Internship at Autodesk

"Tracking Model Training with Mlflow"

Submitted by:

Neeli Krishna Dheeraj **PES1201800182**

Under the guidance of

Gopi Krishna Nuti
Lead Data Scientist,
Autodesk

6 months

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

Date: 15 April 2021

Mr. Neeli Krishna Dheeraj
Room No. 608
New Block
PES Boys Hostel
PES University
Banashankari III Stage (560085)
Bangalore, Karnataka, India

Dear Mr. Neeli Krishna Dheeraj

On behalf of Autodesk India Private Limited, I am pleased to offer you temporary association in our College Internship Program. Your position will be **"Intern Data Science" (Autodesk Construction Solutions)**. In this position, you will report directly to **Mr Gopi Nuti, Lead Data Scientist.** You will be based at our **Bangalore office**, and will take instructions from your manager. If you accept this offer, your start date will be **Tuesday, 1 June 2021** and subject to termination provision of this Letter, your internship will end no later than **Tuesday, 30 November 2021**.

You shall receive a fixed amount of **INR. 30,000 per month** (subject to applicable tax). Unless authorized in advance by Autodesk, your weekly schedule may not exceed 40 hours per week. Because you will be an Intern, you will not be entitled to any employment benefits that Autodesk offers to its regular employees. However, you will receive abbreviated benefits, which are statutorily required leaves plus sick leave. More information will be provided to you at your orientation.

You will be asked to sign standard agreements requiring you to hold in confidence any proprietary information received as an Autodesk Intern. Further, you will be required to comply with the Autodesk's Code of Conduct and all other HR policies, which you will be advised on upon joining us.

We wish to emphasize that you should not bring with you any confidential or proprietary material of a third party including from your college or any former employer, nor violate, in any way, previous obligations regarding confidentiality at another firm. This letter constitutes a contract for a fixed term of six months. Please note that in case of breach of any company policies, misconduct or unethical behaviors, your internship can be terminated with immediate effect.

Your Internship with Autodesk is conditioned upon your being **currently enrolled as a full-time student in good academic standing pursuing a Bachelors, Masters or PHD level degree from an accredited institution**. As a condition of your internship, you may be required to show Autodesk proof that you meet these criteria.

U.S. Export Controls Laws require that Autodesk obtain a U.S. government export license prior to employing certain persons. The Company determines if a license would be required by reviewing the Export Controls Questionnaire completed by the employee and the results of an abridged background check. This offer is contingent on your timely and fully completing the Questionnaire and the background check. In

AUTODESK. Make anything.

19 July 2021

Neeli Krishna Dheeraj
Room No. 608, New Block
PES Boys Hostel, PES University
Banashankari III Stage (560085), Bangalore, Karnataka, India

Dear Neeli Krishna Dheeraj,

Re: Addendum to Internship Agreement

This letter "Addendum" will serve to update the terms and conditions of your internship ("Internship Agreement"), with Autodesk Asia Pte. Ltd. ("Company") dated **15 April 2021** to reflect the change/s detailed below.

You and Autodesk agree to the following update/s to the Contract:

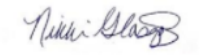
1. Position & Commencement

Paragraph 1 of the Contract shall be replaced in its entirety with the following:

On behalf of Autodesk India Private Limited, I am pleased to offer you temporary association in our College Internship Program. Your position will be "**Intern Data Science" (Autodesk Construction Solutions)**. In this position, you will report directly to **Gopi Krishna Nuti, Lead Data Scientist**. You will be based at our **Bangalore office**, and will take instructions from your manager. If you accept this offer, your start date will be **Tuesday, 1 June 2021** and subject to termination provision of this Letter, your internship will end no later than **Friday, 17 December 2021**.

All other provisions of the Internship Agreement shall remain unchanged.

Warm Regards,

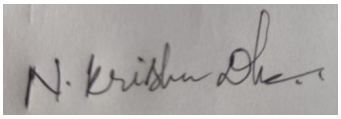


Nikki Glasgo
Global Head of Talent Acquisition
For and on behalf of Autodesk Asia Pte. Ltd.

DECLARATION

We hereby declare that the project entitled **“Tracking model training with Mlflow”** has been carried out at Autodesk Construction Solutions by me under the guidance of Mr. Gopi Krishna Nuti, Lead Data Scientist and submitted in partial fulfillment of the credits for the degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester 1st June to 17th December (7th semester). The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES1201800182 Neeli Krishna Dheeraj



ACKNOWLEDGEMENT

I would like to express my gratitude to my guide Mr. Gopi Krishna Nuti, Lead Data Scientist , Autodesk India Private Limited for his continuous guidance, assistance and encouragement throughout the development of this project.

I am grateful to the internship coordinator Prof. Preethi P, Dept. of Computer Science and Engineering, PES University for organizing, managing and helping out with the entire process.

I take this opportunity to thank Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department.

I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, ViceChancellor, PES University for providing to me various opportunities and enlightenment every step of the way.

Finally, this internship could not have been completed without the continual support and encouragement I have received from my parents and my friends.

During the 6 months of my internship as an Intern in Data Science (ACS) at Autodesk India Private Limited, I gained invaluable experience in technology and corporate culture, which not only groomed me professionally but personally as well. I am thankful to my company for providing me the opportunity to be part of their well organised internship program. I worked in the internal intelligence team. This internship involved a training session for the initial 2 weeks which included technical and professionalism/work culture courses from MyLearning platform and through LinkedIn Learning courses. I was tested on these concepts by asking us to take quizzes on each topic. Once this was finished, I was allocated my project which involved integrating ML models with mlflow, setting up mlflow tracking server in aws, tracking ML model runs in remote mlflow server, and running training jobs in sagemaker with mlflow tracking.

During the project implementation, I was required to learn new technologies and set deadlines. An Agile methodology approach was adopted for the duration of this project. This involved Sync-up calls and bi-weekly meets. The project was built, tested and deployed during these 6 months. It required the use of various services in AWS and mlops tool mlFlow.

I had an amazing learning experience. The friendly work-culture, the support and encouragement provided by everyone at Autodesk especially my

manager, mentor and team-mates added value to my learning. I can proudly and successfully say that I have made new friends and have gained one of the most important initial experiences as a budding software engineer would need to help boost his career with a huge positive impact.

Table Of Contents

Introduction

Brief Introduction about Autodesk

Autodesk is a multinational software corporation that makes software products and services for the architecture, engineering, construction, manufacturing, media, education, and entertainment industries. Autodesk is headquartered in San Rafael, California, and features a gallery of its customers' work in its San Francisco building. The company has offices worldwide. Its U.S. locations are in the states/commonwealths of California, Oregon, Colorado, Texas, Michigan, New Hampshire and Massachusetts. Its Canada offices are located in the provinces of Ontario, Quebec, and Alberta. Around 11,500 and more employees worldwide work at Autodesk. Autodesk aims at changing how the world is designed and made. It spans its resources everywhere to solve challenges big and small. From greener buildings to smarter products to mesmerising blockbusters, Autodesk software empowers innovators to design and make a better world for all. At Autodesk, we don’t believe in waiting for progress, we believe in making it. The software that is made provides customers with the right tools to work, the ability to think flexibly, and the

power to transform what actually needs making. Autodesk became best known for AutoCAD, but now develops a broad range of software for design, engineering, and entertainment and a line of software for consumers. The manufacturing industry uses Autodesk's digital prototyping software including Autodesk Inventor, Fusion 360, and the Autodesk Product Design Suite - to visualise, simulate, and analyse real-world performance using a digital model in the design process. The company's Revit line of software for building information modelling is designed to let users explore the planning, construction, and management of a building virtually before it is built. Autodesk's Media and Entertainment division creates software for visual effects, colour grading, and editing as well as animation, game development, and design visualisation. 3ds Max and Maya are both 3D animation software used in film visual effects and game development. In summary, Autodesk offers the following main products: AutoCAD, Autodesk Construction Cloud, Civil 3D, Fusion 360, InfraWorks, Inventor, Maya, Revit, ShotGrid and 3ds Max.It is Ranked fourth on Barron’s 2021 list of the 100 most sustainable companies and received a perfect score on the HRC Corporate Equality Index in 2021. Autodesk’s FY 2021 Financial Highlights are that total revenue increased 16 percent to $3.79 billion, total billings decreased 1 percent to $4.14 billion, total subscription base grew to 5.27 million users, up 8 percent year over year and current remaining performance obligations increased 16 percent to $2.74 billion.

Internship Project Details

Project Title: Tracking Model training with MLflow

Is it part of another bigger project?

Yes, this is part of ACS intelligence team's MLOps pipeline for continuous training and model tracking .

Project Introduction

This project falls under the jurisdiction of **AUTODESK CONSTRUCTION SOLUTIONS** intelligence team. The project is about the ability to track model training and means to trigger training a new model when new data is available. This project makes the newly available labelled data available for training new model .

Project design details with technologies used

About the Tech used:

This project entailed the use of technologies like **python** as coding language , mlflow services like **mlflow tracking , mlflow ui, mlflow models, mflow model registry** for training and tracking the runs and storing artefacts and aws services like **aws ec2, aws ecr, aws s3 , aws fargate , aws vpc , aws iam, aws cloudformation , aws sagemaker training jobs** for creating mlflow deployment server.

Docker services like **docker images , docker containers** to create images , shell scripting to create **DockerFile** . **Html , css, javascript ,**

bootstrap for front end user interface to run training job and **Flask** to run the web application. **Git** for continuous integration.

To elaborate on the technologies more, following is a brief summary:

Python: Programming language is used for api integration of mlflow and in cloudformation to deploy mlflow server.

Mlflow tracking: All the mlflow runs are tracked , artefacts are stored using Mlflow tracking.

Mlflow UI: The mlflow ui is the user interface to show all the runs and run data which are organised as experiments.

Mlflow Models: These are the models that are saved in each run can be stored as artefact.

Mflow Model Registry: The final ml models used for serving are saved in mlflow Model Registry which also enables versioning of models .

Aws Ec2: Aws ec2 is used to serve the mlflow ui from aws.

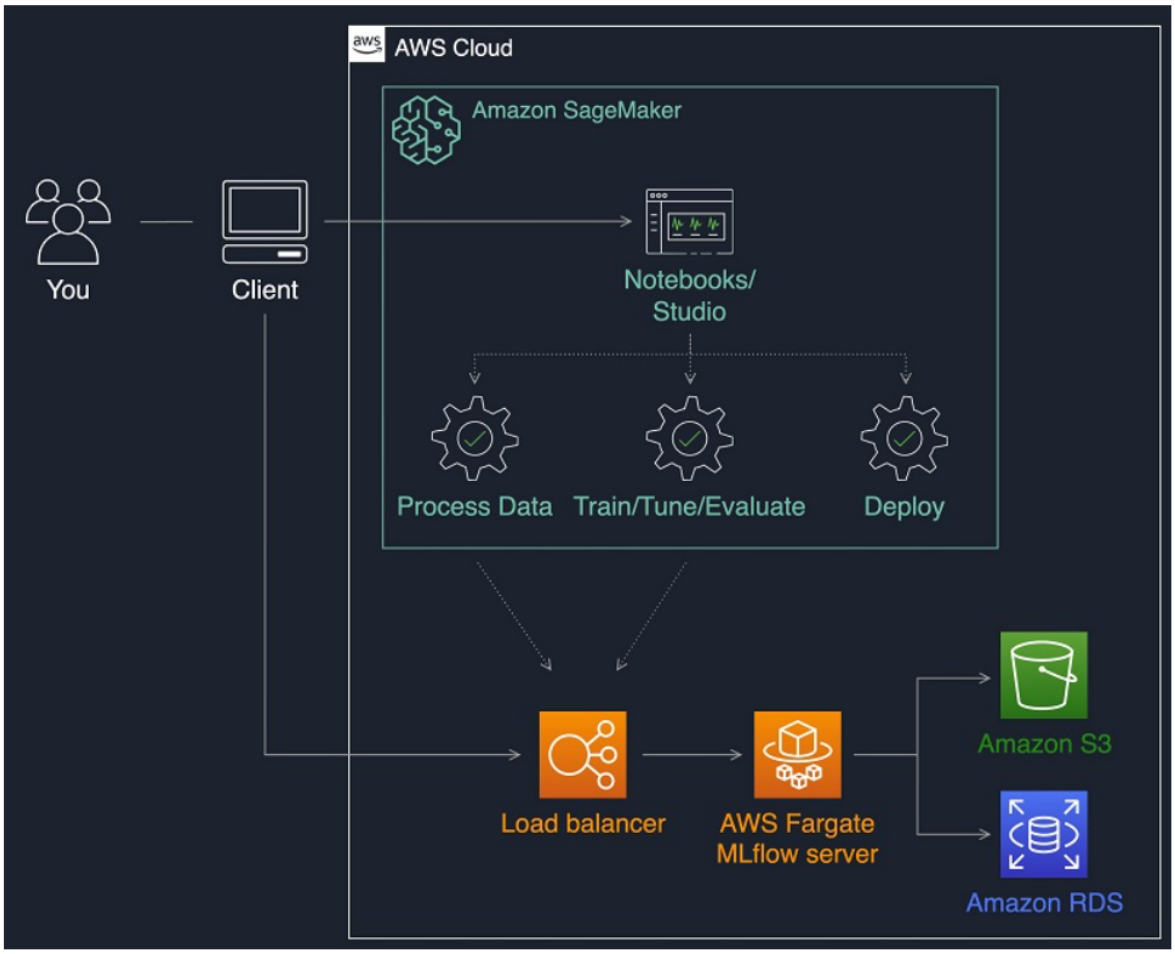
Aws ecr: Docker image built for training job and to deploy mlflow server are registered in aws ecr.

Aws Cloudformation: Aws cloud formation is used to deploy mlflow fargete service in which mlflow ui is hoisted.

Git: Git is a version control software for tracking changes in any set of files, usually used for coordinating work among programmers collaboratively developing source code during software development. It is used for Continuous Integration.

About the project

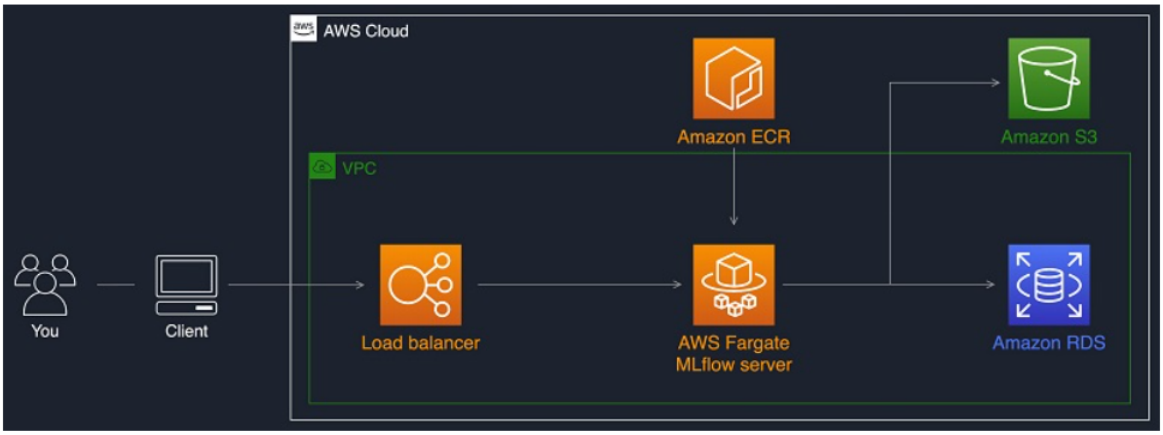
Tracking Model training with MLflow project mainly included three phases:



1. **Logging local training runs to local MLFlow tracking server:**

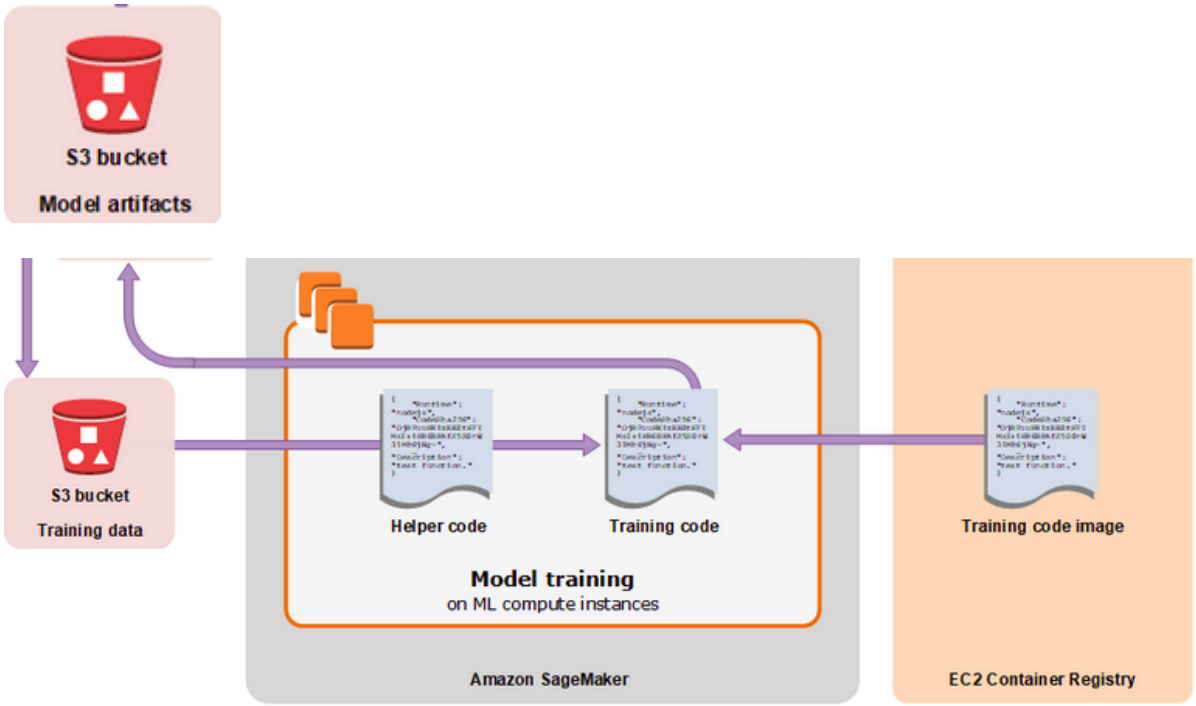
   Already existing ML model is integrated with mlflow python sdk and the run details , parameters , metrics and artefacts are recorded using the mlflow module and run locally.

2. **Logging local training runs to hosted MLFlow tracking server :**



Mlflow server is hoisted in aws fargete services using docker mlflow server image stored in aws ecr and making run data stored in aws s3.

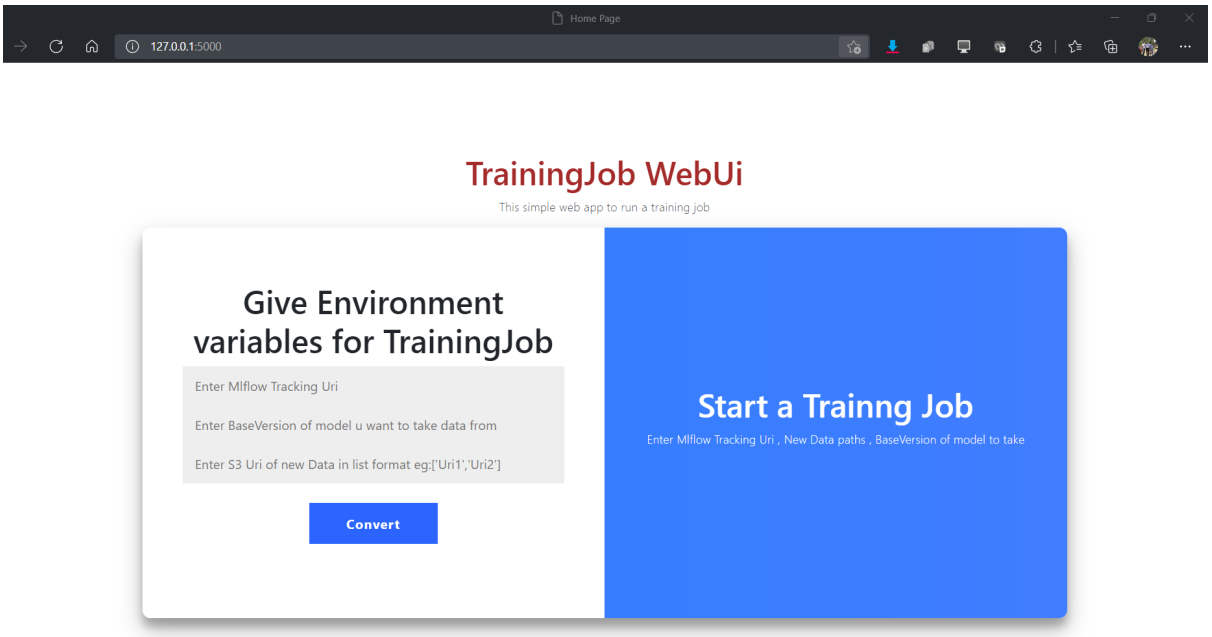
3. **Creating training job:**



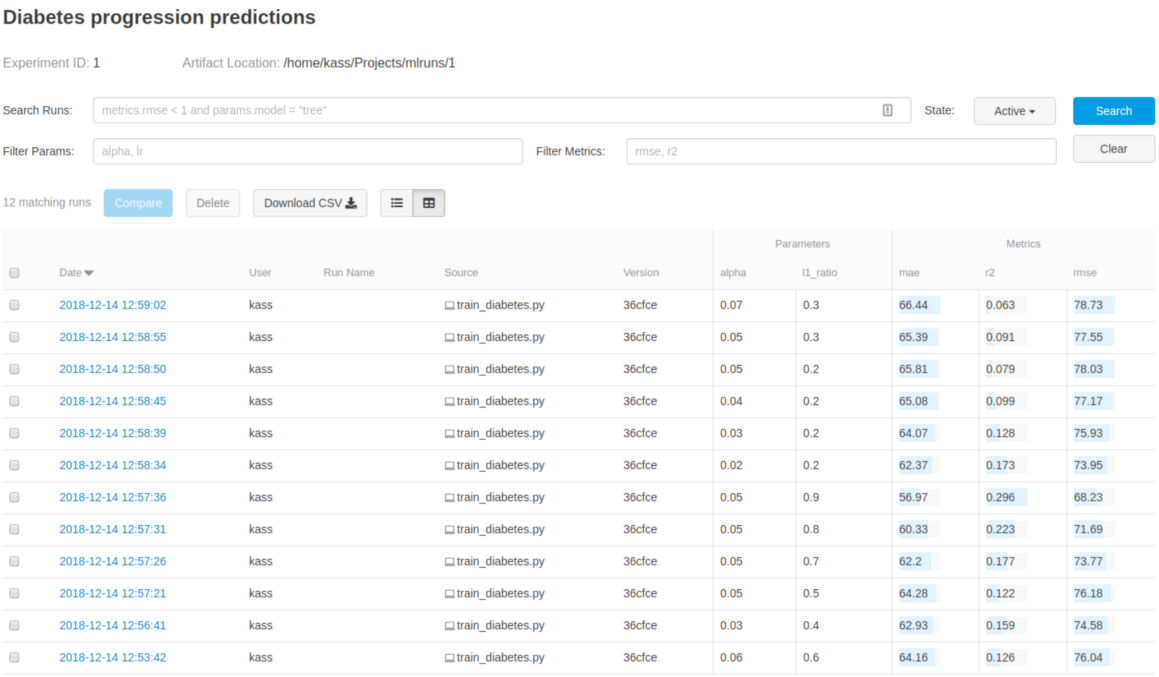
After hoisting the mlflow server we created a sagemaker training job using a self -built docker image which contains all the ML code . The way the model works is it takes s3 path to new data from container

environment variables and the model version that we want to take old data which is stored as an artefact and combine both data and send them for training. The model versioning is achieved by storing the current model as an artefact and registering the model with some version . MLflow also supports the features of keeping models in staging and in production.

Additional step:



As part of an additional step we created a web application using html , css , python, and flask for the demonstration of the project. Though the trigger is implemented as a web service for now it can be implemented as a lambda or can be a part of a bigger training pipeline.



Diabetes progression predictions

Experiment ID: 1 Artifact Location: /home/kass/Projects/mlruns/1

Search Runs: metrics.rmse < 1 and params.model = "tree" State: Active Search

Filter Params: alpha, lr Filter Metrics: rmse, r2 Clear

12 matching runs Compare Delete Download CSV

						Parameters		Metrics		
	Date	User	Run Name	Source	Version	alpha	l1_ratio	mae	r2	rmse
	2018-12-14 12:59:02	kass		train_diabetes.py	36cfce	0.07	0.3	66.44	0.063	78.73
	2018-12-14 12:58:55	kass		train_diabetes.py	36cfce	0.05	0.3	65.39	0.091	77.55
	2018-12-14 12:58:50	kass		train_diabetes.py	36cfce	0.05	0.2	65.81	0.079	78.03
	2018-12-14 12:58:45	kass		train_diabetes.py	36cfce	0.04	0.2	65.08	0.099	77.17
	2018-12-14 12:58:39	kass		train_diabetes.py	36cfce	0.03	0.2	64.07	0.128	75.93
	2018-12-14 12:58:34	kass		train_diabetes.py	36cfce	0.02	0.2	62.37	0.173	73.95
	2018-12-14 12:57:36	kass		train_diabetes.py	36cfce	0.05	0.9	56.97	0.296	68.23
	2018-12-14 12:57:31	kass		train_diabetes.py	36cfce	0.05	0.8	60.33	0.223	71.69
	2018-12-14 12:57:26	kass		train_diabetes.py	36cfce	0.05	0.7	62.2	0.177	73.77
	2018-12-14 12:57:21	kass		train_diabetes.py	36cfce	0.05	0.5	64.28	0.122	76.18
	2018-12-14 12:56:41	kass		train_diabetes.py	36cfce	0.03	0.4	62.93	0.159	74.58
	2018-12-14 12:53:42	kass		train_diabetes.py	36cfce	0.06	0.6	64.16	0.126	76.04

Coding/Implementation details (modules used to achieve the functionality

The Tracking Model training with MLflow project was implemented using Agile methodology where every sprint lasts for 14 days and the project is divided into components.

The different modules were used in the whole project include

- Mlflow
- Boto3
- Aws cli
- Os etc
- Requests

- Python 3.7

First, each feature was developed locally and tested. It was then deployed in the QA environment where further testing was carried out. Upon successful testing and results obtained, the feature was deployed.

Project results/Learning Outcomes

Results:

After demonstrating the project to the manager and team , the reviews were good.This project reduces the task of tracking each run , preparing models and versioning them . with just one simple command whole training process is taken care and results are stored seamlessly.

Learning Outcomes:

After working in the company for 6 months , I can successfully say that I have had an experiential learning and have gained state of the art knowledge as well as insights/practises into the corporate work culture and acquired professionalism to work with fellow team mates.

My learning is vast and includes gaining technical skills that were obtained from the 2 weeks of initial training which focussed on Agile methodology, Automation, Testing and Git. During the course of the project, I learned about the substantial need for Cloud computing and the various services that AWS offers to simplify and accomplish tasks for a software engineer. Apart from the technical skills, I was trained on Security, Privacy, Safety, Code of Business conduct and professional behaviour at the workplace.

Conclusion:

I would like to once again thank Autodesk India Private Limited my manager **Mr. Gopi Krishna Nuti,** and my mentor **Sasikanth Bharadwaj Ammanamanchi** have provided me with such a wonderful opportunity and I am sure this will have a huge positive impact and help boost my career. I was encouraged and was supported throughout the course of my internship by my teammates.

I can say that I have successfully completed my project **Tracking model training with mlflow** and altogether have evolved into an engineer with industry experience. The main takeaway from this internship does not only include the skills acquired but also the network and new friendships that were created during this process.

References:

Autodesk | 3D Design, Engineering & Construction Software

Autodesk - Wikipedia

https://aws.amazon.com/