



# DEMYSTIFYING MLOPS FOR CONTINUOUS DELIVERY OF THE PRODUCT

J. PRAVEEN GUJJAR <sup>a\*</sup> AND V. NAVEEN KUMAR <sup>a=</sup>

<sup>a</sup> CMS Business School, Jain (Deemed-to-be University), Bengaluru, Karnataka, India.

## AUTHORS' CONTRIBUTIONS

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

*Received: 06 December 2021*

*Accepted: 10 February 2022*

*Published: 18 February 2022*

*Original Research Article*

## ABSTRACT

Machine learning is becoming integral part of the development of the product and has its own popularity. Once the model is developed and putting to the deployment stage it has to undergone series of stages. Development team should be constant touch with the operation team to deploy the final product. To make the process smoother MLOps come into existence. MLOps is the acronym for Machine Learning Operations (MLOps). It is the discipline make sure that machine learning model delivery through efficient and repeatable workflows. This paper demystifies the machine learning operations and also focus on the challenges and issues present during the delivery of the machine learning models. In this paper, it has been shown the similarity between DevOps and MLOps during the deployment stage. Result of the paper shows the importance of MLOps for the life cycle of machine learning product development.

**Keywords:** DevOps; MLOps; model preparation; model deployment; SDLC.

## 1. INTRODUCTION

Machine learning and data science are the discipline which can solve many complex real world problems. This discipline can help in transforming industries by provide value in all sectors. Machine learning and data science uses the large datasets to prepare the model such as computer vision, recommended artificial intelligence systems, natural language understanding etc., In-order to provide value to the business predictive models are created to help the users. MLOps uses the principles used by the DevOps to unify the machine learning system [1]. Machine learning system development and machine learning system operations are combined to frame MLOps. In this methodology machine learning process are

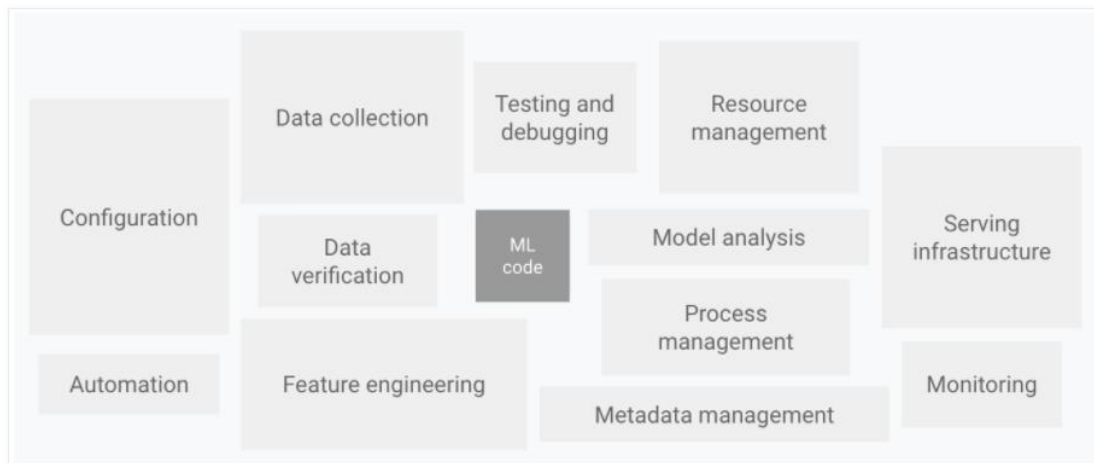
automated and monitored construction of system, model preparation, testing, integration and deployment in an effective manner. This paper demystifies the machine learning operations and also focus on the challenges and issues present during the delivery of the machine learning models. In this paper, it has been shown the similarity between DevOps and MLOps during the deployment stage.

## 2. RELATED WORK

In this section some of the related work of MLOps has been outlined. Fig. 1 shows that machine learning code is a small part in the ecosystem of model development. As in the Fig. 1 shows the elements of machine learning system it is having configuration,

<sup>=</sup>Associate Professor;

\*Corresponding author: Email: dr.praveengujjar@cms.ac.in, gujjarpraveen@gmail.com;



**Fig. 1. Elements of machine learning system**

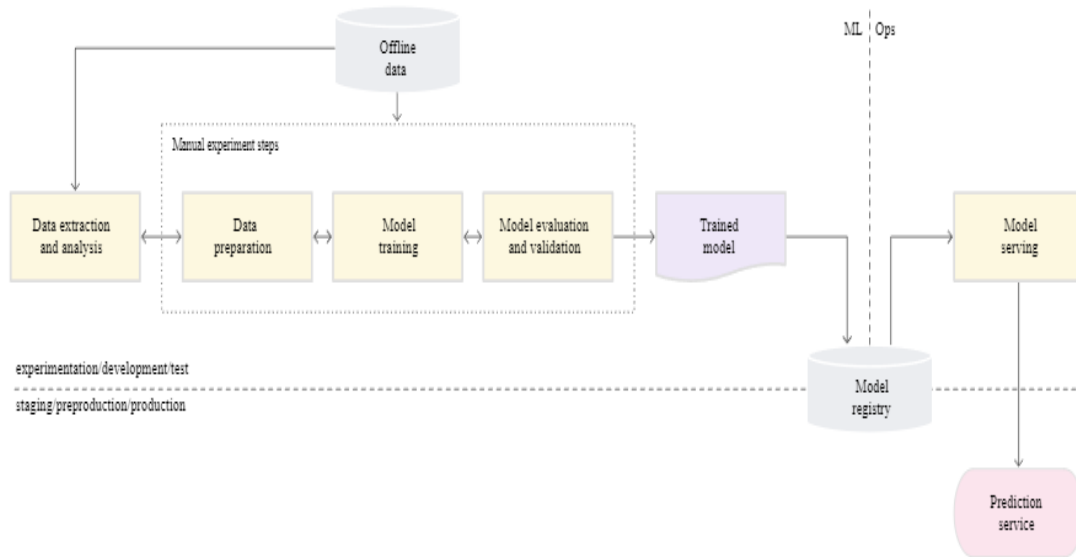
automation, data collection, data verification feature engineering, testing and debugging, resource management, metadata management, service infrastructure etc., In all this elements machine learning code is the small part which enables the model robust and reliable. Alla et al. [1] illustrated how important MLOps especially in the field of data science. Cedric Renggli et al. [2] shows the significance of quality of the data for the MLOps system and also describe the propagation of the data in the different phases of the machine learning model development.

Philipp Ruf et al. [3] illustrated the MLOps tools for connectivity and its role for every aspect of the MLOps cycle. Further, they have shown open source tools. Janis Klaise et al. [4] shown the challenges and issues available in the open source tools. Damnian A. Tamburri [5] describe the challenges and current trends in the sustainability. Machine learning helps in creating more efficient chatbot by understanding the intent of the message and extracting the entities from the user queries. The sentiment [6,7] of the text can also be analyzed with the help of the machine learning technique.

### 3. METHODS

MLOps is having the collection of the tools and techniques for the deployment of the machine learning models for the production. MLOps is the combination of the DevOps and machine learning operations. DevOps is the practices which aims to reduce the software release time by bridging the gap between operations team and software development team [8]. The main component of DevOps principles includes continuous integration and continuous delivery. Continuous integration is also referred as CI which is the practice where software development team integrate written code by the developer so that

they can test the module and make necessary improvement. Continuous integration helps in reduction in the software development life cycle [9]. Continuous deployment is also referred as CD which is the practice where operation team deploy the software to do the necessary configuration and monitor the performance. Continuous integration and continuous deployment helps in the reduction of the overall development cycle and new features may be reached the end user at the earliest [10]. The manual machine learning steps are shows in the figure 2. The machine learning model can be built with the steps as data collection, feature selection, feature engineering, data preparation, model training and model evaluation and finally trained model is prepared. Once the model is trained it has to be deploy the business purpose during this time any changes in the dataset it is going to create unnecessary mal functioning of the model. Hence, MLOps come into existences. Here, operation teams join the data scientist to deploy the developed model. If necessary data scientist may retrain the model to obtain the accuracy in the result. Data science steps for machine learning involves data extraction, data preparation, data analysis, model training, model evaluation, model validation, model serving and model monitoring. The automation of the above mention steps defines the maturity of the machine learning process. The automation of the machine learning gives very simple steps such as input the data set and in the output model is generated. Quality data create the better model and the performance is also get increases [11]. MLflow [12] is a Machine Learning Platform that can be used for partial automation for small- to medium-sized data science teams. The tool has four parts—Tracking, Model, Projects, and registry. Data extraction tools which is also referred as version controls which is managing the different version of data sets in a well-organized and in an accessible way.



**Fig. 2. Manual machine learning steps**

DevOps has two core practices: Continuous Integration (CI) and Continuous Delivery (CD). Continuous Integration is a software practice that focuses on automating the process of code integration from multiple developers. In this practice, the contributors are encouraged to merge their code into the main repository more frequently. This enables shorter development cycles and improves quality, as flaws are identified very early in the process. Similarity between MLOps and DevOps involves the placing of the software in a fault tolerant workflow and repeatable workflow but in the MLOps software is considered as the machine learning component. In this article researcher is demystify the machine learning operations. In this model preparation and deployment place a vital role for the purpose of business. Researcher has used different data processing, modelling and operational tools to perform the MLOps in the real world scenario.

#### 4. RESULTS AND DISCUSSION

Machine learning is becoming integral part of the development of the product and has its own popularity. Once the model is developed and putting to the deployment stage it has to undergone series of stages. Development team should be constant touch with the operation team to deploy the final product. In this article researcher is demystify the machine learning operations. In this model preparation and deployment place a vital role for the purpose of business. Researcher has used different data processing, modelling and operational tools to perform the MLOps in the real world scenario. In this

paper, it has been shown the similarity between DevOps and MLOps during the deployment stage. The different data processing tools, modelling tools and operational tools are listed in table 1, 2 and 3 respectively. In this article only the recent data processing tools, modelling tools and operational tools has been considered.

CI and CD have been adopted as the best practices for software development in recent years. Automating these practices also requires a robust pipeline known as the DevOps pipeline or CI/CD pipeline. The pipeline consists of a series of automated processes that enable software teams to build and deploy a new version of software smoothly. It ensures that the new developments are automatically merged into the software, which is followed up by automated testing and deployment. In the recent past many researcher has focus on the MLOps maturity models and the automated pipelines. There have been many challenges have been identified and lot of work need to be do in-order to overcome those challenges. Major challenges exist in the MLOps is that creation of robust and efficient pipelines to integrate the model and operations. Retraining is considered as the another important challenges because of the data manipulation. When a trained model is deployed it has to predict the unseen data and it may be having to take the raw data to obtain the final result. Hence, retraining in the case of data manipulation has to be done rigors. Another important challenges in the MLOps involves the data monitoring and model monitoring to obtain the result in an accurate way.

**Table 1. Data processing tool**

| Sl.No | Data Processing Tools | Year of Introduction | Remarks                  |
|-------|-----------------------|----------------------|--------------------------|
| 1     | Supervisely           | 2019                 | Data Pre-processing tool |
| 2     | Snorkel               | 2019                 | Data Pre-processing tool |
| 3     | Delta Lake            | 2019                 | Data Versioning          |
| 4     | Segments.ai           | 2020                 | Data Pre-processing tool |
| 5     | LakeFs                | 2020                 | Data Versioning          |
| 6     | Dolt                  | 2020                 | Data Versioning          |

**Table 2. Modelling tools**

| Sl.No | Modelling Tools | Year of Introduction | Remarks             |
|-------|-----------------|----------------------|---------------------|
| 1     | AutoFet         | 2019                 | Feature Engineering |
| 2     | GuildAi         | 2019                 | Experiment Tracking |
| 3     | Feast           | 2019                 | Feature Engineering |
| 4     | ModelDB         | 2020                 | Experiment Tracking |
| 5     | Rasgo           | 2020                 | Feature Engineering |
| 6     | Aim             | 2021                 | Experiment Tracking |
| 7     | HopsWork        | 2021                 | Feature Engineering |

**Table 3. Operational tools**

| Sl.No | Modelling Tools | Year of Introduction | Remarks          |
|-------|-----------------|----------------------|------------------|
| 1     | BentoML         | 2019                 | Model Deployment |
| 2     | MLrun           | 2019                 | Model Monitoring |
| 3     | Seldon          | 2020                 | Model Deployment |
| 4     | KFServing       | 2020                 | Model Deployment |
| 5     | Arize           | 2020                 | Model Monitoring |
| 6     | Cloudera        | 2020                 | end-to-end use   |
| 7     | Cortex          | 2021                 | Model Deployment |
| 8     | Deep checks     | 2021                 | Model Monitoring |
| 9     | Aporia          | 2021                 | Model Monitoring |

## 5. CONCLUSION

In this article, researcher presented MLOPs and DevOps concept and its challenges in the real world scenario. In conclusion, MLOps is the most efficient way to incorporate ML models in production. Every year more enterprises use these techniques and more research has been made in the area. But MLOps maybe has a different usage. In addition to the application of ML models in production, a fully mature MLOps system with continuous training can lead us to more efficient and realistic ML models. Further, choosing the right tools for each job is a constant challenge. This paper demystifies the machine learning operations and also focus on the challenges and issues present during the delivery of the machine learning models. In this paper, it has been shown the similarity between DevOps and MLOps during the deployment stage.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

1. Alla S, Adari SK. What is mlops? in Beginning MLOps with MLFlow. Springer. 2021;79–124.
2. Renggli C, Rimanic L, Gurel NM, Karla B, s, Wu W, Zhang C, Zurich E. A data quality-driven view of mlops. 2021;2. [Online]. Available:<https://arxiv.org/abs/2102.07750v1>
3. Ruf P, Madan M, Reich C, Ould-Abdeslam D. Demystifying mlops and presenting a recipe for the selection of open-source tools. Applied Sciences. 2021;11(9):8861. [Online]. Available:<https://www.mdpi.com/2076-3417/11/>

- 19/8861/htmhttps://www.mdpi.com/2076-3417/11/19/8861
4. Klaise J, Looveren AV, Cox C, Vacanti G, Coca A. Monitoring and explainability of models in production. 2020;7. [Online]. Available:https://arxiv.org/abs/2007.06299v1
  5. Tamburri DA. "Sustainable mlops: Trends and challenges," Proceedings - 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC. 2020; 9:17–23.
  6. Praveen Gujjar, Prasanna Kumar HR. Sentimental analysis for running text in email conversation. International Journal of Computer Science and Engineering (IJCSE). 2020;9(4):67-68.
  7. Manjunatha T, Praveen Gujjar. Performance analysis of indian information technology companies using DuPont model. IUP Journal of Management Research. 2018; 17(4):7-14.
  8. Gift N, Deza A. Practical MLOps: Operationalizing machine learning models. O'Reilly Media, Inc; 2020.
  9. Treveil M, Team D. Introducing mlops how to scale machine learning in the enterprise. 2020;185. [Online]. Available:https://www.oreilly.com/library/view/introducing-mlops/9781492083283/
  10. Adari SASK. Beginning mlops with mlflow deploy models in aws sagemaker, google cloud, and microsoft azure; 2021. [Online]. Available:https://doi.org/10.1007/ 978-1-4842-6549-9
  11. Armbrust M, Das T, Sun L, Yavuz B, Zhu S, Murthy M, et al. "Delta lake," Proceedings of the VLDB Endowment. 2020;13(8):3411–3424. [Online]. Available:https://dl.acm.org/doi/abs/10.14778/3415478.3415560
  12. MLflow. MLflow. Available:https://mlflow.org/ (Accessed on 10 February 2022).