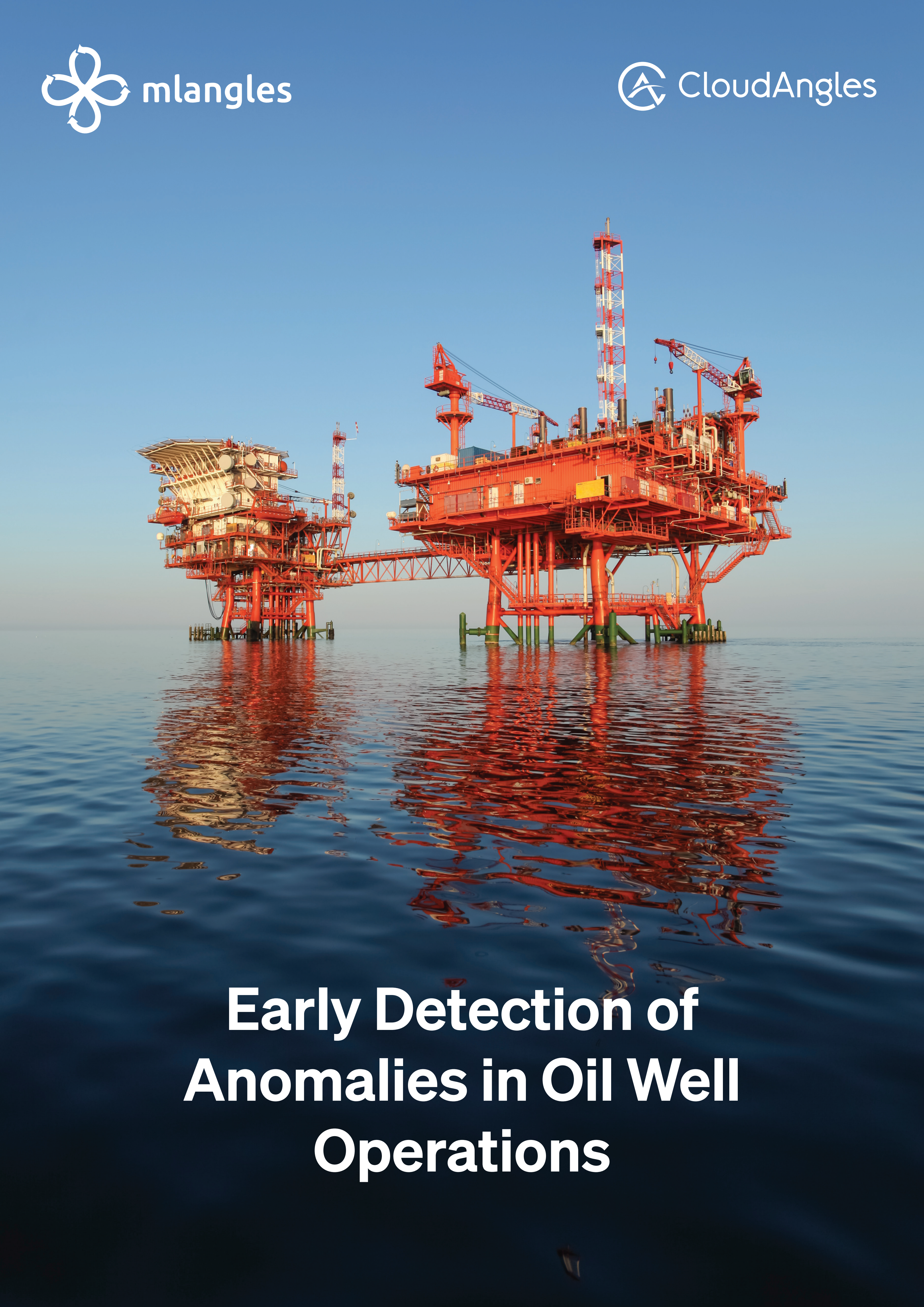


mlangles

CloudAngles

Early Detection of Anomalies in Oil Well Operations

About mlangles Predictive AI

mlangles is a comprehensive AI platform designed to manage the lifecycle of data and models, offering streamlined solutions for every stage of the process.

Through its Predictive AI component, mlangles provides a suite of tools to navigate efficiently through each phase of AI project development, encompassing data engineering, development, deployment, and monitoring. It facilitates continuous integration, continuous deployment, continuous training, continuous monitoring (CI-CD-CT-CM), enabling enterprises to effectively manage their AI initiatives.

Objective of the Use Case

The goal of this use case is to develop a machine learning model to accurately classify various undesirable events in oil wells, minimizing production losses, environmental risks, and maintenance costs.

Overview of Dataset and Use Case:

This dataset comes from a publication "A realistic and public dataset with rare undesirable real events in oil wells". It is important to note that this is also a synthetic dataset containing real, simulated, and hand-drawn instances. The initial data file contained 1,984 CSV files, each file representing an 'instance'. Each instance contained the following features: timestamp, P-PDG, P-TPT, T-TPT, P-MON-CKP, T-JUS-CKP, P-JUS-CKGL, T-JUS-CKGL, QGL, and class. Aside from timestamp and class, the remaining features are pressure and temperature data received from sensors on the oil well production line. The class feature refers to the event type. There are a total of 9 types of events/classes:

0 – Normal,

1 - Abrupt increase of BSW,

2 - Spurious closure of DHSV,

3 - Severe slugging,

4 - Flow instability,

5 - Rapid productivity loss,

6 - Quick restriction in PCK,

7 - Scaling in PCK,

8 - Hydrate in production line.

Working of Use Case:

STEP 1: Data Pipeline

Install Dependencies: Installed tsfresh package.

Extract and Process Data: Loaded data from the source (source was in S3 bucket). Also went through some preliminary data cleanup, to make data smaller and easier to handle. Dropped features P-JUS-CKGL, T-JUS-CKGL, and QGL because there were too many null values to justify any other action. Also dropped rows in class feature containing event 7 (Scaling in PCK) because there were too few instances of that event.

Extract Features: Split data into time windows of 5 minutes, and then performed feature extraction using tsfresh package. This condensed each time window into a single data entry containing 'extracted features' statistically describing the initial time window.

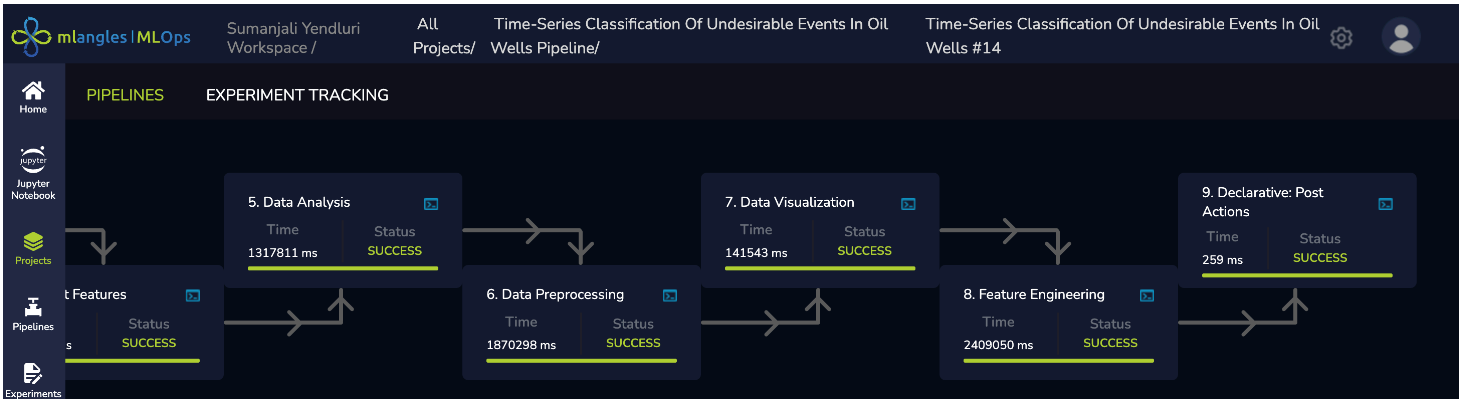
Data Analysis: Performed an analysis of the data to get an overview of the extracted data and determine whether there were any issues with the data.

Data Preprocessing: Dropped any null values

Data Visualization: Contains two plots: first, a correlation plot showing the relation between the original features and class; second, a pie chart displaying the distribution of events.

Feature Engineering: Performs normalization of data via StandardScaler.

mlangles CloudAngles



An overview of the data pipeline.

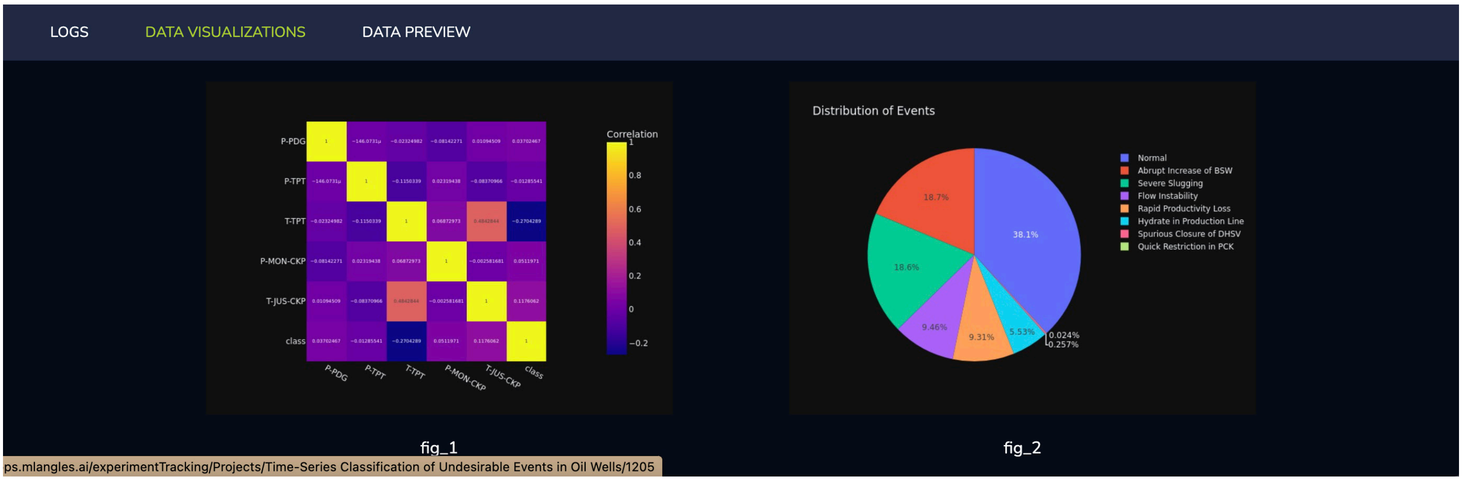
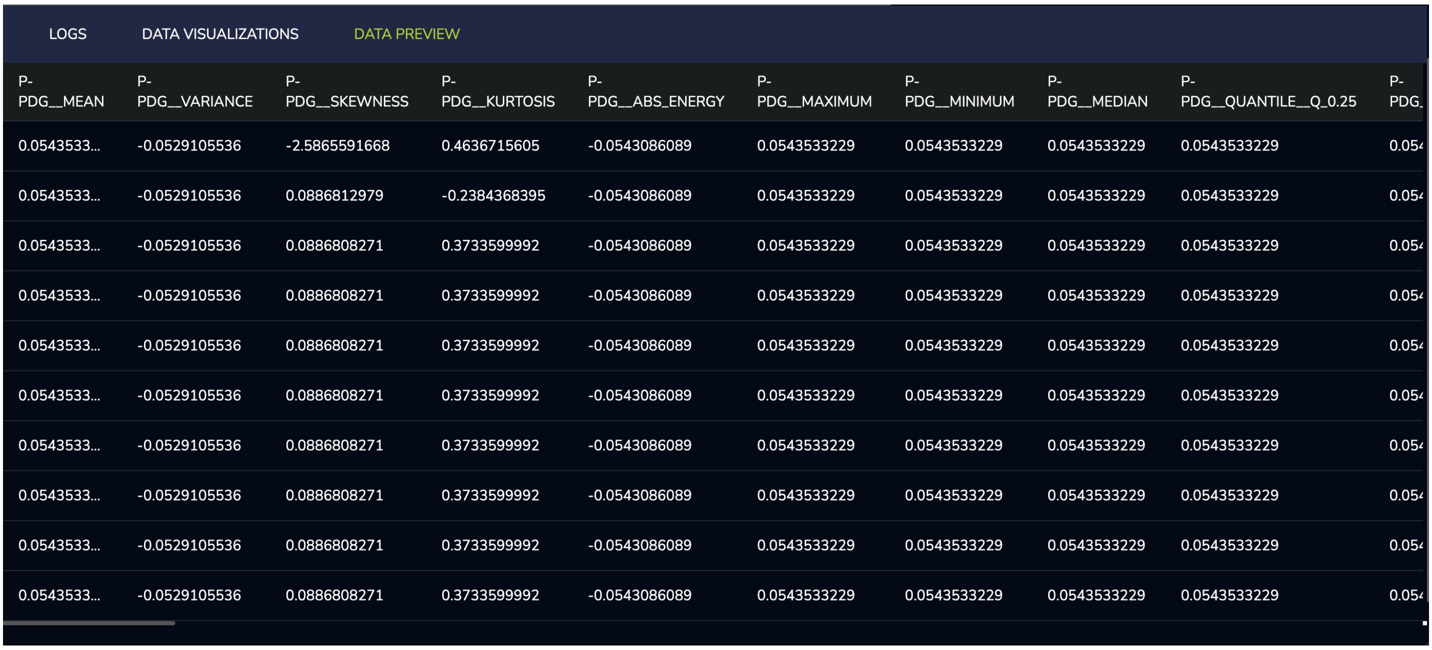


Fig_1 is a correlation plot comparing the correlations of the non-extracted features (original features).

Fig_2 is a pie chart displaying the distribution of events (Event 7 not included).

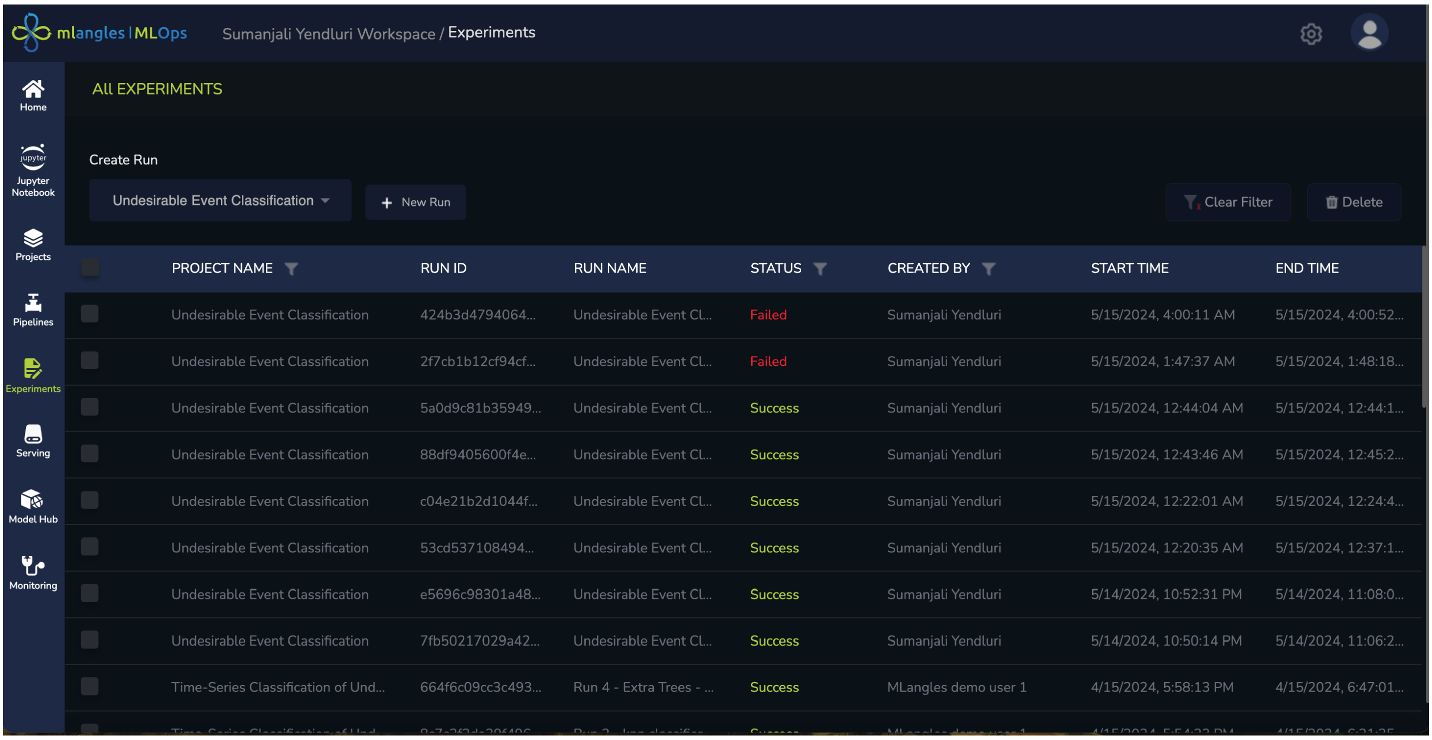


LOGS | DATA VISUALIZATIONS | DATA PREVIEW

P-PDG__MEAN	P-PDG__VARIANCE	P-PDG__SKEWNESS	P-PDG__KURTOSIS	P-PDG__ABS_ENERGY	P-PDG__MAXIMUM	P-PDG__MINIMUM	P-PDG__MEDIAN	P-PDG__QUANTILE__Q_0.25	P-PDG_
0.0543533...	-0.0529105536	-2.5865591668	0.4636715605	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886812979	-0.2384368395	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054
0.0543533...	-0.0529105536	0.0886808271	0.3733599992	-0.0543086089	0.0543533229	0.0543533229	0.0543533229	0.0543533229	0.054

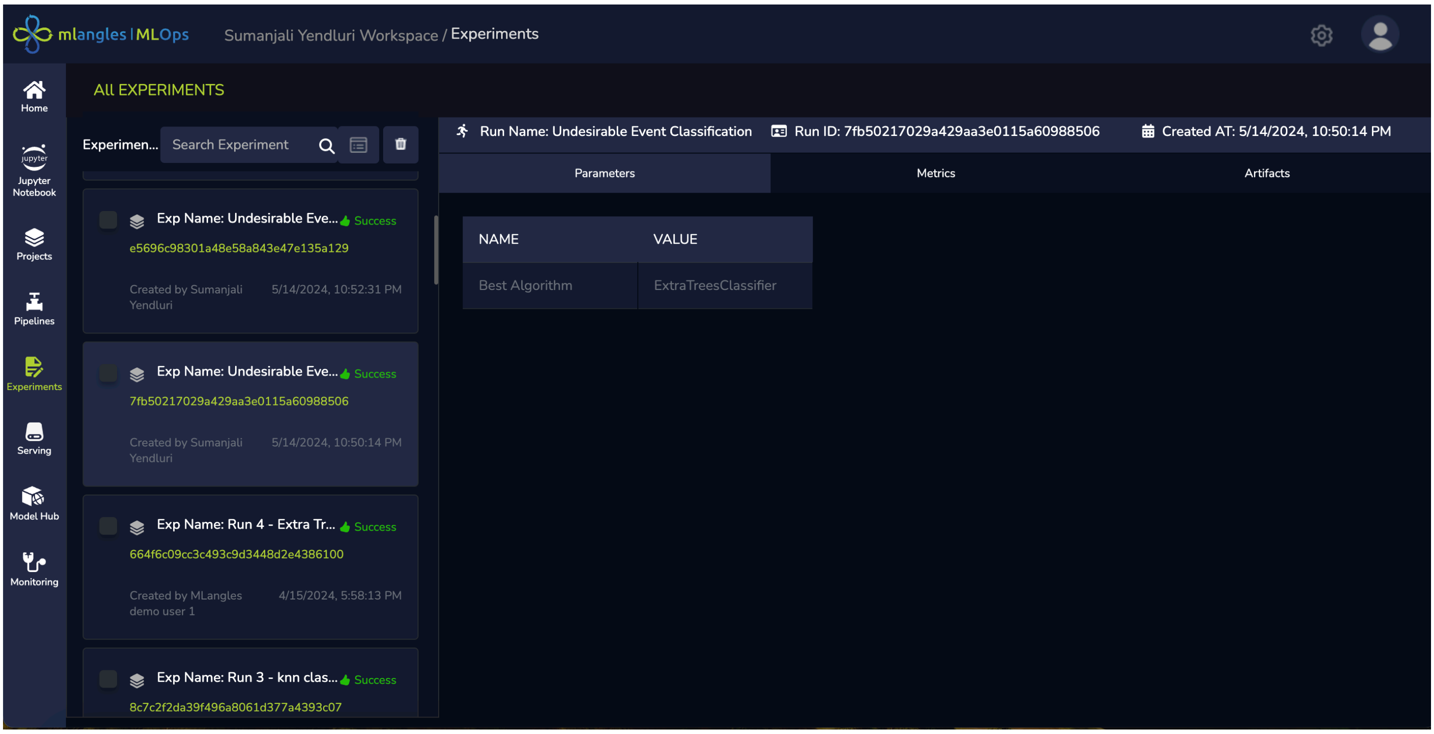
Preview of the data after the data pipeline has been run to completion.

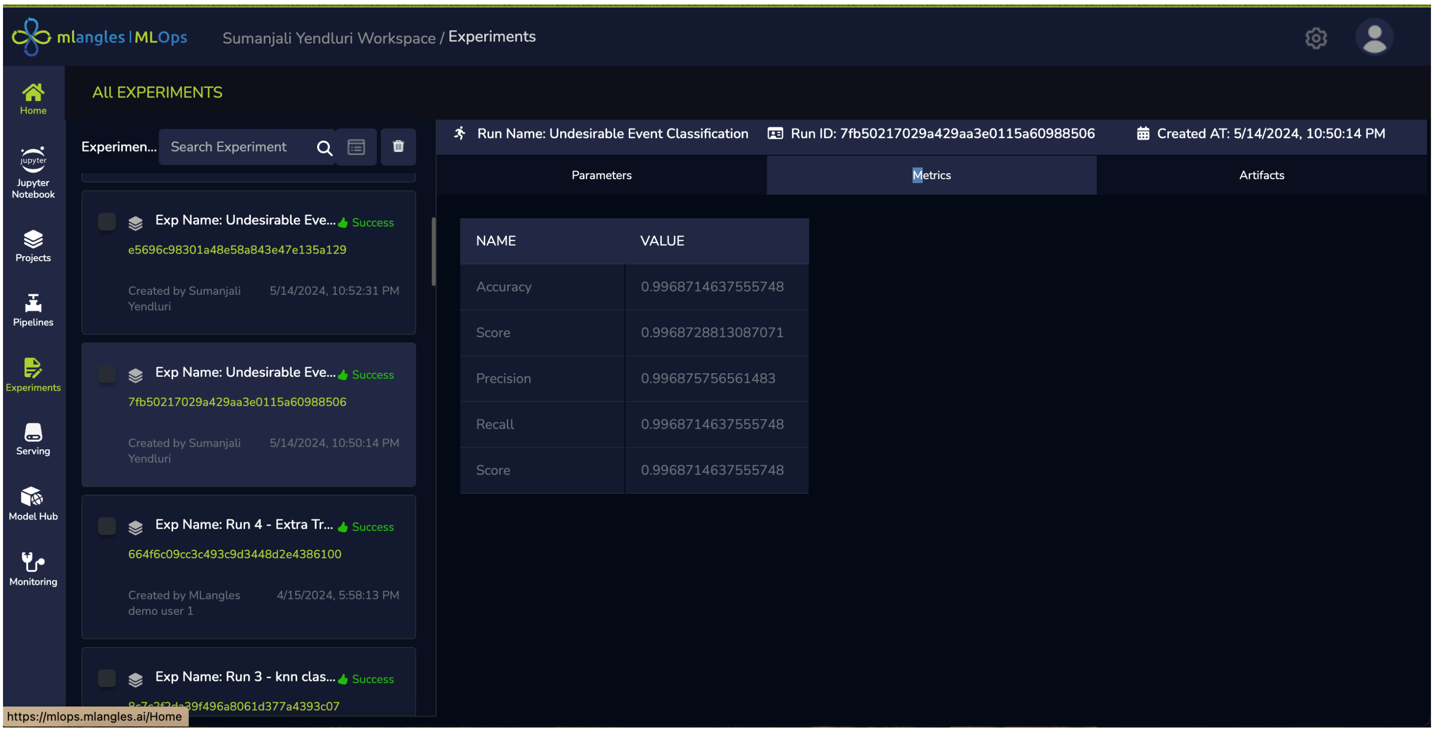
STEP 2: Experiment Tracking



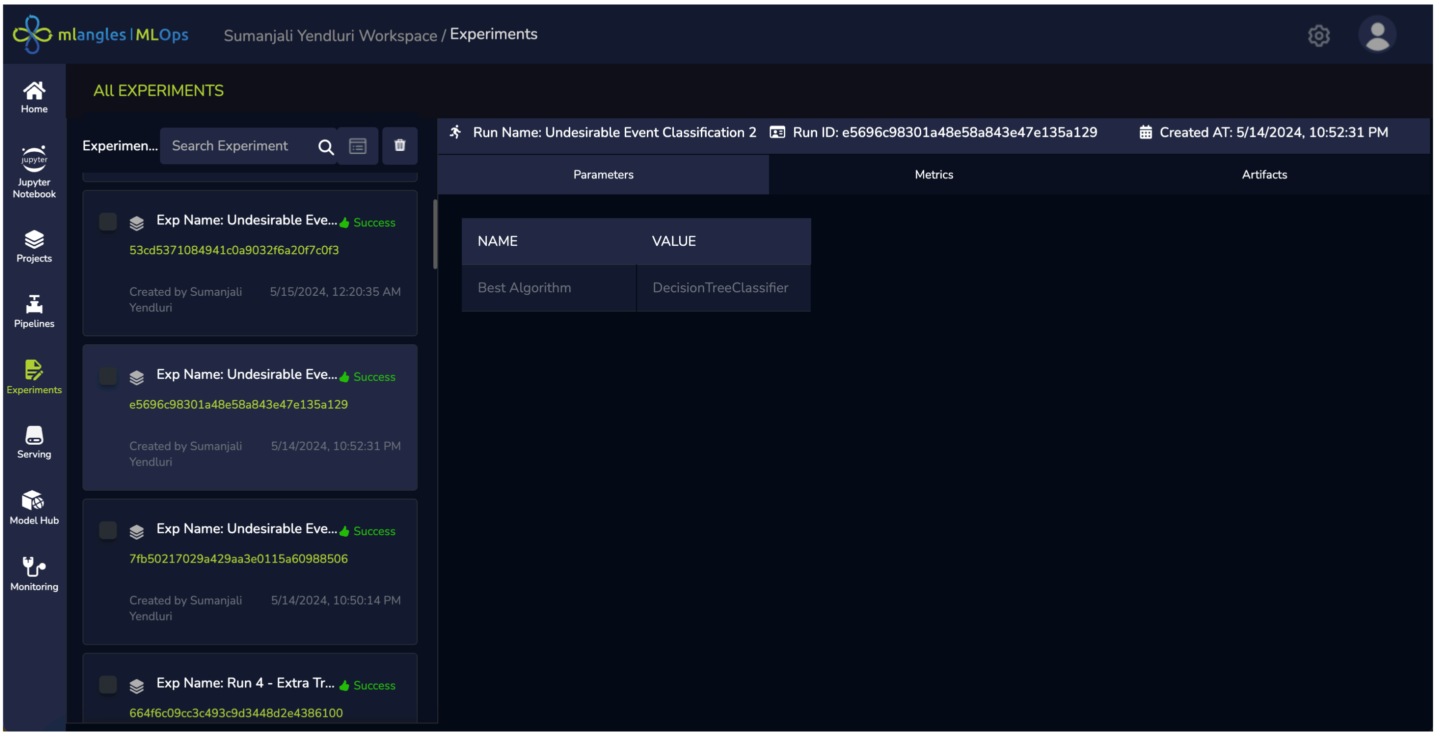
PROJECT NAME	RUN ID	RUN NAME	STATUS	CREATED BY	START TIME	END TIME
Undesirable Event Classification	424b3d4794064...	Undesirable Event Cl...	Failed	Sumanjali Yendluri	5/15/2024, 4:00:11 AM	5/15/2024, 4:00:52...
Undesirable Event Classification	2f7cb1b12cf94cf...	Undesirable Event Cl...	Failed	Sumanjali Yendluri	5/15/2024, 1:47:37 AM	5/15/2024, 1:48:18...
Undesirable Event Classification	5a0d9c81b35949...	Undesirable Event Cl...	Success	Sumanjali Yendluri	5/15/2024, 12:44:04 AM	5/15/2024, 12:44:1...
Undesirable Event Classification	88df9405600f4e...	Undesirable Event Cl...	Success	Sumanjali Yendluri	5/15/2024, 12:43:46 AM	5/15/2024, 12:45:2...
Undesirable Event Classification	c04e21b2d1044f...	Undesirable Event Cl...	Success	Sumanjali Yendluri	5/15/2024, 12:22:01 AM	5/15/2024, 12:24:4...
Undesirable Event Classification	53cd537108494...	Undesirable Event Cl...	Success	Sumanjali Yendluri	5/15/2024, 12:20:35 AM	5/15/2024, 12:37:1...
Undesirable Event Classification	e5696c98301a48...	Undesirable Event Cl...	Success	Sumanjali Yendluri	5/14/2024, 10:52:31 PM	5/14/2024, 11:08:0...
Undesirable Event Classification	7fb50217029a42...	Undesirable Event Cl...	Success	Sumanjali Yendluri	5/14/2024, 10:50:14 PM	5/14/2024, 11:06:2...
Time-Series Classification of Und...	664f6c09cc3c493...	Run 4 - Extra Trees - ...	Success	MLangles demo user 1	4/15/2024, 5:58:13 PM	4/15/2024, 6:47:01...

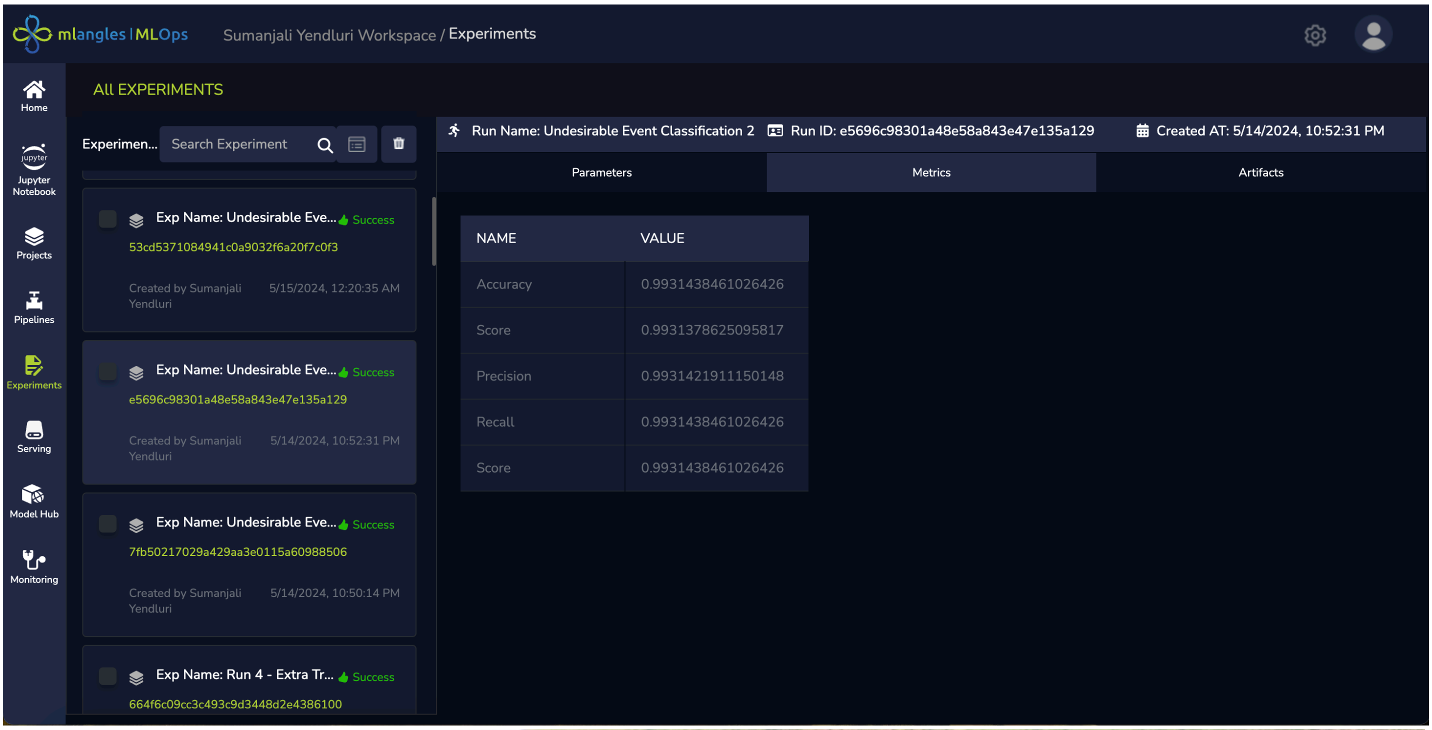
This use case is a time-series classification problem, we have used a feature extraction approach which allows us to use typical supervised classification models to classify the dataset. The following models were used: Ada Boost Classifier, Bernoulli NB, Decision Tree Classifier, Dummy Classifier, Extra Trees Classifier, Gradient Boosting Classifier, KNeighbors Classifier, Linear Discriminant Analysis, Linear SVC, and Logistic Regression. The best classifier was Extra Trees Classifier with an accuracy of 99.68%.

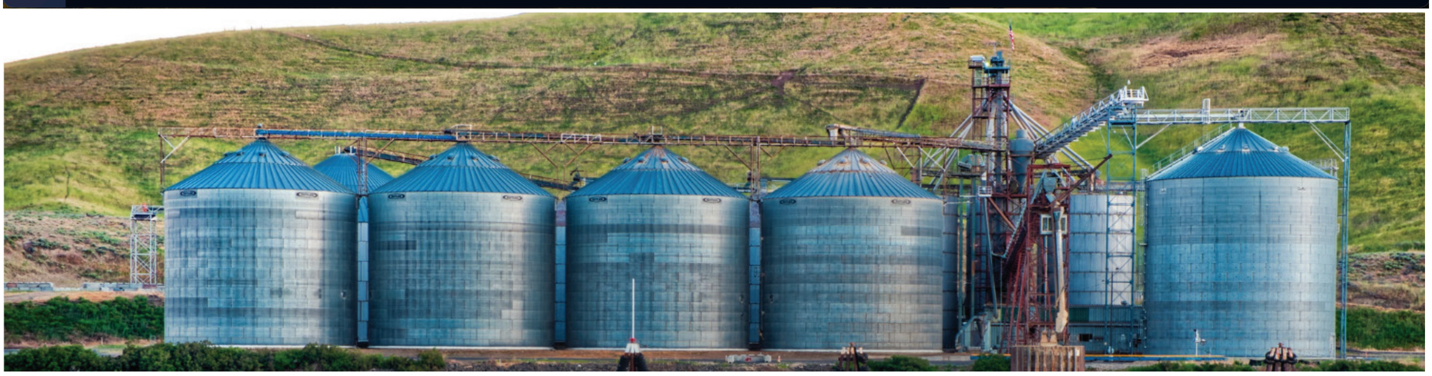


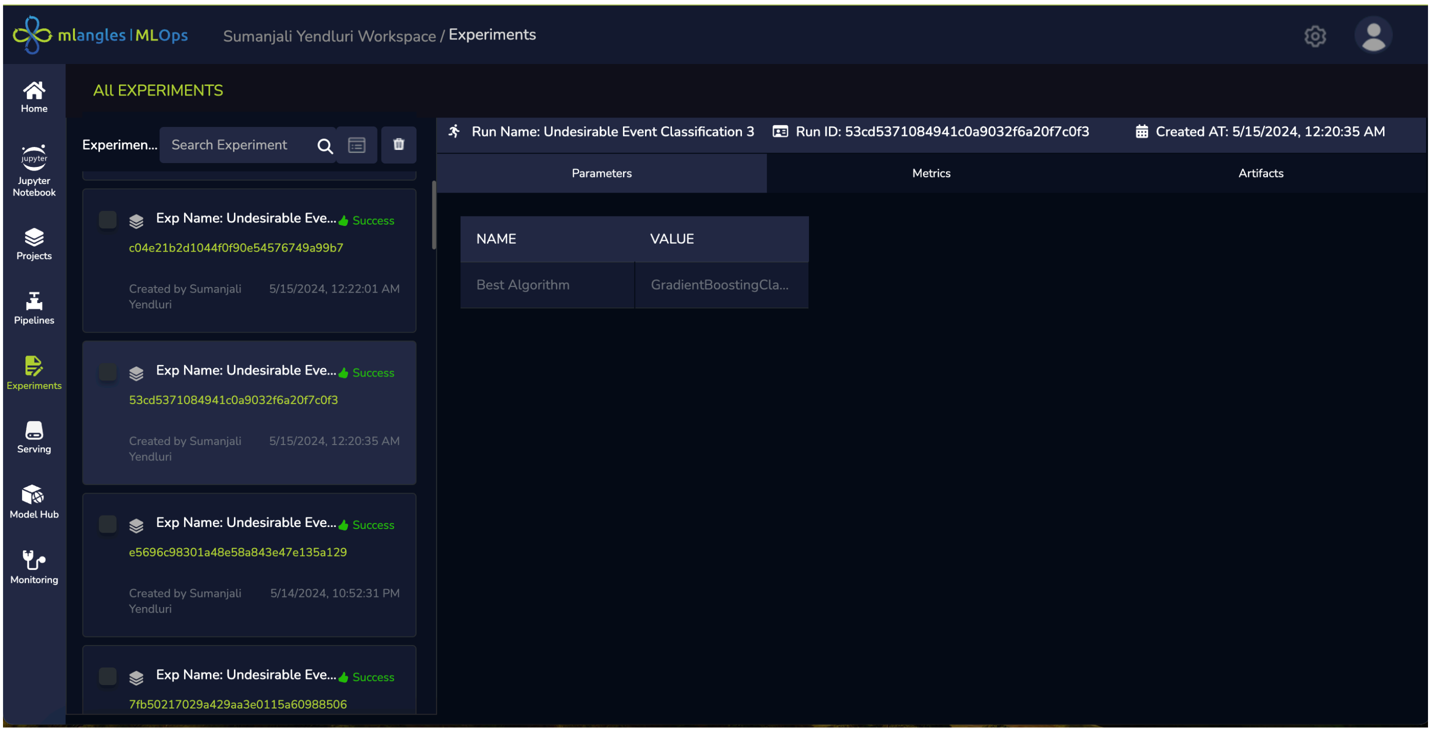


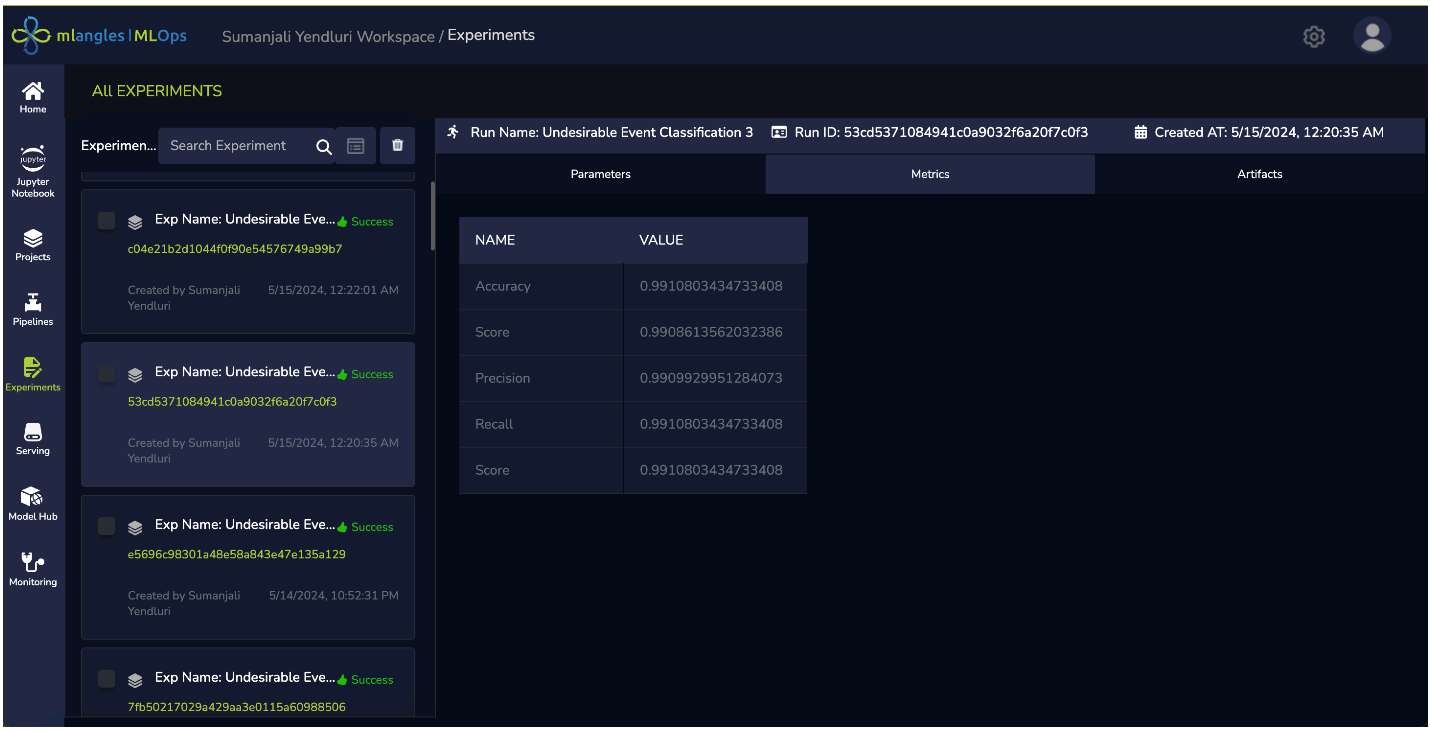
This performance was followed very closely by the Decision Tree and Gradient Boosting Classifiers.


mlangles | MLOps
Sumanjali Yendluri Workspace / Experiments
Home
Jupyter Notebook
Projects
Pipelines
Experiments
Serving
Model Hub
Monitoring
All EXPERIMENTS
Experimen...
Search Experiment
Exp Name: Undesirable Eve... Success
53cd5371084941c0a9032f6a20f7c0f3
Created by Sumanjali Yendluri 5/15/2024, 12:20:35 AM
Exp Name: Undesirable Eve... Success
e5696c98301a48e58a843e47e135a129
Created by Sumanjali Yendluri 5/14/2024, 10:52:31 PM
Exp Name: Undesirable Eve... Success
7fb50217029a429aa3e0115a60988506
Created by Sumanjali Yendluri 5/14/2024, 10:50:14 PM
Exp Name: Run 4 - Extra Tr... Success
664f6c09cc3c493c9d3448d2e4386100
Run Name: Undesirable Event Classification 2
Run ID: e5696c98301a48e58a843e47e135a129
Created AT: 5/14/2024, 10:52:31 PM
Parameters
Metrics
Artifacts
NAME
VALUE
Best Algorithm
DecisionTreeClassifier


mlangles | MLOps
Sumanjali Yendluri Workspace / Experiments
Home
Jupyter Notebook
Projects
Pipelines
Experiments
Serving
Model Hub
Monitoring
All EXPERIMENTS
Experimen...
Search Experiment
Exp Name: Undesirable Eve... Success
53cd5371084941c0a9032f6a20f7c0f3
Created by Sumanjali Yendluri 5/15/2024, 12:20:35 AM
Exp Name: Undesirable Eve... Success
e5696c98301a48e58a843e47e135a129
Created by Sumanjali Yendluri 5/14/2024, 10:52:31 PM
Exp Name: Undesirable Eve... Success
7fb50217029a429aa3e0115a60988506
Created by Sumanjali Yendluri 5/14/2024, 10:50:14 PM
Exp Name: Run 4 - Extra Tr... Success
664f6c09cc3c493c9d3448d2e4386100
Run Name: Undesirable Event Classification 2
Run ID: e5696c98301a48e58a843e47e135a129
Created AT: 5/14/2024, 10:52:31 PM
Parameters
Metrics
Artifacts
NAME
VALUE
Accuracy
0.9931438461026426
Score
0.9931378625095817
Precision
0.9931421911150148
Recall
0.9931438461026426
Score
0.9931438461026426

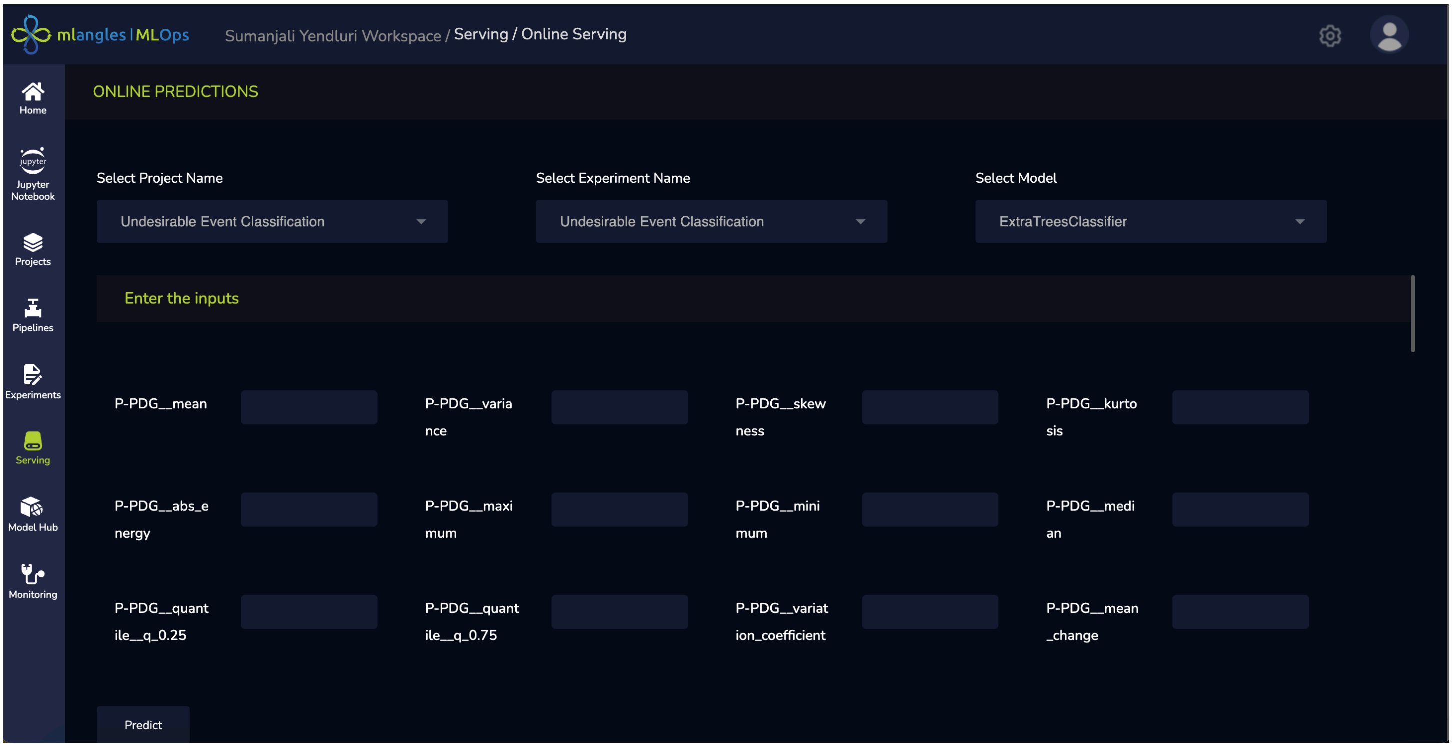






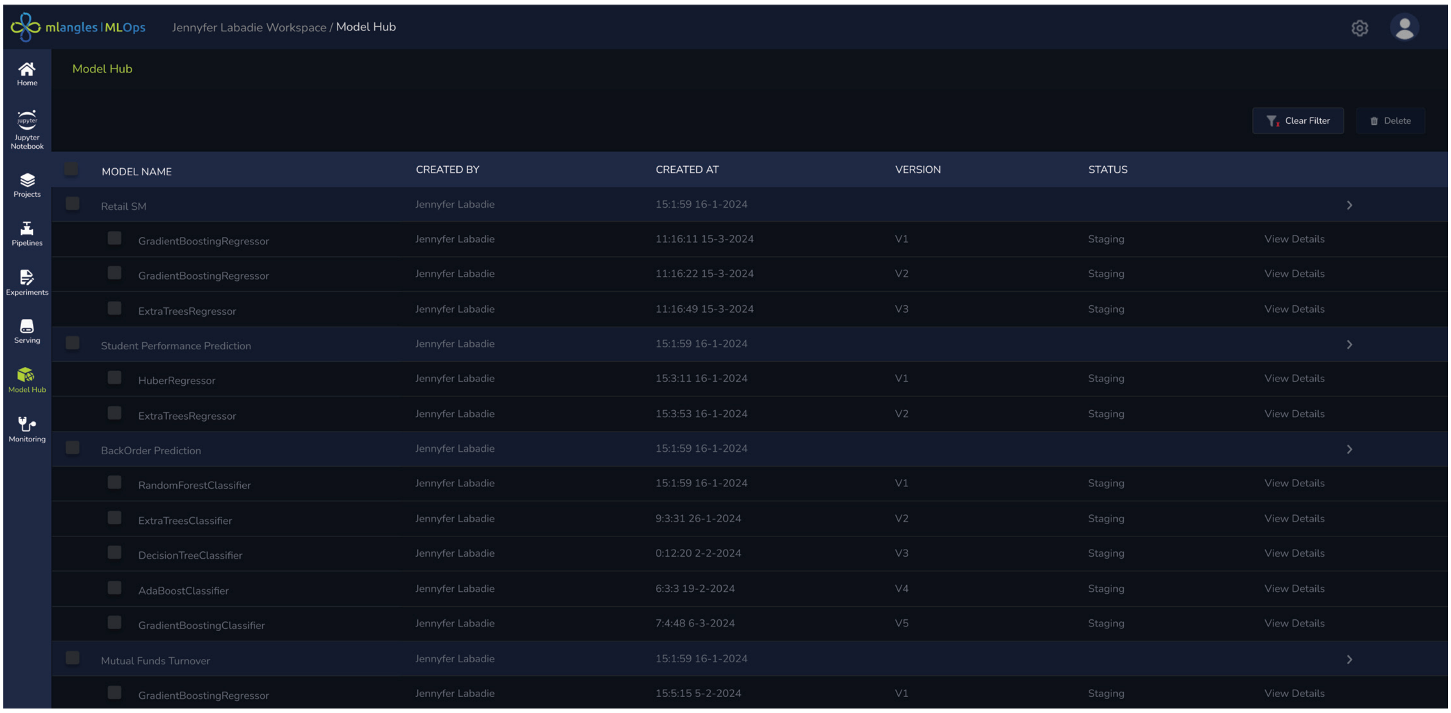
STEP 3: Serving

Serving is not possible with this dataset due to feature extraction. Feature extraction forces the original data set to go through heavy data preprocessing leaving it unrecognizable. The values that the serving needs entered are manipulated values that summarize a time-window (5 minutes in this case) from the original dataset.



Model Hub:

- Trained models are uploaded to the model hub, whereupon deployment, a REST API endpoint is automatically generated.
- Data is transmitted to this endpoint as a request, triggering the model to execute a prediction and return the output as the response to the request.



Conclusion

While we can see that feature extraction is an excellent method for time-series classification in classifying/detecting undesirable events in oil walls, it does have its drawbacks. One, the data goes through heavy data preprocessing before it can be trained for models. Two, the heavy data preprocessing leaves it difficult to use serving. As a result, we are looking into time-series classification with deep learning. While deep learning models are more complex to work with, there will be minimal data preprocessing to deal with and serving will become quite simple. Furthermore, deep learning models will not require us to split data into time windows and then condense those windows into single rows of data. Instead, due to how deep learning models already work with time-series data, we can classify data one second at a time, and past histories (time data) will theoretically reflect on the classifications better.

The use case demonstrates the effectiveness of machine learning for classifying undesirable events in oil well operations. By transitioning to deep learning models, we can potentially achieve real-time anomaly detection, leading to significant improvements in operational efficiency and safety.

To setup a demo

Info.mlangles@cloudangles.com

Visit: www.mlangles.ai