MACHINE LEARNING AUTOMATIONS PIPELINE WITH CI/CD

Surabhi jhariya[1], Kajal Gayakwad[2], Abhilasha Chandel[3], Bhuvneshwari Ninave[4], Vishakha Yadav[5], Guide: Prof. Vinay Sahu[6]

[1-6]Dept. of computer science Engineering, SHRI BALAJI INSTITUTE OF TECHNOLOGY AND MANAGEMENT BETUL, MP
---***---

ABSTRACT: Automated machine learning (AutoML) spans a fairly wide range of functions that can be thought of as being included within a machine learning pipeline. An AutoML "solution "may include tasks such as data preprocessing, feature engineering, algorithmic selection, algorithmic architecture search, and hyperparameter tuning, or some subset or variation of these specific tasks. Thus, automated machine learning can now be considered anything from performing a single task, such as automated feature engineering, through a fully automated pipeline, from data preprocessing to feature engineering, to algorithm selection.

It may take your valuable time, manpower and resources to build a machine learning or deep learning model properly. This paper prepares automation for machine learning and in machine learning or deep learning, it involves manually changing the model several times in search of a model with the most effective accuracy. This paper describes how to integrate devops automation tools Jenkins Continuous Integration (CI) / Continuous Deployment (CD) build pipeline and container technology with Docker Machine Learning to reduce the cost time manpower and resources of data scientists, MLops is a combination of machine learning and operations that is used to collaborate and communicate between data scientists and operations professionals to help manage production life cycles.

**Key Words**:   Machine Learning, CI, CD, Pipeline, Jenkins, Docker, Git.

1. INTRODUCTION

traditional programming or software engineering in diploma are monitoring is quite iterative traditional programming or software engineering in deployment process in was developed in developed basically you have one as the development phase and the other is the operation space in in the development phase 2 plan code build and test and then move on to continuous integration face very new release deploy operate and monitor and then comes a continuous feedback loop and this look continuous for devops.
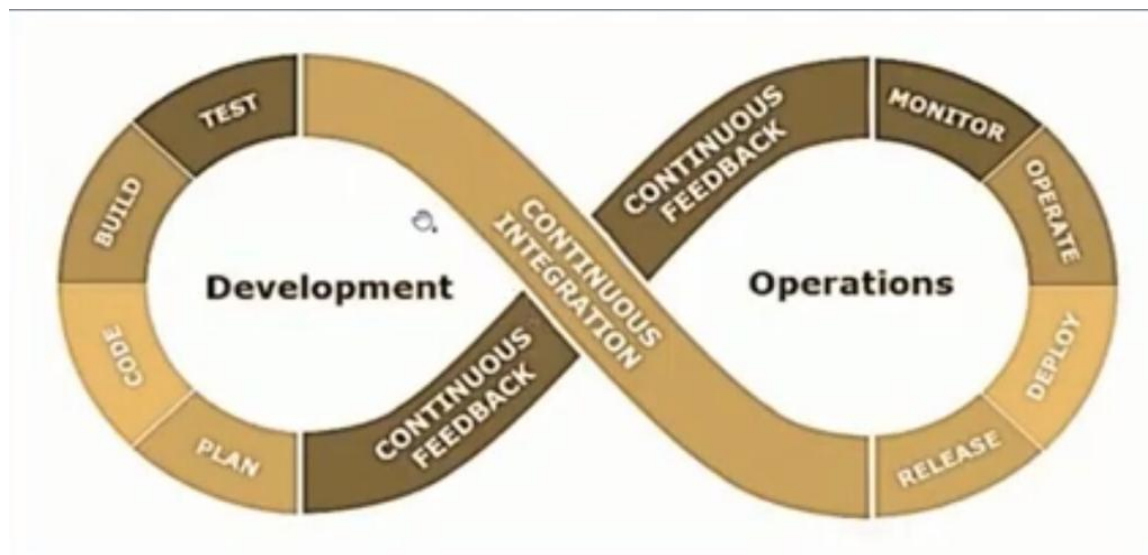
now when it comes to machine learning are monitoring is quite iterative traditional programming or software engineering in deployment process in was developed in developed basically you have one as the development phase and the other is the operation space in in the development phase plane code build and test and then move on to continuous integration face very new release deploy operate and monitor and then comes a continuous feedback loop and this look continuous for devops now

when it comes to machine learning and industry ml lifecycle followed by most projects with an industry where in the first step is your business understanding you project and objective lete your stakeholder comes to you and they say that the required something in the in any x y z domain letter for example Automobile two men they want and output from the regenerated by the car then you understand the business objective from the from your stakeholder and pieces that understanding you try and gather data or else if you have a set of predefined set of data already you perform data understanding in that you explore the data and you perform data selection a set of predefined set of data already you perform data understanding in that used the data and you perform data selection wherein you choose which of these important and useful for you then once you are done with that state you go for data preparation very in feature engineering comes into existence due to heavy feature engineering to prepare your data to be paid by the model in the modelling face what you do is you predict potential customers or any other use cases using your ml algorithm and post for the link that comes and algorithmic evaluation where in test your order as well as Matrix fast for the model evaluation that comes and l letter fees if you are not happy with the evaluation of you see that there are scope of improvement you again back to the business understanding please read in urethra the process and go back and see if there are any additional requirements or if you have understood the previous requirement collect or not and then follow the for the steps which is data in the Senate in the preparation Modelling and then again come to evaluation and see if you are final evaluation is is desired and and this is what you wanted or not and then post that once you're happy with the result you go for the deployment faced when you perform a user acceptance testing and other testing automation suppressive.

Industry ML Lifecycle -   DM Framework
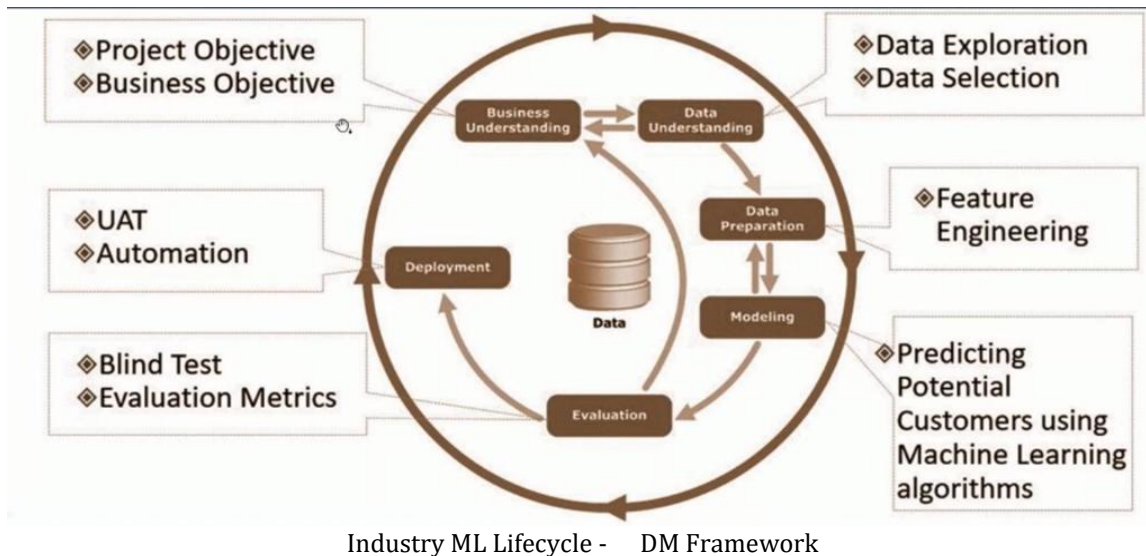
2. LITERATURE REVIEW

In literature we can find studies which shows Devops CI/CD pipelines with machine learning applications   DevOps is a well-established set of practices based around Continuous Integration, Continuous Deployment (CI/CD) and infrastructure. DevOps tools are famous for making software development projects to speed up their work in developments.machine learning uniquely handles the volume, velocity and variety of data that is generated using new delivery Process and using the next generation of atomized Composable, and scaled out applications. DeVops tools also allow engineers to complete different tasks Modern machine learning algorithms and frameworks make it   easy to develop models that can make accurate prognosis. MLOps lives within a CI/CD framework Advocated by DevOps as a proven way to roll out quality code updates at frequent intervals. However, machine learning expands the integration stage with data and model validation, while delivery addresses complexities of machine learning deployments.The primary focus of DevOps is to create a sustainable infrastructure for applications that make them highly scalable. DevOps can use ML to analyze normal patterns user volumes, resource utilization, transaction throughput etc.Nowadays, Machine learning plays a great role in analytics. These tools generally process & identify threats.The main focus of DevOps to improve value in   IT industries.

Cloud technology gives recourses and infrastructures for the development, deployment of applications and software, creating a runtime environment in a minimal time. Hence cloud acts as a facilitator for continuous development   and   integration which   makes   the   resources   available   as   needed   for   the   general application life cycle management.   With cloud computing companies don't have to worry about   infrastructure   thus   any   tool   needed   for development can be acquired on time hence it fastens the development process and facilities delivery of the products fast and in time. As well as

the updates can be done fast and efficiently. DevOps is made simpler with the availability of cloud recourses such as the team can collaborate.

2.1 Docker

   Docker is a tool which create virtual machine. Docker written in 'go' language. It is a tool that perform operating system level virtualization, also known as contanerization. It is a set of platform as a service that was os level virtualization whereas VMware was hardware level virtualization. Docker was container on the host os to run application it allows applications to use some Linux Kernal as a system on the host computer rather then creating a whole virtual os. It is mainly use to deploy web server.Docker containers are used to wrap a piece of software in a complete system that contains everything needed to run code, runtime, system tools, system libraries anything that can be installed on a server using cloud Technologies.
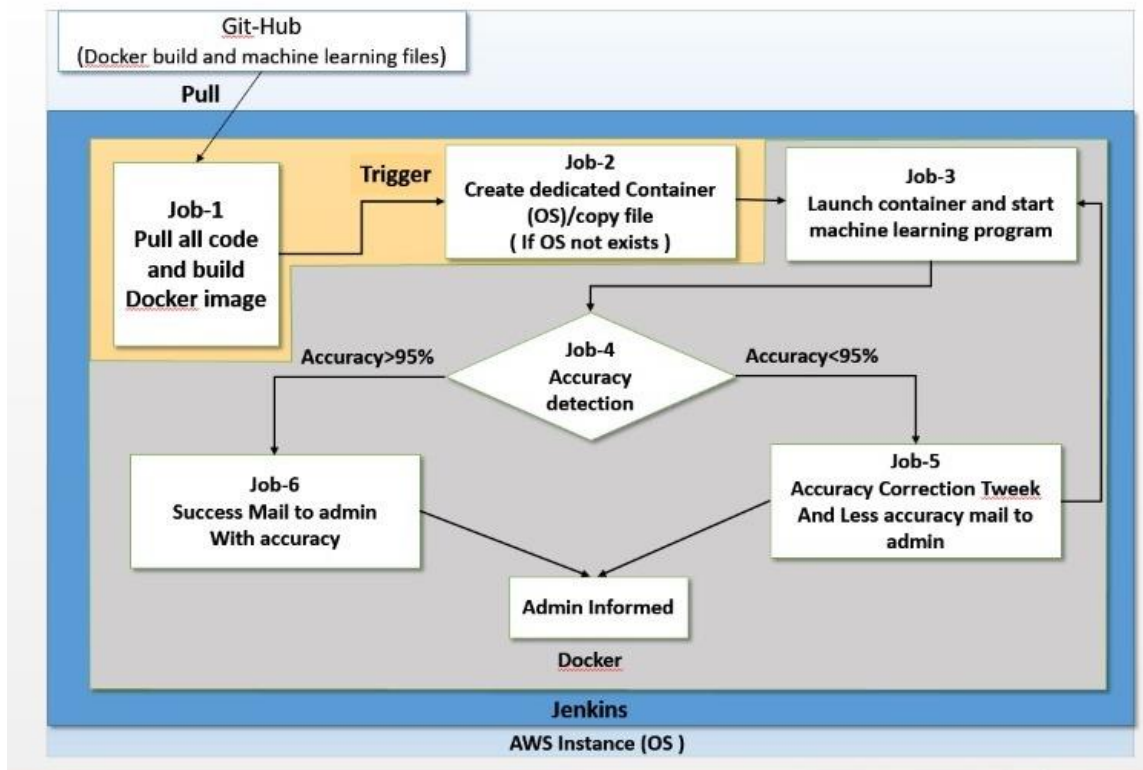
2.2 Jenkins

Jenkins is an open-source automation tool written in Java. Jenkins achieves Continuous Integration with the help of plugins. Plugins allow the integration of Various DevOps stages. If you want to integrate a particular tool, you need to install the plugins for that tool.It is a server-based application and requires a web server like Apache Tomcat.

2.3 Github

Github is a code hosting platform for collaboration and version control. github is a website and cloud based service that helps developers to store and manage their as well as track and control their code.It is a web based platform used for version control. It provides a web based graphical interface and also provide access control and several collaboration feature.

3. METHODOLOGY

First there will be git hub account where all the data will be stored then job 1 will run in jenkins and download the complete data and create a proper image file like job 1 will complete jenkins one then that will trigger job 2 in which jo is downloaded There is a docker file in it from git hub, with the help of that docker file an image will be created here job 1 will be linked to job

2 whose job will be to launch the operation system related to machine learning and copy the code which is uploaded from git hub   As soon as this process is completed then job 3 will be triggered from job 3 of machine learning training model build and as soon as job 3 model is completed then it will trigger job 4 there will be accuracy deletion from job 4 when accuracy is deleted then that   One condition remains if accuracy is above given limit then model will be successfully built and if accuracy is below given limit then job will fail and then after accuracy deletion process of accuracy correction will be done in which python file will run.

CRISP-DM is a process made up of six different phases. These include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment.

Business understanding phase: Bussiness understanding 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

Continuous Integration:-Continuous integration (CI) is practice that involves developers making small changes and checks to their code.

Requirements for Continuous integration :-

1. Source code management tools like Git, github etc.

2. Run the machine learning code

3. Test and validate the code

4. Build the container image

5. 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

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 decencies 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.

4. CONCLUSION

We present machine learning integration with devops CI/CD to improve the performance. CRISP-DM is one of the best machine learning methodology for the data to solve industries problem also we found that the machine learning needs a lot of resources and manpower automated machine learning with the help of devops to produces good results in bussiness, marketing and many industries also waste as compared to manual machine learning. Machine leaning makes integration deployment easier. It will be improve the business value. Machine leaning model life cycle is different from actual software development. It requires a lot of data collection, data cleaning, feature selection, devops provid continuous integration and continuous deployment. In this learning development and operational team are working in many areas like manufacturing, marketing, healthcare and many others industries but they are not achieve greater efficiency because they are not integrate machine learning with devops continuous deployment. In this type of study shows that the integration of machine learning with devops both are the case development and operational team work together process and produce a very good result in data science industry. Machine learning is useful also helpful for future uses.

5. REFERENCES

[1] G. Bou Ghantous and Asif Gill, "DevOps: Concepts practices tools benefits and challenges", PACIS2017, 2017.

[2] DevOps provides practices to bridge the gap between 'Dev' and 'Ops' and improve team collaboration [1], [2] in the overall context of agile software development [15], [20].

[3] In agile software development context, the objective of DevOps approach adoption is to improve cooperation and between Dev and Ops [2], [11].

[4] Georges Bou Ghantous and Asif Qumer Gill, "DevOps Reference Architecture for Multi-Cloud IoT Applications", 2018 IEEE 20 th Conference on Business Informatics (CBI), 2018, [online] Available:

[5] DevOps provides practices to bridge the gap between 'Dev' and 'Ops' and improve team collaboration [1], [2] in the overall context of agile software development [15], [20]

[6] The mentioned concepts could be achieved by applying DevOps practices and using DevOps tools [1], [9].

[7] https://en.wikipedia.org/wiki/DevOps

[8] https://www.docker.com/resources/what-container

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

[10] https://www.jenkins.io/doc/

[11] https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-

pipelines-in-machine-learning.

BIOGRAPHIES

1'st Surabhi Jhariya

2nd Kajal Gayakwad

3rd Abhilasha Chandel

4th Bhuvneshwari Ninave

5th Vishakha Yadav