A Fog Computing Prototype

Course Project for Big Data Analytics — Winter 2019

Ali Alizadeh Mansouri Marco Sassano

April 15, 2019

Concordia University

Introduction

What is Fog Computing?

Internet of Things (IoT)

- 50 billion devices (sensors, smart phones, smart cities, healthcare, smart vehicles, ...)[5]
- 100s of terabytes towards petabytes per day
- limited computing resources

What is Fog Computing?

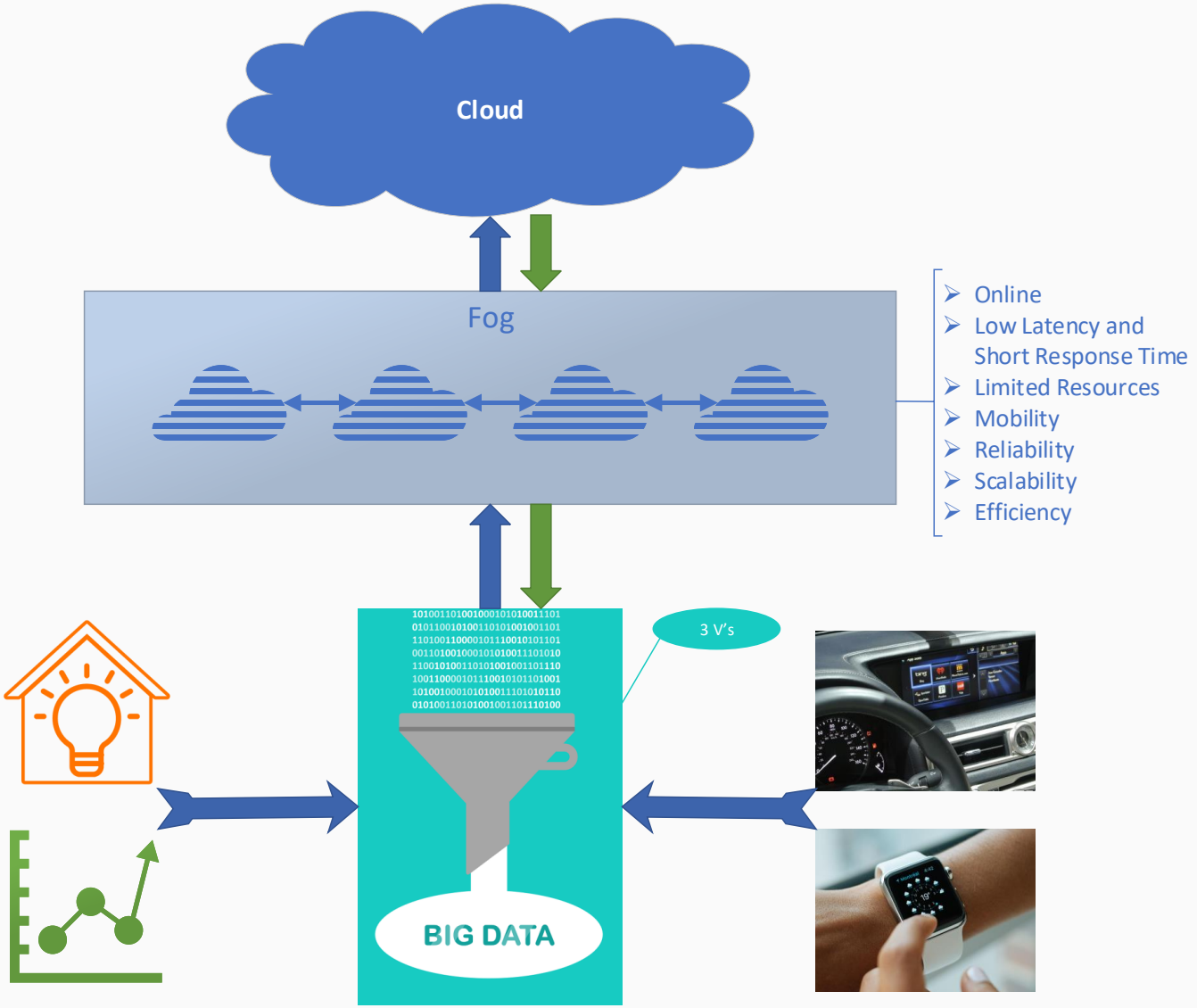
Internet of Things (IoT)

- 50 billion devices (sensors, smart phones, smart cities, healthcare, smart vehicles, ...)[5]
- 100s of terabytes towards petabytes per day
- limited computing resources

Cloud computing

- Flexible economy
- Scalability
- Adaptation to varieties of computational demands
- Suffers from high latency

A Fog Computing environment



Problem

Problem Specification

Early detection of epilepsy seizures using EEG timeseries data

Inspired by Diab Abdulgalil et. al. [4]

- Stream processing on the edge
  - Lower accuracy, but very high response time
- Clustering in the cloud
  - Longer latency and response time, but more accurate

Dataset

- Original dataset[2]: 500 individuals, each with 4097 data points for $23.5s$
- The UCI Machine Learning Repository version [1]:
  $23 \times 500 = 11500$ records, each record contains 178 data points (columns) for 1 second
- Restructured to `(time, value)` tuples to be used as a stream
  $\Longrightarrow 11500 \times 178 = 2\,047\,000$ tuples (windows of 178)
- 5 classes $\longrightarrow$ 2 classes (binary)

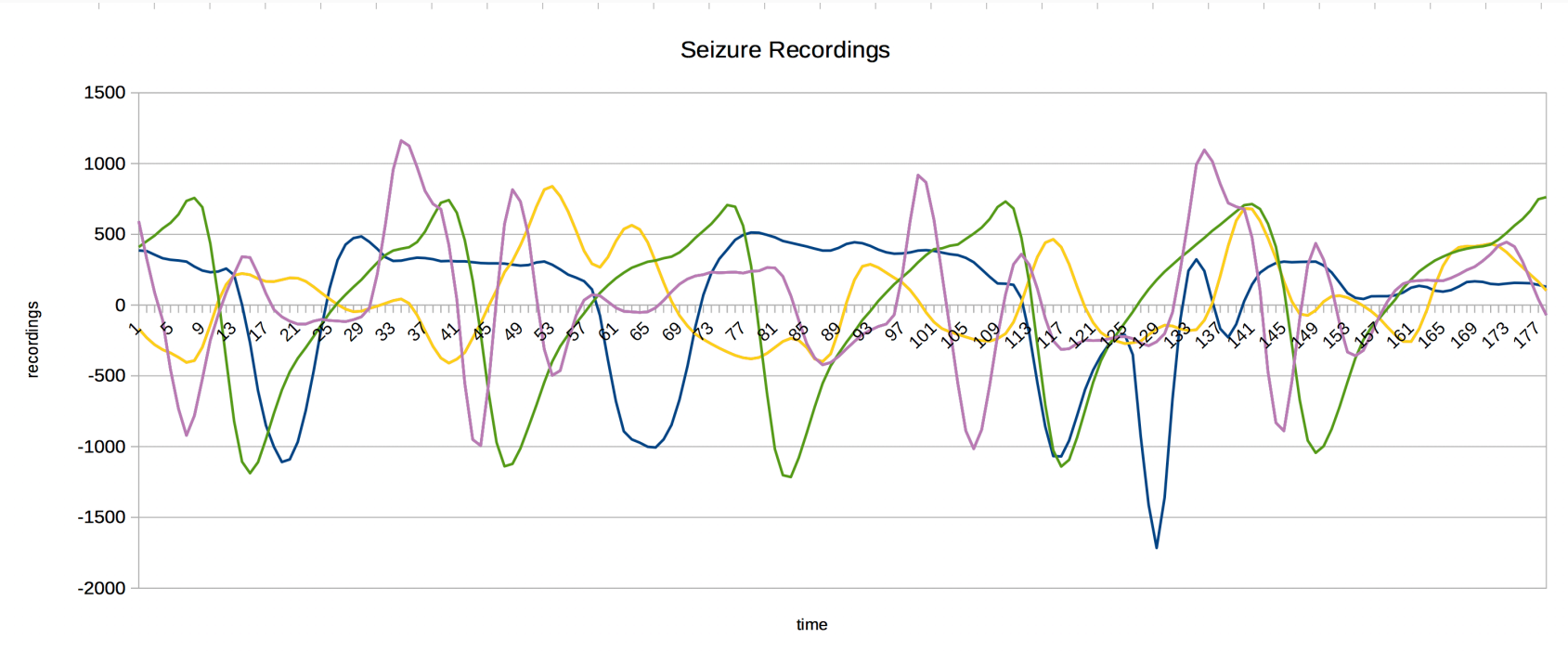
	Total	Positive	Negative
Training	7 000	1 392	5 608
Test	4 500	908	3 592
	11 500	2 300	9 200

Methodology

Stream Filtering

Goal Filter the stream for out-of-range tuples/anomalies

- light-weight on resources
- fast



Different statistical measures: variance, skewness, kurtosis [3]

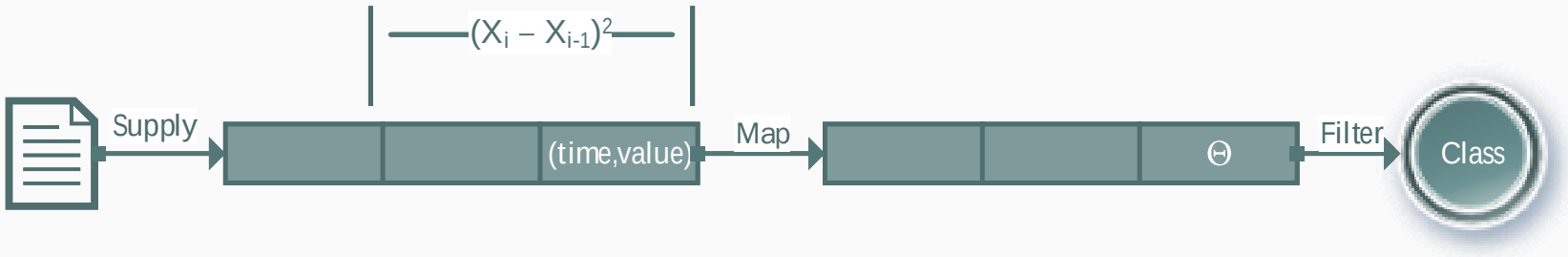
Selected measure

$$\Theta = \sqrt{\sum_{i=2}^{n} (X_i - X_{i-1})^2}$$

Stream Processing at the Edge

Apache Edgent: Programming model and runtime (Java)

- **Supply** reads the dataset and supplies it as a stream
- **Map** maps the tuples to the selected measure
- **Analysis** classifies the out-of-range Θ values as seizures



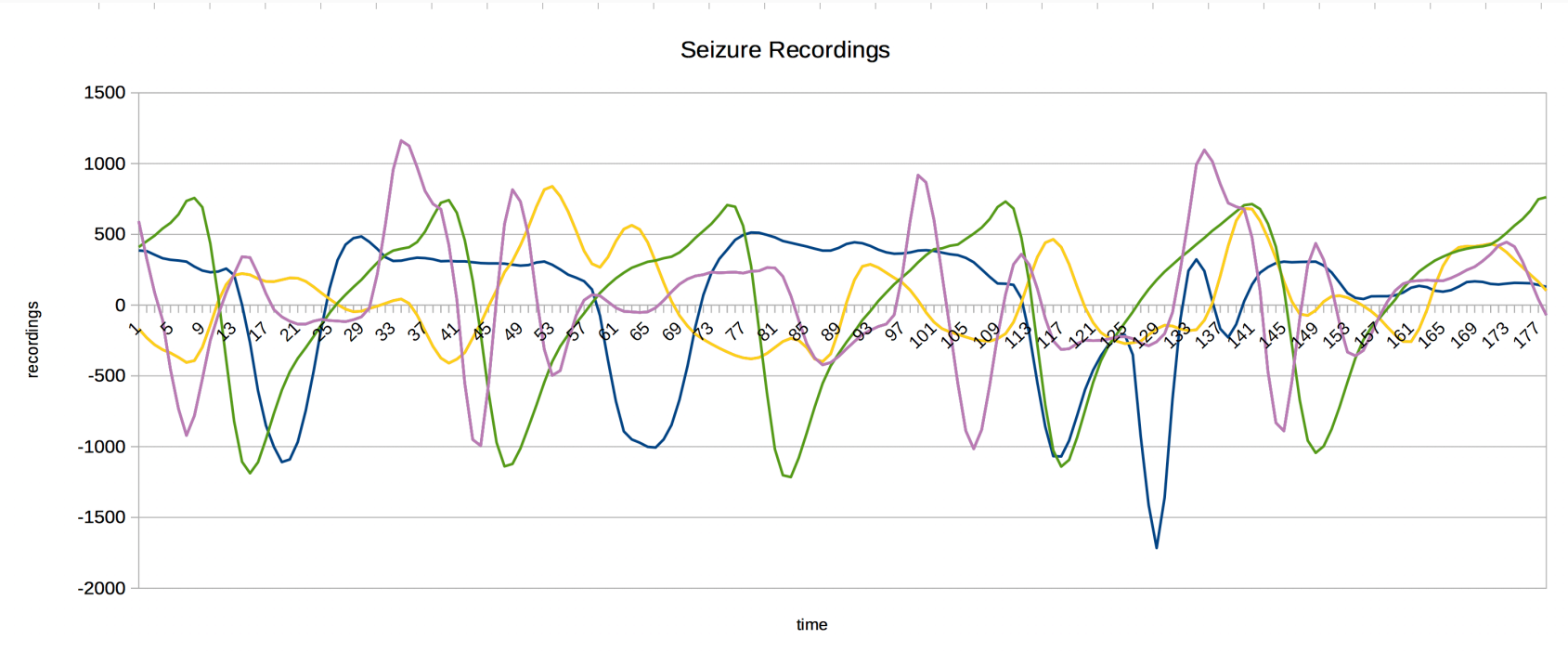
Results


TRAINING DATA
Accuracy
F1 score
θ
185.1589793
186
187
188
200
225
250
275
300
325
350
375
400
425
450
475
500
510
525
550
Accuracy
F1 score
Recall
Precision


TEST DATA
Accuracy
F1 score
θ
185.1589793
186
187
188
200
225
250
275
300
325
350
375
400
425
450
475
500
510
525
550
Accuracy
F1 score
Recall
Precision

Inspirations

- The EGADS anomaly detection library by Yahoo[6] uses clustering as basis for most of its
- Carney et. al.[3]



We tried:

- Euclidean distance
- Transform data entries into 178 dimensions

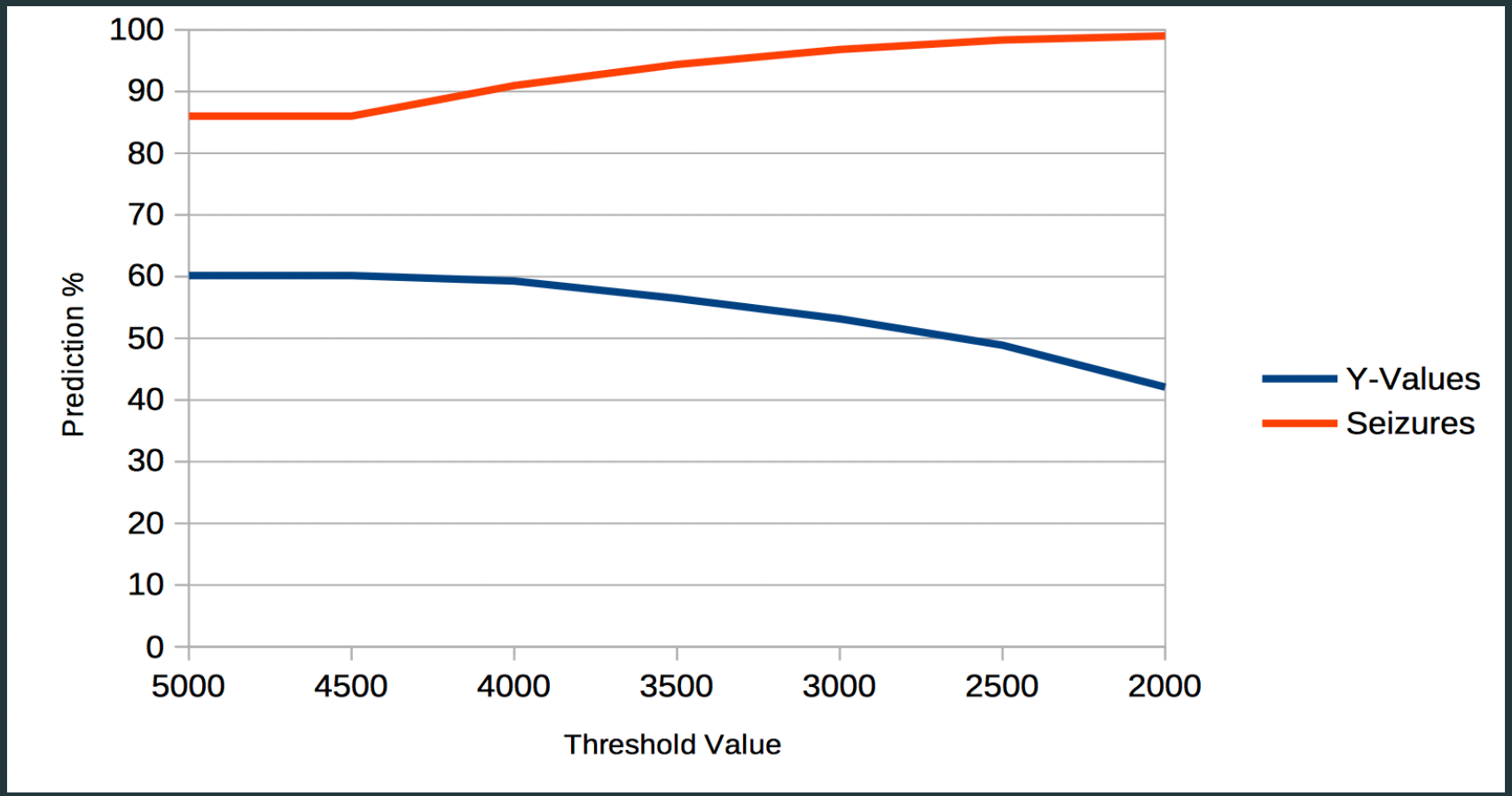
This failed due to different time intervals

Selected distance measure

$$s = \sum_{i=2}^{n} \left| X_i - X_{i-1} \right|$$

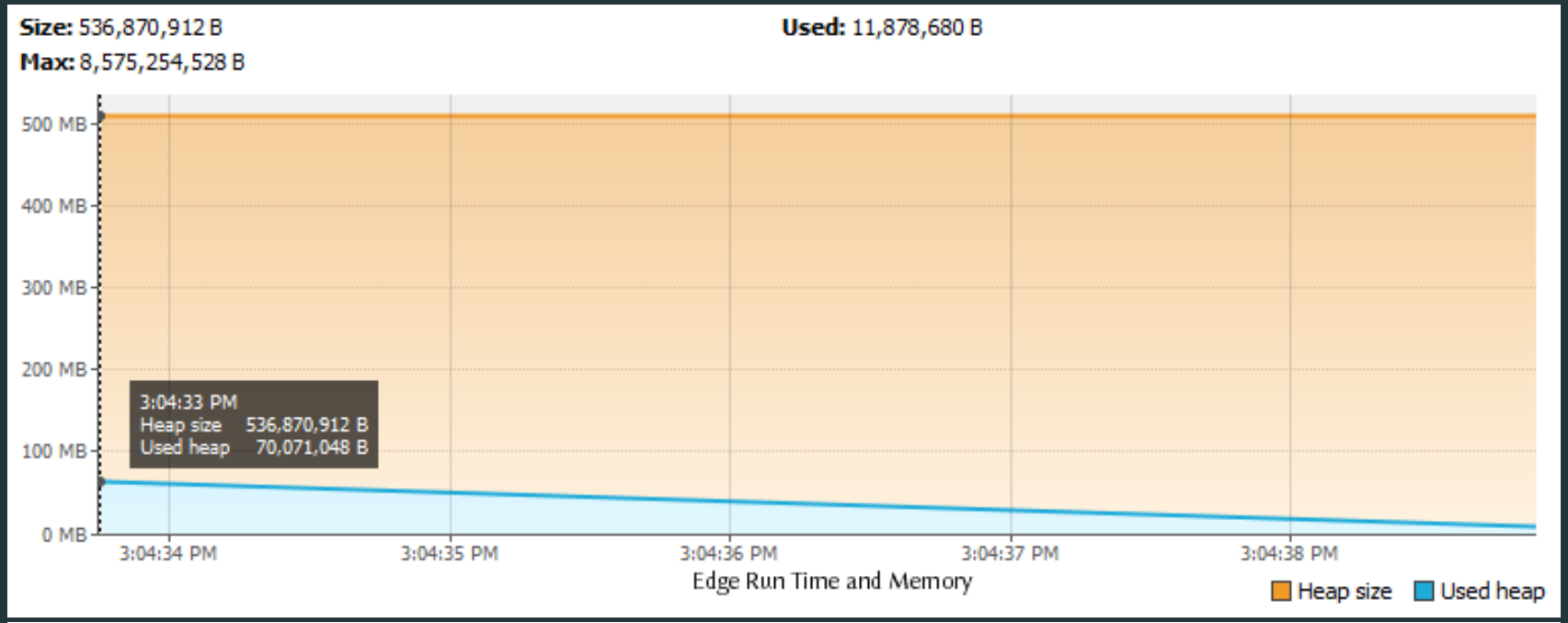
Apache Spark (pyspark/Python)

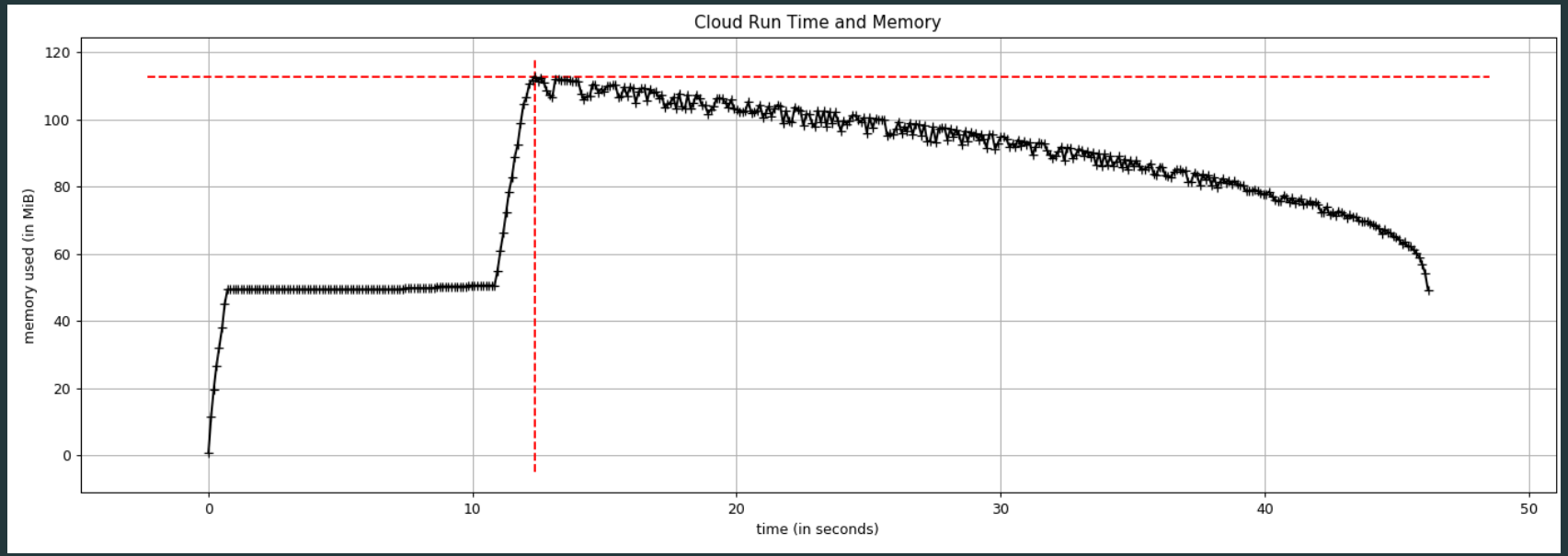
Results


100
90
80
70
60
50
40
30
20
10
0
Prediction %
5000
4500
4000
3500
3000
2500
2000
Threshold Value
Y-Values
Seizures

Conclusion

Performance Comparison





Limitations and Future Work

- Experiment with more than one sensor
- Introduce a publish-subscribe platform such as Apache Kafka, IBM Watson IoT Hub, etc. (e.g. using MQTT as the protocol)

Thank you!

References 1

Epileptic seizure recognition data set.
`https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition`.
Accessed: 2019-04-15.

R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger.
Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state.
Phys. Rev. E, 64:061907, Nov 2001.

References ii

P. R. Carney, S. Myers, and J. D. Geyer.

Seizure prediction: methods.

Epilepsy & behavior : E&B, 22 Suppl 1(Suppl 1):S94–S101, dec 2011.

H. Diab Abdulgalil.

A Multi-Tier Distributed fog-based Architecture for Early Prediction of Epileptic Seizures.

PhD thesis, University of Waterloo, 2018.

D. Evans.

The Internet of Things: How the Next Evolution of the Internet is Changing Everything.

Technical report, 2011.

References iii


N. Laptev, S. Amizadeh, and I. Flint.
Generic and scalable framework for automated time-series anomaly detection.
In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1939–1947. ACM, 2015.