

Dr. Radu Tudoran

# Flink Forward China 2018

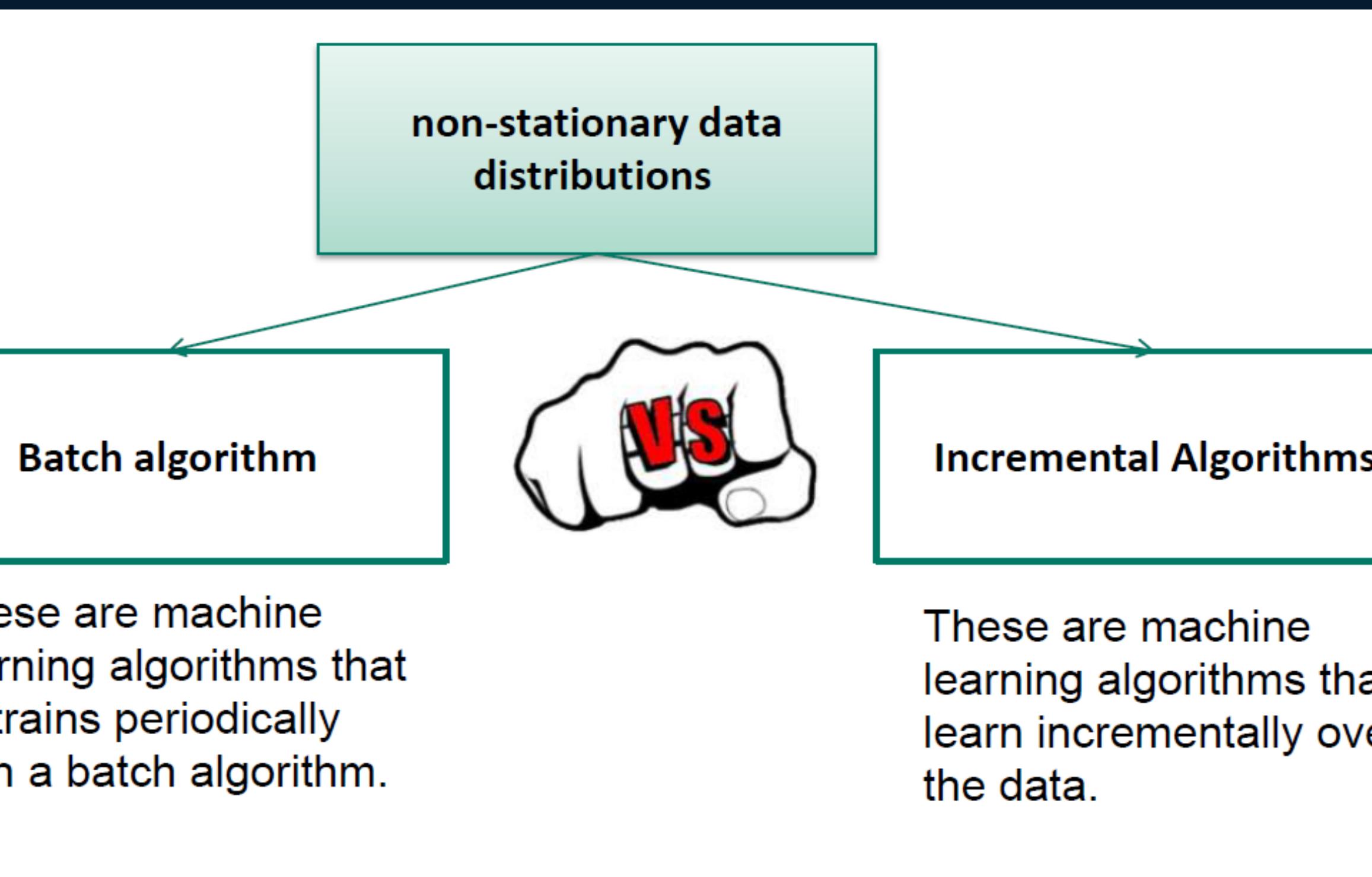
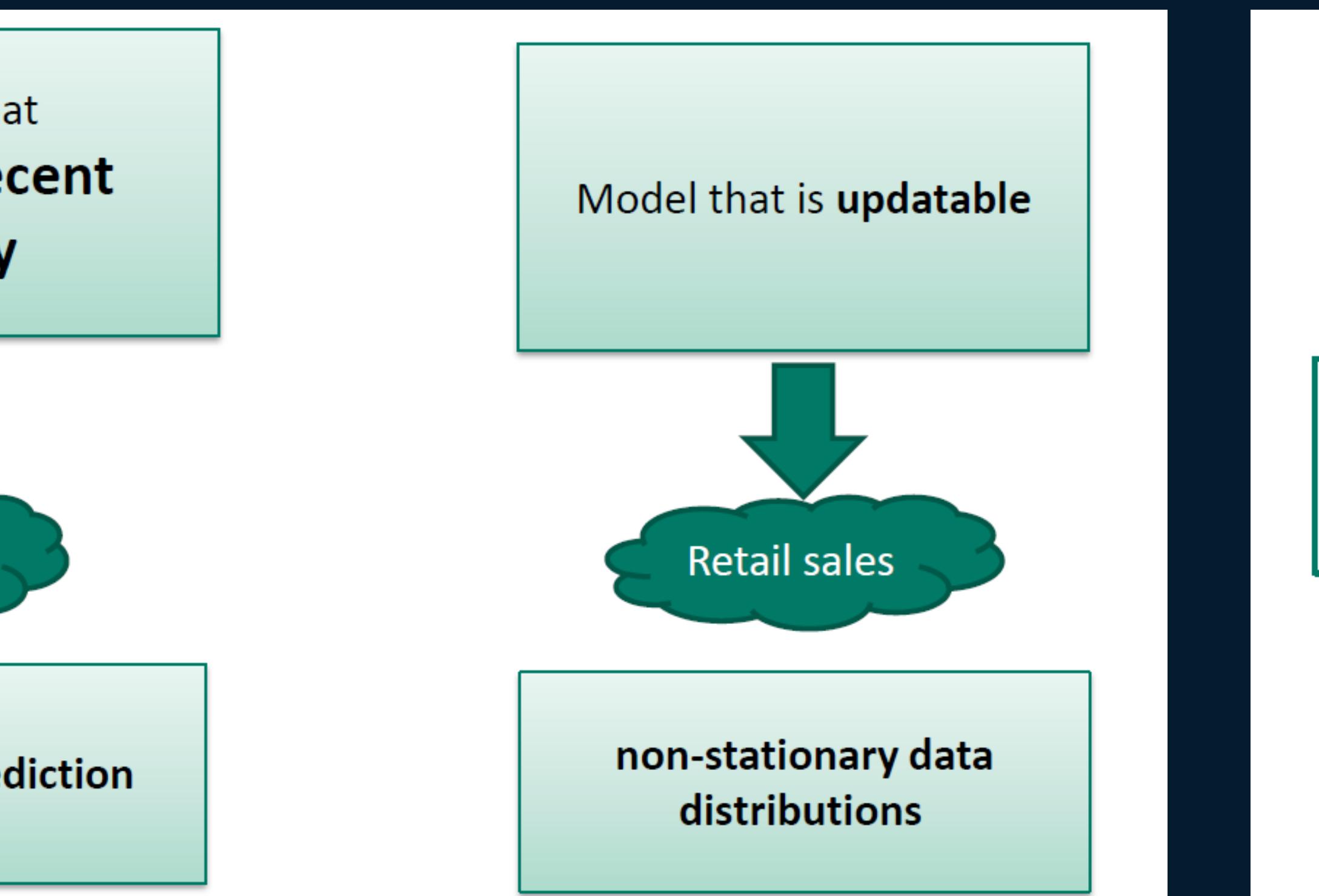
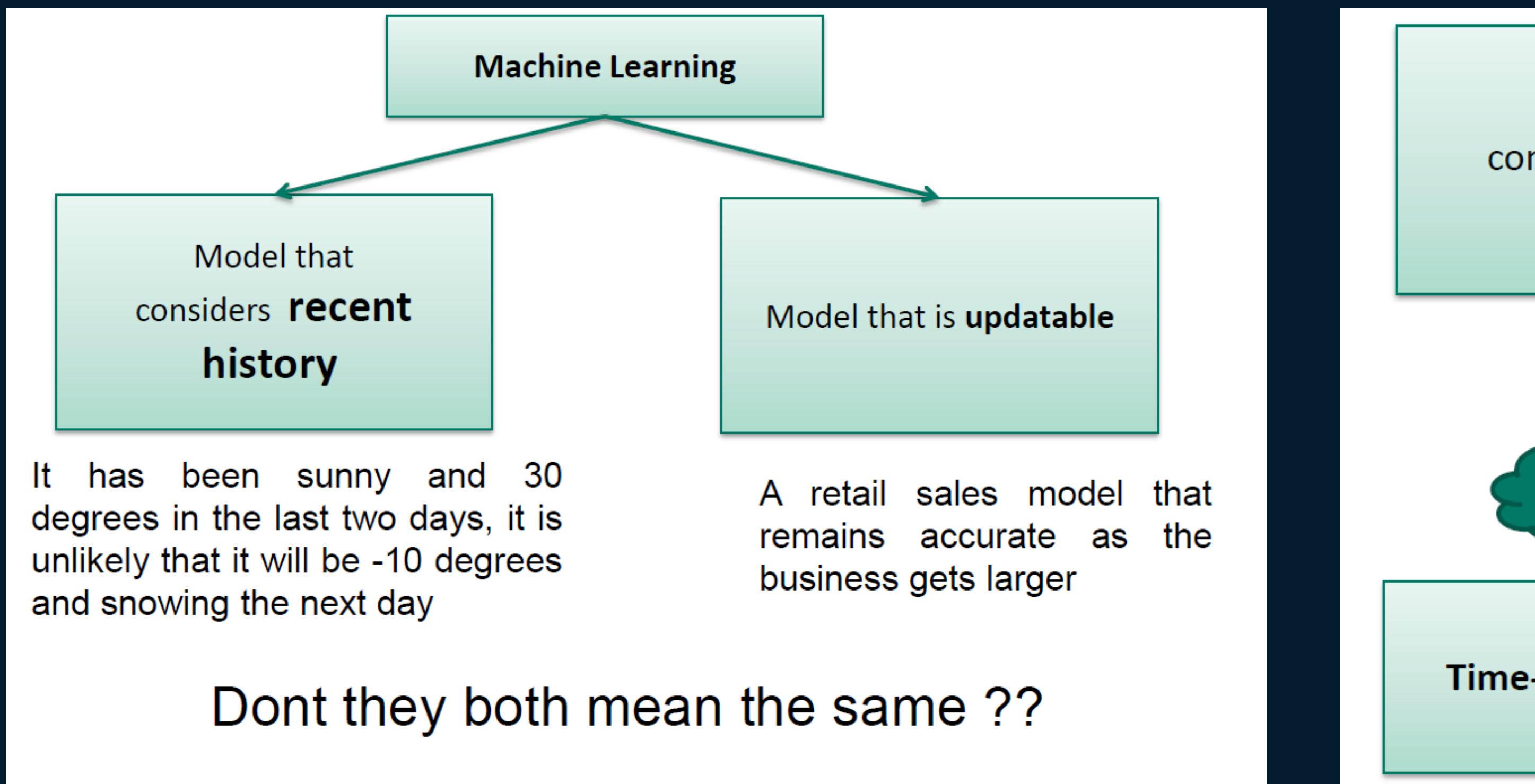




# Flink Forward China 2018

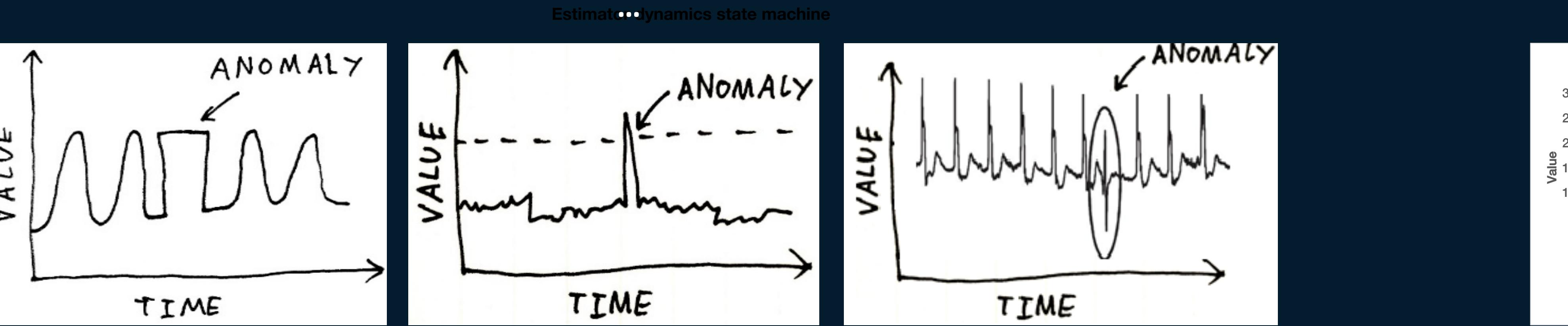
In this AI era and continuous incoming data, we need new platform and algorithmic approaches to enable to learn and take smart decisions in real time by continuous adaptation.

# What is Online ML?

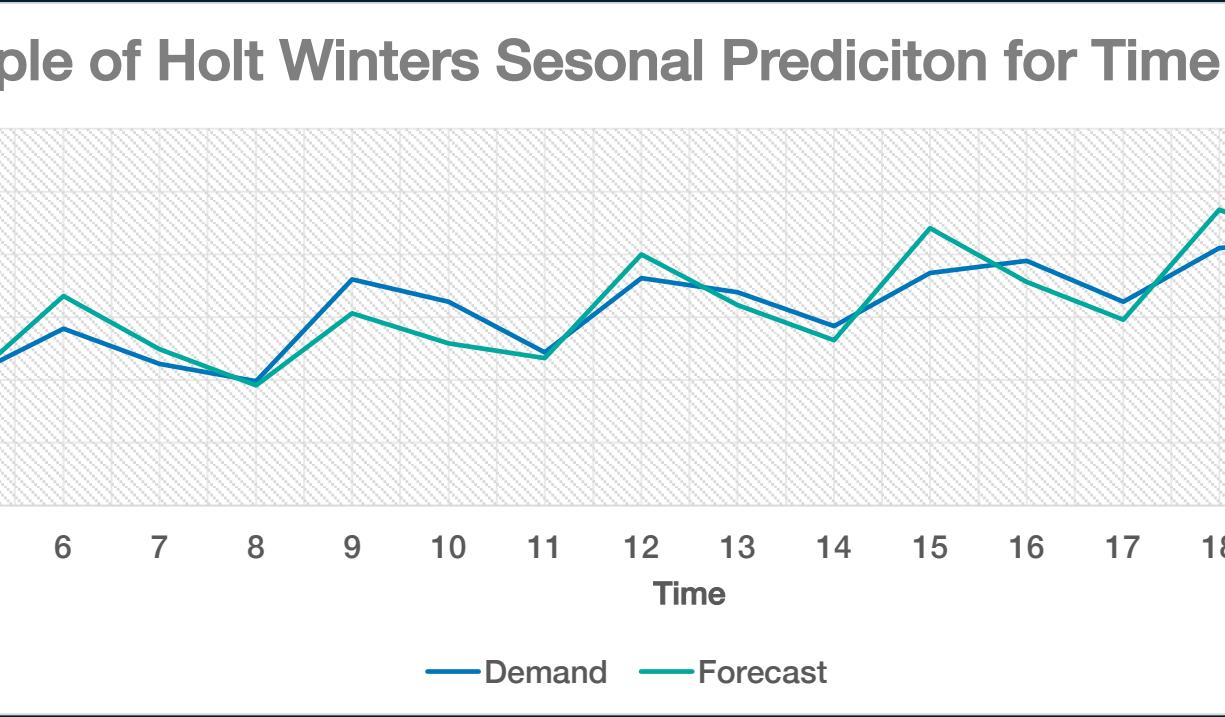


# Business Scenarios for Online ML

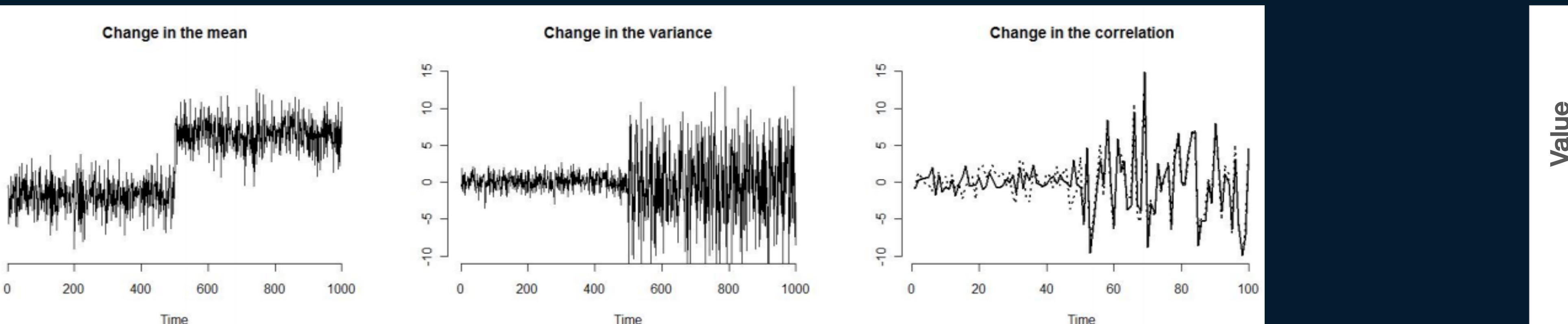
Anomaly Detection and Prediction for: IoT, Smart Factory, Smart Manufacturing, SafeCity



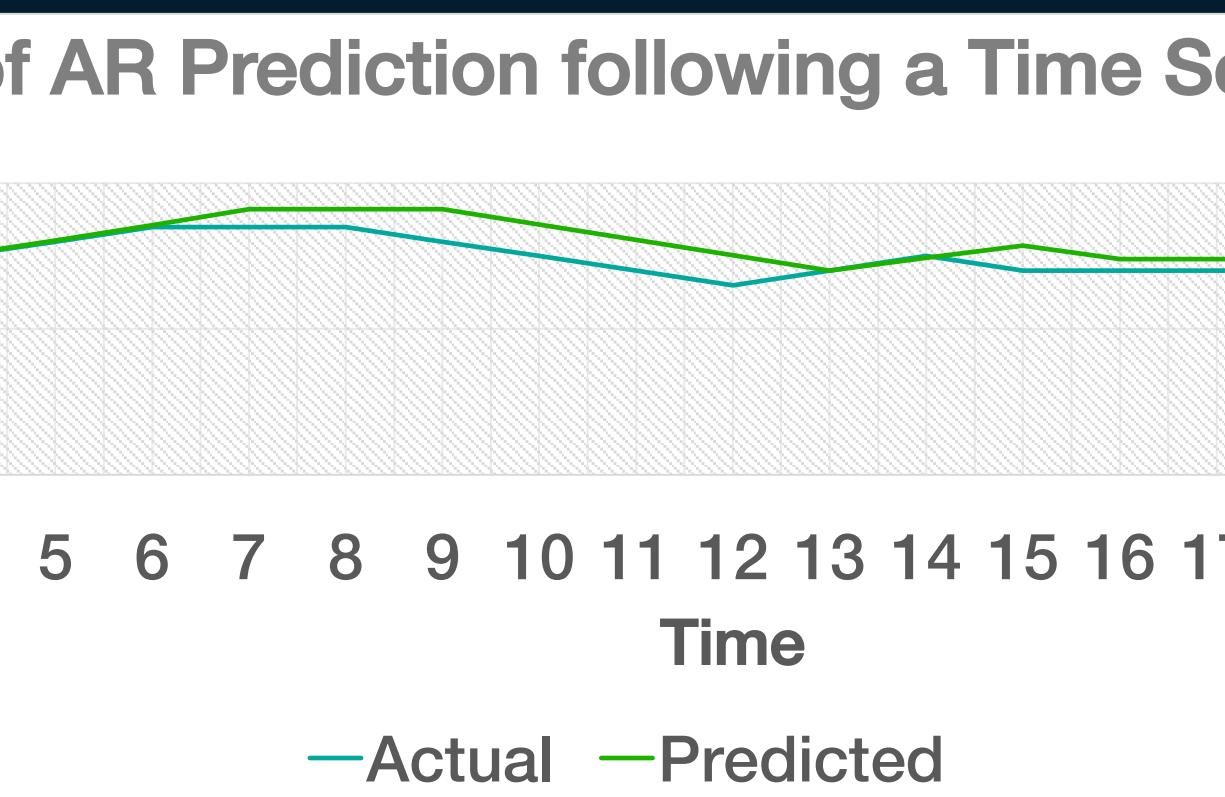
Forecast of values in real time by modeling the sesonality of time series samples



Distribution Detection and Prediction for: Smart Factory, Smart Manufacturing, Advertising...

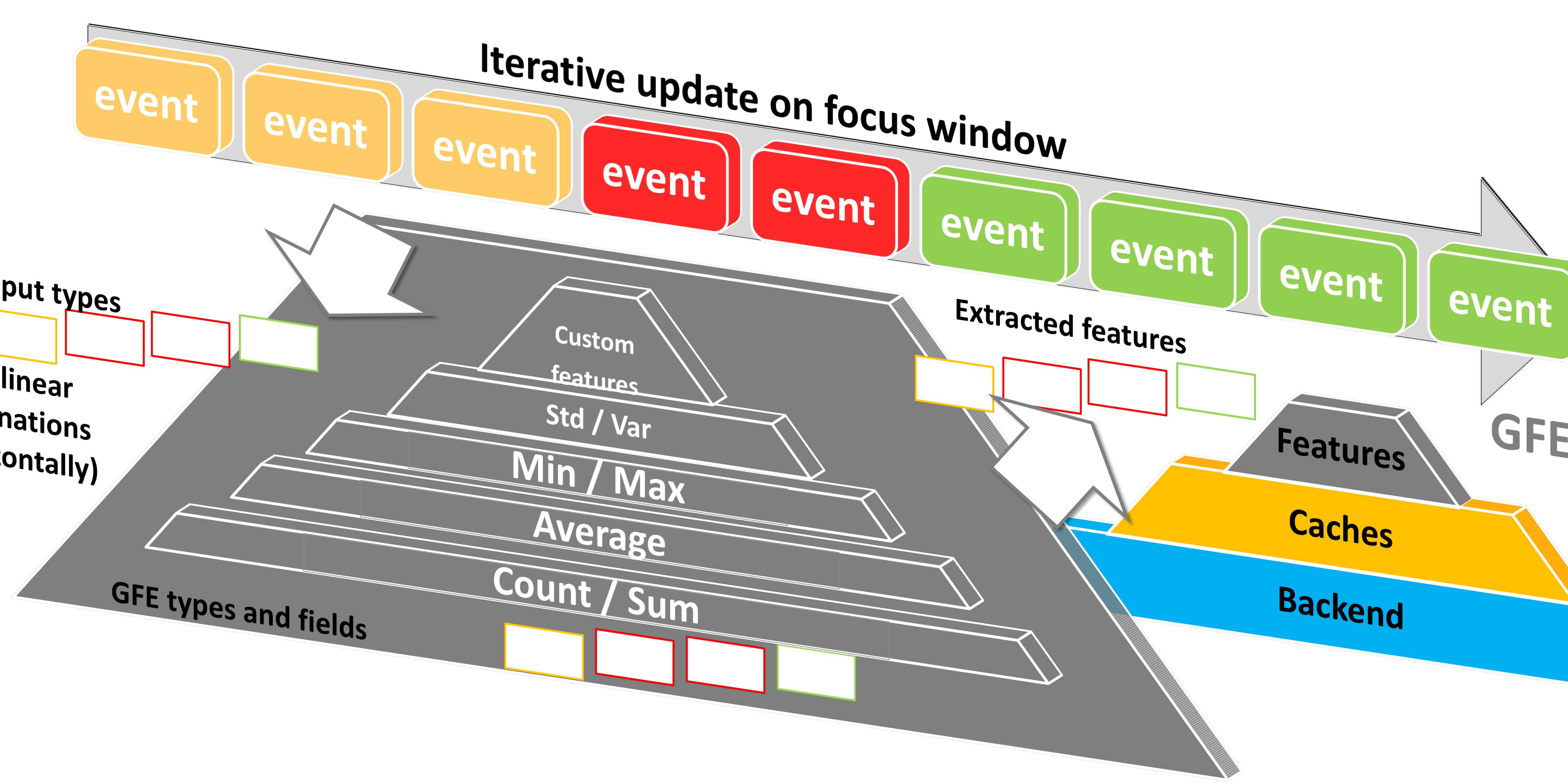


Forecast of values in real time by modeling the time series samples

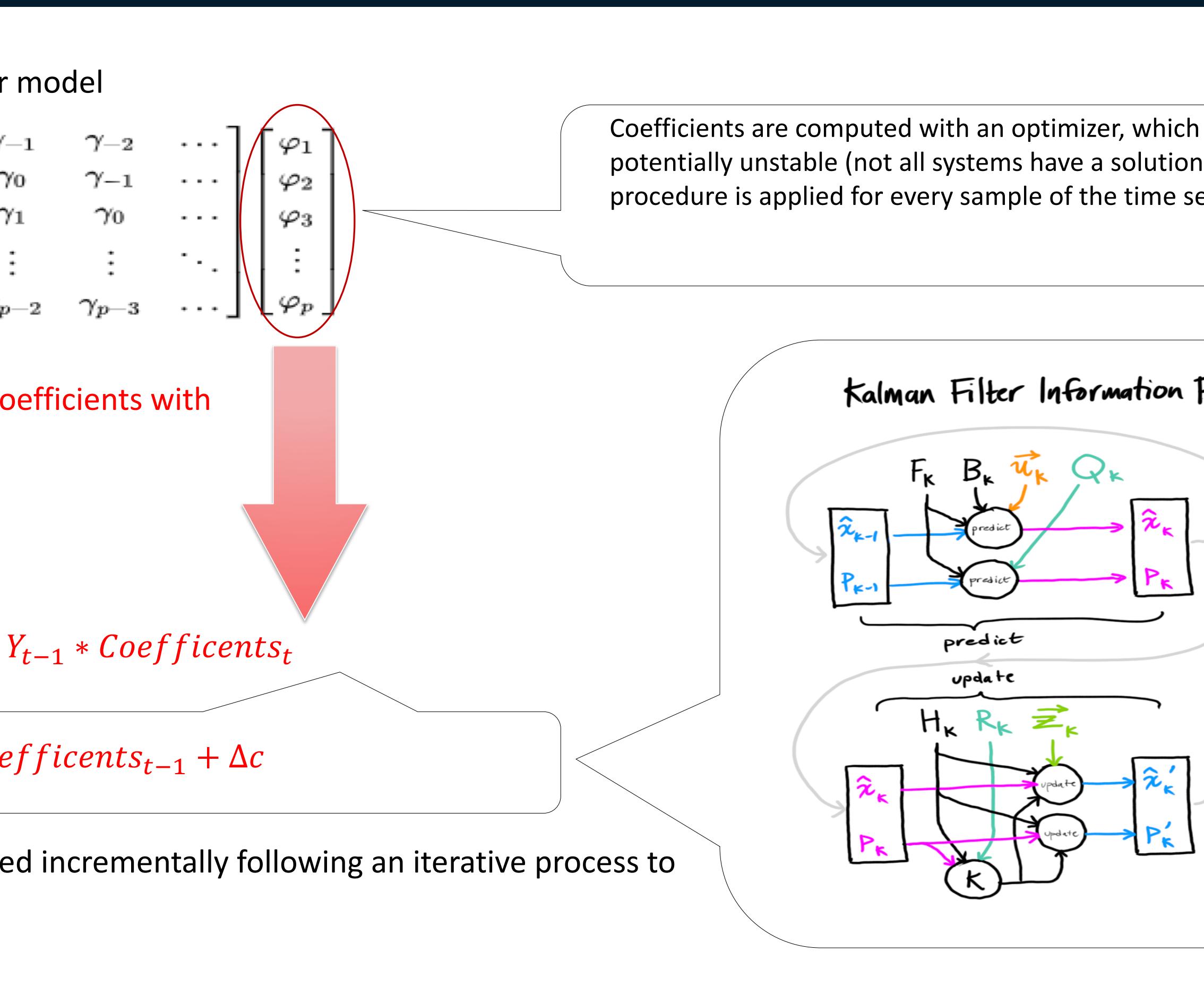
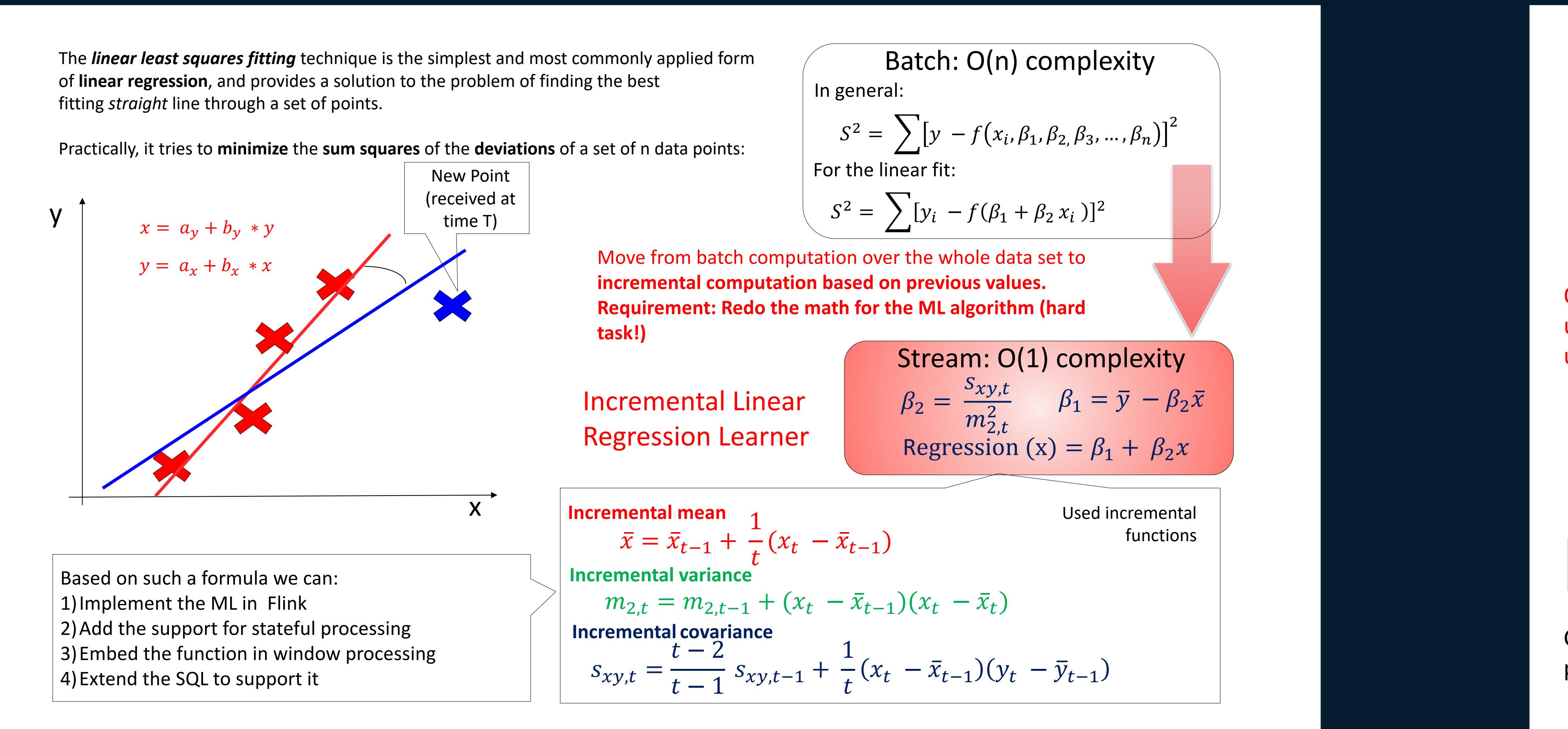


- Design Principles:**
- Incremental computation
  - Fixed size memory
  - Constant to sub-linear time complexity

# Stream Platform for Online ML



# Stream (Redesigned) Algorithms for Online ML



# Online ML Enabled in (Stream) SQL

```
SELECT
    COVAR(c, e) OVER (ORDER BY rowtime RANGE BETWEEN INTERVAL '10' HOUR PRECEDING AND CURRENT ROW) AS cov,
    LLS_PRED(c, e) OVER (ORDER BY rowtime RANGE BETWEEN INTERVAL '10' HOUR PRECEDING AND CURRENT ROW) AS pred,
    LLS_REG(c, e) OVER (ORDER BY rowtime RANGE BETWEEN INTERVAL '10' HOUR PRECEDING AND CURRENT ROW) AS reg
FROM stream
```

```
SELECT
    ARGMIN(e, LOG(e)) OVER (ORDER BY proctime RANGE BETWEEN INTERVAL '30' SECOND PRECEDING AND CURRENT ROW) AS aminE,
    ARGMAX(e, LOG(e)) OVER (ORDER BY proctime RANGE BETWEEN INTERVAL '30' SECOND PRECEDING AND CURRENT ROW) AS amaxE
FROM stream
```

```
SELECT
    CENTROID(ARRAY[c,e], 1.0) OVER (ORDER BY proctime RANGE UNBOUNDED PRECEDING) AS centroid,
    CLUSTER_CENTROIDS(ARRAY[c,e], 1.0) OVER (ORDER BY proctime RANGE UNBOUNDED PRECEDING) AS centroids
FROM MyTable
```

```
SELECT c,
    CASE WHEN SRF_UNSUP(ARRAY[c]) OVER (ORDER BY proctime RANGE BETWEEN INTERVAL '300' SECOND PRECEDING AND CURRENT ROW) > 0.8
        THEN 'anomaly'
        ELSE 'not anomaly'
    END
FROM MyTable
```

```
SELECT b,
    HOLT_WINTERS(b, 5) OVER (ORDER BY rowtime ROWS BETWEEN 5 PRECEDING AND CURRENT ROW) AS hw1,
    HOLT_WINTERS(b, 5, 2) OVER (ORDER BY rowtime ROWS BETWEEN 5 PRECEDING AND CURRENT ROW) AS hw2,
FROM MyTable
```

```
SELECT b,
    AR_PRED(b) OVER (ORDER BY rowtime ROWS BETWEEN 5 PRECEDING AND CURRENT ROW) AS ar,
    ARMA_PRED(b) OVER (ORDER BY rowtime ROWS BETWEEN 5 PRECEDING AND CURRENT ROW) AS arma,
    ARIMA_PRED(b) OVER (ORDER BY rowtime ROWS BETWEEN 5 PRECEDING AND CURRENT ROW) AS arima
FROM MyTable
```

```
SELECT
    NEAREST_FUZZY_CENTROID(ARRAY[c,e], 3, 10, 2, 0.01, \genericEuclideanDistance\, \weightedEuclideanAvg\
    ARRAY[ARRAY[1.0,1.0],ARRAY[3.0,3.0], ARRAY[5.0,5.0]], ARRAY[10,20,15]) OVER
    (ORDER BY proctime RANGE BETWEEN INTERVAL '60' MINUTE PRECEDING AND CURRENT ROW) AS probableCentroid,
FROM MyTable
```

```
SELECT
    KMEANS_CLUSTER_CENTROID(ARRAY[c,e], 3, 10, \genericEuclideanDistance\, \genericEuclideanAvg\ ) OVER
    (ORDER BY proctime RANGE BETWEEN INTERVAL '60' MINUTE PRECEDING AND CURRENT ROW) AS KMeansCentroid,
FROM MyTable
```

40 new Online ML SQL functions added already.

# Online DeepLearning with SQL

**Enabled Stream DL serving via SQL based on multiple engines:**

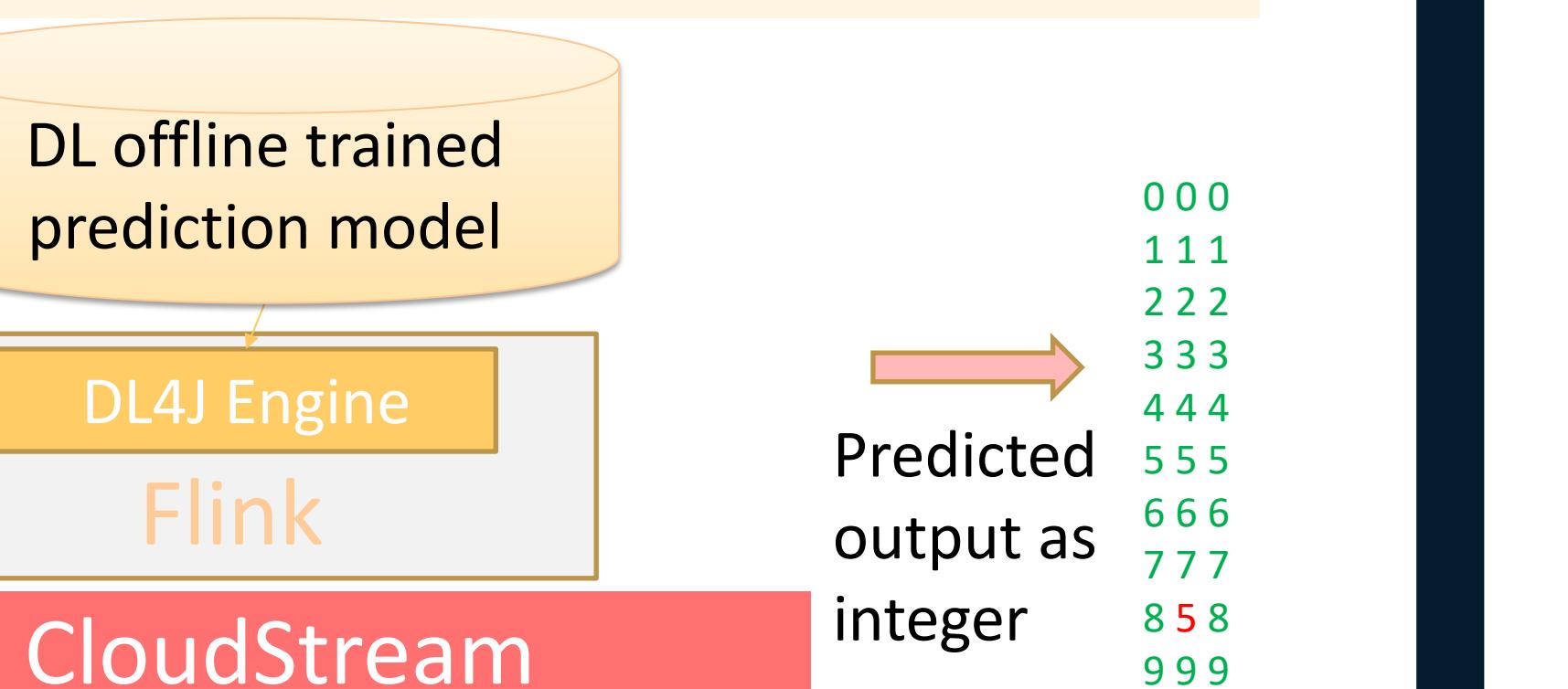
- **DL4J** (importing also models from TensorFlow using Kera)
- **TensorFlow native support**

## DL Online Serving for Image

Application scenario supported

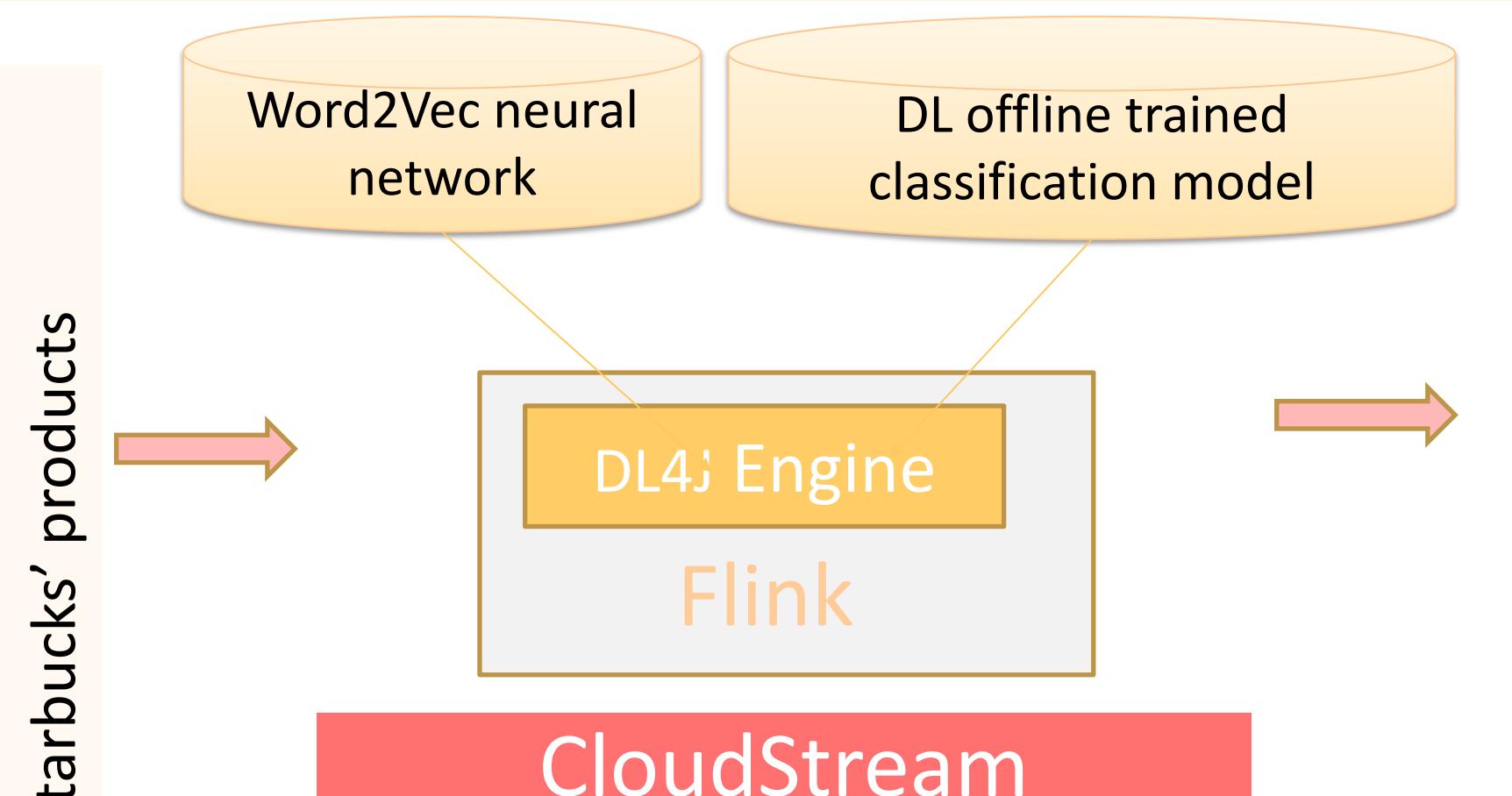
- Text classification (using Word2Vec)
- Image classification
- Handwriting digits prediction

```
SELECT DL_IMAGE_MAX_PREDICTION_INDEX(  
    image, /* input image under byte[ ] format*/  
    modelConfigFilePath,  
    weightsHdf5File  
) as prediction  
FROM MyTable
```

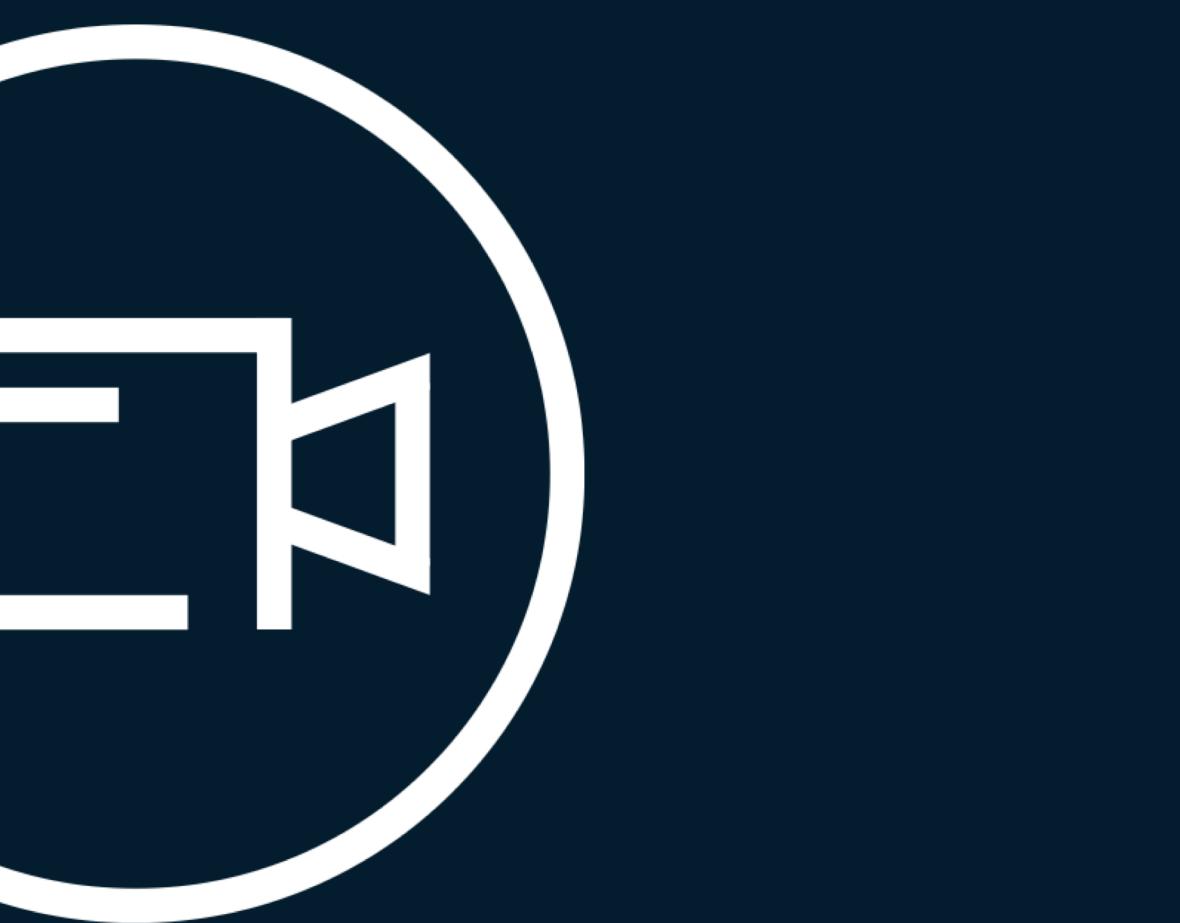


## DL Online Serving for Text

```
SELECT headlines  
FROM MyTable  
WHERE  
    DL_TEXT_MAX_PREDICTION_INDEX(  
        headlines, /* Text input data*/  
        word2VecModelFile, /* Word2Vec neural network file path*/  
        predictionModelFile, /* Pre-trained prediction model path*/  
        dl4jModel /* True if DL4J model, false otherwise*/  
) = 2 /* Bollywood category*/
```



FLINK  
FORWARD



THANKS

