Explainable AI for ML Ops

Sandeep Pathak, Harsh Shah, Zihao Zhou

pathak.sandeep@tcs.com, s.harsh4@tcs.com, zhou.zihao@tcs.com

Analytics & Insights

1. Executive Summary

This white paper explains the significance of blending two emerging technologies in AI/ML – Explainable AI (XAI) and Machine Learning Operations (ML Ops) and demonstrates a focused use case that derives value from leveraging XAI to enhance ML Ops.

The white paper starts by laying out the "growing pains" problems that enterprises are encountering to scale AI. We highlight the relatively low maturity of post-production ML processes thus exposing enterprises to reputation, compliance and hence financial risk. We give a historical perspective on the rise of AI. Subsequent gain in mindshare of ML Ops is explained. XAI as a solution is introduced. After a brief explanation on XAI, we delve into the experimental setup and detail the results which demonstrate the potential of using XAI to enhance ML Ops. We end the white paper by reiterating the benefits, opportunities and market potential and finally some recommendations.

2. The Problem

ML and the "Last Mile" Problem

The rapid adoption of AI has also brought to the forefront the myriad growing pains that mirror the issues impacting enterprise software years ago – development across multiple teams, time to market, product quality, scaling up, monitoring, to name a few. Similar issues impact AI as it strives to scale up and across. There are products addressing the data engineering spectrum of the AI pipeline with promising commercial entrants like Snowflake. Vendors like Dataiku, H2O and cloud platform offerings like SageMaker are helping 'democratize' AI by offering 'operationalization' of ML models also known as "ML Ops" capabilities and end-to-end UI based centralized platforms to meet the needs of all stakeholders from data scientists to business analysts. This helps mitigate the 'friction' from data cleansing, model training, model management to model deployment, monitoring and feedback.

The focus changes from technical challenges to business value in post-production. Are the models making the predictions and recommendations as expected? How is the notion of 'model drift' handled? What about 'data drift' – changes to profile of data sets either slow or abrupt or data 'poisoning' attacks. Are the model monitoring processes capable of alerting before it is "too late"? The notion of 'late' implies models too slow to adapt to changing data profiles which means lost business opportunities. It also means not realizing negative impacts of predictions thus exposing enterprises to business and reputational risk. There are many instances where AI predictions have gone awry (or 'rogue') and they were pulled back after negatively impacting the use case (Analytics India, 2017; Simonite 2020).

ML Ops solutions, as of now, focus on improving the ML 'experience', democratizing AI to non-data scientists. There is a gap in focus on the "last mile" aspect of the AI/ML lifecycle – ML Monitoring.

Keeping Tabs on the Model

ML practices offer the notion of measuring 'model drift' i.e. a shift in the prediction power of the model. Standard measures like precision, recall, accuracy, AUC etc. can be used to keep track of model performance and can indicate when a model needs to be retrained. But this does not offer traceability all the way back to features, data, and reasoning. There is a very real possibility that the measures will not raise alarms while there has been a shift in the properties of data. The pandemic is perhaps the biggest trigger that has caused upheaval in many patterns in socio-economic behavior and data. Industries like mortgage lending had to keep consistent lending practices but respond to changes in conditions (e.g., how to do home inspections remotely?).

It is thus important that ML Monitoring be expanded to include capabilities that enable 'deep' monitoring i.e., not stop at validating labels but trace deviations back to features and data.

Explainable AI for Model Monitoring

The branch of AI called Explainable AI (XAI) holds promise here. As the name implies, XAI research and resulting frameworks like – ELI5, LIME, SHAP help determine feature importance and impact in the model results. A XAI framework will indicate which feature contributed how much in the model's predicted result (Google, n.d.). This is a direct use of XAI. This ability of XAI can be extended to improve ML monitoring. In instances where the standard measures fail to indicate a significant drift in the model, XAI can be used to indicate 'feature drift' and help identify a change in constituency of ML results.

This is even more important in the COVID era where pre COVID, COVID and post COVID conditions are resulting in significant changes to data characteristics.

3. The History

AI/ML Maturity

The explosive growth in AI/ML adoption is fueled by many factors - adopters seeking to assimilate AI into their DNA seek to get benefits in key areas like reducing costs, generate better customer insights and intelligence, reduce friction in customer experience (Algorithmia, 2020). While there is no dearth of issues in this AI/ML "wish list" that corporations aspire to address, the ML maturity of many organizations needed to bring this wish list to fruition is very much a work in progress. Very few organizations are at the highly mature end of the spectrum. The high maturity is typically explained in many ways (Algorithmia, 2020):

- Models in production for a long duration (years)
- Clear ML strategy alignment between management and analytics teams
- Traceability between investments and benefits from ML model predictions or recommendations i.e. a good handle on AI/ML ROI
- Mature data and machine learning pipeline
- Established operationalized processes around deployment and monitoring of AI models
- Model governance
- Maturity to buy third-party models and customize/train to enterprise needs (e.g. ResNet)

Rise of ML Ops

However, enterprises are much behind this high maturity level with only about half of them with any models in production (Algorithmia, 2020). Also just deploying to production is not enough. It will matter how the deployment is done, how soon is it done and how the models are being monitored. The 'how' and 'soon' aspects of ML development are addressed by instituting an extension of DevOps software engineering principles to encompass data and models used in ML; also known as "ML Ops". ML Ops strives to extend the aspirations of software engineering to machine learning (McKnight, 2020), namely:

- Faster turnaround time all the way from data cleansing to model deployment
- Facilitate faster rate of experimentation and adoption
- Governance, assurance of data quality and ethical AI

DevOps has also matured in the area of monitoring. Many frameworks like OpenShift offer monitoring or work seamlessly with tools ideal for monitoring (Witzenman, 2017). Open-source tools like Grafana have become popular to create dashboard to visualize metrics. Compared to the DevOps, ML Ops tools landscape is very much a work in progress and not surprising, given the fairly recent spike in AI/ML usage after being dormant for decades. The spectrum of ML Ops space has a few promising open-source candidates on one end and then native support by the leading cloud vendors. The commercials off-the-shelf (COTS) space is still sparse. In the open-source space, frameworks like Kubeflow, ML Flow and recently launched ZenML show promise and are being leveraged (Vizard, 2021). In the cloud landscape the three prominent vendors – AWS, Azure and Google Cloud all have stated support for ML Ops via proprietary offerings incorporated into their platform offerings – e.g. ML Ops via SageMaker on AWS. In the commercial space, there are products being launched at a brisk pace – Neptune.ai, Fiddler.ai that offer monitoring. Also, ML development platforms like Dataiku, H2O and Databricks focus on enabling the ML development to deployment pipeline i.e. a subset of the holistic ML Ops end-to-end flow.

Overall, the MLOps market is still in its early stages, with technology solutions emerging only in the last year or two for effective model management. Growing pains and maturity notwithstanding, the market potential for ML Ops is pegged to reach $4 Billion (products and services) in a span of a few years and ML Ops will be an intrinsic part of any organization's AI strategy (Schmeltzer, 2020).

ML Ops in Postproduction

In the previous section we talked about the rapid evolution of ML Ops akin to the rapid adoption of DevOps to address similar issues –industrialization of AI/ML to offer rapid turnaround, reduce time to market, reliability and operations predictability.

Machine Learning poses its own unique challenges in terms of "Business as Usual (BAU)". Unlike software where the logic is deterministic, a ML model is highly dependent on data. Significant changes in macro conditions (COVID) will likely change characteristics of data. This will cause models to decay and "drift" away from their initial training and hence expected behavior requiring retraining (Schmeltzer, 2020). Organizations deploying models into production need to extend their notion of ML Ops from deployment to incorporate continuous model monitoring. Monitoring metrics should indicate a trigger to retrain the model and address downstream risks.

4. The Solution

Explainable AI 101

The rapid adoption of AI has also brought with it a surge in aspects of AI like safety, fairness or bias and interpretability or explainability. The increasingly black-box nature of complex models makes it difficult to test in the traditional notion of software testing, though opinion is divided on this. The model-based decisions still need to enable enterprises to abide by regulations and ensure fairness in decision making. If the decisions of these models are 'explainable' then it brings transparency.

Explainable AI refers to techniques that helps to make AI solutions human understandable (Doshi-Velez & Kim, 2017). XAI techniques add trustworthiness, auditability and credibility to our algorithmic solutions. This has taken on significance as standards and regulations are being developed, especially in EU to report explainability and demonstrate fairness in AI (European Parliament, 2019).

With the rise and demand of Explainable AI there have been many frameworks developed to advance our understanding of ML predictions (Analytics India, 2019; DeepFindr, 2020). These frameworks are categorized by their ability to do local (explains each observation/row as required) and global (explains the importance of model features in general) interpretability of ML outcomes. Some libraries are:

- SHAP – SHapley Additive exPlanations: game theory principles to explain individual as well as global predictions.
- LIME – Local Interpretable Model Agnostic Explanations: explain individual predictions
- What-If Tool – Visualizer for Tensorflow data
- ELI5 – Explain Like I am 5: for simpler models
- AIX360 – AI Explainability 360: new library. Extensible.
- Skater – new library for local and global explanations.

All of these techniques have specific advantages. We chose to work with SHAP considering its accuracy, compute optimization, open-source framework and variety of explainers offered for different machine learning algorithms making SHAP the most popular XAI library currently in use. Note that material is only recently made available which detail and compare XAI frameworks (Molnar, 2021).

SHAP is based on Shapley values using coalitional game theory to distribute payouts from a game. Like game theory, the goal of SHAP is to explain the prediction of an instance 'X' by computing the contribution of each feature prediction (Lundberg & Lee, 2017). SHAP offers explainers that can support local as well as global explainability based on Shapely values. The explainers are developed specifically for each or a family of ML model types. SHAP authors proposed KernelSHAP an alternative, kernel-based (can be used with any machine learning algorithm) estimation approach for Shapely values inspired by local surrogate methods. They also proposed TreeSHAP (optimized for tree-based algorithms), an efficient estimation approach for tree-based models.

Explainability and ML Monitoring

Our proposition is to extend usage of XAI not only for debugging and exploration but also for real time monitoring as it can help indicate deviations from model prediction to data reality. Question becomes, how can we implement it? What are the correct indicators pointing the change or 'poisoning' in the test data?

We were able to get a substantial indication about change in the test data using these two approaches:

1. Interpreting Local explainability for similar type of users over time
2. Tracking SHAP loss for each iteration of prediction that takes place

The flow of our experiment is as follows: Upon model deployment, we will monitor metrics from our model output that suggest how well the model is performing. We will show how the same metric might not indicate changes in data. At this point, SHAP loss metric will be utilized to show how it can play a vital role during model monitoring. We will discuss more about both of these approaches in our "Experiment" section.

Setup

Dataset

For this case study we decided to use a dataset from Kaggle that contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

Data Source - https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset

There are 25 variables in the dataset:

- ID: ID of each client
- LIMIT_BAL: Amount of given credit in NT dollars (includes individual and family/supplementary credit
- SEX: Gender (1=male, 2=female)
- EDUCATION: (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
- MARRIAGE: Marital status (1=married, 2=single, 3=others)
- AGE: Age in years
- PAY_1: Repayment status in September, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, 8=payment delay for eight months, 9=payment delay for nine months and above)
- PAY_2: Repayment status in August, 2005 (scale same as above)
- PAY_3: Repayment status in July, 2005 (scale same as above)
- PAY_4: Repayment status in June, 2005 (scale same as above)
- PAY_5: Repayment status in May, 2005 (scale same as above)
- PAY_6: Repayment status in April, 2005 (scale same as above)
- BILL_AMT1: Amount of bill statement in September, 2005 (NT dollar)
- BILL_AMT2: Amount of bill statement in August, 2005 (NT dollar)
- BILL_AMT3: Amount of bill statement in July, 2005 (NT dollar)
- BILL_AMT4: Amount of bill statement in June, 2005 (NT dollar)
- BILL_AMT5: Amount of bill statement in May, 2005 (NT dollar)
- BILL_AMT6: Amount of bill statement in April, 2005 (NT dollar)
- PAY_AMT1: Amount of previous payment in September, 2005 (NT dollar)
- PAY_AMT2: Amount of previous payment in August, 2005 (NT dollar)
- PAY_AMT3: Amount of previous payment in July, 2005 (NT dollar)
- PAY_AMT4: Amount of previous payment in June, 2005 (NT dollar)
- PAY_AMT5: Amount of previous payment in May, 2005 (NT dollar)
- PAY_AMT6: Amount of previous payment in April, 2005 (NT dollar)
- Default_Target: Default payment (1=yes, 0=no)

Exploratory data analysis

Before our experiment we took the following EDA steps:

- Missing value check
- Numerical and categorical variables segregation

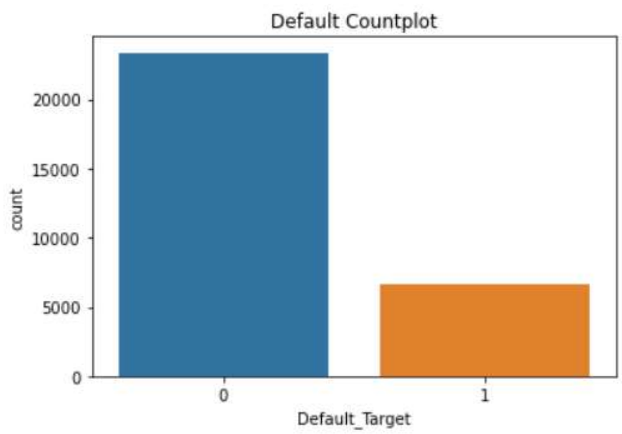
- Plotting numerical feature distribution
- Bi-variate analysis with dependent variable
- Dropping duplicate rows/columns
- Outlier treatment
- Multivariate correlation check

Here are some highlights from our detailed exploratory data analysis,

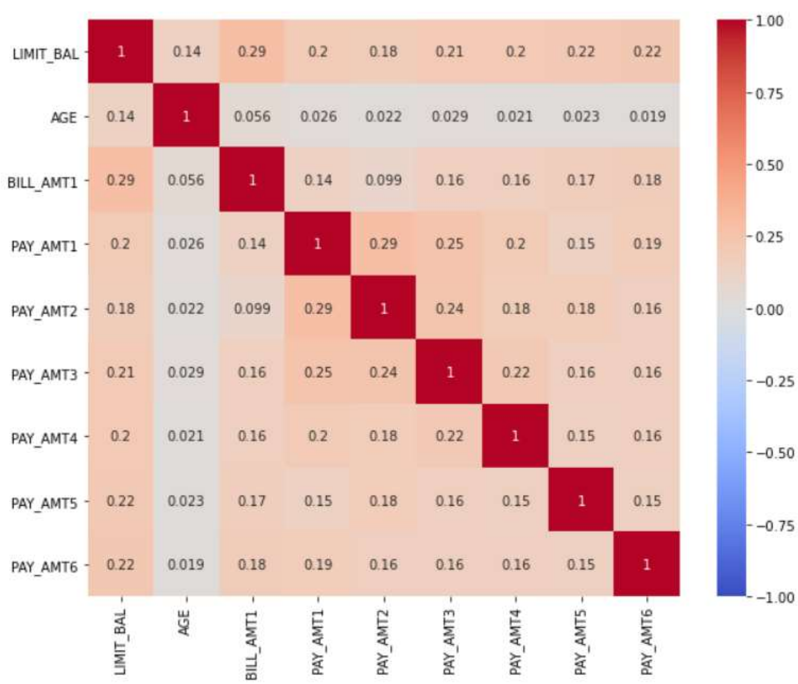
Missing values

None of the columns have any missing values and data set is complete.

Target variable Count plot



Heatmap



	LIMIT_BAL	AGE	BILL_AMT1	PAY_AMT1	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6
LIMIT_BAL	1	0.14	0.29	0.2	0.18	0.21	0.2	0.22	0.22
AGE	0.14	1	0.056	0.026	0.022	0.029	0.021	0.023	0.019
BILL_AMT1	0.29	0.056	1	0.14	0.099	0.16	0.16	0.17	0.18
PAY_AMT1	0.2	0.026	0.14	1	0.29	0.25	0.2	0.15	0.19
PAY_AMT2	0.18	0.022	0.099	0.29	1	0.24	0.18	0.18	0.16
PAY_AMT3	0.21	0.029	0.16	0.25	0.24	1	0.22	0.16	0.16
PAY_AMT4	0.2	0.021	0.16	0.2	0.18	0.22	1	0.15	0.16
PAY_AMT5	0.22	0.023	0.17	0.15	0.18	0.16	0.15	1	0.15
PAY_AMT6	0.22	0.019	0.18	0.19	0.16	0.16	0.16	0.15	1

Conclusions based on the EDA

- There are some unknown and undocumented categories in variables MARRIAGE, EDUCATION which needs to be combined with OTHERS
- As per the Data Definition PAY_* variables start with -1(e.g., Pay duly). However, there are values -2 and 0 which also needs to be assumed as Pay Duly
- Many customers are having very high LIMIT BALANCE compared to rest of the customer base which will be treated as outliers.
- Heatmap clearly suggests extremely high CORRELATION between variables BILL_AMT*, which either needs to be dropped or converted to a Principal component
- Target variable is imbalanced in the ratio of 1:3.5 (Default: Not Default)

Experiment

We trained and conducted explainability on six different machine learning models ranging from Logistic Regression to Deep Neural Networks. Out of six models, XGBoost had best results according to confusion matrix. For the purpose of model monitoring, the model performance details of all the other models are not relevant. We will continue with XGBoost model from here on.

This experiment will be explained in steps to be able to clearly showcase each step of model monitoring and how we determine change/manipulation in test data.

Load and train the XGBoost model

After training XGBoost model for the unbalanced data set with the ratio of 1:3.5 (default: not default), we logged standard classification metrics like classification report, confusion matrix and AUC score to determine the results.

```
Cross-Validation score for Xgboost:0.7915160194821584
--------------------------------------------------
Predicting the outcomes using Xgboost
..................................................
Classification report using Xgboost

              precision    recall  f1-score   support

           0       0.81      0.87      0.84      5373
           1       0.58      0.47      0.52      2006

    accuracy                           0.76      7379
   macro avg       0.70      0.67      0.68      7379
weighted avg       0.75      0.76      0.75      7379

--------------------------------------------------
Confusion matrix using Xgboost

[[4687  686]
 [1064  942]]
--------------------------------------------------
ROC AUC score using Xgboost

0.670957906106731
```

Train Tree Explainer from SHAP using trained XGBoost model

```
explainer = shap.TreeExplainer(best_model)

shap_values = explainer.shap_values(X_test)
```

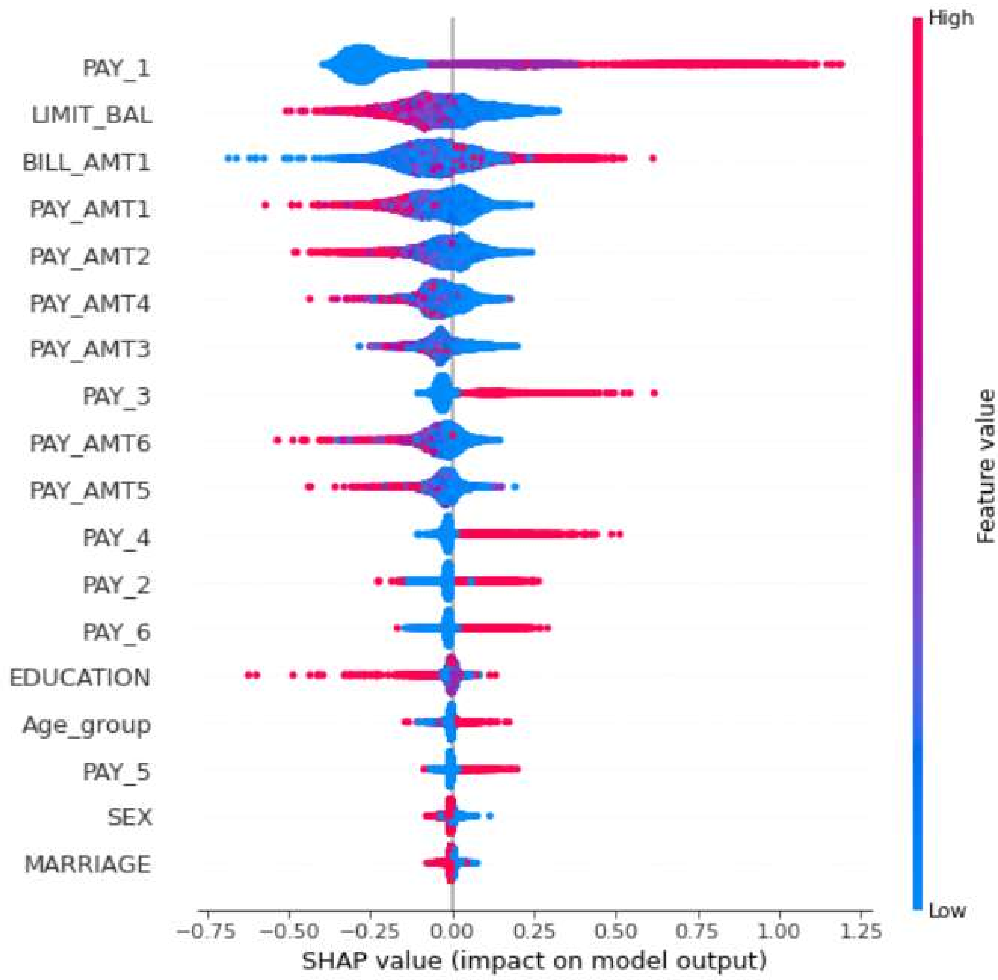
SHAP Summary plot

```
shap.summary_plot(shap_values, X_test)
```

SHAP summary plot gives an overview of which features are most important for a model and the feature effect. The plot is for SHAP values over all samples. The plot shows the following information (Molnar, 2021):

- Feature importance: features are sorted in descending order
- Feature impact: the horizontal location shows if the feature had a higher or lower prediction impact (magnitude).
- Contribution: the color represents the 'push/pull' impact on the prediction i.e. red represents towards (or higher) the prediction and blue represents away (or lower).

This reveals for example that a high **PAY_1** pushes the prediction towards **Default.** It clearly indicates that **PAY_1** and **LIMIT_BAL** are the two most important features which has the highest impact on our predictions.



SOLUTION 1: Local explanation with one observation

For this particular step, we will pull out one particular observation out of our test data as that would explain features for a particular customer.

Observation Features:

	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	PAY_1	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1	PAY_AMT1	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6	Age_group
7330	240000.0	2.0	2.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	71876.0	3100.0	2400.0	2367.0	2435.0	2500.0	3000.0	0.0

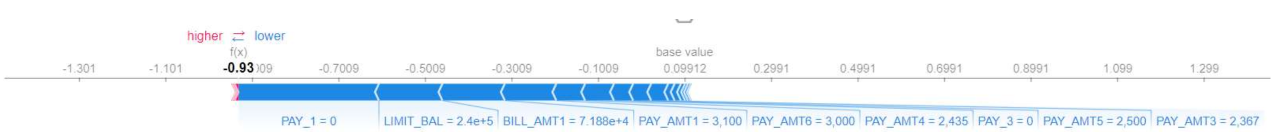
Here, the variable **Pay_1 =2,** which indicates that the customer is 2 months behind in his payments. **Bill_Amt1 = 71876,** meaning the customer's outstanding amount.

Our model in normal scenario's predicted that this customer will **DEFAULT,** and rightly so. The local explanation summary plot looks like the following graph where we can understand how each feature is pushing the prediction either towards **DEFAULT** or **NOT DEFAULT**.

In this plot, **PAY_1** is the most contributing feature pushing the prediction towards **DEFAULT** while **LIMIT_BAL** is pushing towards **NOT DEFAULT.**

Introduce data change by swapping Pay_1 with Pay_6 for this observation

After introducing this change in the observation, we noticed a distinct change. **Pay_1** was still the most significant contributor of the prediction but this time it was pushing the prediction towards **NOT DEFAULT.**



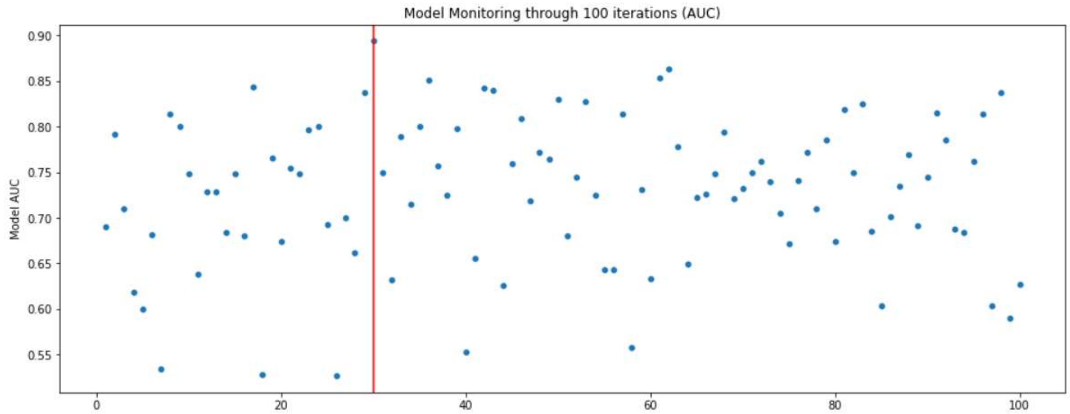
This change in local explainability alerts us to look into the data and investigate further to either inspect the data credibility and quality or to retrain model. This is the first approach among the two to monitor the model using XAI.

SOLUTION 2: Global monitoring: Iterating the model 100 times, introduce the manipulation from 30th iteration

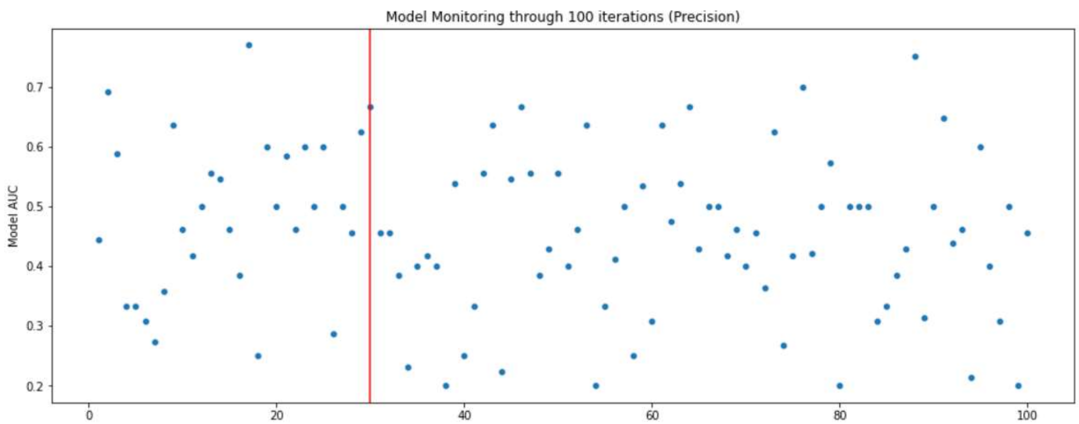
The objective of this step is to prove that primary metrics like **AUC & Precision** might still remain similar after completely swapping two features (**PAY_1, PAY_6)**. Whereas, **SHAP loss** would picture a completely different story indicating the change in data from 30th iteration itself.

We ran the model for 100 iterations, bootstrapping 50 observations on each iteration from the test data. From the 30th iteration, the data change was introduced as explained.

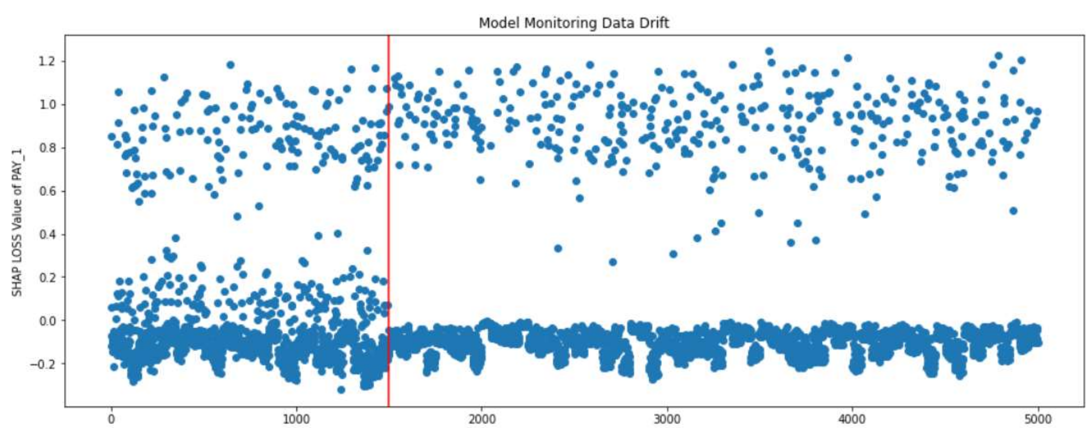
Plot AUC over the iterations:



Plot Precision over the iterations:



Plot SHAP loss over the iterations:



Conclusion

The results from model metrics and distinct change in SHAP loss monitoring clearly demonstrate that while model monitoring would not have indicated that anything is amiss, the XAI monitoring clearly indicates a change in data. This should suggest to the ML Ops team to investigate the test data further to identify either a data quality issue OR the need for retraining the model.

5. The Benefits

In a PwC survey over 67% of business leaders indicated that *"AI and Automation will impact negatively on shareholder trust levels in the next five years"* (PwC, 2017). AI, through deep learning and reinforcement learning is becoming increasingly sophisticated and evolving into an algorithmic black box. Enterprises need to be able to place their belief in the outcomes of these complex models. They need to be able to instill trust in their customers and shareholders. The model outputs also need to comply with regulations to demonstrate fairness and lack of bias. This ability needs to be available not only when models are approved but post-production too.

The need for real-time predictions is increasingly becoming important. In such a context, it is critical that enterprises enhance their ML Ops capabilities with real-time XAI frameworks. The ability to have multiple levels of model behavior validation will prove invaluable. The XAI capability to indicate feature changes will greatly facilitate data quality and validation processes. It will also be an integral part of cyber security where 'data poisoning' attempts can be monitored.

It is thus highly recommended that readers put XAI at the top of their list of AI/ML topics to understand and gain expertise in. The benefits of XAI in compliance are getting mainstream. The potential in ML Ops needs to be understood and exploited.

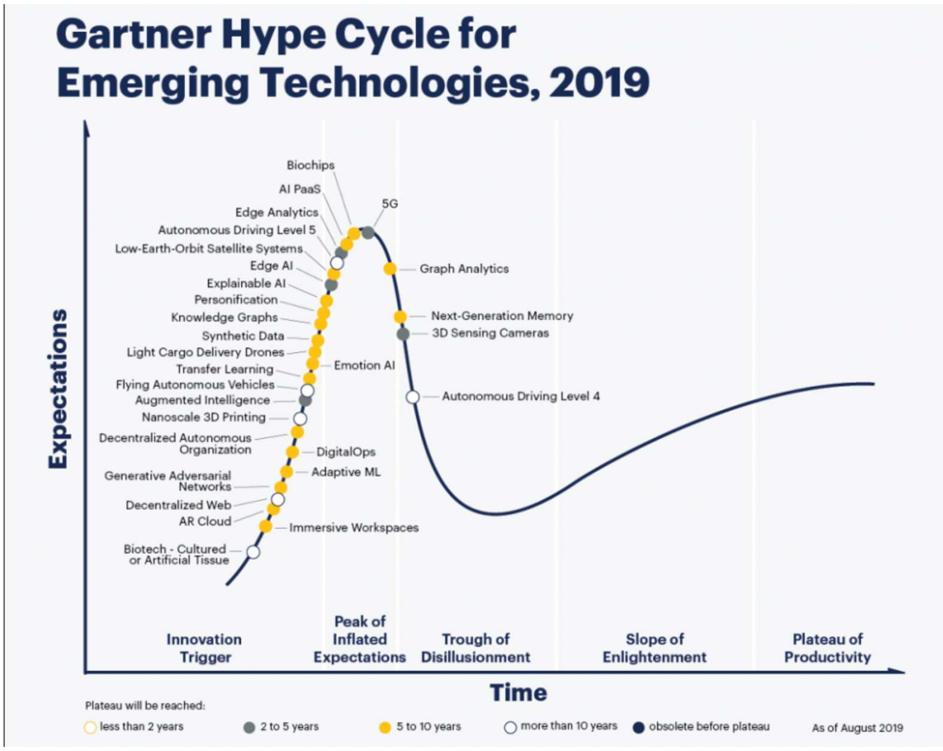
6. The Call-To-Action

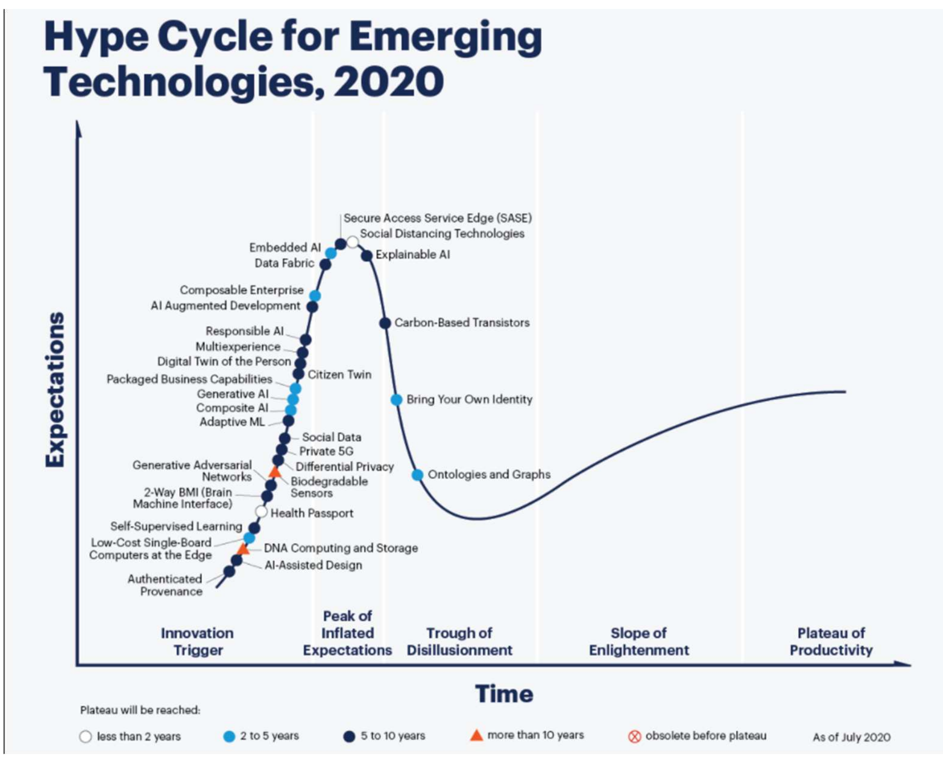
NOTE: *This section is written from a perspective of benefiting TCS A&I.*

AI is on its way to be a trigger for driving GDP worldwide to about $15 trillion by 2030 (PwC, 2017). It is also predicted that 30% of government and large enterprise contracts for AI product and services procurement will require some element of explainable and/or ethical AI (Haranas, 2019). It has already been stated previously in this paper that ML Ops presents a $4 billion opportunity in a few years.

XAI for ML Monitoring, a part of ML Ops finds itself at a highly opportunistic intersection of these two technologies with significant upside potential – Explainable AI and ML Ops.

Explainable AI has also moved up significantly in Gartner's Hype Cycle for Emerging Technologies as shown below (Gartner 2020).


Gartner Hype Cycle for
Emerging Technologies, 2019
Expectations
Biochips
AI PaaS
Edge Analytics
Autonomous Driving Level 5
Low-Earth-Orbit Satellite Systems
Edge AI
Explainable AI
Personification
Knowledge Graphs
Synthetic Data
Light Cargo Delivery Drones
Transfer Learning
Flying Autonomous Vehicles
Augmented Intelligence
Nanoscale 3D Printing
Decentralized Autonomous Organization
Generative Adversarial Networks
Decentralized Web
AR Cloud
Biotech - Cultured or Artificial Tissue
5G
Graph Analytics
Next-Generation Memory
3D Sensing Cameras
Emotion AI
Autonomous Driving Level 4
DigitalOps
Adaptive ML
Immersive Workspaces
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Time
Plateau will be reached:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau
As of August 2019


Hype Cycle for Emerging
Technologies, 2020
Expectations
Secure Access Service Edge (SASE)
Social Distancing Technologies
Explainable AI
Embedded AI
Data Fabric
Composable Enterprise
AI Augmented Development
Carbon-Based Transistors
Responsible AI
Multiexperience
Digital Twin of the Person
Citizen Twin
Packaged Business Capabilities
Generative AI
Composite AI
Adaptive ML
Bring Your Own Identity
Social Data
Private 5G
Differential Privacy
Biodegradable Sensors
Generative Adversarial Networks
2-Way BMI (Brain Machine Interface)
Ontologies and Graphs
Health Passport
Self-Supervised Learning
Low-Cost Single-Board Computers at the Edge
DNA Computing and Storage
AI-Assisted Design
Authenticated Provenance
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Time
Plateau will be reached:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau
As of July 2020

Now that the growth aspirations and opportunities have been made clear, A&I can focus on a few initiatives.

1. Training: An introduction and a hands-on led training need on Explainable AI and ML Ops need to be an integral part of ILP training.
2. Offering: A&I needs to create a clear differentiator for an ML Ops expertise by incorporating the benefits of post-production ML Monitoring with XAI. This could be the seed for an ML Ops accelerator.
3. COIN partnership: Given the sheer potential it is no surprise that there has been a surge of innovation and consequent investment activity in XAI. There are many startups in this space segmented into those focusing on a problem space – Kyndi (audit) and KenSci (healthcare) and those offering a framework style solution – Fiddler and Pachyderm. These are but a few and many more are entering the XAI space (Shaan, 2020). A few of these deserve due diligence as part of our COIN ecosystem. These will have promise to develop a GTM strategy to help us deliver accelerated business outcomes.

7. Acknowledgements

While submission guidelines constrain us to list 3 authors, Sandeep Pathak would like to mention and thank the entire ILP team who worked on this experiment while still undergoing their stipulated training, namely (in alphabetical order) –Aishwarya Prasad Venkatesh, Ehtheshamuddin Mohammed, Eshaan Anirudha Thakar, Josh Jacob. We would like to thank Priyanka Dabadge for helping facilitate this work and Hrishikesh Sherlekar for enabling the initiative symbiotically along with the ILP training program. Sandeep Pathak would like to thank Venkatesan Sukumaran for enabling him the opportunity to explore such exciting AI/ML avenues under the Advisory structure and finally, Santanu Ghosh for graciously supporting him in these endeavors.

8. References

Algorithmia (2020). 2020 state of enterprise machine learning. https://algorithmia.com/state-of-ml

Analytics India (2019). *8 Explainable AI Frameworks Driving A New Paradigm For Transparency In AI*. https://analyticsindiamag.com/8-explainable-ai-frameworks-driving-a-new-paradigm-for-transparency-in-ai/

Analytics India (2017). *8 Real Life Examples When Algorithms Turned Rogue, Causing Disastrous Results*. https://analyticsindiamag.com/8-real-life-examples-algorithms-turned-rogue-causing-disastrous-results/

Analytics Vidhya (2020, March 25). 6 Python Libraries to Interpret Machine Learning Models and Build Trust. https://www.analyticsvidhya.com/blog/2020/03/6-python-libraries-interpret-machine-learning-models

Doshi-Velez F. & Kim, B. (2017). *Towards a Rigorous Science of Interpretable Machine Learning.* https://arxiv.org/abs/1702.08608v2

European Parliament (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

Gartner (2020). *5 Trends Drive the Gartner Hype Cycle for Emerging Technologies, 2020.* https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/

Gilpin, L., Bau, D., Yuan, B., Bajwa, A., Specter, M. & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. https://arxiv.org/abs/1806.00069

Google (n.d) *Explainable AI*. Retrieved Feb 2nd 2020 from https://cloud.google.com/explainable-ai/

Haranas, M. (2019, October 28). *Gartner's Top 10 Technology Trends For 2020 That Will Shape The Future.* CRN. https://www.crn.com/slide-shows/virtualization/gartner-s-top-10-technology-trends-for-2020-that-will-shape-the-future/8

Lundberg, S. & Lee, S. (2017). *A Unified Approach to Interpreting Model Predictions*. 31st Conference on Neural Information Processing Systems. https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

McKnight W. (2020, January 21). *Delivering on the Vision of MLOps: A maturity-based approach.* GigaOm. https://gigaom.com/report/delivering-on-the-vision-of-mlops/

Molnar, C. (2021). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.* https://christophm.github.io/interpretable-ml-book/

Price-Waterhouse Coopers (PwC) (2017). *Explainable AI: Driving business value through greater understanding.* https://www.pwc.co.uk/audit-assurance/assets/pdf/explainable-artificial-intelligence-xai.pdf

Schmelzer, R. (2020, March 8). *The Emergence of ML Ops.* Forbes. https://www.forbes.com/sites/cognitiveworld/2020/03/08/the-emergence-of-ml-ops/?sh=2949ac8d4698

Shaan (2020, October 24). Explainable AI is gaining momentum among VCs and startups. https://medium.com/the-blue-stars/explainable-ai-is-gaining-momentum-among-vcs-and-startups-445906869d01

Vizard, M (2021, January 25). *Open Source Platforms Vie with IT Vendors for Management of MLOps.* IT Business Edge. https://www.itbusinessedge.com/blogs/it-unmasked/open-source-platforms-vie-with-it-vendors-for-management-of-mlops.html

Simonite, T. (2020, October 26). *How an Algorithm Blocked Kidney Transplants to Black Patients.* Wired. https://www.wired.com/story/how-algorithm-blocked-kidney-transplants-black-patients/

Witzenman, M (2017, August 24). *Monitoring OpenShift: Three Tools for Simplification.* Red Hat. https://www.openshift.com/blog/monitoring-openshift-three-tools