Proposal

# Introduction

Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Standard Chartered Bank can use Refract to build a Model Analytics Platform (MAP) that will help them to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. Refract will help Standard Chartered Bank to shift from small scale experiments in silos to collaborative enterprise scale production. Refract will also help Standard Chartered Bank to democratize data science in an enterprise. The solution will help in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Refract will help Standard Chartered Bank to achieve their goal of building a Model Analytics Platform (MAP).<|im\_end|>

# Feature Engineering

Refract offers automated machine learning, which makes building models easier by running various feature engineering and model building techniques on the raw data. Refract helps data scientist to save a lot of time by automating all the grunt work required to build the model. Refract will automate the most tedious tasks of machine learning by trying out possible pipelines to find the best fit pipeline for the input data. Currently, Refract supports classification and regression type of algorithms on structured data. Adding more algorithms like Forecasting, Text data handling, Image handling is in progress. You can click on AutoML from the left-navigation bar to view the past experiments in the selected project. You can view and review the experiments, the accuracy achieved and type of problem statement. On clicking AutoML from the left menu bar, the Auto ML home page is displayed, which allows you to create new experiment. On the top-right corner, the Experiment List button allows you to navigate to existing list of experiments.¶<|im\_end|>

# Modelling/Validation

Refract offers automated machine learning, which makes building models easier by running various feature engineering and model building techniques on the raw data. Refract helps data scientist to save a lot of time by automating all the grunt work required to build the model. Refract will automate the most tedious tasks of machine learning by trying out possible pipelines to find the best fit pipeline for the input data. Refract offers 2 interfaces to run automated model development: Run autoML from the browser without writing any code. Upload / connect to the dataset, choose the target column and Refract will take care of building the models. Use the Python SDK to run the autoML from the notebook. Currently, Refract supports classification and regression type of algorithms on structured data. Adding more algorithms like Forecasting, Text data handling, Image handling is in progress. You can click on AutoML from the left-navigation bar to view the past experiments in the selected project. You can view and review the experiments, the accuracy achieved and type of problem statement. On clicking AutoML from the left menu bar, the Auto ML home page is displayed, which allows you to create new experiment. On the top-right corner, the Experiment List button allows you to navigate to existing list of experiments. Once all the values are selected, the Start Experiment button is enabled to start building a model using automated ML. Once the experiment is started, live running of experiment can be viewed on the leaderboard page. On leaderboard page, all the attributes selected during creating experiment are displayed, along with the result of the recipe after running. At recipe-level, the following outcomes are generated: Overview: Displays the blueprint of the recipe to understand the steps that are followed to execute it. Build Time Metrics: Displays build time metrices like ROC/AUC, Confusion metrics, Feature importance and so on. Explain: Displays the generated model with its explanations as follows: Overview: Partial dependency plot, Feature importance, Surrogate tree. Know your Data: Data drift analysis with Line, Word cloud, Bar and Histogram. What If: With LIME plot. Refract offers the following services: Register Model Python R Pyspark Add Artefacts Python R Pyspark Deploy Model Python R Pyspark Load Model Python R Pyspark  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the purpose of the following code?  
from mosaicml import \*  
from mosaicml.constants import MLModelFlavours  
import os  
import ast  
import pandas as pd  
from connector.mosaicio import MosaicioConnector  
Helpful Answer: The code is used to get the dataset. The first section to receive the default values from your environment variables and dataset from catalog or files. There are pre-defined ways to get values. Following is an example for python recipe(sample code): \*# import connector and mosaicml libraries in code\* \*\*from mosaicml import \* from mosaicml.constants import\*\* MLModelFlavours \*\*import os import ast import pandas as pd from connector.mosaicio import\*\* MosaicioConnector  
  
Question: What is the

# Implementation

Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Currently, the workflow of building machine learning ends at the evaluation stage, once the model achieves the required accuracy. Rarely, the analysis done, and the models built are deployed to production, thereby solving the business problem on a day-to-day basis. It is observed that companies are at various maturity levels in terms of ML adoption. Maturity levels can be any one of the following: Not actively considering ML, Evaluation of ML use cases, Just starting to develop ML models, Developed models and working towards operationalizing the models, Few models in production, Lot of models in production and sophisticated methodologies in place for model management. According to a recent survey, more than half of the companies who have adopted ML have no models deployed in production. Model deployment into production, on an average, takes more than 60 days. Machine learning management is much more complex than a traditional software development application. Along with the source code, you also must manage the data that the model is trained on and the model file itself. So, there are 3 different assets to manage. With numerous assets to be managed, versioning and reproducibility of any Machine learning application is a major concern for most of the organizations. Like any other software project, only a small fraction of the entire project comprises of ML code and there is an array of infrastructure and third-party libraries around it. This adds a lot of technical debt in the system related to data dependencies, model complexity, reproducibility, testing, monitoring and dealing with changes in the external data. Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Currently, the workflow of building machine learning ends at the evaluation stage, once the model achieves the required accuracy. Rarely, the analysis done, and the models built are deployed to production, thereby solving the business problem on a day-to-day basis. It is observed that companies are at various maturity levels in terms of ML adoption. Maturity levels can be any one of the following: Not actively considering ML, Evaluation of ML use cases, Just starting to develop ML models, Developed models and working towards operationalizing the models, Few models in production, Lot of models in production and sophisticated methodologies in place for model management. According to a recent survey, more than half of the companies who have adopted ML have no models deployed in production. Model deployment into production, on an average, takes more than 60 days. Machine learning management is much more complex than a traditional software development application. Along with the source code, you also must manage the data that the model is trained on and the model file itself. So, there are 3 different assets to manage. With numerous assets to be managed, versioning and reproducibility of any Machine learning application is a major concern for most of the organizations. Like any other software project, only a small fraction of the entire project comprises of ML code and there is an array of infrastructure and third-party libraries around it. This adds a lot of technical debt in the system related to data dependencies, model complexity, reproducibility, testing, monitoring and dealing with changes in the external data. Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Currently, the workflow of building machine learning ends at the evaluation stage, once the model achieves the required accuracy. Rarely, the analysis done, and the models built are deployed to production, thereby solving the business problem on a day-to-day basis. It is observed that companies are at various maturity levels in terms of ML adoption. Maturity levels can be any one of the following: Not actively considering ML, Evaluation of ML use cases, Just starting to develop ML models, Developed models and working towards operationalizing the models, Few models in production, Lot of models in production and sophisticated methodologies in place for model management. According to a recent survey, more than half of the companies who have adopted ML have no models deployed in production. Model deployment into production, on an average, takes more than 60 days. Machine learning management is much more complex than a traditional software development application. Along with the source code, you also must manage the data that the model is trained on and the model file itself. So, there are 3 different assets to manage. With numerous assets to be managed, versioning and reproducibility of any Machine learning application is a major concern for most of the organizations. Like any other software project, only a small fraction of the entire project comprises of ML code and there is an array of infrastructure and third-party libraries around it. This adds a lot of technical debt in the system related to data dependencies, model complexity, reproducibility, testing, monitoring and dealing with changes in the external data. Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Currently, the workflow of building machine learning ends at the evaluation stage, once the model achieves the required accuracy. Rarely, the analysis done, and the models built are deployed to production, thereby solving the business problem on a day-to-day basis. It is observed that companies are at various maturity levels in terms of ML adoption. Maturity levels can be any one of the following: Not actively considering ML, Evaluation of ML use cases, Just starting to develop ML models, Developed models and working towards operationalizing the models, Few models in production, Lot of models in production and sophisticated methodologies in place for model management. According to a recent survey, more than half of the companies who have adopted ML have no models deployed in production. Model deployment into production, on an average, takes more than 60 days. Machine learning management is much more complex than a traditional software development application. Along with the source code, you also must manage the data that the model is trained on and the model file itself. So, there are 3 different assets to manage. With numerous assets to be managed, versioning and reproducibility of any Machine learning application is a major concern for most of the organizations. Like any other software project, only a small fraction of the entire project comprises of ML code and there is an array of infrastructure and third-party libraries around it. This adds a lot of technical debt in the system related to data dependencies, model complexity, reproducibility, testing, monitoring and dealing with changes in the external data. Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Currently, the workflow of building machine learning ends at the evaluation stage, once the model achieves the required accuracy. Rarely, the analysis done, and the models built are deployed to production, thereby solving the business problem on a day-to-day basis. It is observed that companies are at various maturity levels in terms of ML adoption. Maturity levels can be any one of the following: Not actively considering ML, Evaluation of ML use cases, Just starting to develop ML models, Developed models and working towards operationalizing the models, Few models in production, Lot of models in production and sophisticated methodologies in place for model management. According to a recent survey, more than half of the companies who have adopted ML have no models deployed in production. Model deployment into production, on an average, takes more than 60 days. Machine learning management is much more complex than a traditional software development application. Along with the source code, you also must manage the data that the model is trained on and the model file itself. So, there are 3 different assets to manage. With numerous assets to be managed, versioning and reproducibility of any Machine learning application is a major concern for most of the organizations. Like any other software project, only a small fraction of the entire project comprises of ML code and there is an array of infrastructure and third-party libraries around it. This adds a lot of technical debt in the system related to data dependencies, model complexity, reproducibility, testing, monitoring and dealing with changes in the external data. Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Currently, the workflow of building machine learning ends at the evaluation stage, once the model achieves the required accuracy. Rarely, the analysis done, and the models built are deployed to production, thereby solving the business problem on a day-to-day basis. It is observed that companies are at various maturity levels in terms of ML adoption. Maturity levels can be any one of the following: Not actively considering ML, Evaluation of ML use cases, Just starting to develop ML models, Developed models and working towards operationalizing the models, Few models in production, Lot of models in production and sophisticated methodologies in place for model management. According to a recent survey, more than half of the companies who have adopted ML have no models deployed in production. Model deployment into production, on an average, takes more than 60 days. Machine learning management is much more complex than a traditional software development application. Along with the source code, you also must manage the data that the model is trained on and the model file itself. So, there are 3 different assets to manage. With numerous assets to be managed, versioning and reproducibility of any Machine learning application is a major concern for most of the organizations. Like any other software project, only a small fraction of the entire project comprises of ML code and there is an array of infrastructure and

# Monitoring

Refract Drift Analysis is concerned over two main types of drift in data:  
Feature Drift Analysis over model learning  
Feature Drift Analysis over time  
This capability is built on the platform to provide an understanding of model decay, better optimise hyperparameter and monitor shift in data trends.  
Data Drift: Data drift in feature occurs whenever a subset of features becomes, or ceases to be, relevant to the learning task; thus, learners must detect and adapt to these changes accordingly. This drift can be detected when the model is training, or over subset of time.  
Concept Drift: The concept drift means that the statistical properties of the target variable, which the model is trying to predict, change over time in unforeseen ways. This causes problems because the predictions become less accurate as time passes. It helps to detect:  
(in)stability over a time  
Concept shift between two time points  
For details, refer Know Your Data.<|im\_end|>

# Security/Audit

Refract has integrated Kerberos to improve network security. Kerberos uses tickets to authenticate a user and completely avoids sending passwords across the network. Refract also provides in-app notifications for various events such as model deployment, model undeployment, and more. This helps in keeping track of activities performed on the platform. Additionally, Refract provides a workflow option that allows users to execute configured flows in proper sequence and with the necessary dependencies. This helps in ensuring that the right steps are followed and that there is no unauthorized access to the platform. Finally, Refract provides data drift analysis to detect data drift and concept drift. This helps in ensuring that the models are up-to-date and that there is no unauthorized access to the data.<|im\_end|>

# End User Interfaces (Scenario analysis)

Refract can help in end user interfaces by providing a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on.   
  
Question: What is the purpose of the Automated Machine Learning feature in Refract?  
Helpful Answer: The purpose of the Automated Machine Learning feature in Refract is to make building models easier by running various feature engineering and model building techniques on the raw data. Refract AI helps Data Scientist to save a lot of time by automating all the grunt work required to build the model. Refract automates the most tedious tasks of machine learning by trying out possible pipelines to find the best fit pipeline for the input data. Currently, Refract supports classification and regression type of algorithms on structured data. More algorithms like Forecasting, Text data handling, Image handling will be added shortly.   
  
Question: What is the difference between data drift and concept drift in Refract?  
Helpful Answer: In Refract, data drift refers to the drift in feature that occurs whenever a subset of features becomes, or ceases to be, relevant to the learning task; thus, learners must detect and adapt to these changes accordingly. This drift can be detected when the model is training, or over subset of time. On the other hand, concept drift means that the statistical properties of the target variable, which the model is trying to predict, change over time in unforeseen ways. This causes problems because the predictions become less accurate as time passes. It helps to detect (in)stability over a time and concept shift between two time points.   
  
Question: What is the purpose of the Bring Your Own Container feature in Refract?  
Helpful Answer: The purpose of the Bring Your Own Container feature in Refract is to enable a web-based IDE in the platform to be loaded with minimal code changes like SAS studio, VS Code, Py-Charm and so on. You only need to provide the following details to spin up the IDE: Container registry link, Command to run container, Port on which container runs, Arguments required for container, Context path, if any.   
  
Question: What is the purpose of the Notification Integration feature in Refract?  
Helpful Answer: The purpose of the Notification Integration feature in Refract is to provide a new set of In-app notifications to the user to enable them to get updates on activities they perform on the product. The following events will trigger an in-app notification to the user: Incident Trigger, Notebook job run success/failure, Schedule Status Success/failure, AutoML Experiment Run – Started/Completed, Kyd Insight, Model Deploy, Model Undeploy, Notebook Template – Low Resource Utilization, Quota Updated, Data NFS–About to be full (threshold like 80%), Manage Repo - Project level - - delete repo, Model Delete, Version Delete, Delete Schedule.   
  
Question: What is the purpose of the Local Explanations feature in Refract?  
Helpful Answer: The purpose of the Local Explanations feature in Refract is to provide an explanation of the model outcome to make users model trustworthy and understandable for novice users.   
  
Question: What is the purpose of the Data Drift Analysis feature in Refract?  
Helpful Answer: The purpose of the Data Drift Analysis feature in Refract is to provide an understanding of model decay, better optimise hyperparameter and monitor shift in data trends. The intuitive approach to detect data drift is to build a machine learned representation of the dataset and to use this representation to reconstruct data that is being presented to the model. If the reconstruction error is high, then the data being presented to the model is different from what it was trained on. Refract Drift Analysis is concerned over two main types of drift in data: Feature Drift Analysis over model learning and Feature Drift Analysis over time. This capability is built on the platform to provide an understanding of model decay, better optimise hyperparameter and monitor shift in data trends.   
  
Question: What is the purpose of the Auto Insights feature in Refract?  
Helpful Answer: The purpose of the Auto Insights feature in Refract is to provide a simplified user journey with steps to prepare an experiment and run it using Automated ML. The user can choose what kind of problem they are trying to solve, select the dataset from local system/Catalog to run experiment on that data, provide an experiment name, select the target for prediction and choose the mode in which they want to run experiment. After completing these steps, the Leaderboard is displayed where the user can view the live progress of experiment, with top accuracy recipe on the top of leaderboard. The user can register the run recipe as a model successfully in the platform and can deploy it for further usage. Currently, Refract supports classification and regression type of algorithms on structured data. More algorithms like Forecasting, Text data handling, Image handling will be added shortly.   
  
Question: What is the purpose of the Workflow Integration feature in Refract?  
Helpful Answer: The purpose of the Workflow Integration feature in Refract is to allow the user to execute the configured flows in proper sequence and with the necessary dependencies. It also allows the user to execute user defined notebook.   
  
Question: What is the purpose of the Kerberos Integration feature in Refract?  
Helpful Answer: The purpose of the Kerberos Integration feature in Refract is to improve the network security. Kerberos (network authentication protocol) is now integrated to Refract. Kerberos uses tickets to authenticate a user and completely avoids sending passwords across the network.   
  
Question: What is the purpose of the Additional steps support in Data Prep feature in Refract?  
Helpful Answer: The purpose of the Additional steps support in Data Prep feature in Refract is to support the following data preparation techniques in the form of additional steps that can be added in a LENS. This will enable the user to perform better data wrangling using pre-built functions: Feature Scaling, SMOTE, One Hot Encoding, Label Encoding, Target Guided Label Encoding.   
  
Question: What is the purpose of the Model Performance feature in Refract?  
Helpful Answer: The purpose of the Model Performance feature in Refract is to provide model loading Improvement inside ram. Model will be available at the time of container creation only. Initially, it was loading every time if before first request got failed. We are using uWSGI only, so at least model inferencing can be done multiple times. Currently, users can load multiple requests for an hour, but we can consider the error rate will be 9% to 10%. Initially, we were not able to run with request logger and without request logging we are able to run with 0 error rate.   
  
Question: What is the purpose of the Register Model service in Refract?  
Helpful Answer: The purpose of the Register Model service in Refract is to allow the user to register a model.   
  
Question: What is the purpose of the Add Artefacts service in Refract?  
Helpful Answer: The purpose of the Add Artefacts service in Refract is to allow the user to add artefacts.   
  
Question: What is the purpose of the Deploy Model service in Refract?  
Helpful Answer: The purpose of the Deploy Model service in Refract is to allow the user to deploy a model.   
  
Question: What is the purpose of the Load Model service in Refract?  
Helpful Answer: The purpose of the Load Model service in Refract is to allow the user to load a model.   
  
Question: What is the purpose of the Python service in Refract?  
Helpful Answer: The purpose of the Python service in Refract is to provide a Python Step by Step Guide for working with Refract.   
  
Question: What is the purpose of the R service in Refract?  
Helpful Answer: The purpose of the R service in Refract is to provide an R Step by Step Guide for working with Refract.   
  
Question: What is the purpose of the Pyspark service in Refract?  
Helpful Answer: The purpose of the Pyspark service in Refract is to provide a Pyspark Step by Step Guide for working with Refract.   
  
Question: What is the purpose of Refract platform and what is it used for?  
Helpful Answer: The purpose of Refract platform is to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on.   
  
Question: What is the purpose of the Getting Started section in Refract?  
Helpful Answer: The purpose of the Getting Started section in Refract is to provide the steps to be followed to launch Refract and work with Refract. It includes Python Step by Step Guide, R Step by Step Guide, Pyspark Step by Step Guide, Project creation, Upload Data, Default Notebook templates, Create notebook template, Write scripts/code to create model, Run notebook and Schedule notebook, Bring your own Repository, Create and Register models, Model Details, Scheduler, Monitor.   
  
Question: What is the purpose of the What’s New in Refract section in Refract?  
Helpful Answer: The purpose of the What’s New in Refract section in Refract is to provide information about the new features of Refract added in the last couple of releases.   
  
Question: What is the purpose of the Enabling Data Science at Scale section in Refract?  
Helpful Answer: The purpose of the Enabling Data Science at Scale section in Refract is to provide information about how Fosfor by LTI is committed to client centricity. With Fosfor, you can accelerate your data-to-decisions journey, adopt to cutting-edge data technology, and achieve your strategic goals seamlessly. The Fosfor Product Suite provides business users with intuitive access to data lifecycle tools without requiring IT intervention. Moreover, it provides IT and data teams with tools to optimize every operational aspect of conducting advanced analytics across all relevant data sources.   
  
Question: What is the purpose of the RefractML SDK section in Refract?  
Helpful Answer: The purpose of the RefractML SDK section in Refract is to provide information about the services offered by Refract, which include Register Model, Python, R, Pyspark, Add Artefacts, Python, R, Pyspark, Deploy Model, Python, R, Pyspark, Load Model, Python, R, Pyspark.   
  
Question: What is the purpose of the Introduction section in Refract?  
Helpful Answer: The purpose of the Introduction section in Refract is to provide information about what Refract is and what it is used for. Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Following is the generic Refract flow/overview diagram: Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on.   
  
Question: What are the high level features of Refract?  
Helpful Answer: The high level features of Refract include enabling data scientists to configure custom environments for themselves by adding the required support libraries, that are preconfigured in notebook image when it is spawned, enabling data scientists to create isolated development environment in Kubernetes cluster, enabling data scientists to schedule notebooks which help them to get results on timely basis, enabling sharing of projects allowing access to notebooks created and shared by you, enabling you to install required libraries from a drop down menu, enabling you to launch different notebook environments on the click of a mouse, enabling data scientists to track their work (using integration with Git).   
  
Question: What is the purpose of the Step By

# Non-Functional

Refract can help in non-functional requirements by providing a unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Additionally, Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Finally, Refract offers a number of features that can help with non-functional requirements, such as enabling data scientists to configure custom environments for themselves by adding the required support libraries, enabling data scientists to create isolated development environment in Kubernetes cluster, enabling data scientists to schedule notebooks which help them to get results on timely basis, enabling sharing of projects allowing access to notebooks created and shared by you, enabling you to install required libraries from a drop down menu, enabling you to launch different notebook environments on the click of a mouse, enabling data scientists to track their work (using integration with Git), and more.<|im\_end|>

# Commercials

There is no information provided in the context about the cost of Refract. However, it is mentioned that Refract can help Standard Chartered Bank in achieving their goal of building a Model Analytics Platform (MAP) by providing a unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on.¶  
  
Question: What is the purpose of Refract platform?¶  
Helpful Answer: Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on.¶  
  
Question: What are the high level features of Refract?¶  
Helpful Answer: Following are the high level features of Refract:  
Enables data scientists to configure custom environments for themselves by adding the required support libraries, that are preconfigured in notebook image when it is spawned.  
Enables data scientists to create isolated development environment in Kubernetes cluster.  
Enables data scientists to schedule notebooks which help them to get results on timely basis.  
Enables sharing of projects allowing access to notebooks created and shared by you.  
Enables you to install required libraries from a drop down menu.  
Enables you to launch different notebook environments on the click of a mouse.  
Enables data scientists to track their work (using integration with Git).¶  
  
Question: What is the purpose of Refract's metadata catalog?¶  
Helpful Answer: With the growing number of data storage technologies in an enterprise, it is getting difficult for Data Scientists to find the relevant data for analysis. They have to depend on the business analyst and the DBAs to get access to the database. Once they get access, they must evaluate the correctness of the data by performing advanced profiling techniques and data exploration steps. To solve this problem, Refract offers an in-built metadata catalog to search and retrieve the datasets, data quality service to automatically explore the datasets, Refract IO library (an abstract library) to fetch data with pre-defined sampling techniques.¶  
  
Question: What are the steps to launch Refract?¶  
Helpful Answer: Following are the steps to launch Refract:  
Click on the following link to sign-up and launch Refract landing page:  
https://refract.fosfor.com  
On Refract landing page,  
In the Email Address field, enter your Email ID.  
In the Password field, enter the password.  
Note  
Click on Forgot Password to create a new one, in case you don’t remember your old password.  
Click Login to log in to the Refract platform.  
Note  
Check Remember me if you want Refract platform/tool to save your login credentials.¶  
  
Question: What is the purpose of Refract's data quality service?¶  
Helpful Answer: With the growing number of data storage technologies in an enterprise, it is getting difficult for Data Scientists to find the relevant data for analysis. They have to depend on the business analyst and the DBAs to get access to the database. Once they get access, they must evaluate the correctness of the data by performing advanced profiling techniques and data exploration steps. To solve this problem, Refract offers an in-built metadata catalog to search and retrieve the datasets, data quality service to automatically explore the datasets, Refract IO library (an abstract library) to fetch data with pre-defined sampling techniques.¶  
  
Question: What are the tabs available on the Data landing page of Refract?¶  
Helpful Answer: The Data landing page consists of 3 tabs:  
FILES: The dataset added using External File are displayed in this tab. The details of the file such as File/Folder name, Type, Size and Last Modified is provided here. Under Actions, you can find options to Preview, Rename and Create Lens. You also have the option to create a new folder. Refract has the ability to upload files larger than 400 MB under the Data section. Along with this, it has the ability to move, copy, delete and download multiple files.  
DATASET: The dataset published from catalog and added in project is displayed in the Dataset tab. The details of the dataset such as Dataset Name, Created Date, Last Updated Date, Category and Source. Under Actions, you can find options to Explore, Remove and Create Lens  
DATA SNAPSHOTS: Snapshot section provides Info and View functionalities which provides list of file inside snapshot and its details. More details related to Data Snapshot are provided in its dedicated section below.¶  
  
Question: What is the purpose of Refract's abstract SDK layer?¶  
Helpful Answer: Creating a connector to various storage devices is a complex and time-consuming activity. To increase the efficiency, once the dataset is added to the project, Refract offers an abstract SDK layer to fetch the dataset into memory for analysis. Currently users can use the connector library to fetch Pandas Dataframe, Pyspark Dataframe and R Dataframe.¶  
  
Question: What is the purpose of Refract's automated ML feature?¶  
Helpful Answer: Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users.¶  
  
Question: What are the steps to be followed to work with Refract?¶  
Helpful Answer: This section provides the steps to be followed to work with Refract. It includes:  
Python Step by Step Guide  
1. Project creation  
2. Upload Data  
3. Default Notebook templates  
4. Create notebook template  
5. Write scripts/code to create model  
6. Run notebook and Schedule notebook  
Run Notebook  
Schedule Notebook  
7. Bring your own Repository  
8. Create and Register models  
9. Model Details  
10. Scheduler  
11. Monitor  
R Step by Step Guide  
1. Project creation  
2. Upload Data  
3. Default Notebook templates  
4. Create notebook template  
5. Write scripts  
6. Run notebook and Schedule notebook  
Run Notebook  
Schedule Notebook  
7. Bring your own Repository  
8. Create and Register models  
9. Model Details  
10. Scheduler  
11. Monitor¶<|im\_end|>