

# MACHINE LEARNING INTEGRATION WITH DEVOPS

Surbhi Maam

School of electronics and electrical engineering  
Lovely professional university jalandhar, India

Rededdy vijay Kumar reddy

School of computer science engineering  
Lovely professional university jalandhar, India

Ravipati Mohan Venkatesh

School of computer science engineering  
Lovely Professional University Jalandhar, India

Shikari Sairaj

School of computer science engineering  
Lovely professional university jalandhar, India

Ravipati Satya Sai Rama Krishna

School of computer science engineering  
Lovely Professional University Jalandhar, India

## **ACKNOWLEDGEMENT:**

This entire project was successfully Completed with the valuable support and Guidance of Surabi Ma'am We truly Grateful to have their support and guidance for completion of the work we are thankful to Lovely Professional University Phagwara for their support during the working period.

## **ABSTRACT:**

This paper presents devops automation for machine learning. In Machine Learning or Deep Learning, data scientists must to change the model several times to seek out the most effective accuracy model manually. This took a plenty of your time, manpower and resources for making a machine learning or deep learning model precisely. This paper describes How to integrate Devops automation tool Jenkins Continuous integration (CI)/Continuous Deployment (CD) build pipeline and container technology Docker with Machine Learning to minimize the cost time manpower and resources of data scientists, MLops is a combination of machine learning and operations which is used to collaborate and communicate between data scientists and operations professionals to help managed production life cycle.

## **KEYWORDS:**

Machine Learning, Devops, CI, CD, Pipeline, Jenkins, Docker .

## **INTRODUCTION:**

In Data science world Machine learning is core approach to resolve the important life problems today's big data world ML plays vital role to produce best solutions to the customer Now a days Machine Learning is buzz word but why the proportion of AI models created but not put in production environment is quite 90% with massive investment in data science teams platforms and infrastructure the quality of ai projects is dramatically increasing together with the no of missed opportunities unfortunately most projects not showing the business needs business introducing new risks that require to be managed single technology won't help to unravel the issue today industry need machine learning with devops . MLops is solution to any or all problems in AI world .MLops delivers the capabilities that data science and it ops teams must to work together to deploy monitor and manage machine learning model in production environment and to manipulate their use in production environment MLops brings the most effective of repetitious development involved

in training machine learning models and scalable and manageable model deployment it's currently missing puzzle within the enterprise AI world

Machine learning operations (MLOps) is DevOps for machine learning processes Mlops helps data scientists to cooperate and increase the speed of delivery and quality of development models through monitoring affirmation and governance of machine learning models. MLOps supports the data science life cycle even as DevOps supports the application development life cycle.

As such, MLOps is predicted on the principles of DevOps. The goal of adopting MLOps is to push more efficient experimentation that will result in faster development and deployment of models.

Devops is a combination of Development and operations methodology that brings developers and operators work together DevOps life-cycle devops automate one or more phases of devops life cycle Coding, building, testing, releasing, configuring, and monitoring.

**Coding phase:** this phase include development of code developer write the code push the code into source code management tools like github, gitlab etc.

**Building phase:** This phase automatically create environment in step with the wants as an example team create Docker image for particular Machine learning model

**Testing phase:** This phase includes continuous testing tools, team develop separate environment for testing the code and it gives the standard of the code various testing techniques includes in this phase. They are

1. Unit testing
2. Integration testing
3. Configuration testing
4. Performance testing

**Releasing phase:** this phase consists of releasing plan and automation. The team will decide how to release the product into the market.

**Configuring phase:** this phase consists of automatic configuration management with the help of infrastructure as code team will create the scripts according to the requirement script automatically launch everything. As an example they launch running environment, data bases etc.

**Monitoring phase:** this phase continuously monitors the product whenever production environment goes down it notifies the developer it tells about what number users are using the merchandise at particular point of your time.

### **Pipeline:**

To achieve CI/CD developers create "pipelines", which demonstrate how to automatically build, test, release and configure software release. The pipeline as an event that tells how to execute the steps in sequential manner. If any step fails it will stop entire pipeline and it gives feedback to the developer. CI/CD supports small commits and large number of releases per day.

Phase	Tools
Build	Docker, Gradle.
Test	pytest, Cypress.
Release	Jenkins, Flux.
Configure	Terraform, Ansible.
Monitor	Sentry, Prometheus.

Creating CI/CD pipelines consists of different tools that are created to solve different tasks of the devops lifecycle. Tools that help to run CI/CD pipeline on host OS are Github actions and Travis, tools that help to run CI/CD pipeline on own machines are Jenkins or argoCD. These tools provide a configuration language to run steps and tools as a pipeline. For every single release, there is a master plan to roll back the previous version in case of an updated version fails. An easy solution for roll back is running the previous version through CI/CD pipeline.

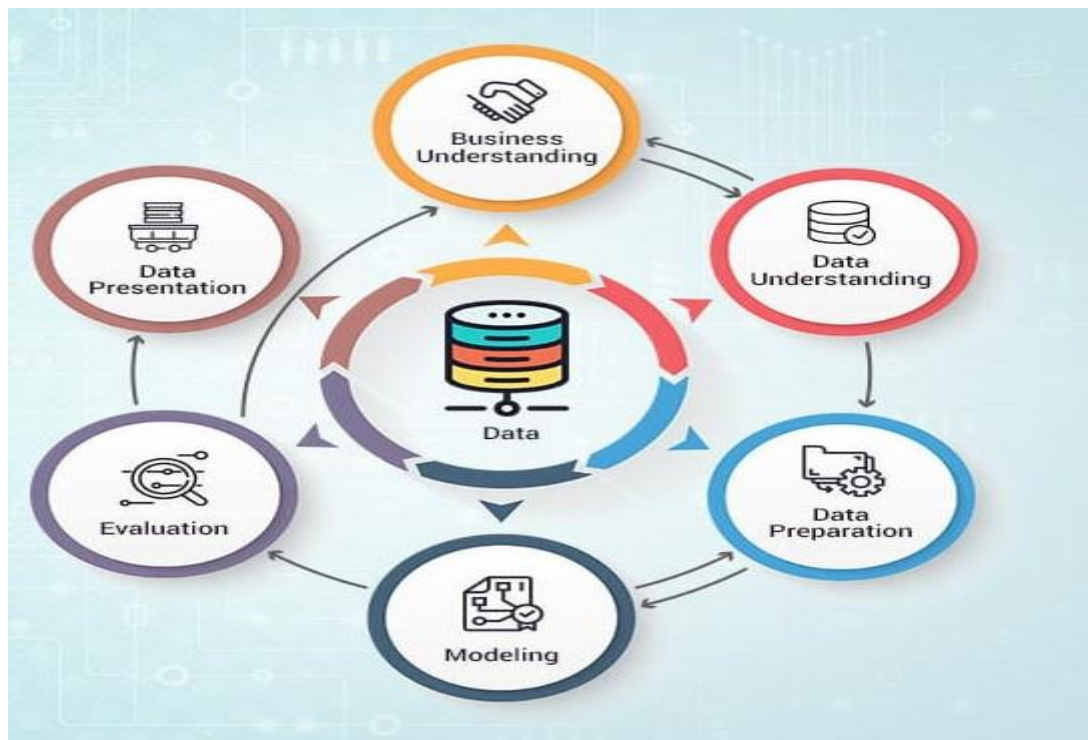
A container is built from an image which contains a set of instructions to build the container with the help of Docker file. We can create a container image, an image contains operating system and software to run particular services. A container runs in a container runtime, a process which manages the lifecycle of a container. With the help of container technology, we can launch the operating system in one second. Unexpectedly, if the production environment goes down, container technology launches the environment quickly as compared to virtualization. In dynamic and large scale environment, we have to change machines and services frequently. Virtually, it is impossible to provide high availability. Container orchestration provides highly-available and robust service systems.

## **LITERATURE REVIEW:**

In literature, we can find studies which show how to apply machine learning with devops. Machine learning with operations (MLOPS) using tools and platforms to automate machine learning models. The authors develop continuous integration for machine learning to allow users to provide a wide range of integration in order to achieve continuous improvement. One of the most critical challenges are data collection, data extraction, and data cleaning. Authors create an application life cycle model based on MLOPs to optimize the manufacturing process. The study shows how to apply Devops CI/CD pipelines with machine learning applications. Several researchers Karamitsos, Virmani, Erich have agreed that Agile transformation is essential to improve the efficiency of the companies. MLOps fills the gap between business users and development teams. They highlight the devops principles and guidelines to adapt continuous integration and continuous deployment, which results in an increment in the development process and improve the quality. Many applications using machine learning techniques now a days, but which is not enough to produce great results when we integrate with devops. It gives great efficiency in any field like healthcare, safety, etc. The main goal of devops is to create cross functional teams. Both operational and development tools work together. The entire goal of devops is to improve the business value in IT industry, it produces best results in agile world with the help of continuous integration and continuous deployment. Devops breaks the hurdle between operations and development and they collaborate both machine learning and devops, Machine learning+Devops (MLOPS).

## **METHODOLOGY:**

The industry Standard Process for Data processing(CRISP-DM)could be a process model with six phases that describes the data science life cycle.it helps to implement machine learning projects It is published in 1999 it's one among the foremost common methodology for data science projects.



CRISP DM Methodology has six steps

**Business understanding phase:** during phase specialized to understand the objectives and needs of the project.

**Data understanding phase:** This phase contains initial data for experimental analysis this phase focus identify collect and analyses the information to succeed in the project goal

**Data preparation phase:** during this phase data preparation takes place steps include in this process are feature extraction, data cleaning, data reduction, data selection, and transformation.

**Modeling Phase:** This phase select the machine learning model based on the requirement.

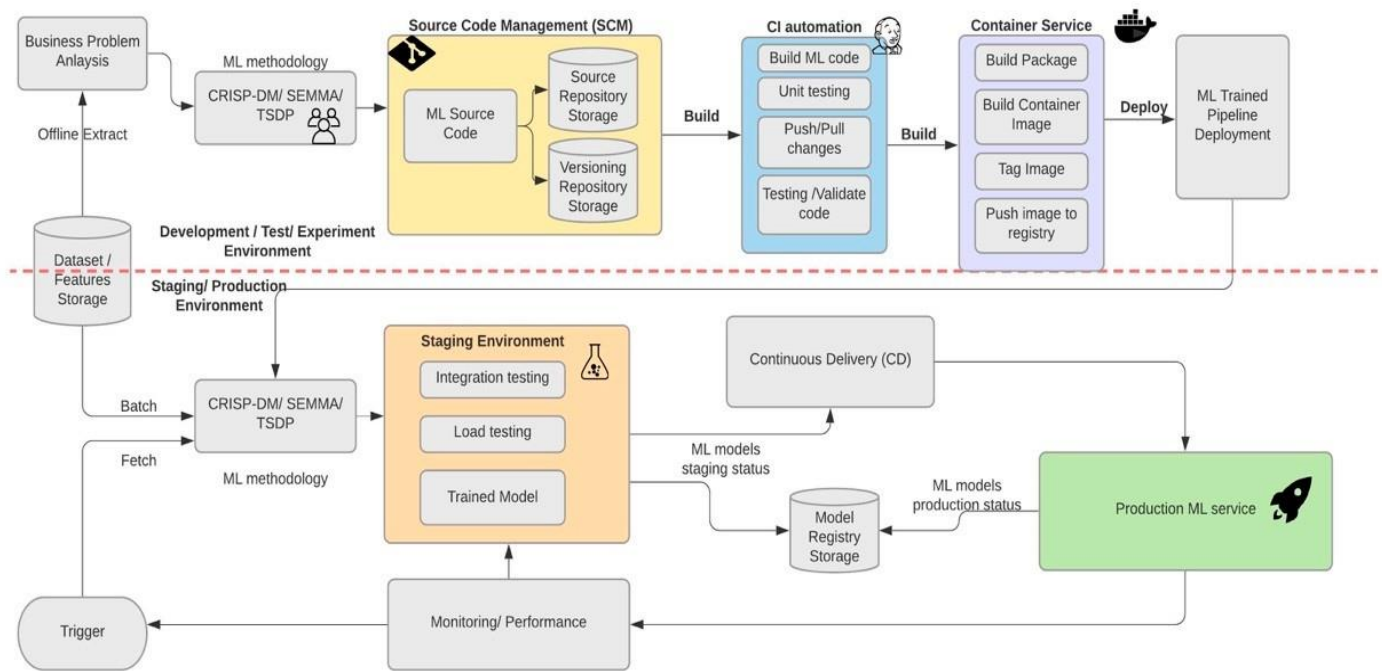
**Evaluation Phase:** This phase looks which model connects the business requirement. And also a review decide whether the business requirement is achieved or not.

**Deployment Phase:** This final phase contains four tasks

1. Plan deployment.
2. Plan monitoring and maintenance.
3. Produce final report.
4. Review project.

### Machine learning automate pipeline with CI/CD

The main purpose of this approach is continuous training and testing the machine learning model with the help of pipeline the term mlops combines various principles of Continuous integration and continuous deployment to automate the machine learning pipeline The process of automatic model retrain in machine learning is possible with the 2 elements of devops such as 1) continuous integration 2) continuous deployment the automated CI/CD helps the data scientists in feature engineering model architecture and hyper parameters data scientists implement lot of things like building the model ,push the code into github, build deployment environment etc.



Main elements required for the implementation of machine learning CI/CD pipeline

- 1) Business problem analysis
- 2) Dataset features and storage
- 3) ML methodology
- 4) CI components
- 5) CD components
- 6) Automated ML triggering
- 7) Model registry storage
- 8) Monitoring and performance
- 9) Production ML service

First of all we understand the business problem and How to solve this business problem using technologies dataset features and storage collect the features related to problem, In machine learning creating model is not big deal collecting right features is major problem in today's machine learning world accuracy always depends on the right features.

**Machine Learning methodology:** This is the major step in mlops pipeline during the analysis of data we use the methodology CRISP methodology is one of the best and common methodology for Automated machine learning CI/CD pipeline

**Continuous integration:** In this step build the code run the machine learning models. outcomes of this stage is pipeline components In machine learning creating the code is not big deal but training the machine learning models requires a lot of resources and manpower with the help of continuous integration we can do this things with minimum resources and minimal efforts

Requirements for Continuous integration

- 1.Source code management tools like Git, github etc.
- 2.Insertion/deletion takes place in repository automatically update and deploy in production environment
- 3.Run the machine learning code



4. Test and validate the code
5. Build the container image
6. push the container image into repository

**Continuous delivery:** outputs of Continuous integration pipeline components deployed in the production/staging environment the output of this step is testing machine learning model. Jenkins is one of the most famous open source tool for continuous deployment it is written in java language this tools works with SCM tools like github

The components of continuous deployment

- 1) Production environment:** First we have to push machine learning model into the production environment it contains all the required dependencies to run that machine learning code output of this stage is testing machine learning model.
- 2) Model registry storage:** output of production environment is testing machine learning we have to put that machine learning model into the model registry storage.
- 3) Automated trigger:** If there is any change in code or commit in repo trigger automatically fire and they will initiate new build output of this model updated machine learning model pushed in the production environment.
- 4) Performance monitoring:** This is one of the most important step for data scientists performance monitoring contains lot of considerations like how much resources used, efficiency of the model, throughput, and features available to the users.
- 5) Monitoring resources:** In this step continuously observe the resources of the system like RAM CPU storage etc. lot of resource monitoring tools available in the market like zabbix windows task manager.

## CONCLUSION:

In this paper we present Machine learning integration with Devops CI/CD principles to improve the performance and business values .we present use case of Machine Learning methodology to solve the industry problem CRISP CM is one of the best machine learning methodology used in Data science projects. We found that manual machine learning needs a lot of resources and manpower and it lags the business organization. Automated machine learning with devops improves a lot of things in today's data science industry it produces great results in business, marketing and also it produce less waste as compared to manual machine learning. It makes deployment and integration easier.

Machine learning model lifecycle is different from actual software development .it requires a lot of things data collection data cleaning, feature selection, setup environment after deployment monitoring and maintenance takes place ,Machine learning alone can't do these things effectively, Devops provide continuous integration and continuous deployment principles for these type of problems. development and operational teams are working in many areas like manufacturing industry, marketing industry ,healthcare industry but they are not achieve greater efficiency because they are not integrate machine learning with devops continuous development. This study shows that integration of machine learning with devops (MLOPs) In this case both development and operational team work together to optimize the process and produce great results in data science industry

## REFERENCE:

- [1] A research paper on Application of Devops in the improvement machine learning process
- [2] A research paper on Applying DevOps Practices of Continuous Automation for Machine Learning
- [3] Beginning MLOps with MLFlow Deploy Models in AWS SageMaker, Google Cloud, and Microsoft Azure by Sridhar AllaSuman and Kalyan Adar
- [4] Data Science Solutions on AzureTools and Techniques Using Databricks and MLOps by Julian Soh and Priyanshi Singh

[5] Master's thesis Master's Programme in Data Science Designing an open-source cloud-native MLOps pipeline by Sasu Mäkinen March 12, 2021

[6]<https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning>

[7]<https://www.infoworld.com/article/3271126/what-is-cicd-continuous-integration-and-continuous-delivery-explained.html>

[8] <https://www.datascience-pm.com/crisp-dm-2/>

[9] <https://en.wikipedia.org/wiki/DevOps>

[10] <https://www.docker.com/resources/what-container>

[11] <https://www.redhat.com/en/topics/devops/what-cicd-pipeline>

[12] <https://www.jenkins.io/doc/>

