**SUMMER INTERNSHIP REPORT**

By: **Soham Chaudhari**

**201070025**

**B.Tech in Computer Engineering**

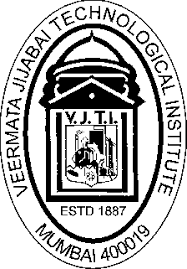
Under the Supervision of:

**Dr. S. G. Bhirud**



Company: **HDFC BANK**

Duration: 5th July, 2023 – 5th August, 2023



DEPARTMENT OF COMPUTER ENGINEERING AND INFORMATION TECHNOLOGY

**VEERMATA JIJABAI TECHNOLOGICAL INSTITUTE**

(An Autonomous Institute Affiliated to Mumbai University)

(Central Technological Institute, Maharashtra State) Matunga, MUMBAI – 400019

**CERTIFICATE**

This is to certify that Soham Chaudhari (201070025) is a student of B. Tech. Computer Engineering has successfully completed his internship at HDFC BANK for a duration of two months to our satisfaction.

|  |  |  |
| --- | --- | --- |
| **Internal Supervisor** | **Industrial Supervisor** | **Head of Department** |
| **Dr. S. G. Bhirud** | **Mr. Ravikumar Kamma** | **Dr. Vijay Sambhe** |
| Department of CE and IT | Manager, CRAIN Team | Department of CE and IT |
| VJTI, Mumbai | HDFC Bank Limited, Mumbai | VJTI, Mumbai |

Place: Senapati Bapat Marg,

Lower Parel, Mumbai

Date: 5th August, 2023

**APPROVAL SHEET**

The Final Internship Report by Soham Chaudhari is satisfactory and approved for Degree of Bachelor of Technology.

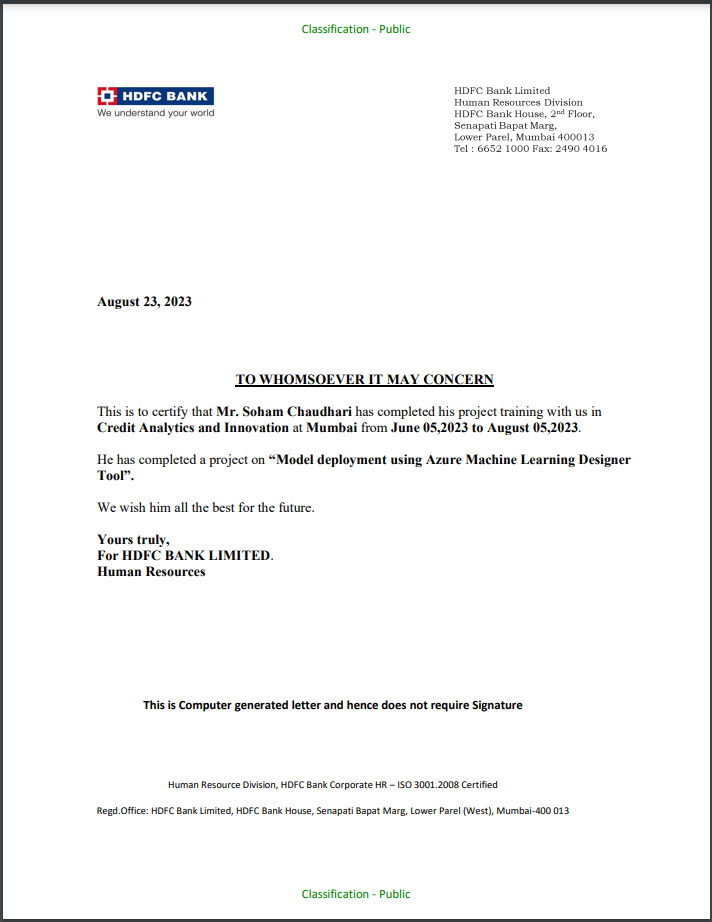
|  |  |  |
| --- | --- | --- |
| **Internal Supervisor** | **Industrial Supervisor** | **Head of Department** |
| **Dr. S. G. Bhirud** | **Mr. Ravikumar Kamma** | **Dr. Vijay Sambhe** |
| Department of CE and IT | Manager, CRAIN Team | Department of CE and IT |
| VJTI, Mumbai | HDFC Bank Limited, Mumbai | VJTI, Mumbai |

Place: Senapati Bapat Marg,

Lower Parel, Mumbai

Date: 5th August, 2023

**INTERNSHIP CERTIFICATE**

****

**STATEMENT OF CANDIDATE**

I, Soham Chaudhari (201070025) state that the work embodied in this Internship Report forms my own contribution to work at HDFC BANK as an intern. The report reflects the work done during the period of internship from 5th June 2023 to 5th August 2023. No part of this work has been used by me for the requirement of another degree except where explicitly stated in the body of the text and the attached statement. I assert that statements made, and conclusions drawn are an outcome of my internship work and are true to my knowledge.

I further certify that:

1. The work contained in the report is original and has been done by me under the supervision of my supervisor.
2. I have followed the guidelines provided by the college in writing the report.
3. Wherever I have used materials (text, data, theoretical analysis equations, codes, figures, etc.) from other sources, I have given due credit to them in their port by giving their details in the references.

Name: Soham Chaudhari Roll No: 201070025

Date: 5th August, 2023 Place: Mumbai

**TABLE OF CONTENTS**

1. Acknowledgement 7
2. About HDFC BANK 8
3. Abstract 9
4. Introduction 10
5. Objective of Work 11
6. Weekly Overview 12
7. Technology Learned 14
8. Project Details 15
9. Conclusion 20
10. References 21

**ACKNOWLEDGEMENT**

I would like to express my sincere gratitude and appreciation for the invaluable opportunity I was provided with during my enriching internship experience with HDFC BANK this summer. This internship has been a remarkable journey of growth, learning, and professional development, and I am truly honored to have been a part of the HDFC BANK team.

First and foremost, I extend my heartfelt appreciation to Dr. N. P. Gulhane, the head of VJTI TPO, the HR department at HDFC BANK, and all the support staff who played a pivotal role in ensuring a smooth and productive internship experience.

I would also like to express my gratitude to my internship mentor, Dr. S. G. Bhirud, for his unwavering support and guidance throughout the internship period.

My deep thanks go to all the individuals who have guided and supported me during this internship. I am sincerely grateful to my supervisors and mentors, Mr. Ravikumar Kamma, Mr. Harsh Verma, Mr. Abhijeet Patil, Mr. Bhupesh Sable, and Mr. Akshay Orpe for their invaluable guidance, mentorship, and encouragement. Their insights and expertise have been instrumental in shaping my understanding of the industry and enhancing my skills. I am also thankful for the opportunity to meet many wonderful people who guided me through this internship period.

I would like to extend my appreciation to the entire HDFC BANK team for their warm welcome and collaborative spirit. The camaraderie and teamwork I experienced during my time here have left a lasting impact on me.

I am thankful for the challenging tasks I had the opportunity to work on, which not only expanded my skill set but also deepened my understanding of the banking sector. The exposure to real-world scenarios and the chance to contribute to meaningful projects have been truly enriching.

My time at HDFC BANK has been an invaluable chapter in my professional journey, and I am excited to carry the lessons, experiences, and relationships I have gained forward with me. Once again, thank you for this incredible opportunity.

Sincerely,

Soham Chaudhari

**ABOUT HDFC BANK**

HDFC Bank, India's largest private sector bank, stands as a venerable pillar of the Indian financial landscape, boasting a legacy dating back to its establishment in 1994. With a steadfast commitment to delivering excellence, the bank has evolved into a stalwart institution, providing a wide spectrum of financial services that cater to the needs of individuals, businesses, and institutions alike.

The institution's offerings encompass an impressive range, including retail and commercial banking, mortgage lending, investment banking, and wealth management. By embracing technological advancement, HDFC Bank ensures that its services remain easily accessible through an extensive network of branches and ATMs, supplemented by intuitive digital banking solutions. This dedication to innovation underscores the bank's responsiveness to the ever-evolving needs of its customers.

HDFC Bank's core values of integrity, customer focus, and community engagement drive its operations. Through its responsible banking practices, the institution has fostered a culture of accountability, ensuring that its customers' financial well-being takes precedence. This commitment to ethical conduct and customer satisfaction distinguishes HDFC Bank as an institution that prioritizes long-term relationships and mutual growth.

Recognizing the importance of community, HDFC Bank actively engages in philanthropic endeavours, striving to make positive contributions to society. These efforts underscore the bank's commitment to social responsibility and its role in promoting economic vitality and inclusivity in the communities it serves.

In conclusion, HDFC Bank's enduring legacy, commitment to innovation, ethical foundation, and community engagement position it as a cornerstone of the Indian financial sector. Through its unwavering dedication to positive values, the bank continues to empower individuals and businesses, fostering financial well-being and contributing to the overall prosperity of the nation.

**Vision, Mission and Values**

HDFC Bank’s mission is to be a world-class Indian bank. We have a two-fold objective: first, to be the preferred provider of banking services for target retail and wholesale customer segments. The second objective is to achieve healthy growth in profitability, consistent with the bank’s risk appetite.

The bank is committed to maintaining the highest ethical standards, professional integrity, corporate governance and regulatory compliance. HDFC Bank’s business philosophy is based on five core values: Operational Excellence, Customer Focus, Product Leadership, People and Sustainability.

**ABSTRACT**

During the initial phase of our internship program, all interns were convened for an introductory meeting, where the Credit Analytics and Innovation team at HDFC BANK provided us with a comprehensive overview of the organization and its objectives. Following this introduction, we were organized into teams, aligning our assignments with our respective skills and interests, and each of us was paired with a dedicated mentor.

In the initial two weeks of my internship, my focus was dedicated to familiarizing myself with Azure Machine Learning Studio. I embarked on this journey by diligently following tutorials provided by my mentor. This period was primarily characterized by a deep dive into Azure Machine Learning Studio, as well as fostering a continuous channel of communication with my mentor. This dialogue was instrumental in addressing any questions or concerns that arose and ensuring regular updates on my progress.

As I transitioned into the second month of my internship, a pivotal milestone was reached. I was entrusted with the responsibility of delivering a demonstration presentation to the esteemed Data Science team at HDFC BANK. This presentation showcased the tasks and accomplishments I had achieved up to that point. The feedback I received from this knowledgeable group proved to be invaluable in shaping the direction of my subsequent work.

Over the following two weeks, I concentrated my efforts on a specific task that required building a Machine Learning (ML) pipeline using the designer tool. Given the substantial size of the dataset, the primary emphasis during this phase was on feature selection. I was not only successful in constructing the ML pipeline but also in deploying the ML model to both Batch and Real-time endpoints. To conclude this phase, I presented the results of this task to the Senior Vice President of HDFC BANK's Credit Analytics Team, a moment that underscored the significance of my work.

In the concluding week of my internship, I had the unique opportunity to work under the guidance of a different mentor. The task assigned to me was to construct a Notebook runner step for the Azure ML notebook. This experience provided me with a holistic view of diverse aspects within the organization and enhanced my skill set. On the final day of my internship, I formally handed over Task II to my mentor, concluding this enriching chapter of my professional development.

This internship at HDFC BANK has not only been a journey of learning but also an opportunity to make meaningful contributions to the field of credit analytics and innovation. It has equipped me with valuable insights and skills that will undoubtedly shape my future endeavours.

**INTRODUCTION**

Throughout my tenure at HDFC BANK, I was fortunate to partake in a profoundly transformative internship experience that revolved around a project at the crossroads of Credit Analytics and Innovation. In the sections that follow, I will delve into the specifics of this project and its implications, showcasing how it allowed me to leverage cutting-edge technologies and advanced analytical techniques to drive meaningful enhancements within the organization.

At the core of my internship lay a deep immersion into the realms of Data Analytics and Machine Learning. Guided by my mentors, I had the unique opportunity to harness the capabilities of Azure Machine Learning Studio, an advanced platform. The project was meticulously designed to optimize and streamline various processes within HDFC BANK, ultimately contributing to the institution's growth and success.

Over the course of my internship, I ventured into the dynamic domains of data science, machine learning, and innovation, collaborating closely with the esteemed team at HDFC BANK. My role encompassed a diverse array of responsibilities, spanning from data exploration and model development to the design and deployment of machine learning pipelines. Additionally, I had the honour of presenting my findings and contributions to senior leadership, including the Senior Vice President of HDFC BANK's Credit Analytics Team.

In the following sections, I will provide an in-depth overview of the project, offering insight into my role in its execution, the technology stack I employed, the challenges I encountered, and the significant outcomes that transpired during my internship at HDFC BANK. This experience not only enriched my knowledge but also afforded me the opportunity to make a tangible impact within the financial sector.

**OBJECTIVES OF WORK**

* **Career exploration and alignment:** Gaining insights into various career paths within my field of study, evaluating their alignment with my interests, skills, and values, and making informed decisions about my future career goals.
* **Integration of theory and practice:** Applying theoretical knowledge and concepts gained through coursework to real-world projects and tasks, identifying gaps between theory and practice, and developing strategies for bridging these gaps.
* **Skills development and assessment:** Identifying and developing new skills relevant to my field of study, assessing my proficiency in these skills, and obtaining feedback from supervisors and mentors on my performance.
* **Professional growth and development:** Gaining exposure to professional norms and expectations, developing my communication, teamwork, and problem-solving skills, and enhancing my overall professional conduct.
* **Contribution to the organization:** Demonstrating my ability to make valuable contributions to the organization through my work on projects and tasks, and gaining an understanding of the impact of my work on the organization's goals.

**WEEKLY OVERVIEW**

|  |  |
| --- | --- |
| **Week** | **Tasks** |
| 1st Week | * General Introductory Meeting with the Credit Analytics and Innovation Team with Mentor Assignment. * Setting Up Citrix Workspace (Virtual Desktop). * Task Assignment |
|  | |
| 2nd Week | * Kanjur Marg Office Visit (On Monday, 12th June) * Got Azure Machine Learning Subscription * Task I: Create a Notebook in Azure ML and Integrate it with Azure DevOps. * Got an error related to Azure ML Jobs (Azure Subscription Issue). |
|  | |
| 3rd Week | * Installed a new Azure Subscription on a Personal Device. * Setting Up Azure Account. * Airoli Office Visit (On Thursday, 22nd June) * Deployed First Azure ML Model using Azure ML Notebook and MLFlow Library. |
|  | |
| 4th Week | * Deployed Azure ML Model using Azure ML Designer Tool. * Integrated Azure ML Studio with Azure DevOps Repository. |
|  | |
| 5th Week | * Updated Task-I:  1. Fetch data using ADLS (Azure Data Lake Storage). 2. Deploy the model with a single click. 3. Deploy model to Real-time endpoint and Batch endpoint.  * Task I Demo Meet with HDFC BANK Data Science Team. * Feedback and Updated Task I: Deploy model to Batch Endpoint and Real-time endpoint using Azure Designer tool. |
| 6th Week | * Completed Task I with all updates. * Final Updates in Task I:  1. Use a dataset having 200 Features and 50,000+ Rows.  * Perform Feature Engineering. |
|  | |
| 7th Week | * Completed Task I. * Created Report and Final demo Presentation for the same. * Task II: To create the Notebook Runner Step (AML Designer pipeline using notebook) of previously created Python notebooks. |
|  | |
| 8th Week | * Task I Demo Meet with HDFC BANK’s Senior Vice President of Credit Analytics Team. * Task II handover. * Task I Report Submission. |

**TECHNOLOGY LEARNED**

**Azure Machine Learning**

Azure Machine Learning is a cloud-based platform that empowers data scientists and machine learning experts to efficiently create, train, deploy, and manage machine learning models at scale.

* **Azure Data Lake Storage (ADLS)**

Azure Data Lake Storage is a scalable and secure data lake solution designed for storing and analysing large volumes of data in the cloud.

* **Azure Notebooks & Azure Designer Tool:**
* Azure Notebooks: Azure Notebooks provides a cloud-based, interactive development environment tailored for data science and machine learning tasks.
* Azure Designer Tool: Azure Designer is a visual interface that simplifies the process of building and deploying machine learning models without the need for extensive coding.

**Azure DevOps**

Azure DevOps is a comprehensive suite of development tools and services aimed at streamlining software development and deployment processes. It encourages collaboration and automates workflows between development and operations teams.

**Python Libraries**

1. **NumPy:** NumPy is a fundamental library for numerical computing in Python. It provides support for working with large, multi-dimensional arrays and matrices, as well as a wide range of mathematical functions to operate on these arrays.
2. **Pandas:** Pandas is a powerful data manipulation and analysis library for Python. It offers data structures and functions for efficiently cleaning, transforming, and analysing data.
3. **MLFlow:** MLFlow is an open-source platform designed to manage the machine learning lifecycle. It assists in tracking experiments, packaging code into reproducible runs, and sharing and deploying models.
4. **Machine Learning Libraries:** This category typically includes various Python libraries and frameworks specifically tailored for machine learning tasks, such as scikit-learn, TensorFlow, PyTorch, and more. These libraries provide tools for training, evaluating, and deploying machine learning models.

**PROJECT DETAILS**

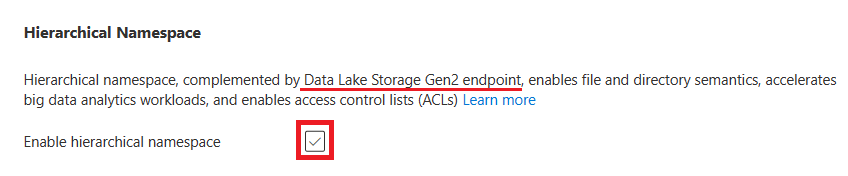
**Project:** Model Deployment Using Azure ML Designer Tool

**Dataset Used:** Customer Transaction Data (to predict Customer will buy a Credit card or not). Binary Classification.

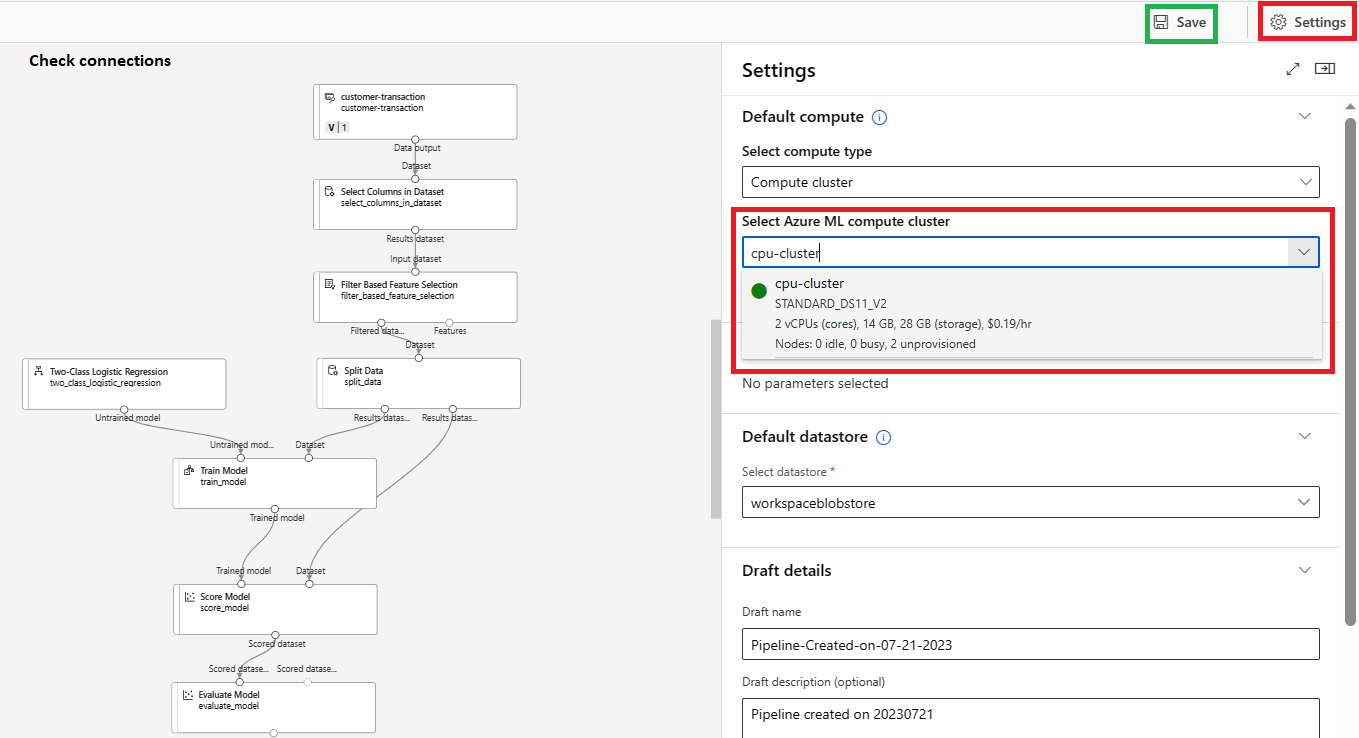
**Project Objective:** The project entails several key tasks, including retrieving data from Azure Data Lake Storage (ADLS), constructing a machine learning model, conducting feature engineering, training that model, deploying it to Batch and Real-time Endpoints, and making the deployed model accessible for use through the Azure Machine Learning Studio.

**Solution Methodology**

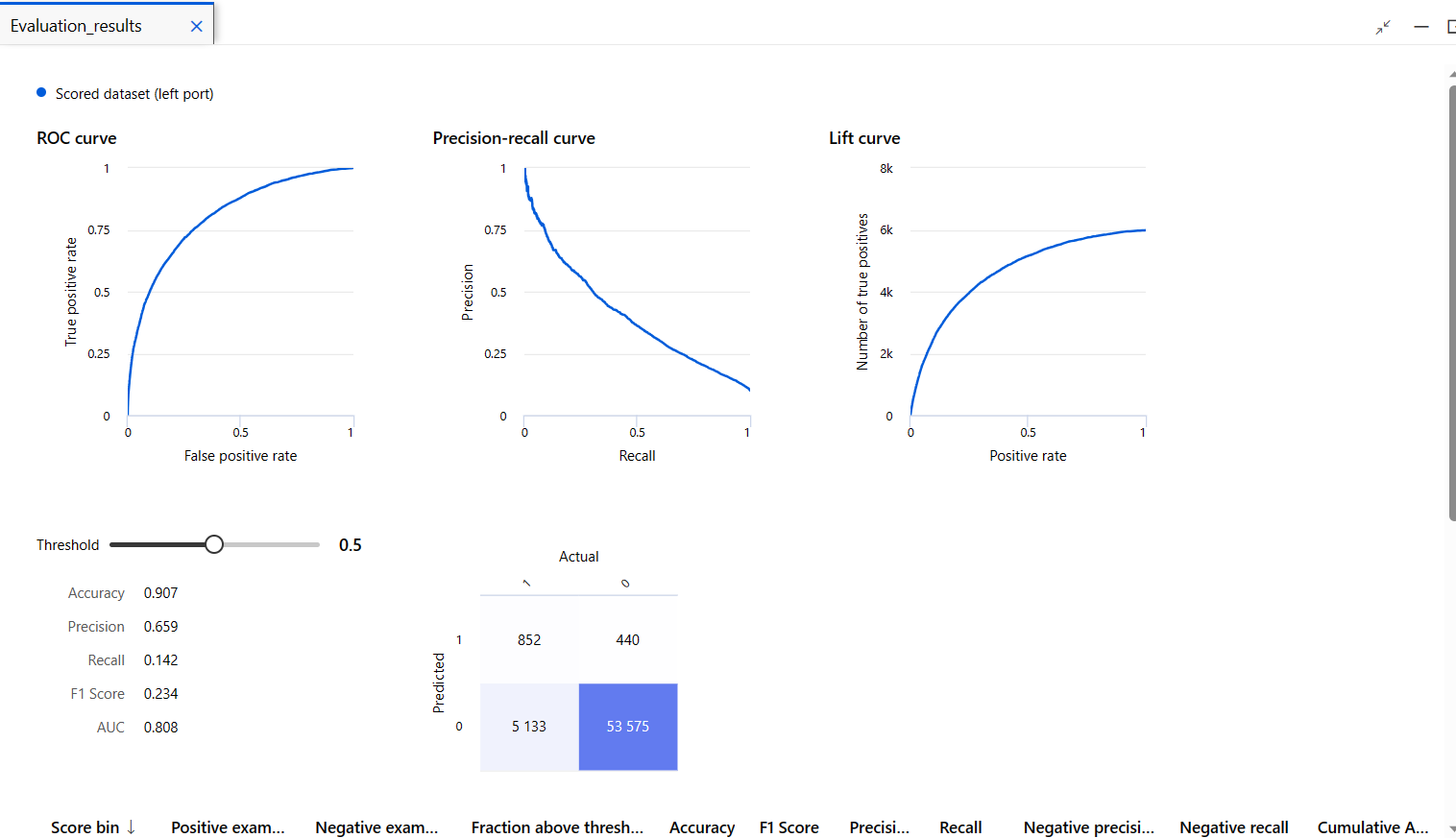
1. Create an Azure Storage Account (ADLS)
   1. Once ADLS gets created we can upload our dataset inside the ADLS.
   2. Further we can create Data Asset inside Azure ML Studio to use this data in Designer Pipeline.



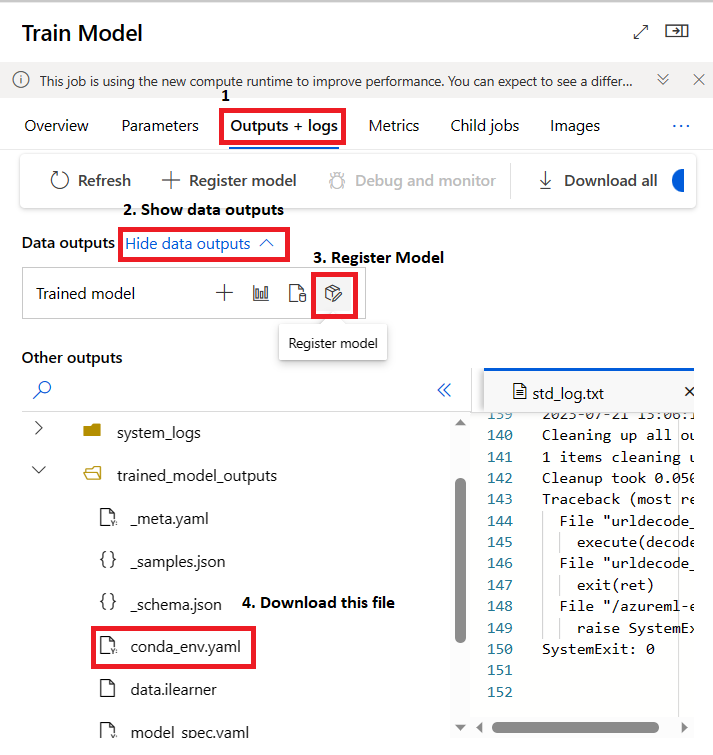
1. Create the Data Asset
   1. To use the dataset in ADLS we must create the data asset in Azure ML Studio so that we can use that dataset or file for training or testing purposes.
   2. Changes made inside ADLS get reflected inside the data asset created.
   3. While creating the data asset, select the ADLS path where the data file is present and Click on Review and Create
2. Create a Designer Model
   1. Azure ML Designer tool is a no-code platform.
   2. To create a model inside the designer tool we can drag and drop different components required to build our model.
   3. To run the designer pipeline, we require the Compute Cluster.
   4. Once we SUBMIT the pipeline the Job gets created where we can evaluate our model



1. Evaluate Job



1. Register model
   1. Once the Job gets executed successfully and we are satisfied with the results obtained from the job, we can register the model.
   2. After registration check for score.py and conda\_env.yaml files and download them because we will be using them for Batch-Deployment.



1. Create an Environment for deployment
   1. Environment is required for deployment; we can create this by using conda\_env.yaml file.
   2. Click on the Environment tab, create a new environment, and copy-paste the conda\_env.yaml file code in the .yaml file.
2. Deployment to Batch endpoint
   1. Create a data asset folder and pass the data in batches inside that folder.
   2. Deploy the registered model to the batch endpoint and once the deployment is successful create the Job inside the batch deployment.
   3. Use the score.py and Environment created in the previous steps.
   4. Create a batch deployment Job and use the data asset in which the data is present in batches.
   5. The output of this job will be stored in predictions.csv.
3. Deployment to Real-time endpoint
   1. We can deploy the model to a real-time endpoint with the help of custom score.py.
   2. But rather than creating score.py, we can create an inference pipeline with the help of Job.
   3. Then after successful submission of that pipeline we can deploy that pipeline.
   4. When we deploy this pipeline, it registers the model and deploys the pipeline to the Real-time endpoint. Hence, we can deploy the model without writing any scoring code.
   5. Once the deployment state becomes Healthy, we can consume this model.

**Limitations**

1. **Library Installation**

* In Jupyter Notebook we can install libraries with the help of the pip install command but inside the designer tool, we cannot directly execute the pip install command and install any library.
* Solution: Download the .whl file of the library required, then extract the required packages from that file, Zip all required packages, upload the zipped file inside data assets of the Azure ML Studio, Use the zipped data asset as component in designer tool and then we can execute the python script with all necessary libraries.

1. **Deploy again after modification**

* If some component gets changed or dataset gets updated or if the pipeline is modified then we have to follow the complete procedure to deploy the model. Cannot automate this process.
* Solution: -

1. We can automate this process with the help of Azure Notebooks, by writing all the code for model training, registering and deployment and then we can execute all cells to deploy the model.(We can use MLFlow/AutoML libraries to automate the process.)
2. We can use Azure DevOps Pipelines to build and deploy the model. In Azure DevOps for every single commit on the main branch, the pipeline gets executed automatically and deploys the model. Although building the pipeline in Azure DevOps is more difficult than inside Azure ML Studio once the pipeline is built one can monitor the commits, check for modifications and automate the complete process. (Limitation: Need to follow some folder structure.)­­­­­­

**Benefits**

1. **No-Code Platform**

We can perform almost every operation inside our Azure ML Designer tool which is needed for model training.

1. **ADLS**

* Can use external Azure Data Lake Storage to store large datasets and use it inside Azure Designer.
* Modifications in data present inside ADLS will automatically modify the data present inside Azure ML Data asset.
* No need to store the files on our local machine.

1. **Batch-Deployment**

If the test data size is very large then we can divide the data into batches, pass the data inside the folder in Azure ML Data Asset and further we can create a Batch Deployment Job to get output as a .csv file.

1. **Inference Pipeline**

Rather than creating score.py file for custom deployment to real-time endpoint one can create an inference pipeline for easy deployment to the real-time endpoint.

**CONCLUSION**

My eight-week internship at HDFC Bank was an invaluable experience that allowed me to gain hands-on experience in the field of machine learning. I worked on a variety of projects using Azure Machine Learning, Azure Data Lake Storage, Azure DevOps, and Python libraries such as NumPy, Pandas, MLFlow, and scikit-learn.

One of my most significant takeaways was the opportunity to learn about the machine learning lifecycle. I gained experience in collecting, cleaning, and transforming data, as well as training, evaluating, and deploying machine learning models. I also learned how to use Azure DevOps to automate the software development and deployment process.

Another important takeaway was the importance of collaboration. I worked closely with a team of experienced data scientists and engineers to develop and implement machine learning solutions. This experience taught me how to communicate effectively, share ideas, and work together to achieve common goals.

I am grateful for the support and mentorship I received from the team at HDFC Bank. I learned a great deal about machine learning and the financial industry during my internship. I am now well-positioned to pursue a career in machine learning or data science.

During my internship at HDFC Bank, I developed and deployed Azure Machine Learning models using Azure ML Studio, Azure ML Notebook, Azure ML Designer Tool, and other Azure services. I also integrated Azure DevOps for automation and Azure Data Lake Storage for data storage. Additionally, I performed feature engineering on a dataset with 200 features and 50,000+ rows. Finally, I created a report and presented my work to the team.

Overall, my internship at HDFC Bank was a valuable experience that gave me the skills and knowledge I need to pursue a career in machine learning. I am grateful for the support and mentorship I received from the team at HDFC Bank, and I am excited to see what the future holds.

**REFERENCES**

<https://learn.microsoft.com/en-us/azure/machine-learning/?view=azureml-api-2>

<https://github.com/Soham0779/AzureML-Designer-Deployment>

<https://www.youtube.com/playlist?list=PLiQS6N-W1p3m9squzZ2cPgGdH5SBhjY6f>

<https://towardsdatascience.com/azure-machine-learning-service-part-1-an-introduction-739620d1127b>