Machine Learning in the Real World
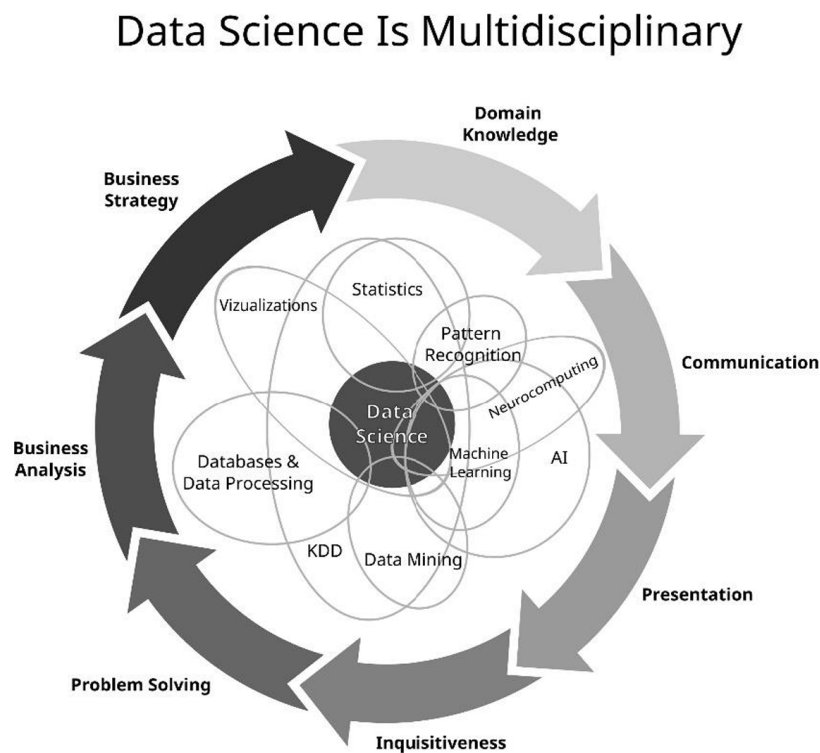
Stylianos Kampakis

Centre for Blockchain Technologies, University College London, UK

INTRODUCTION

Data science cannot be disentangled from organizational demands. A famous graph (Figure 1) produced by Brendan Tierney (2012) demonstrates how data science is integrated within an organization.

Figure 1. A concise summary of what data science is (Tierney, 2012)

This well-known diagram shows that data science is surrounded by a cycle of "soft skills" and business-related terms like "business strategy." What this diagram demonstrates is that data science is an applied discipline. Data science aims to bring impact within an organization.

When machine learning is seen in isolation, it can be treated as an academic exercise, and it is acceptable for the researcher to deal with questions such as convergence bounds and metrics like categorical cross-entropy. On the other hand, when machine learning is used in practice, it is called data science, and

M

research questions are essential only when converted into revenue. Metrics like the RMSE are replaced with KPIs, and the end-to-end process needs to translate into tangible impact for an enterprise.

The successful application of data science requires a set of steps within the organization.

Data science and machine learning are often taught in isolation as if the successful practice of those disciplines simply is about implementing algorithms. However, experience has demonstrated that the successful implementation of data science requires covering other aspects, like data strategy.

This chapter aims to assist data scientists and machine learning practitioners in engaging with the larger picture.

BACKGROUND

The first step is the successful design and execution of a data strategy. Data strategy refers to the long-term plan and goals of a business with regard to its data. Data strategy encapsulates business strategy, data governance, and data science and requires the participation of business leaders and technical experts.

The second step is the successful choice and implementation of a data science process. A data science process represents a structured approach to data science and machine learning projects. For example, methodologies like AGILE (Dingsøyr et al., 2012) and SCRUM (Hossain et al., 2011) have helped structure and define the work of software developers. Similar methodologies have appeared in the last few years in data science, whose requirements are different from the requirements of traditional software development.

The third step is to consider recent developments in machine learning and AI, like MLOps, which streamline machine learning training, deployment, and execution. MLOps stands for machine learning operations and is an acronym borrowed from Devops. Devops is defined on the official website of Amazon AWS (one of the pioneers of this field) as:

DevOps is the combination of cultural philosophies, practices, and tools that increases an organization's ability to deliver applications and services at high velocity: evolving and improving products at a faster pace than organizations using traditional software development and infrastructure management processes. This speed enables organizations to better serve their customers and compete more effectively in the market. (Amazon Web Services, n.d.)

In a similar spirit, MLOps combined a set of philosophies and practices to streamline the development, testing, and deployment of machine learning algorithms.

DATA STRATEGY: AN OVERVIEW

Companies must be able to manage vast quantities of data to succeed in the current business environment. Unfortunately, despite many companies instituting data-management and chief data officer roles, they are still playing catch-up.

Studies show that most businesses use less than half of their structured data in the decision-making process and use less than 1% of their unstructured data. Furthermore, over 70% of employees can access data they should not, and analysts spend 80% of their time discovering and preparing data (DalleMule & Davenport, 2017). Unfortunately, as the figures show, most companies do not have a coherent plan

Related Content

Generating an Artificial Nest Building Pufferfish in a Cellular Automaton Through Behavior Decomposition

Thomas E. Portegys (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-12).*

www.irma-international.org/article/generating-an-artificial-nest-building-pufferfish-in-a-cellular-automaton-through-behavior-decomposition/233887

Demystifying Federated Learning for Blockchain: A Case Study

Pawan Whig, Arun Veluand Pavika Sharma (2022). *Demystifying Federated Learning for Blockchain and Industrial Internet of Things (pp. 143-165).*

www.irma-international.org/chapter/demystifying-federated-learning-for-blockchain/308117

A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques

Ramanujam Elangovanand Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 39-56).*

www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127

Autoencoder Based Anomaly Detection for SCADA Networks

Sajid Nazir, Shushma Pateland Dilip Patel (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 83-99).*

www.irma-international.org/article/autoencoder-based-anomaly-detection-for-scada-networks/277436

YOLO v5 and Faster R-CNN Performance Evaluation of Solid Waste in Object Detection Application

 Pradeep Kumar T., Lilly Florence M.and Fathima G. (2023). *Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications (pp. 288-304).*

www.irma-international.org/chapter/yolo-v5-and-faster-r-cnn-performance-evaluation-of-solid-waste-in-object-detection-application/316761