



**BITS Pilani**  
Pilani | Dubai | Goa | Hyderabad

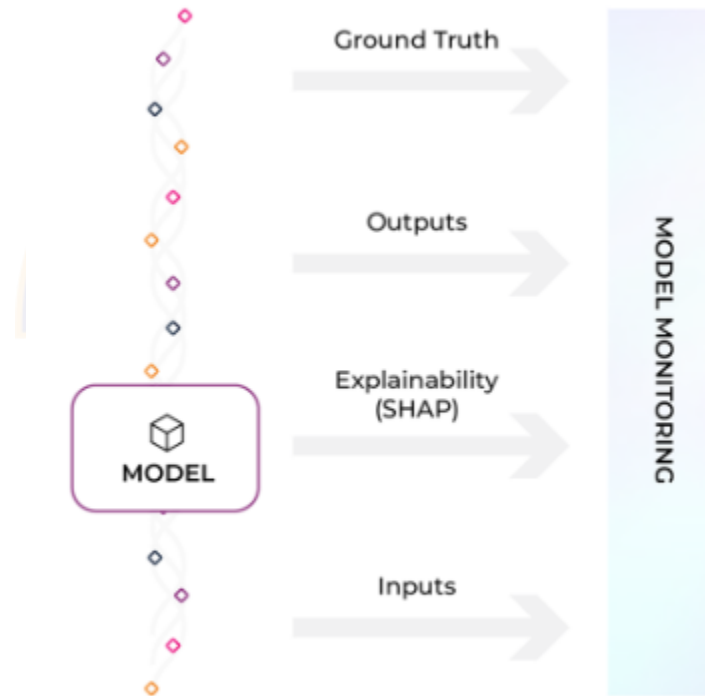
# Model Monitoring

**Pravin Y Pawar**

Adapted from ["What is Model Monitoring?"](#)

# What is Model Monitoring?

- A series of techniques used to detect and measure issues that arise with machine learning models
  - Once configured, monitors fire when a model metric crosses a threshold
- Areas of focus for monitoring include
  - model performance,
  - data quality,
  - drift detection
  - and embedding analysis



# Why Is Model Monitoring Important?

- A lot can go wrong with a production model
  - navigating model issues can be challenging for even the most seasoned machine learning (ML) practitioner
- With model monitoring,
  - can immediately know when issues arise in machine learning models
  - better insights empower data scientists and ML engineers to pinpoint where to begin further analysis
- Model Monitoring Improves Business Outcomes
- Despite the multitude of problems a model can encounter in the real world,
  - over half of ML teams lack a reliable way to proactively surface when something is going wrong with a model in production
  - Many rely on
    - batch in-house solutions or dashboards that may not catch issues in time
    - tools that are not purpose-built for machine learning
- In an era where ML models relied on to increase profitability and even save lives, it's clear that better model monitoring is critical.

# Model Performance Management

- Model performance indicates how model performs in production
  - Measure model performance with an evaluation metric,
  - which can be evaluated with daily or hourly checks
  - on metrics such as accuracy, recall, precision, F1, MAE, MAPE, and more
  - model type directs which performance metrics are applicable to model
- Performance monitor recommendations:
  - Performance monitors measure performance based on an evaluation metric
  - Can use performance metrics to compare model behavior between different environments
  - Use those insights to drill into the root cause of performance degradation
  - Important to look at the performance of models across various cohorts and slices of predictions



# Drift Monitoring

- Drift monitors measure distribution drift, which is the difference between two statistical distributions
- Models are trained with polished data to represent production environments
  - it's common for real-world production data to deviate from training parameters over time
  - need to measure drift to identify if models have grown stale, have data quality issues, or if there are adversarial inputs in model
- Drift monitor recommendations:
  - To detect drift over time, set baseline using training data to identify how model changes between features, predictions, and actuals
  - To detect short-term drift, set baseline using historical production data (i.e. two weeks).
  - With proactive monitoring, detecting drift should be easy with automatic alerts
  - Bulk-create monitors with multiple baselines, view feature performance at a glance, access a historical view of drift, and access the distribution view associated with drift metric

# Data Quality Monitoring

- ML models rely on upstream data to train and make predictions
  - Data is commonly collected from multiple systems, vendors, or can be owned by another team,
  - making it difficult to ensure always have high-quality data
- Model health depends on high-quality data that powers model features!
  - Important to immediately surface data quality issues to identify how data quality maps to model's performance
- Data quality monitors help identify key data quality issues
  - such as cardinality shifts, data type mismatch, missing data, and more
- Data quality monitor recommendations:
  - Use data quality monitors to detect shifts in upstream data and alert underlying changes
  - Configure data quality monitors to detect data issues like change in cardinality and change in percent of missing values

# Monitoring Unstructured Data

- Most companies building computer vision (CV) or natural language processing (NLP) models lack a window into their models
  - are performing in production, running the risk of models impacting earnings or acting in unfair ways
- Since deep learning models rely on human labeling teams,
  - identifying new patterns in production and troubleshooting performance can be tricky
- CV and NLP Model Monitoring Recommendations:
  - Visualizing and monitoring embeddings — vector representations of data where linear distances capture structure
  - By monitoring embeddings, ML teams can proactively identify when their unstructured data is drifting



# Thank You!

In our next session: