Stop Blaming the Algorithm: It's Your Feature Store!

A Guide to Monitoring your Feature Store

Claire Longo, Head of Post Sales ML Solutions Engineering, Arize AI

FEATURE STORE SUMMIT 2023

Organized by HOPSWORKS

What does Feature Failure Look Like?

Features can fail silently

1

Vendor Data Shifting

A Data Scientist's model relies on a feature engineered from data purchased from a vendor.

The vendor changes the schema of their data!

The model breaks!

2

Statistical Drift

A Data Scientist's model uses location data as a feature.

The company adds another service area. The feature now shows new location values never seen by the model.

The model breaks!

3

training/serving skew

A Data Scientist creates a model in a notebook.

A ML Engineer deploys that model.

The online and offline feature transformations don't match.

The model breaks!

Types of Monitoring Across IT

System / Infrastructure

DESCRIPTION

- Infra/App timing as the base of monitoring
- App & system response time issues
- Tracing & troubleshooting response time

PERSONAS

- Software Engineer
- DevOps Engineer

Data

DESCRIPTION

- Tables as the base of monitoring
- Monitoring data changes
- Schema monitoring

PERSONAS

- Data Engineer
- Data Architect

Machine Learning

DESCRIPTION

- Models are the base of monitoring
- Distributions vs baselines, model version, SHAP analysis and performance
- Deep model performance analysis vs data

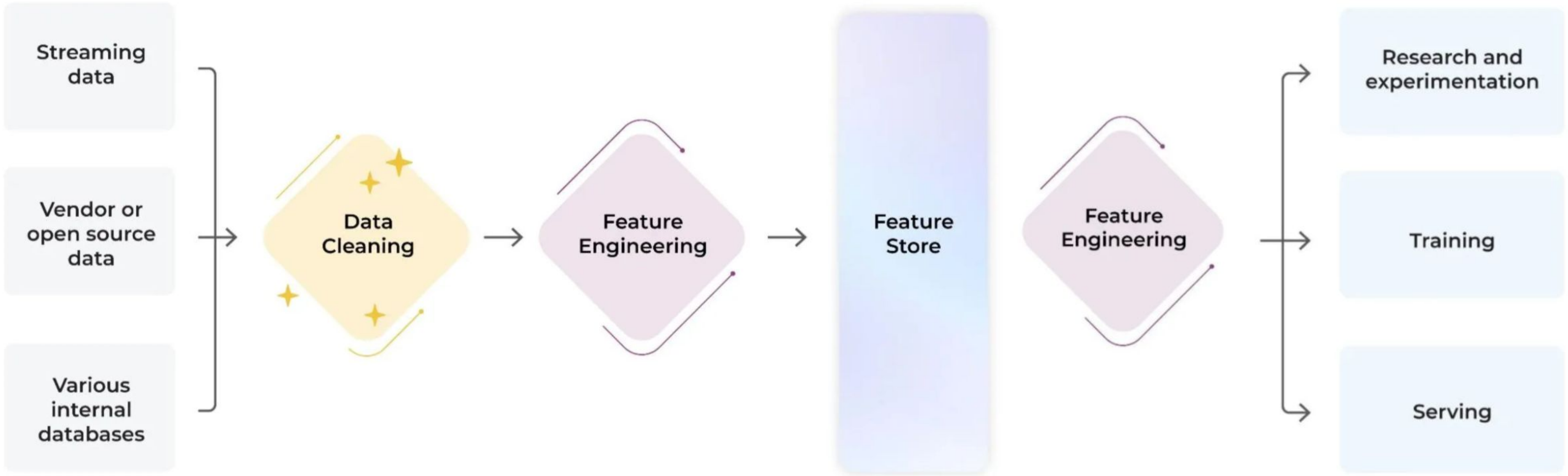
PERSONAS

- ML Engineer
- Data Scientist

FEATURE STORE SUMMIT 2023

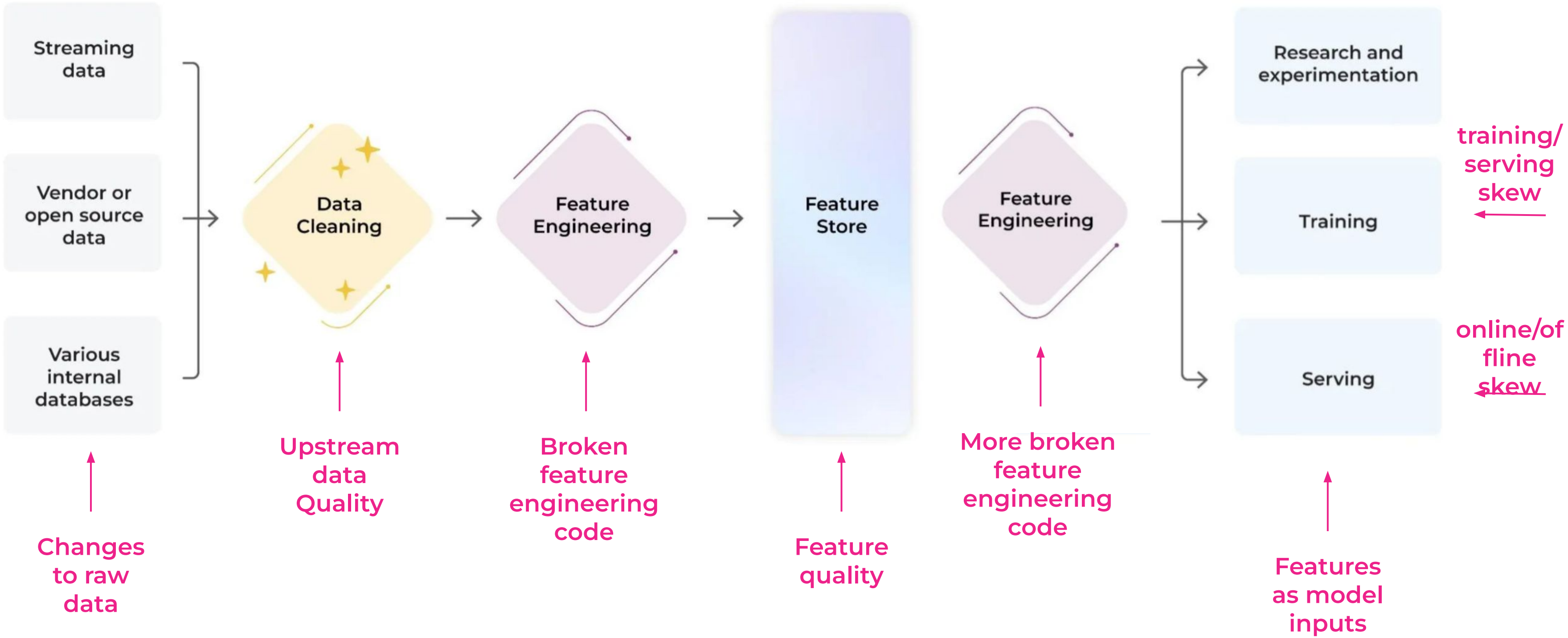
A Day in the Life of a Feature

The feature data lifecycle, from raw data to inference



Where Should I Monitor Features?

Answer: Everywhere!



But wait…

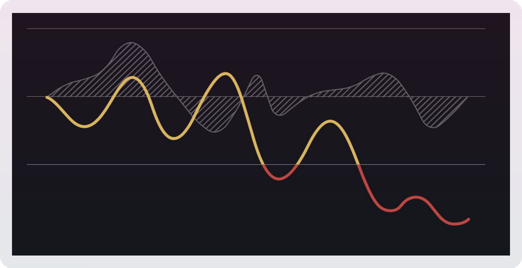
The Feature Store team does not own the upstream data sources, or the downstream modeling projects.

This data can shift under our feet!

So how do we use the Feature Store Monitoring to solve this?

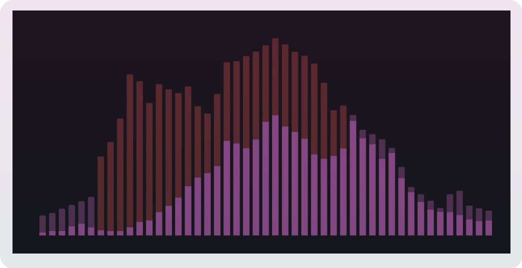
What Should I Monitor?

Monitors **proactively track** specific metrics over time and alert you when they cross a threshold.



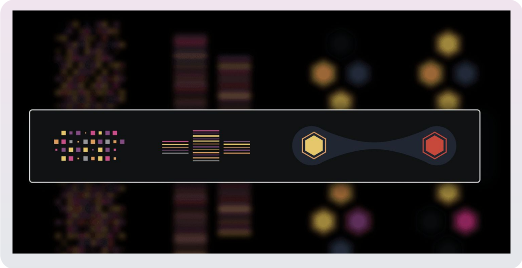
Data Quality Monitors
Ensure high quality upstream data

Data Quality Metrics: Percent empty, quantiles, sum, count, average, etc.



Tabular Drift Monitors
Detect Statistical Distribution Changes over lifetime of a Feature

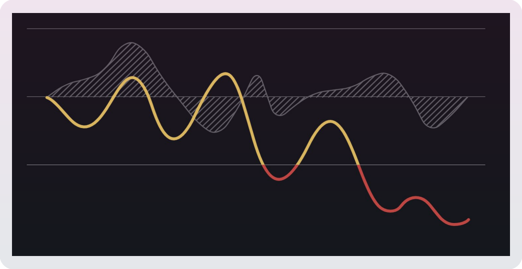
Drift Metrics: PSI, KL Divergence, JS Distance, KS Statistic



Embeddings Drift Monitors
Detect Data Distribution Changes in unstructured data Features (NLP, CV)

Performance Metrics: Euclidean Distance, Cosine Similarity

Data Quality Monitoring



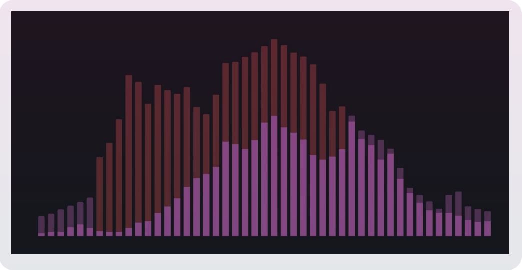
Data Quality Monitors
Ensure high quality upstream data

Data Quality Metrics: Percent empty, quantiles, sum, count, average, etc.

Examples

- Catching missing values
- Catching new categorical values
- Catching outliers
- Catching Online/offline skews

Tabular Data Drift Monitoring



Tabular Drift Monitors
Detect Statistical Distribution Changes over lifetime of a Feature

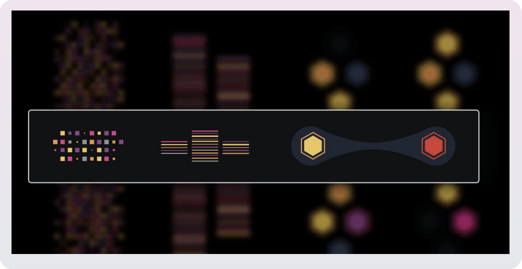
Drift Metrics: PSI, KL Divergence, JS Distance, KS Statistic

Drift compares a current data distribution to a reference distribution.

Examples

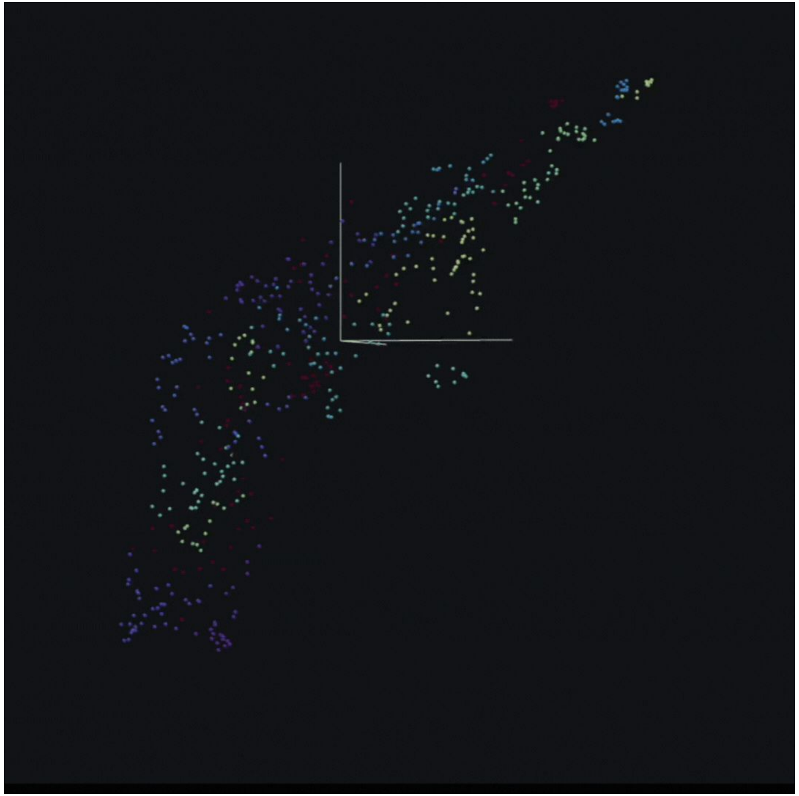
- Sudden drift
- Gradual drift

Unstructured Data Drift Monitoring



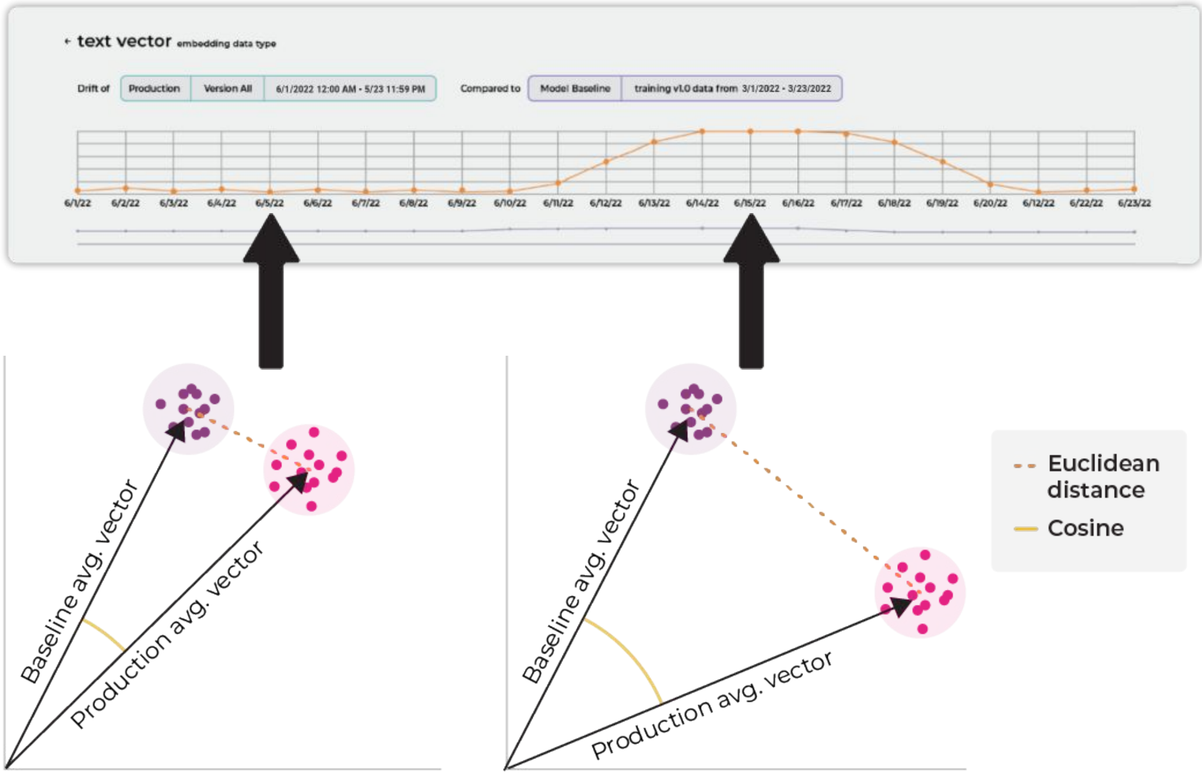
Embeddings Drift Monitors
Detect Data Distribution Changes in unstructured data Features (NLP, CV)

Performance Metrics: Euclidean Distance, Cosine Similarity



Embedding Drift

Measure drift as the distance between the primary dataset's centroid and the baseline



FEATURE STORE SUMMIT 2023

What does Feature Monitoring Look Like?

Features can fail silently

1

Vendor Data Shifting

Monitor for Data Quality issues in the Feature Store

✅

2

Statistical Drift

Monitor for Tabular Data and Embedding Drift

✅

3

online/offline skew

Monitor for Custom Data Quality metrics for data skew

✅

Thank you!

Questions?