SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY — INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Seamless Synergy: Unifying Local Development and Cloud Execution in Machine Learning DevOps

Baraa Alnassan

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY — INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Seamless Synergy: Unifying Local Development and Cloud Execution in Machine Learning DevOps

Nahtlose Synergie: Vereinheitlichung von lokaler Entwicklung und Cloud-Ausführung in DevOps für maschinelles Lernen

Author:	Baraa Alnassan
Supervisor:	Jens Großklags
Advisor:	Derui Zhu (TUM), Kaio Giurizatto Utsch (IT4IPM)
Submission Date:	September 1st 2024

I confirm that this bachelor’s thesis is my own work and I have documented all sources and material used.

Munich, September 1st 2024 Baraa Alnassan

Acknowledgments

Abstract

Artificial Intelligence (AI) is a branch of computer science that aims to create systems capable of performing tasks that would typically require human intelligence. These tasks include learning and adapting to new information, understanding human language, recognizing patterns, solving problems, and making decisions. AI can be categorized into two main types: narrow or weak AI, which is designed to perform a specific task, and general or strong AI, which can perform any intellectual task that a human being can do. AI has various applications, such as self-driving cars, virtual personal assistants, and image recognition systems.
Machine Learning (ML) is a subcategory of AI. It is based on algorithms trained for decisions making that automatically learn and recognize patterns from data.
Machine Learning Operations (MLOps), is a practice that combines machine learning, data engineering, and DevOps practices to streamline the deployment and management of machine learning models. MLOps aims to bridge the gap between data science and DevOps, ensuring seamless collaboration and efficient deployment of ML models. It involves various components, such as model versioning, data pipeline management, model monitoring and observability, testing strategies for models, deployment orchestration, and choosing the right tools. Effective collaboration between data scientists and DevOps engineers is essential for successful MLOps practices.

This Bachelor thesis aims to highlight the difficulties of transitioning from local development environment to cloud based code execution in production environment in machine leaarning operations, and tries to find solutions to closeing the gap between local development and successful code execution on the cloud.

Contents

1 Introduction

ML is about the development of computer systems that can learn and adapt by following explicit instructions or any human interference. And it uses special algoriths and statistical models to analyze and predict the outcome from a given pattern of data. It was born to allow computers to learn and control their environment.
In today's rapidly evolving landscape of machine learning and software engineering, the integration of ML operations (MLOps) plays a crucial role. MLOps serves as the critical bridge between data science and DevOps practices, ensuring seamless collaboration and efficient deployment of ML models. As organizations increasingly rely on ML for decision-making, it becomes imperative to address the challenges associated with transitioning from local development to production environments.
To address these challenges the goal of this Thesis is to try to bridge the gap between local development and successful code execution on the cloud, in other words, how to make ML processes automated and operationalized so that more ML proof of concept can be brought into production.
To overcome those challanges, I conduct a mixed-method research endeavor to identify important principles of MLOps, carve out functional core components, highlight the roles necessary to successfully implement MLOps, and derive a general architecture for ML systems design. In combination, these insights result in a definition of MLOps, which contributes to a common understanding of the term and related concepts.
The remainder of this thesis is constructed as follows. I will first elaborate on the necessary foundations and related work in the field. Next, I will give an overview of the utilized methodology, consisting of a literature review, a tool review, and an interview study. I then present the insights derived from the application of the methodology and conceptualize these by providing a unifying definition. I conclude the paper with a short summary, limitations, and outlook.

2 Background

Before diving into the details of the main objective of the thesis, it is important to understand some background knowledge. In this chapter i will introduce some necessary detailed background of the rise of ML, and how the concenpt of MLOps came to be.

2.1 The Rise of Machine Learning

2.1.1 The Birth of Machine Learning (1950s)

The foundations of machine learning were laid in the 1950s by computer scientists and mathematicians who sought to develop algorithms that could enable computers to learn from data [Jac22; RVK24; Naq20]. During this era, several key milestones shaped the field. Such as the Perceptron Algorithm - one of the earliest breakthroughs in machine learning - by Frank Rosenblatt [Van86], was an early form of an artificial neural network, it became the building block for future advancements. Where he formulated a serie of machines, each serves to interduce a new concept. It aimed to mimic the way biological neurons process information.

2.1.2 Resurgence and New Techniques (1980s-1990s)

An AI winter started in the 1970s, that was described as a chain reaction much the same as a nuclear winter, that began with pessimisum in the AI community, followed by cutback in fundings, which in turn resulted in end of serious research. After the AI winter, machine learning was resurrected in the 1980s and 1990s. New techniques brought to the light, such as Neural Network (NN) where Researchers were inspired by the human brain's structure, according to Geoffrey Hinton [McD] , often referred to as the "father of Deep Learning,", who made significant contributions during this time.

2.1.3 The Rise of Big Data (2000s)

The 2000s witnessed the rise of big data and the availability of vast amounts of data for training machine learning models. This era witnessed the emergance of new techniques to be able to deal with vast amount of data, such as, Data-Driven Approaches [TR23; AA23], where machine learning algorithms learned patterns directly from large datasets, and Supervised Learning Dominance, where models learn from labeled examples. Techniques like regression and classification gained prominence.

2.1.4 Deep Learning and Beyond (2010s)

The 2010s marked a transformative period for machine learning, where the term Deep learning was first introduced , it it was defined as a subfield of machine learning. other powerful methods were perfected in this era such as Convolutional Neural Networks (CNN) for image recognition and recurrent neural networks for sequential data revolutionized various domains.

2.1.5 State-of-the-Art Machine Learning (Present and Future)

Nowadays machine learning has made far-reaching changes in various applications across industries. Machine learning is applied in healthcare, fraud detection, speech recognition, autonomous vehicles, recommendation systems, and more (Figure 2.1 shows some applications of machine learning). It revolutionizes industries and improves efficiency. However, researchers continue to push boundaries, aiming for AI systems that understand context, and are looking forward for the exciting possibilities ahead.

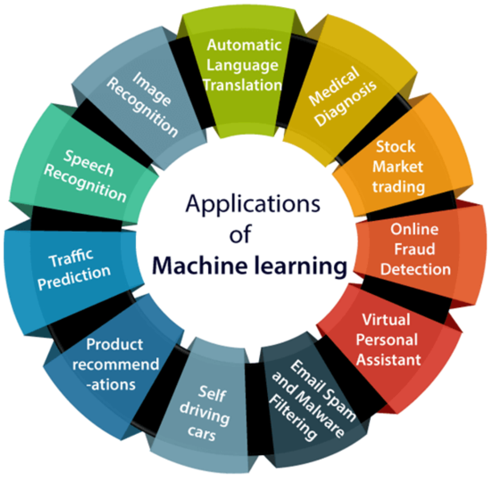

Figure 2.1: Few applications of machine learning.

2.2 The Evolution of MLOps

With the rise of machine learning, so did the need for efficient operations around it. This led to the evolution of MLOps MLOps is an engineering practice that aims at making the process of deploying machine learning models more efficient, reliable, and maintainable. It uses Continuous Integration and Continuous Deployment (CI/CD) and machine learning models to rationalize the monitoring, deployment and maintenance of machine learning systems.

2.2.1 Key Components of MLOps

As the datasets to train machine learning models got bigger and more complex, the realization for the need of ML life cycle grow, as the old techniques weere slow and difficult to scale. Data scientists worked together with IT teams to create and develop an assembly line for each step of training machine learning models. In this chapter I will discuss few steps of michine learning life cycle that are crucial for MLOps process.

- **Data Collection**: Gathering and refining relevant data set from different sources to prepare it for modeling is one of the key components of ML life cycle.
- **Building and training**: This is the step where models are created and trained using data, that were prepared in the previous step. It involves selecting appropriate algorithms and preprocessing the data.

And more as shown in Figure 2.2.

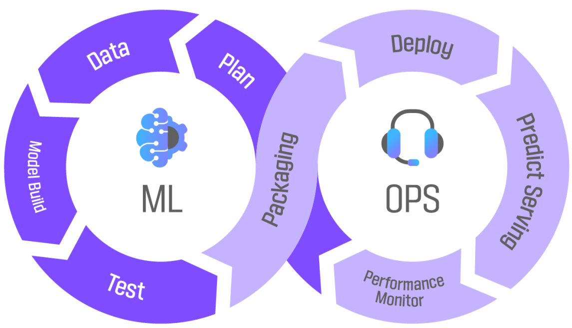

Figure 2.2: machine learning life cycle.

2.2.2 Comparing MLOps and DevOps

DevOps is a method which aims at bringing together software development and operations, it shortens the system's development life cycle and provides a continu delivery to a high software quality.
MLOps does, however, borrow from the DevOps principles of a rapid, continuous approach to writing and updating applications, in that they both have a code-validate-deploy loop. But the MLOps life cycle contains additional steps, that are necessary for building and training the machine learning model as shown in Figure 2.3. The aim in both cases is to take the project to production more efficiently with faster fixes, faster releases and ultimately, a higher quality product that boosts customer satisfaction, whether that's software or machine learning models.

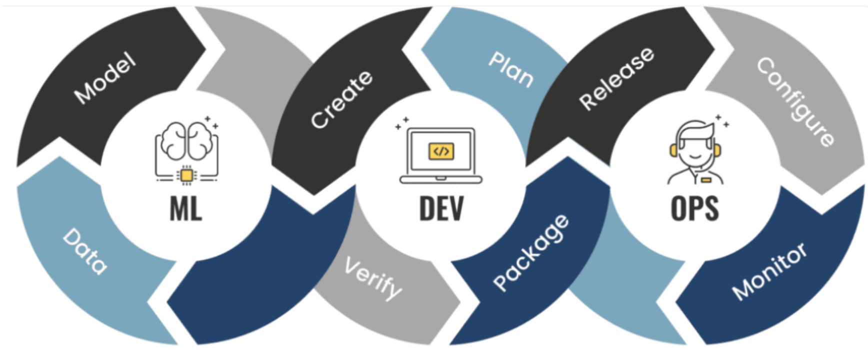

Figure 2.3: The transition from DevOps to MLOps life cycle.

In this thesis, I delve into these aspects, aiming to bridge the gap between local ML implementation and successful code execution in production systems. Let us embark on this journey toward efficient MLOps practices.

3 The Gap: WHy Does It Exists

Before diving into the details of bridging the gap between local code development and cloud execution in MLOps, it is essential to understand how the gap exists in the first place. In this chapter, I will explore a brief explanation of why there is a gap between the two environments.

3.1 Why the Gap Exists

There is a notable distinction between running code locally on a developer's machine and deploying it to a cloud server for execution. This disparity can pose certain challenges in terms of infrastructure, environment, configuration and more, which can make it challenging to ensure seamless functionality across both environments.
I would like to take this opportunity to examine a few of the reasons that have been presented. It is important to consider these reasons in a fair and impartial way:

- Local development environments often have different configurations, dependencies, and resources than cloud environments. This can lead to inconsistencies and incompatibilities when deploying code to the cloud.
- Cloud environments often have different security, compliance, and scalability requirements than local environments. This can lead to additional complexity and overhead when deploying code to the cloud.
- Cloud environments often have different monitoring, logging, and debugging tools than local environments. This can make it difficult to diagnose and resolve issues that arise in the cloud.

In the coming chapters, I will discuss different approaches to overcoming these difficulties and how they can be implemented in practice and put a spurt in the development and debugging process on the cloud.

4 Bridging the Gap

4.1 Containerization

4.2 Hybrid Approaches

4.3 Cloud Execution

5 Evaluation

6 Conclusion

Abbreviations

ML Machine Learning

AI Artificial Intelligence

MLOps Machine Learning Operations

NN Neural Network

CNN Convolutional Neural Networks

CI/CD Continuous Integration and Continuous Deployment

List of Figures

List of Tables

Bibliography

[AA23] F. Akram and A. Abbas. "Data-driven Decisions: Exploring the Power of Machine Learning and Analytics." In: (Nov. 2023).

[Jac22] A. Jacinto. *Machine Learning History*. 2022. URL: https://www.startechup.com/blog/machine-learning-history/.

[McD] M. McDonough. https://www.britannica.com/biography/Geoffrey-Hinton.. Accessed: 2024-5-7.

[Naq20] A. Naqvi. "Rise of Machine Learning." In: (Aug. 2020), pp. 51–67. DOI: 10.1002/9781119601906.ch4.

[RVK24] C. Rahal, M. Verhagen, and D. Kirk. "The rise of machine learning in the academic social sciences." In: *AI & SOCIETY* 39.2 (Apr. 2024), pp. 799–801. ISSN: 1435-5655. DOI: 10.1007/s00146-022-01540-w.

[TR23] M. Torkjazi and A. K. Raz. "Data-Driven Approach with Machine Learning to Reduce Subjectivity in Multi-Attribute Decision Making Methods." In: *2023 IEEE International Systems Conference (SysCon)*. 2023, pp. 1–8. DOI: 10.1109/SysCon53073.2023.10131094.

[Van86] C. Van Der Malsburg. "Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms." In: *Brain Theory*. Ed. by G. Palm and A. Aertsen. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 245–248. ISBN: 978-3-642-70911-1.