elastic

Machine Learning for the Elastic Stack

Hwanggon Kim
Solutions Architect
hwanggon.kim@elastic.co

핵심 1

어떤 머신 러닝(Machine Learning)???

Image Classification **Recommendations**

Autonomous cars Voice Recognition Predictive Medicine

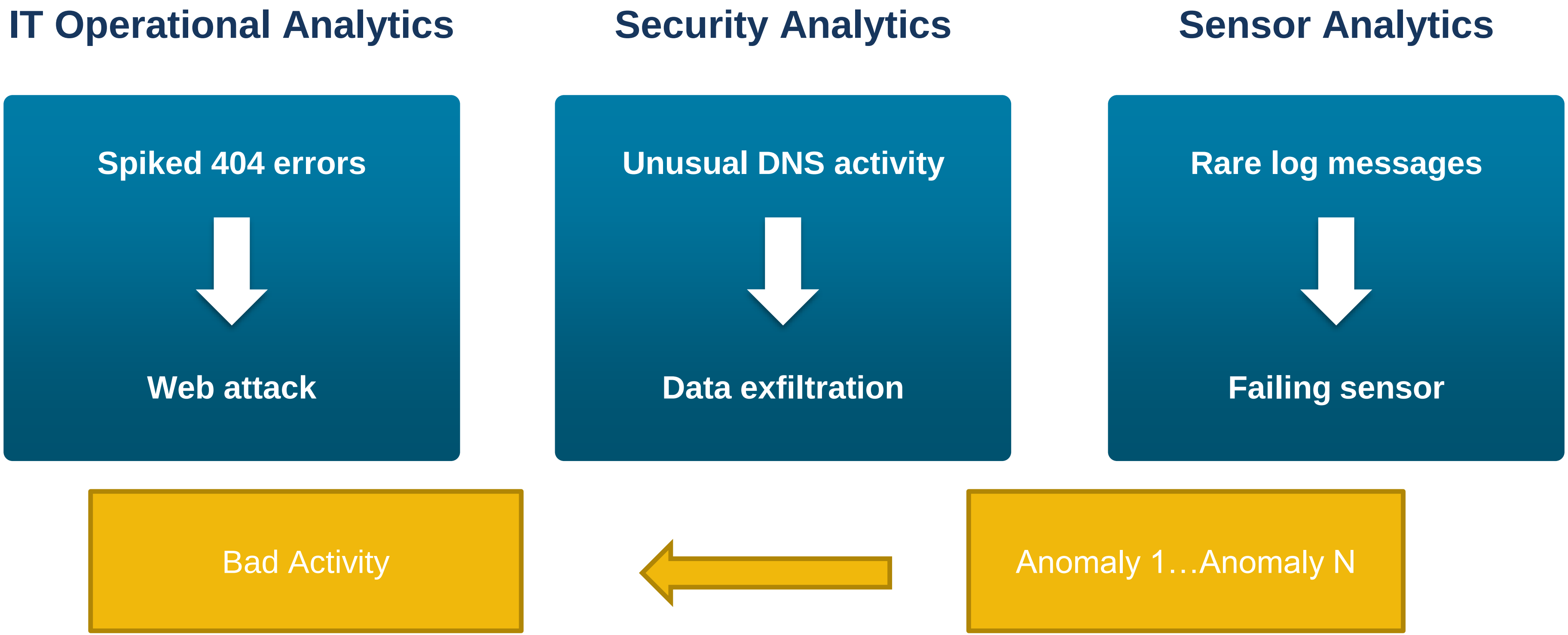
Fraud detection **Anomaly Detection**

Learn to Rank Speech Recognition

Language Translation **Entity Resolution**

시계열(Time-series) 데이터를 학습해서 이상 징후(anomaly) 탐지

“시계열(Time-Series) 데이터의 이상 징후 탐지”에 특화



IT Operational Analytics

Spiked 404 errors

Web attack

Security Analytics

Unusual DNS activity

Data exfiltration

Sensor Analytics

Rare log messages

Failing sensor

Bad Activity

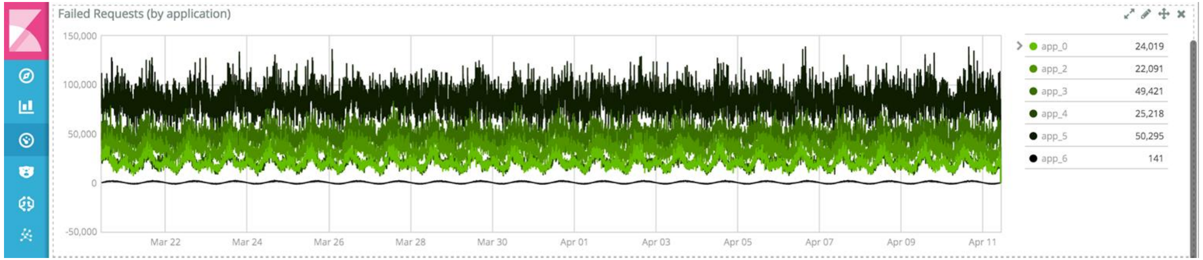
Anomaly 1…Anomaly N

Visual inspection is not practical

Detecting (noteworthy) anomalies is hard!

- Data is complex, high dimensional, fast moving
- Human inspection is not practical
- Easy to miss things

어느 부분에 이상 징후가 있는 지 식별할 수 있는가.



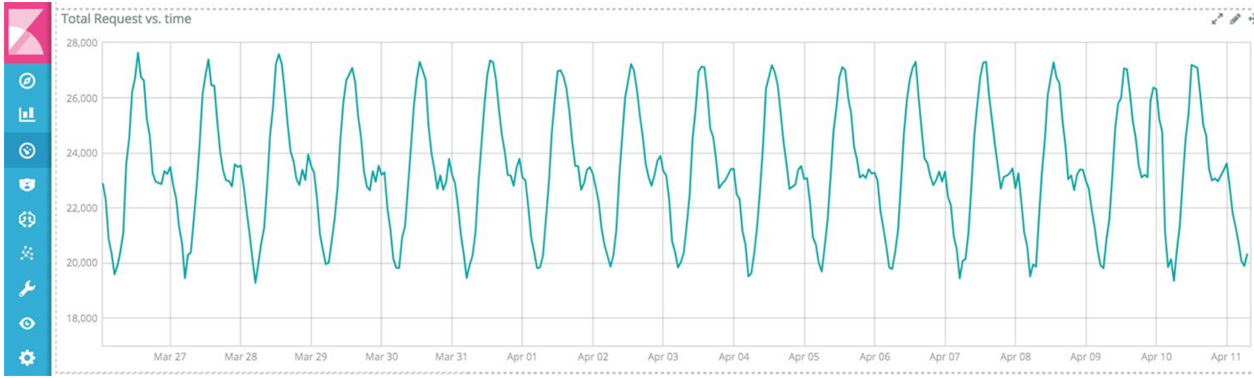
● app_0	24,019
● app_2	22,091
● app_3	49,421
● app_4	25,218
● app_5	50,295
● app_6	141

Rule-based alerts are insufficient

Detecting (noteworthy) anomalies is hard!

- Defining “normal” via static thresholds is hard
- Rules don’t evolve with data / infrastructure
- Rules can be bypassed

어떤 임계치(Threshold)가 가장 적절한 값인가.



3가지 타입의 이상 징후를 탐지 (Important)

핵심 2

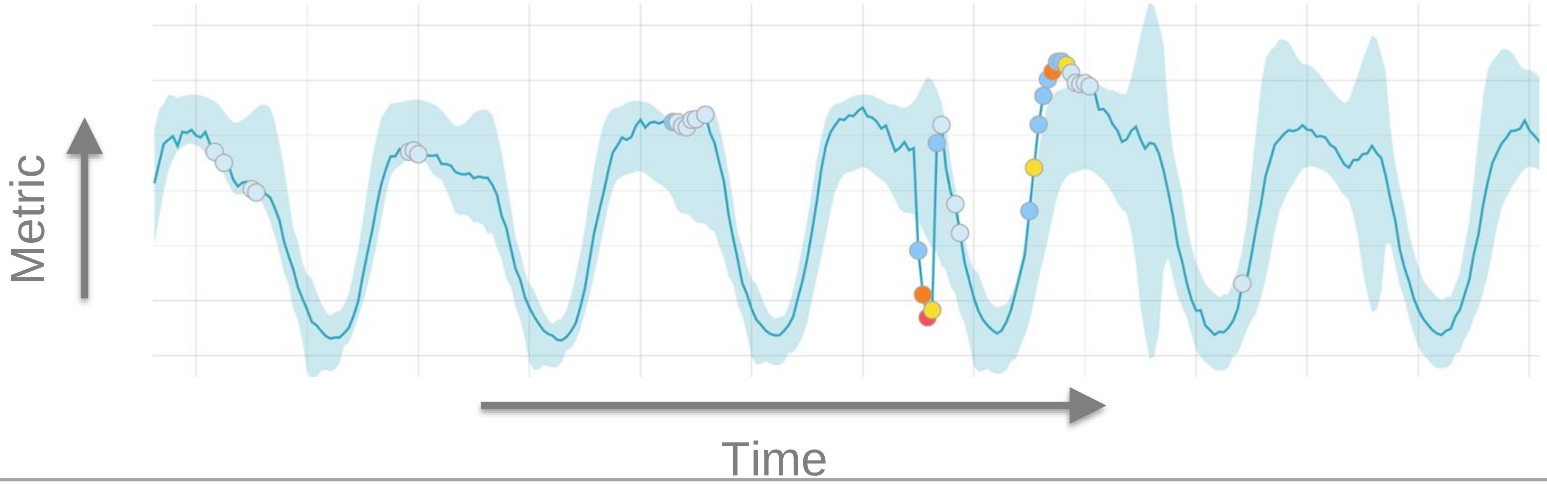
- Time series - 과거와 다른 행동 패턴 (by)
- Profiling - Outliers in population (using entity profiling) : 비슷한 다른 것들에 비교해서 다른 행동 패턴 (over)
- Rare / unusual rates in “categories” of events : 보기 드문 행동 패턴 (rare)

* 몇 십년 경험을 가진 시스템 아키텍트/관리자 및 보안 전문가의 노우하우(Know-How)를 시뮬레이션

Time – 싱글 메트릭(Single Metric)

- Single (univariate) time series

Example: *Is there unusual traffic on website ?*

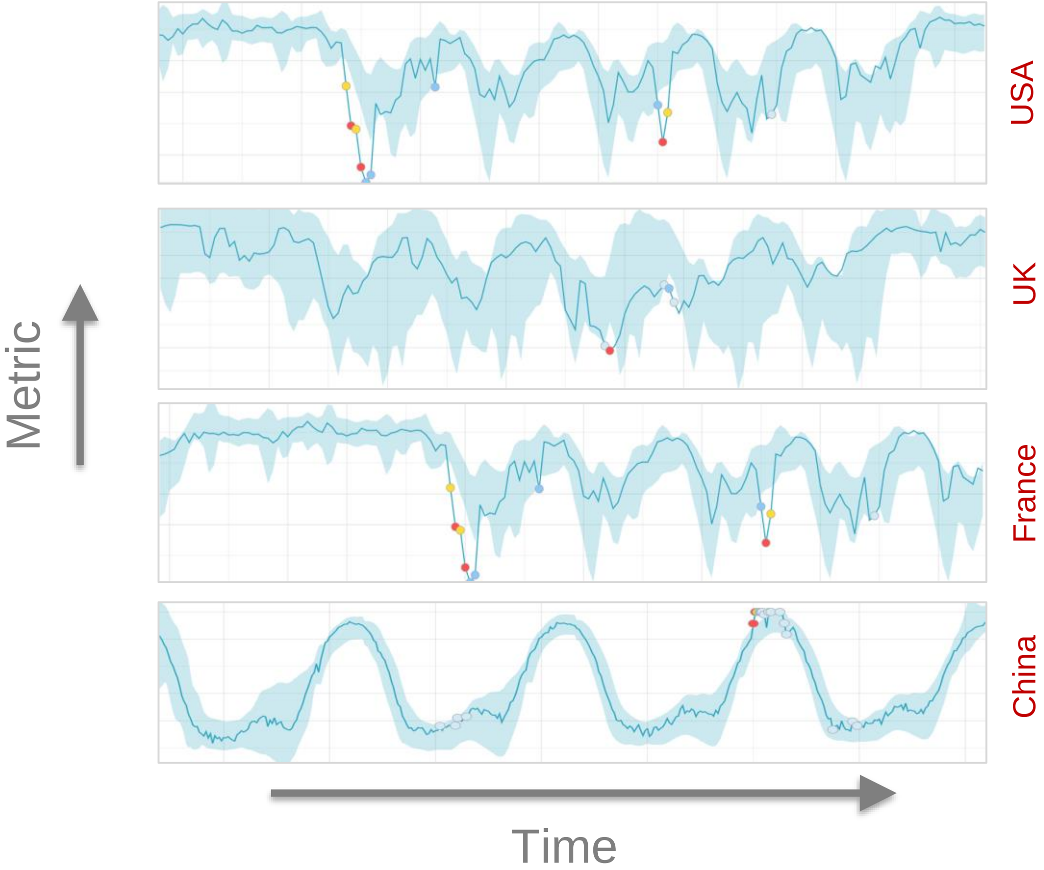


Time – 멀티 메트릭(Multi-Metric)

- Multiple time series
  - Multiple metrics
  - Single metric split by a field;
- Each series modeled independently

Example:

Is there unusual web activity from any country?



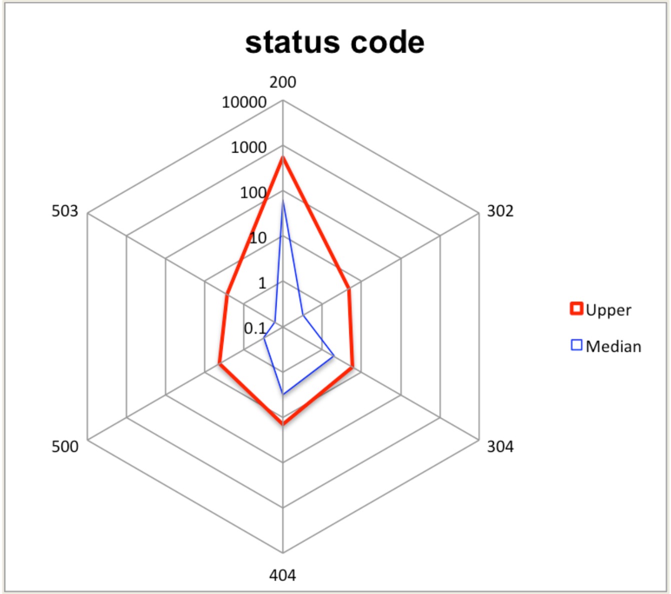
Profiling - Outliers in population (1)

- Create a profile for a “typical” entity (server, user, IP, etc.) in a population
- Detects entities (outlier) that deviate from the typical profile

Example*:*

- *Which IP address is not like the others?*

(indication of a bot / attacker)



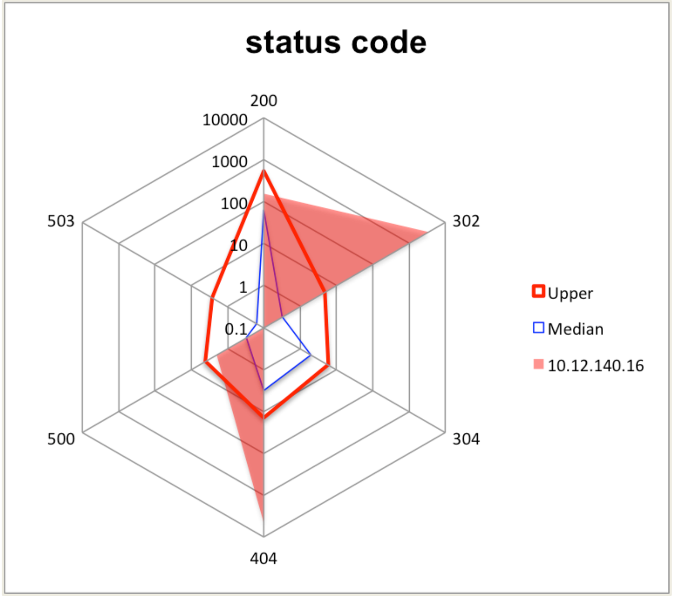
Profiling - Outliers in population (2)

- Create a profile for a “typical” entity (server, user, IP, etc.) in a population
- Detects entities (outlier) that deviate from the typical profile

Example:

- *Which IP address is not like the others?*

(indication of a bot / attacker)

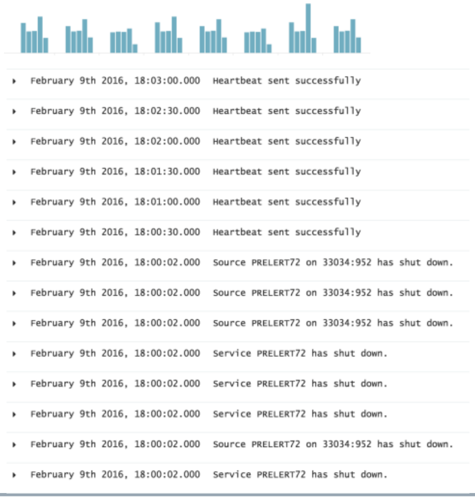


Rare – Unusual Events (via log categorization)

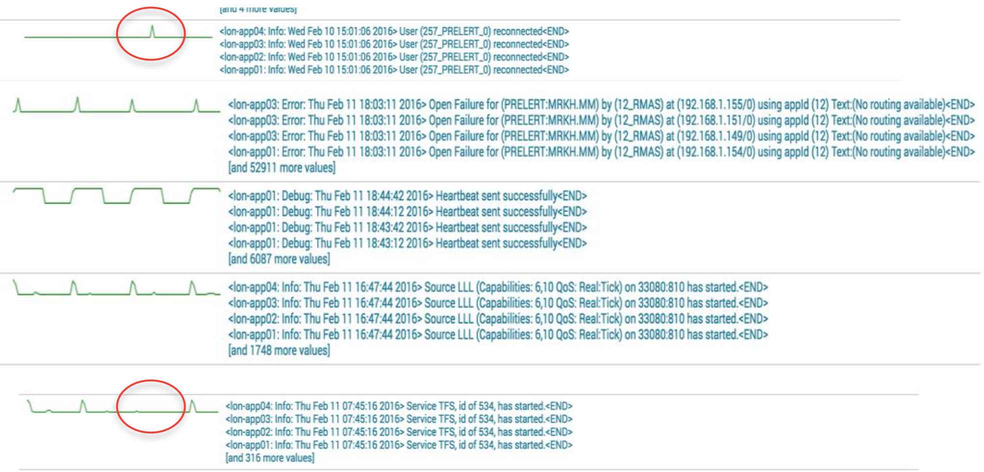
- Classify raw messages into groups based on similarity
- Models frequencies of each message category over time
- Spot anomalous in message groups

Example*:*

- *Do my application logs contain unusual messages*



```
▸ February 9th 2016, 18:03:00.000  Heartbeat sent successfully
▸ February 9th 2016, 18:02:30.000  Heartbeat sent successfully
▸ February 9th 2016, 18:02:00.000  Heartbeat sent successfully
▸ February 9th 2016, 18:01:30.000  Heartbeat sent successfully
▸ February 9th 2016, 18:01:00.000  Heartbeat sent successfully
▸ February 9th 2016, 18:00:30.000  Heartbeat sent successfully
▸ February 9th 2016, 18:00:02.000  Source PRELERT72 on 33034:952 has shut down.
▸ February 9th 2016, 18:00:02.000  Source PRELERT72 on 33034:952 has shut down.
▸ February 9th 2016, 18:00:02.000  Source PRELERT72 on 33034:952 has shut down.
▸ February 9th 2016, 18:00:02.000  Service PRELERT72 has shut down.
▸ February 9th 2016, 18:00:02.000  Service PRELERT72 has shut down.
▸ February 9th 2016, 18:00:02.000  Service PRELERT72 has shut down.
▸ February 9th 2016, 18:00:02.000  Source PRELERT72 on 33034:952 has shut down.
▸ February 9th 2016, 18:00:02.000  Service PRELERT72 has shut down.
```



```
[and 4 more values]
<lon-app04: Info: Wed Feb 10 15:01:06 2016> User (257_PRELERT_0) reconnected<END>
<lon-app03: Info: Wed Feb 10 15:01:06 2016> User (257_PRELERT_0) reconnected<END>
<lon-app02: Info: Wed Feb 10 15:01:06 2016> User (257_PRELERT_0) reconnected<END>
<lon-app01: Info: Wed Feb 10 15:01:06 2016> User (257_PRELERT_0) reconnected<END>

<lon-app03: Error: Thu Feb 11 18:03:11 2016> Open Failure for (PRELERT:MRKH.MM) by (12_RMAS) at (192.168.1.155/0) using appId (12) Text:(No routing available)<END>
<lon-app03: Error: Thu Feb 11 18:03:11 2016> Open Failure for (PRELERT:MRKH.MM) by (12_RMAS) at (192.168.1.151/0) using appId (12) Text:(No routing available)<END>
<lon-app03: Error: Thu Feb 11 18:03:11 2016> Open Failure for (PRELERT:MRKH.MM) by (12_RMAS) at (192.168.1.149/0) using appId (12) Text:(No routing available)<END>
<lon-app01: Error: Thu Feb 11 18:03:11 2016> Open Failure for (PRELERT:MRKH.MM) by (12_RMAS) at (192.168.1.154/0) using appId (12) Text:(No routing available)<END>
[and 52911 more values]

<lon-app01: Debug: Thu Feb 11 18:44:42 2016> Heartbeat sent successfully<END>
<lon-app01: Debug: Thu Feb 11 18:44:12 2016> Heartbeat sent successfully<END>
<lon-app01: Debug: Thu Feb 11 18:43:42 2016> Heartbeat sent successfully<END>
<lon-app01: Debug: Thu Feb 11 18:43:12 2016> Heartbeat sent successfully<END>
[and 6087 more values]

<lon-app04: Info: Thu Feb 11 16:47:44 2016> Source LLL (Capabilities: 6,10 QoS: Real:Tick) on 33080:810 has started.<END>
<lon-app03: Info: Thu Feb 11 16:47:44 2016> Source LLL (Capabilities: 6,10 QoS: Real:Tick) on 33080:810 has started.<END>
<lon-app02: Info: Thu Feb 11 16:47:44 2016> Source LLL (Capabilities: 6,10 QoS: Real:Tick) on 33080:810 has started.<END>
<lon-app01: Info: Thu Feb 11 16:47:44 2016> Source LLL (Capabilities: 6,10 QoS: Real:Tick) on 33080:810 has started.<END>
[and 1748 more values]

<lon-app04: Info: Thu Feb 11 07:45:16 2016> Service TFS, id of 534, has started.<END>
<lon-app03: Info: Thu Feb 11 07:45:16 2016> Service TFS, id of 534, has started.<END>
<lon-app02: Info: Thu Feb 11 07:45:16 2016> Service TFS, id of 534, has started.<END>
<lon-app01: Info: Thu Feb 11 07:45:16 2016> Service TFS, id of 534, has started.<END>
[and 316 more values]
```

DEMO 1 – Single / Multi Metrics and basic concept