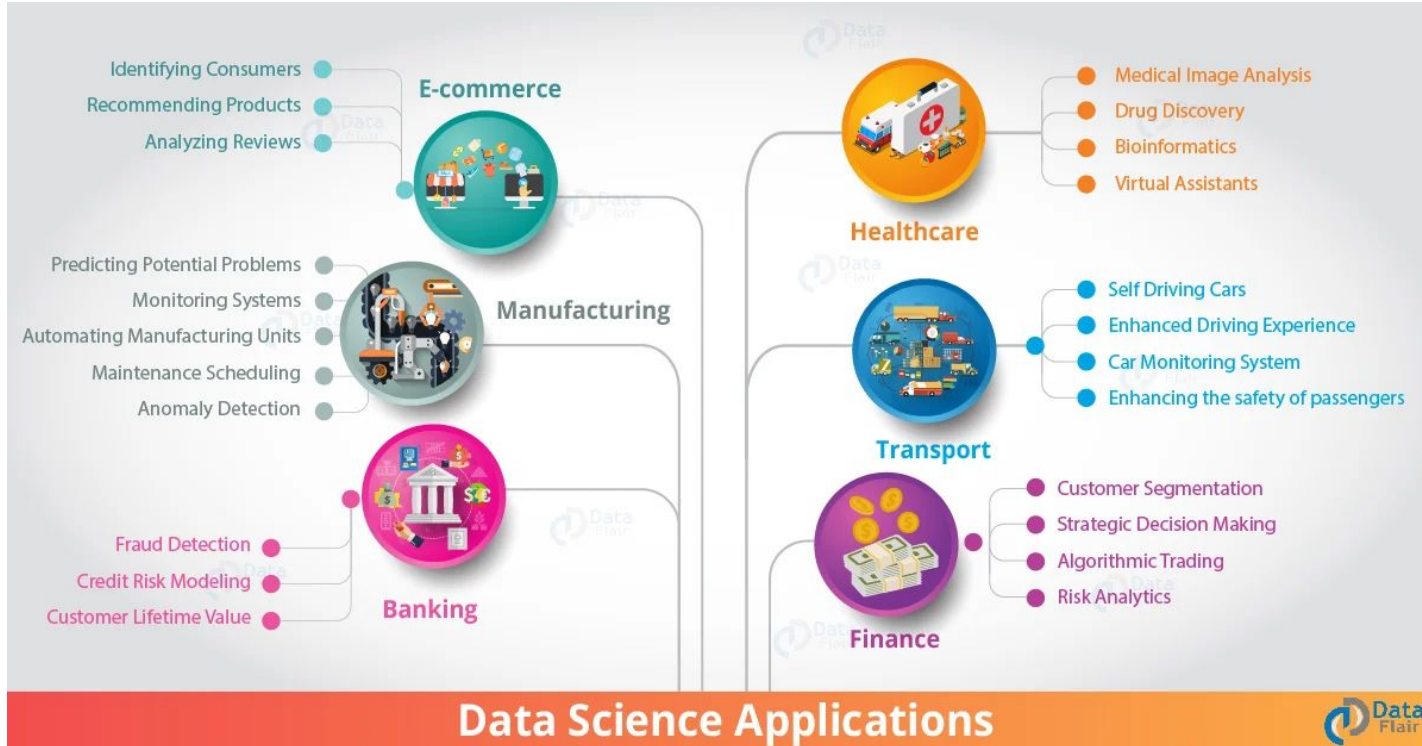5. Customer Acceptance

Data Roles on Data Lifecycle

More info

Data Roles on Data Lifecycle

Data Science Use Cases

Data Science Applications

What is a Database

A database is a collection of data that is organized in a structured way. It can be accessed or stored in a computer system and managed through a Database Management System (DBMS). Databases are used to store and manage data for various applications such as websites, mobile apps, and enterprise systems.

Relational Database

- also known as SQL database
- based on relational model and has *predefined* schema
- suited for *structured data*
- use **SQL** (Structured Query Language) for querying

NoSQL Database

- "not-only" SQL, or non-relational database
- *flexible* schema
- suited for *semi-structured / unstructured* data
- might or might not use **SQL**

Python Data Science Stack

Data Science Toolbox