ThirdEye: LinkedIn’s Business-wide Monitoring Platform

Strata Data Conference, NY

Akshay Rai
Senior Software Engineer

26th Sept 2019
Agenda

1. What is ThirdEye?
2. ThirdEye @ LinkedIn
3. Anomaly Detection & Analysis
4. Overview of Entity Monitoring
Anomaly Detection

Root-Cause Analysis

MTTD (Mean time to detect)

MTTR (Mean time to restore)
“A self-service platform enabling anyone to efficiently identify and investigate deviations in business and operational metrics.”

ThirdEye
ThirdEye @ LinkedIn
Business-wide monitoring platform

**Teams**
Major teams across LinkedIn leverage ThirdEye

50+

**Smart Alerts**
Self-Serve onboarded with guidance

>2k

**Time Series**
Thousands of time-series monitored automatically

>100k
LinkedIn Abusers

Login attacks,
Member scraping,
Fake account signups,
Registration attacks etc.

“Recent metric spikes in ThirdEye suggested on-going attack is likely”
AI Model Monitoring

Bug in Tracking events

=> Affected AI Models

=> Further impacted Revenue
Anomaly Detection & Analysis
Challenges solved by ThirdEye

Let’s deep dive into each one of these sections

**Anomaly Detection**
Automated monitoring of metrics for outliers

**Root Cause Analysis**
Collaborative root-cause identification from data
Challenges solved by ThirdEye
Let’s deep dive into each one of these sections

**Anomaly Detection**
Automated monitoring of metrics for outliers

**Root Cause Analysis**
Collaborative root-cause identification from data
Why is Anomaly Detection Difficult?

Scalability

- Manual monitoring of all relevant metrics is infeasible
- Manual configuration of alerts for 10,000+ metrics/dimensions is infeasible
- Point-solutions for each individual domain is wasteful

Performance

- Simple threshold and percentage-based rules are not enough
- Too many false alarms will mask the important issues
- Tuning of alerts and algorithms difficult for non-ML engineers
- Adapting gracefully to changes in systems and business is hard
Algorithm-Driven Anomaly Detection

• Learn the trends and impacts of seasonality
• Adapt to persistent change
• Tolerate errors and noise
• Find relevant outliers
• Take user inputs
Integrated Workflow for Detection

- Real time detection on fine-grained data
- 1-click user feedback for sensitivity tuning
- Summary dashboards and report generator
Anomaly Detection for Time Series Data

Metrics support

• Data-Source Connectors
  Pinot, MySQL, Presto

• Metric Granularity
  Minute, Hourly & Daily granularity

• High-dimensional data points

Detection Support

• Rules
  Thresholds, Week-over-week, Duration

• Algorithms
  Sign Test, Scan Statistic, Spline Regression

• Plug-Ins
  Custom logic and modules
Challenges solved by ThirdEye

Let’s deep dive into each one of these sections

Anomaly Detection
Automated monitoring of metrics for outliers

Root Cause Analysis
Collaborative root-cause identification from data
Why is Root Cause Analysis Difficult?

- **Human knowledge** → modelled relationships
- **Human attention** ← data from 20+ sources

*Iterative process* - human guiding + machine processing
Collaborative Analysis Dashboard

Seamless integration of analysis events across LinkedIn
Heatmap Analysis

- Overview of distributions and changes
- Interactive slice-and-dice
Automatic Cube Analysis

- Contribution ranking
- iDiff operator

Typical monitoring workflow
Typical Monitoring Flow

Configure Alert

Receive Alerts

Investigate Alert
Configure Alert

• What to Detect?
• Whom to Notify?
Receive Alerts

ACTION REQUIRED ON THIRDEYE ALERT

We have detected an anomaly on the metric some_critical_metric between Sep 20, 04:00 and Sep 20, 07:00 (PDT)

View anomaly on ThirdEye

Metric: some_critical_metric
Alert: Alert_critical_metric_upDown7pct_3hour (Edit Settings)

Something is fishy about this critical metric. If this alert fires then most likely ...

<table>
<thead>
<tr>
<th>Start / Duration</th>
<th>Dimensions</th>
<th>Current</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep 20, 04:00 PDT</td>
<td></td>
<td>32,847</td>
<td>28,080 (+16.98 %)</td>
</tr>
</tbody>
</table>

3 hours (investigate)

Useful Links
ThirdEye User Guide
**Investigate Alert!**

### Anomaly #139099641 Alert
**someCriticalMetric_upDown7pct_3hour**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Time &amp; Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>current</td>
<td>Fri, Sep 20 2019, 4:00 am PDT - Fri, Sep 20 2019, 7:00 am PDT (3 hours)</td>
</tr>
<tr>
<td>predicted</td>
<td>28.1k (+17%)</td>
</tr>
<tr>
<td>WoWw</td>
<td>30.2k (+9%)</td>
</tr>
<tr>
<td>wWwW</td>
<td>30.3k (+8%)</td>
</tr>
<tr>
<td>wWwwW</td>
<td>28.8k (+14%)</td>
</tr>
<tr>
<td>wWwwW</td>
<td>0.00 (+1000%)</td>
</tr>
</tbody>
</table>

**Is this an anomaly?**
- Yes - unexpected
- Expected temporary change
- Expected permanent change
- No change observed
- Not reviewed yet

**Feedback!**
Entity Monitoring Overview
Why Entity Monitoring?

- Intrinsic relationships across metrics
- Meaningful to alert at the scale of the system
- Too many metrics to monitor
- Lots of noise from anomalies

detectionName: feed_feature_composite_alert
description: Critical features in terms ...

type: COMPOSITE_ALERT
alerts:
  - type: METRIC_ALERT
    name: coverage_health
    metric: feature_count
    dataset: feed_feature_monitoring
    rules:
      - detection:
        - name: algorithm_based_detection
          type: ALGORITHM
          ...
    - type: METRIC_ALERT
      name: distribution_health
      metric: feature_value_mean
      dataset: feed_feature_monitoring
      rules:
        - detection:
          - name: detection_rule_1
            type: ALGORITHM
            ...

grouper:
  - type: GROUPER
    name: coverage_or_distribution_grouper
    params:
      expression: coverage_health OR distribution_health
ThirdEye is Open Source!

We welcome your contributions!

Details:
Blogs: https://tinyurl.com/thirdeyeBlogs
Slack: https://tinyurl.com/thirdeyeSlackChannel
Docs: https://tinyurl.com/thirdeyeDocs
GitHub: https://tinyurl.com/master-thirdeye
Open Source!

Out-of-the-box support

Visualization / APIs

• Collaborative Dashboards
• Powerful analysis tools
• Once-click feedback

Pluggable Detection & Notification Modules

• Rules & Filters
• ML-Algorithm Library
• Flexible Alerting

Anomaly Detection Engine

Root-Cause Analysis Engine

Data Source Connectors

• Metric & Event correlation
• Dimension Heatmap
• Auto-cube algorithm

Pinot, MySQL & Presto connectors

• Holiday Calendar
• Jira Integration

Pluggable Analysis Modules
Open source: What’s next?

- Deep Learning Infrastructure
- “Entity” as a first class citizen
- Event driven anomaly detection
- Data quality & availability monitoring
Thank you

Join the Slack Channel: https://tinyurl.com/thirdeyeSlackChannel