Mphasis
The Next Applied

Machine Learning & IT Operations to Manage End-to-End Machine Learning Life Cycles

Whitepaper by Dr. Jai Ganesh, SVP & Head - Mphasis NEXT Labs | Dr. Archisman Majumdar, AVP & Lead - Applied AI, Mphasis NEXT Labs | Saurabh Singh, Sr Manager, Data Science - Mphasis NEXT Labs | Kaushlesh Kumar, AVP - Applied AI, Mphasis NEXT Labs | Abinaya Mahendiran, Assistant Manager - Mphasis NEXT Labs

Contents

1. Introduction 1

2. Key Tenets of Machine Learning Projects 2

3. PACE-ML - An Integrated Approach to Machine Learning Using MLOps 2

4. References 8

1.

Introduction

Machine Learning & IT Operations (MLOps) – a combination of Machine Learning (ML) and IT Operations – focuses on automating and productizing machine learning algorithms through enhanced automation, collaboration and communication between data scientists and information technology professionals. It improves the efficiency and streamlines the management of model selection, reproducibility, versioning, auditability, AI explainability, packaging, re-usability, validation, deployment and monitoring, which helps in building ML models faster and at scale.

There have always been challenges in moving applications from development to operations, which over the years, the software engineers have learnt to tackle through advances in DevOps and adoption of collaboration tools, versioning, testing, automation and new development practices. However, this is not the case with large scale ML projects that involve continuous model build, train, deployment and retrain. Therefore, automated, repeatable and scalable best practices are needed to manage them.

Currently, the ML tools are focused mostly on the management, versioning auditability governance and control of code. Some of the key challenges in the adoptions of AI/ML solutions are as follows:

- Articulation of business case to implement the AI solution
- Non-availability of quality data and inability to handle changes in data
- Time taken for integration with the current model of business, process, delivery & infrastructure
- Cost of solution viz. skills, infrastructure
- Confidence in the solutions with the limited generalization and black box model
- Legal, compliance and regulatory concerns

As a response to these challenges, the best practices as well as tools and platforms have been standardized under the banner of MLOps. The value of MLOps in the ML project life cycle can be summarized as below:

1. Collaboration among the data science, software development and operations/delivery teams
2. Control over data, code, algorithms and models with versioning, tracking and auditability
3. Checks on the quality of solutions facilitated by easy debugging and interpretability of models and model performance (drift detection)
4. Continuous integration and deployment of models in the software development, and delivery environments leveraging automated pipelines and feedback loops

In this paper, we highlight the need of new practices across the ML life cycle ranging from data tagging to the deployment of ML models in production. We discuss approaches, methodologies,

frameworks and tools needed to tackle the ever-increasing complexities. We also introduce **PACE-ML,** which is Mphasis MLOps Framework & Methodology for automated continuous end-to-end machine learning.

2.

Key Tenets of Machine Learning Projects

Skills: The teams should consist of data scientists, data architects, data engineers and business analysts who focus on data analysis, model development, experimentation and visualization.

Model Training & Development: ML development is experimental and different from traditional programming. The teams work on model features, algorithms and configurations iteratively, as model auditing, data & model versioning and reproducibility are critical. Automated model training is a key need as the teams should be able to create a training pipeline that runs across several machines and can be reused by others.

Collaboration: Developing a successful ML system requires collaboration across multiple groups in an iterative manner. Developers should be able to leverage past knowledge and results of experiments across versions, ensure peer-to-peer sharing and branch out new variants of experiments.

Testing: Testing an ML system involves validating input data, model quality and performance, explainability, infrastructure, pipeline integration, API and data drift. Model reuse is different from software reuse, as models must be tuned based on input data/scenario.

Deployment: ML systems may need a multi-step pipeline to automatically retrain and deploy models. Therefore, the teams should be able to create deployment pipelines that runs across several machines and can be reused by others. It should support model portability across a variety of platforms, and be able to monitor and know when to retrain given scenarios such as data drift.

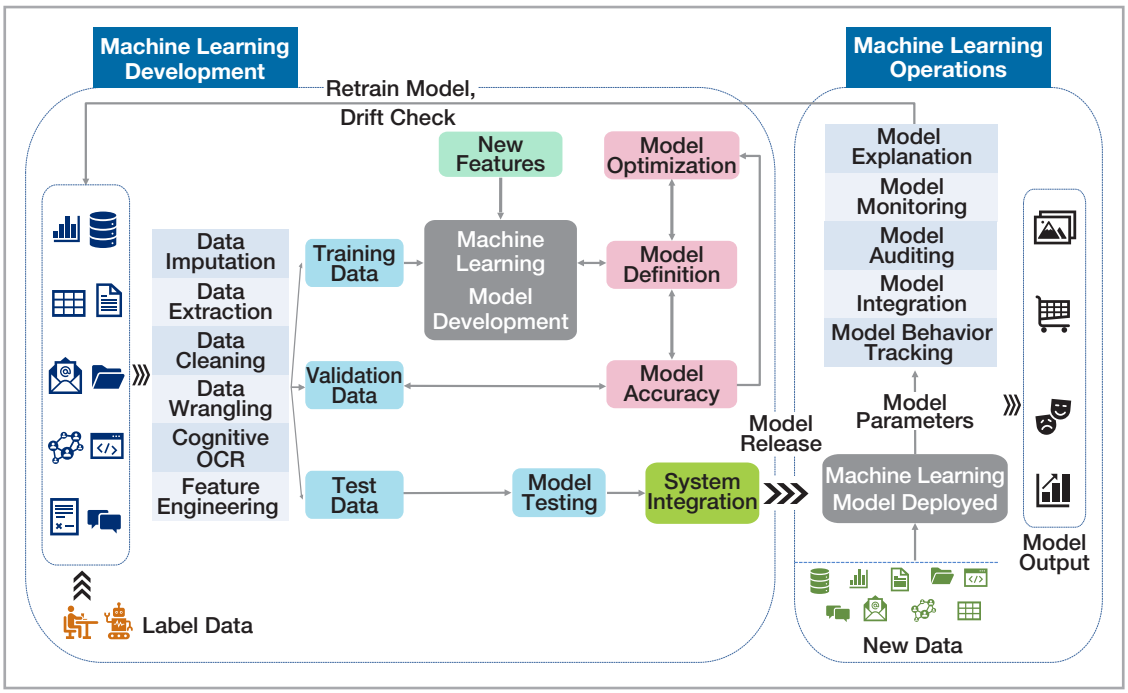
Production: The performance of ML models can be affected not only due to suboptimal coding, but also due to constantly evolving data profiles. In other words, models can decay in more ways than conventional software systems and this degradation must be considered. Therefore, the teams need to track summary statistics of the given data and monitor the online performance of the model to send notifications or roll back when values deviate from what's expected.

3.

PACE-ML - An Integrated Approach to Machine Learning Using MLOps

PACE-ML *(Pipeline for Automated Continuous End-to-End – Machine Learning)* is Mphasis' framework for machine learning development and deployment, built on MLOps principles to facilitate a set of practices and activities which enable data scientists and IT operations to collaborate. A combination of our proprietary tools and methodologies along with best-in-class

third-party as well as open-source tools, PACE-ML automates multiple stages in the pipeline, accelerating the life cycle of development, deployment and productionizing of ML algorithms. The framework uses workflows, collaboration platforms and tools to improve model selection, reproducibility, versioning, auditability, explainability, packaging, re-usability, validation, deployment and monitoring.



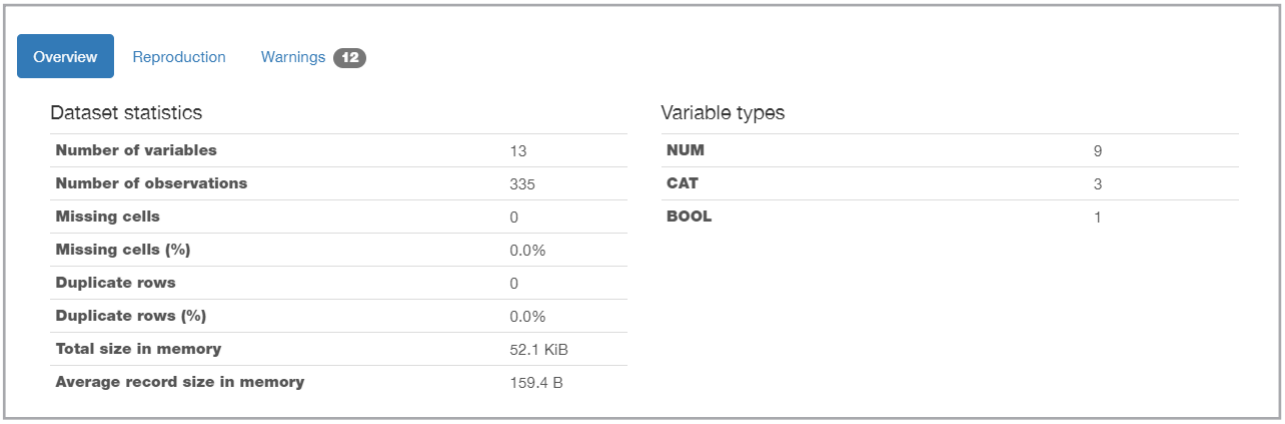
PACE-ML Architecture

Key Features

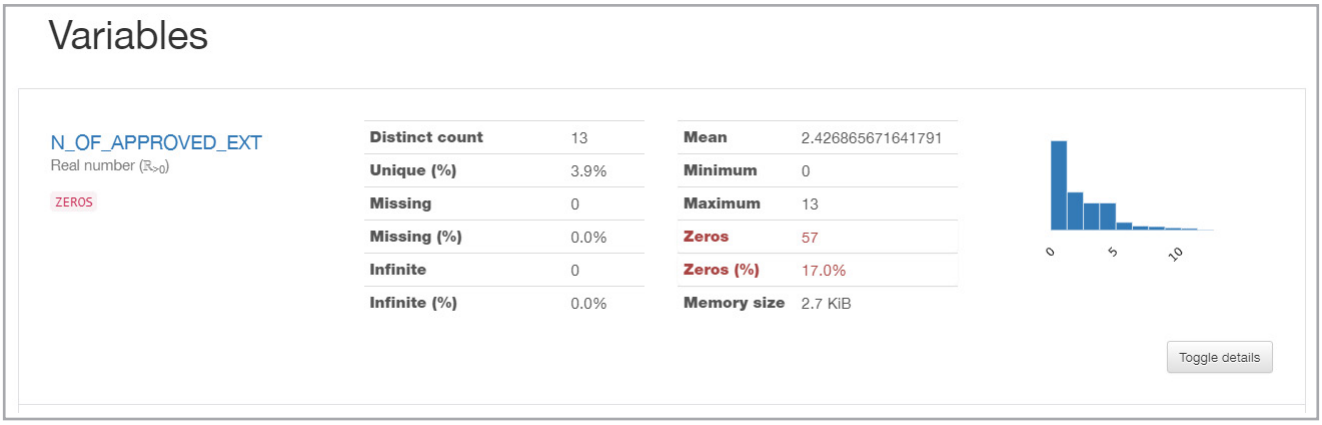
Below are some of the key features of PACE-ML which powers successful ML projects:

Data Preparation, Feature Engineering and Version Control: This includes importing, validating and cleaning, munging and transformation, normalization, staging and feature engineering. Given the inconsistency across data sources and various time periods, input check for validity and automation of data input checks and notifications is important.

Feature engineering, driven more by the business use context and less by data, is the process of transforming raw data gathered from clients into features that represent the underlying structure of problems to the ML models. Once data is gathered, PACE-ML follows a due-diligence process to ensure viability of the project and assess its risk. It utilizes the open source tool DVC to keep track of ML processes and file dependencies in the simple form of git-like commands. The data is stored in cloud locations (like S3) and changes are tracked in the system using data version control.

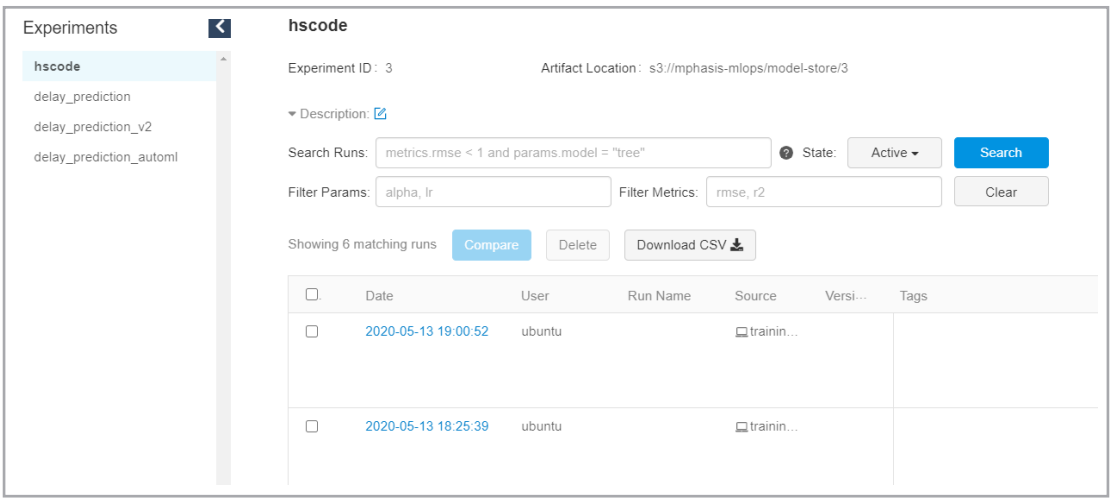


Feature Engineering



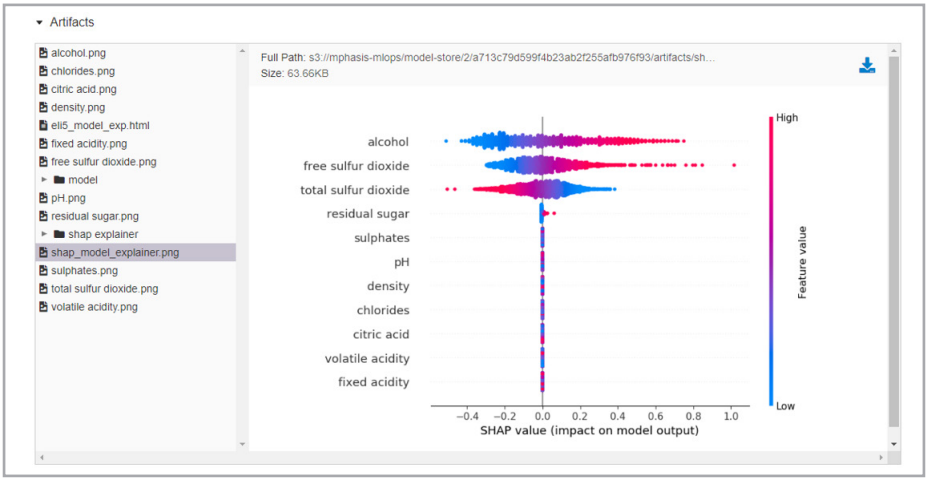
Feature Engineering

Collaboration among Stakeholders: Using collaborative workspaces (like Jupyter) for model development and workflows (like MLflow), PACE-ML ensures seamless collaboration among the stakeholders involved.



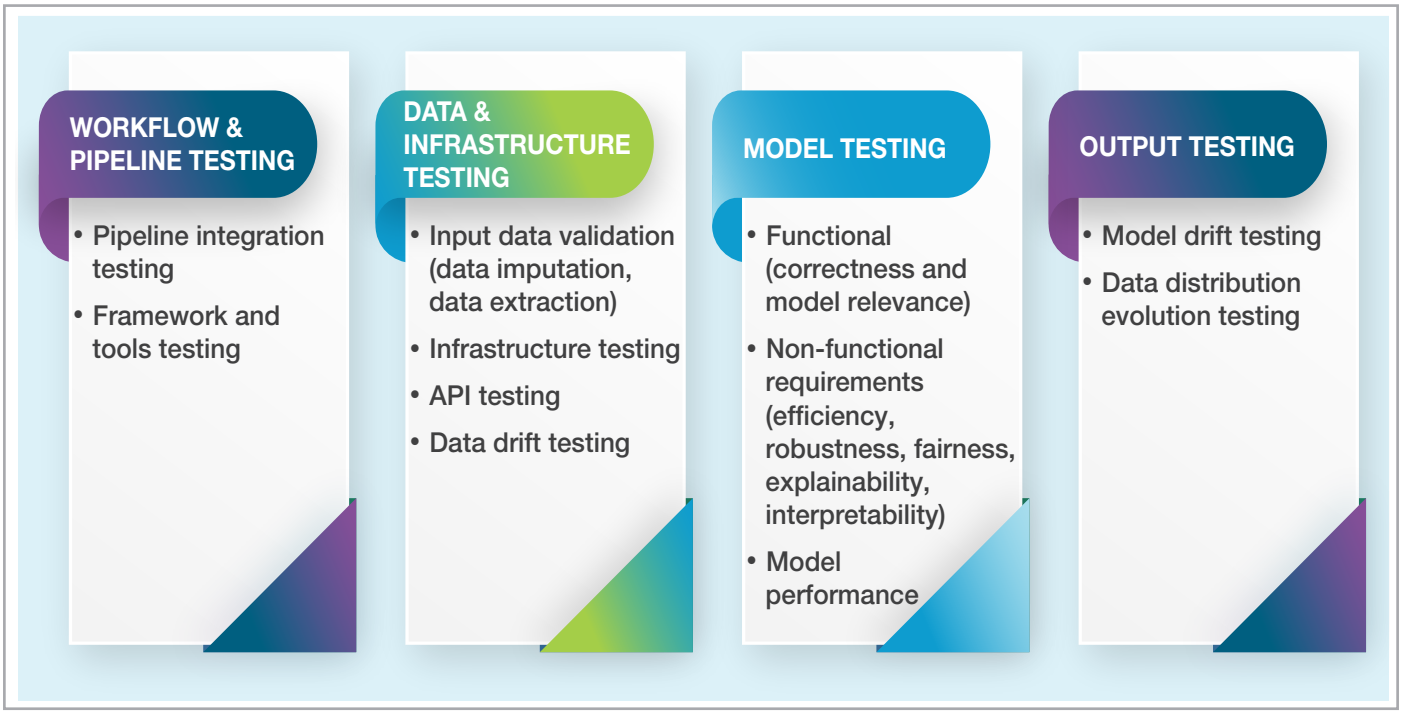
Collaboration among Stakeholders

Model Explainability: One of the primary concerns with the use of AI models is their inherent complexity and black box nature. Apart from this, since the models are generated primarily by learning the existing data, the models have an inherent risk of incorporating the procedural biases of the creators of the data. Most of the times, these biases are unwarranted and difficult to identify. PACE-ML incorporates model explainability as its core feature, which is fully integrated with the model development and auditability pipelines. The explainability is provided at both the model level (features learned and their importance from the model perspective) and the prediction level (why the model is predicting the outcome for a specific case), which help the developers and users in identifying issues and/or biases.

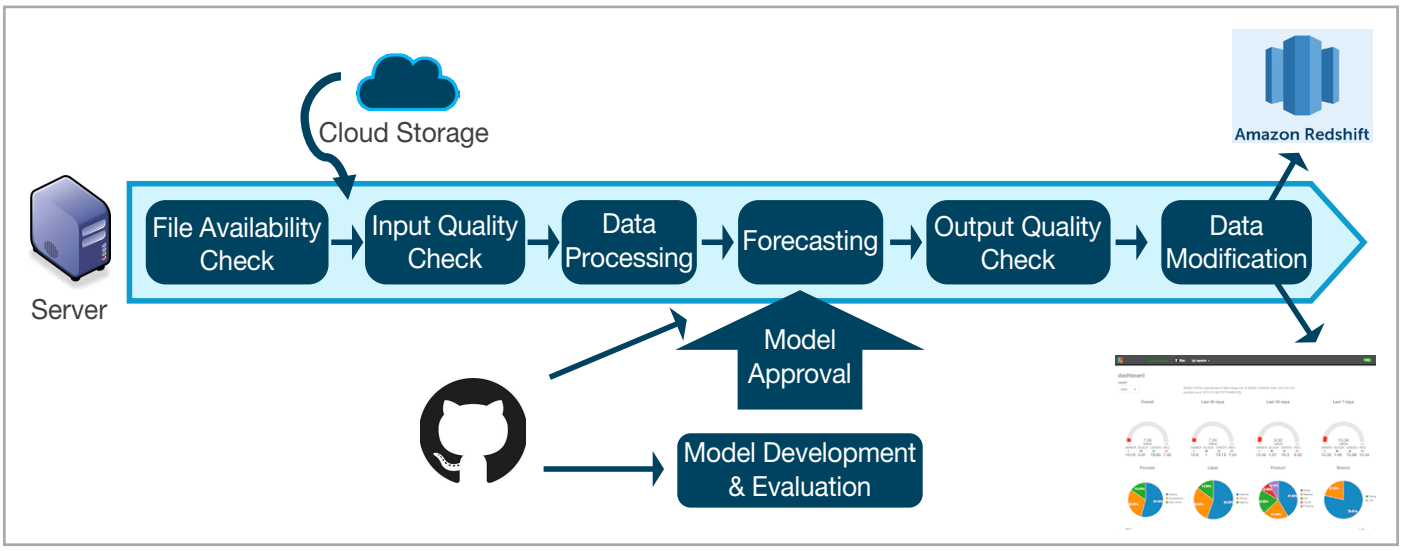


PACE-ML: ML Model Explainability

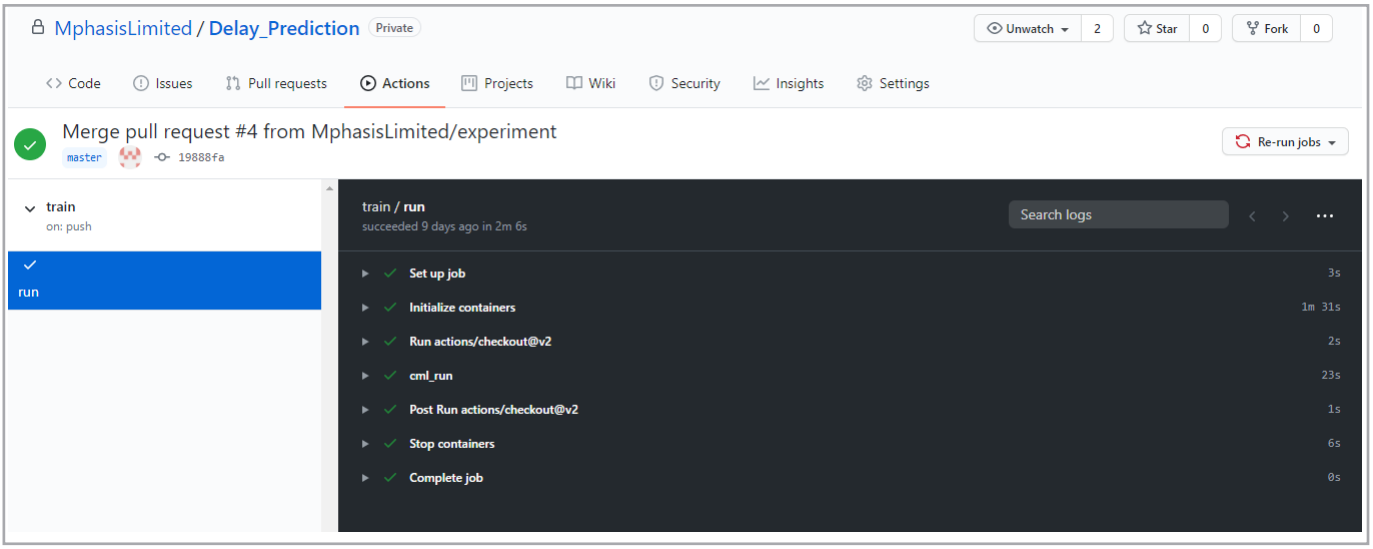
Automated Build, Testing and Deployments: It is important to prepare disaster recovery plan before deploying machine learning models to ensure swift recovery in any unforeseen circumstance. The platform utilizes managed pipelines (e.g., AWS Sagemaker) for the deployment and creating the endpoints. Further, rapid one touch deployment is accelerated using custom scripts and Git Actions. These result in reduced number of touchpoints/handoffs to cut off cycle time and risks. The entire process from development to deployment of models is organized as pipelines which are then automated.



PACE-ML: Testing

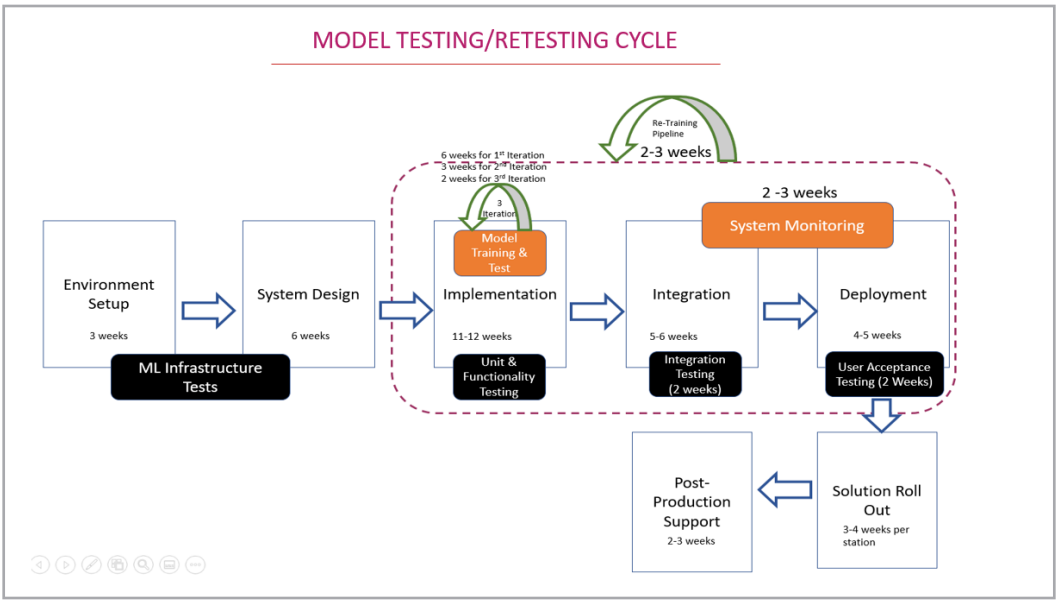


PACE-ML: ML Deployment

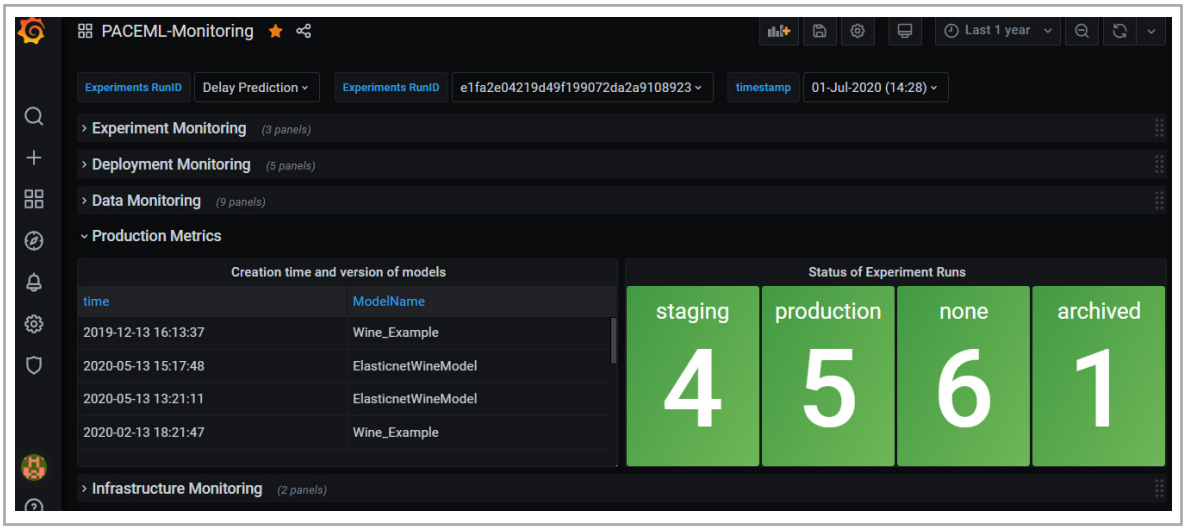


PACE-ML: CI-CD using Git Actions

Model Monitoring: Using appropriate tools, PACE-ML provides monitoring for both the infrastructure and model performance at scale. The graphs related to the load, uptime, performance, security, as well as performance comparisons between challenger and deployed models are available through intuitive visualizations in the control tower.



PACE-ML: Monitoring Workflow

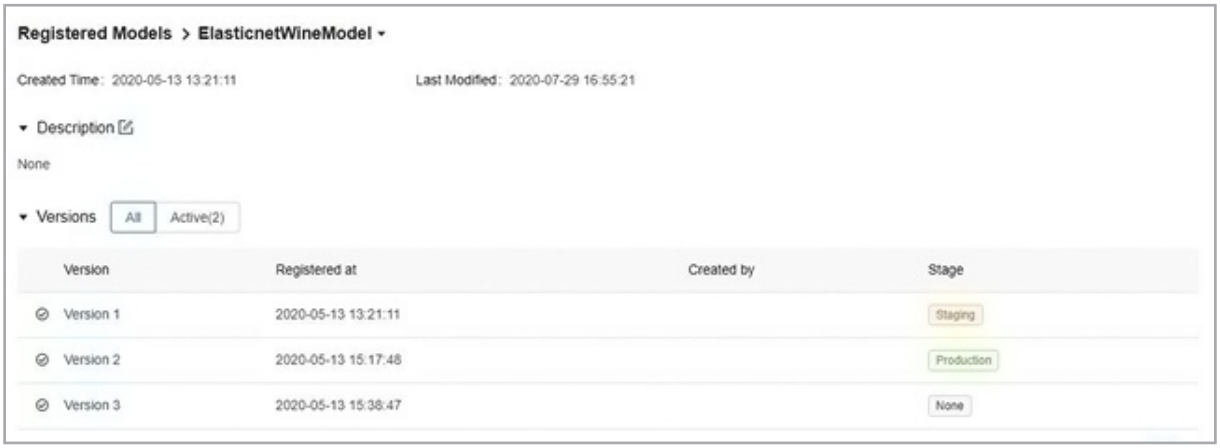


PACE-ML: Dashboard for Monitoring of Models

Drift Monitoring: PACE-ML provides monitoring for data drift through version control of the data, looking for outliers/anomalies in the training data distribution over time. For model drift detection, we provide a continuous monitoring of the model performance by either continual tagging of random samples (e.g., manual tagging of fraud) in the model output or when system feedback is available (e.g., recommender systems) through continuous monitoring of the logs.

PACE-ML: Drift Velocity Monitoring Dashboard

Model Auditability: PACE-ML provides an option for maintaining versions of models and the training data used in that model. Further, the use of explainability in the pipeline ensures that we can also explain the reasons the decisions were taken at that point of time.



PACE-ML: Auditability of Past Models

Key Benefits

PACE-ML offers multi-fold benefits to our clients and their AI/ML projects. Some of these pertain to –

- **Speed**: Faster time-to-market for products and services
- **Efficiency**: Facilitates model development and deployment at reduced effort and time
- **Explainability**: Ensures governance, risk management and compliance through easy debugging of models
- **Effectiveness**: Helps users make accurate decisions

- **Trust**: Increases users' confidence in the system
- **Automation**: Reduces manual interventions and enables continuous delivery with automated model pipeline management
- **Collaboration**: Tracks model, code and data changes, and increases collaboration among teams
- **Scrutability**: Lets the users tell the system that it is wrong
- **Debugging**: Identifies biases and/or defects in the system so that they can be corrected
- **Monitoring**: Monitors and ensures no broken models in production, and ensures faster response to performance issues
- **Cost Optimization**: Reduces cost of development through automation and seamless integration

4.

References

1. A Guidance Framework for Operationalizing Machine Learning for AI, Published 24 October 2018 - ID G00366587
2. Artificial Intelligence (AI) Market Size, Share and Industry Analysis By Component (Hardware, Software, Services), By Technology (Computer Vision, Machine Learning, Natural Language Processing, Others), By Industry Vertical (BFSI, Healthcare, Manufacturing, Retail, IT & Telecom, Government, Others) and Regional Forecast, 2019-2026
3. McKinsey Global Institute, The promise and challenge of the age of artificial intelligence, October 2018 | Executive Briefing
4. Googles rubric for tests, https://developers.google.com/machine-learning/guides/rules-of-ml/
5. The Rise of the Term "MLOps" Properly Operationalized Machine Learning is the New Holy Grail Kyle Gallatin, Sep 18
6. Rules of Machine Learning: Best Practices for ML Engineering, http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdfne%20learning/2019/03/17/how-to-deploy-machine-learning-models/
7. McKinsey flips lid, open sources Kedro machine learning framework, https://devclass.com/2019/06/04/mckinsey-flips-lid-open-sources-kedro-machine-learning-framework/
8. What CI/CD Tool Should I Use? Learn the components of a typical automated CI/CD deployment pipeline and why you need one, https://dzone.com/articles/what-cicd-tool-should-i-use
9. LF AI Foundation Interactive Landscape, https://landscape.lfai.foundation/

10. AI adoption is being fueled by an improved tool ecosystem, https://www.oreilly.com/radar/ai-adoption-is-being-fueled-by-an-improved-tool-ecosystem/

11. https://www.kdnuggets.com/2018/07/devops-data-scientists-taming-unicorn.html

12. https://www.forbes.com/sites/cognitiveworld/2020/03/08/the-emergence-of-ml-ops/#32624c3c4698

13. What's your ML Test Score? A rubric for ML production systems, Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, D. Sculley Google, Inc, 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

14. https://www.seldon.io/introducing-seldon-core-machine-learning-deployment-for-kubernetes/

15. Mlflow.org

16. https://github.com/features/actions

17. Brugman, S. (2020). pandas-profiling: Exploratory data analysis reports in Python. https://github.com/pandas-profiling/pandas-profiling.

Authors

Dr. Jai Ganesh

SVP & Head - Mphasis NEXT Labs

Dr. Jai Ganesh is Product and Service Innovation leader with extensive experience in inventing, conceptualizing, building and commercializing successful technology products and service innovations. Award winning digital transformation and innovation leader with expertise in lab-to-market product and service innovations. Under his leadership, NEXT Labs has created several global award-winning solutions, products and service offerings.

Dr. Archisman Majumdar

AVP & Lead - Applied AI, Mphasis NEXT Labs

Dr. Archisman leads a cross-functional team of Data Scientists and consults Fortune 500 companies on AI and ML implementations. He holds a PhD from the Indian Institute of Management Bangalore (IIMB) in the Quantitative Methods and Information Systems area. His areas of expertise are in machine learning, product management and information systems research.

Saurabh Singh

Sr. Manager, Data Science - Mphasis NEXT Labs

A data geek, Saurabh is part of Mphasis' innovation group – NEXT Labs. He has been working in the corporate world for the last 5 years helping businesses improve their bottom line by implementing machine learning and statistical algorithms. An MBA from IIM Bangalore has helped him to understand business intricacies and offer pragmatic solutions based on the data.

Kaushlesh Kumar

AVP - Applied AI, Mphasis NEXT Labs

Kaushlesh Kumar, in his career spanning more than a decade, has helped solve business problems for clients using innovative and structured approaches. In his current role, Kaushlesh designs, develops and executes solutions for transforming business services leveraging machine learning, data science and process. These initiatives deliver super normal value to the clients. He has a PG Diploma in Management and is a graduate engineer.

Abinaya Mahendiran

Assistant Manager - Mphasis NEXT Labs

Abinaya Mahendiran holds a Master's degree in Computer Science with a specialization in Machine Learning and Deep Learning from International Institute of Information Technology Bangalore (IIIT-B). Her research areas include Natural Language Understanding/Processing, Machine Learning, Deep Learning and MLOps. She has an extensive software engineering and data science experience. At NEXT Labs, she has been building and productionizing NLU/NLP solutions for various clients both on premise and on cloud.

About Mphasis

Mphasis (BSE: 526299; NSE: MPHASIS) applies next-generation technology to help enterprises transform businesses globally. Customer centricity is foundational to Mphasis and is reflected in the Mphasis' Front2Back™ Transformation approach. Front2Back™ uses the exponential power of cloud and cognitive to provide hyper-personalized ($C = X2C^{2}_{TM} = 1$) digital experience to clients and their end customers. Mphasis' Service Transformation approach helps 'shrink the core' through the application of digital technologies across legacy environments within an enterprise, enabling businesses to stay ahead in a changing world. Mphasis' core reference architectures and tools, speed and innovation with domain expertise and specialization are key to building strong relationships with marquee clients. To know more, please visit www.mphasis.com

For more information, contact: marketinginfo.m@mphasis.com

USA
460 Park Avenue South
Suite #1101
New York, NY 10016, USA
Tel.: +1 212 686 6655

UK
1 Ropemaker Street, London
EC2Y 9HT, United Kingdom
T : +44 020 7153 1327

INDIA
Bagmane World Technology Center
Marathahalli Ring Road
Doddanakundhi Village
Mahadevapura
Bangalore 560 048, India
Tel.: +91 80 3352 5000

www.mphasis.com

NR 26/08/20 US LETTER BASL&S385