

# AVFDT: ADAPTIVE VERY FAST DECISION TREE

## PRELIMINARY RESULTS

Eva García-Martín, N. Lavesson, H. Grahn, E. Casalicchio, V. Boeva  
Blekinge Institute of Technology  
{eva.garcia.martin@bth.se}



## INTRODUCTION

### Problem

Machine learning algorithms account for a significant amount of energy consumption in data centers.



### Goal

Reduce energy consumption of online decision trees.

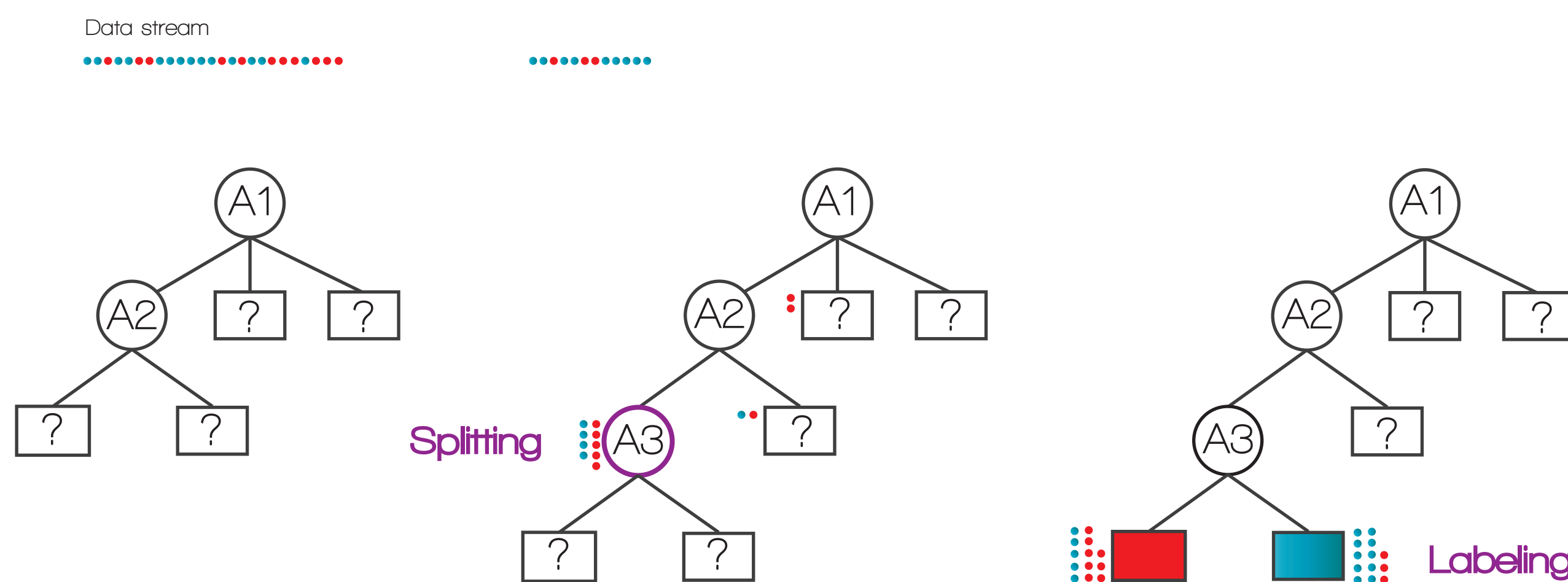


### How?

Adaptive Very Fast Decision Tree (AVFDT).

Extension of the VFDT that uses the  $nmin$  adaptation method.

## Very Fast Decision Tree (VFDT [1])



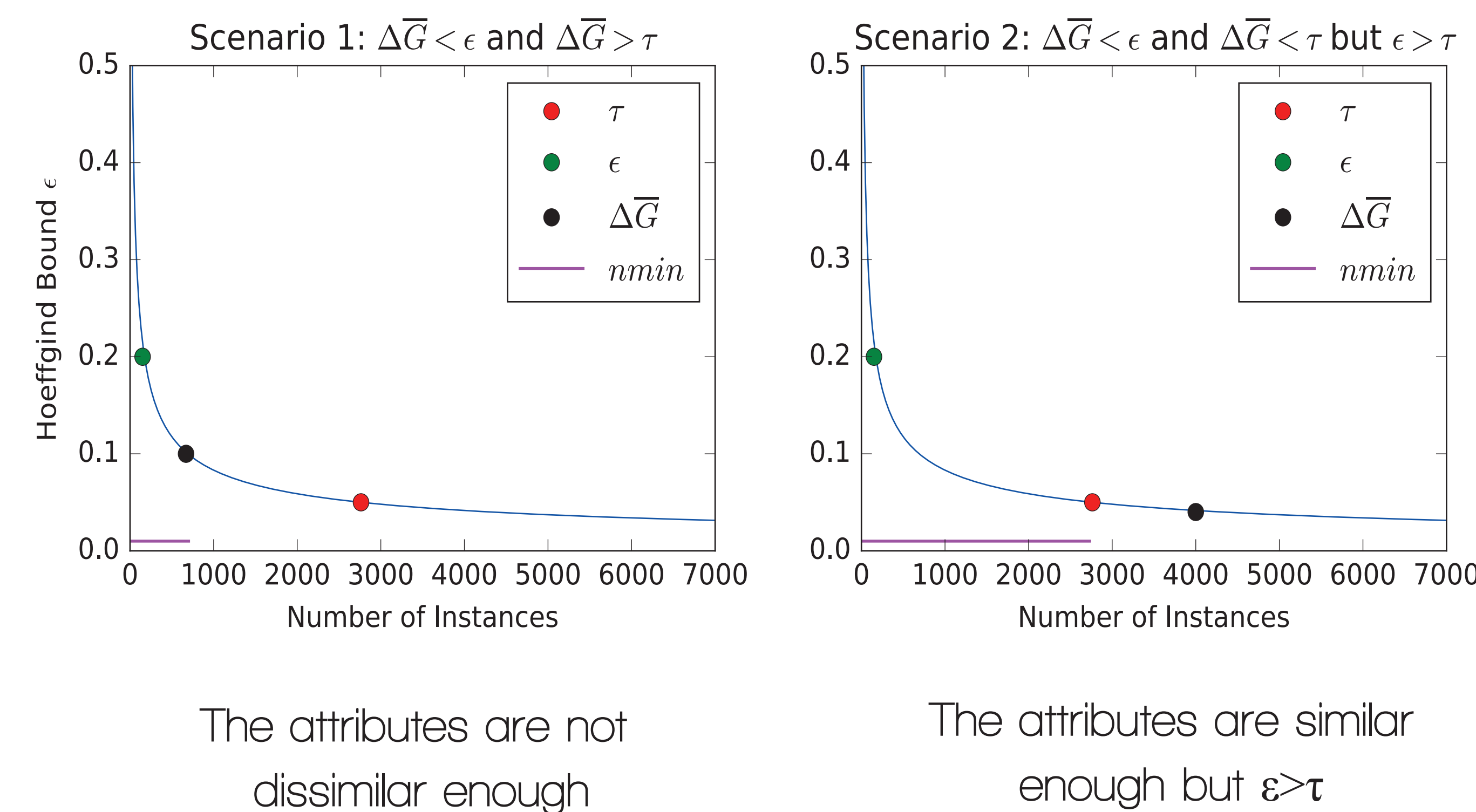
- VFDT builds a tree incrementally
- After  $nmin$  instances are observed at a node, the **best attributes** are obtained (information gain)

- If  $(\Delta G > \epsilon)$  or  $(\Delta G < \epsilon \text{ and } \epsilon < \tau) \rightarrow \text{Split}$
- If no split  $\rightarrow$  A lot of energy is wasted to calculate the best attributes

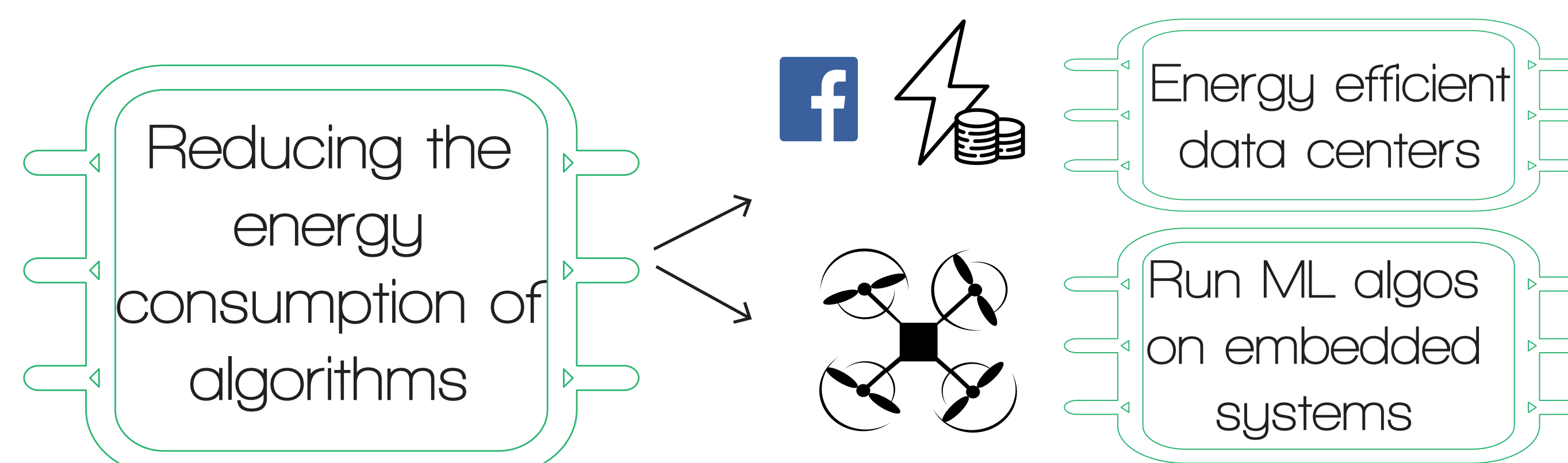
## AVFDT

**Method:** Adaptation of the  $nmin$  parameter for Hoefding Trees.

**How?:** Dynamically adapt the value of  $nmin$  depending on the incoming data. The goal is to set  $nmin$  to a value that ensures a split on the node.



## MOTIVATION

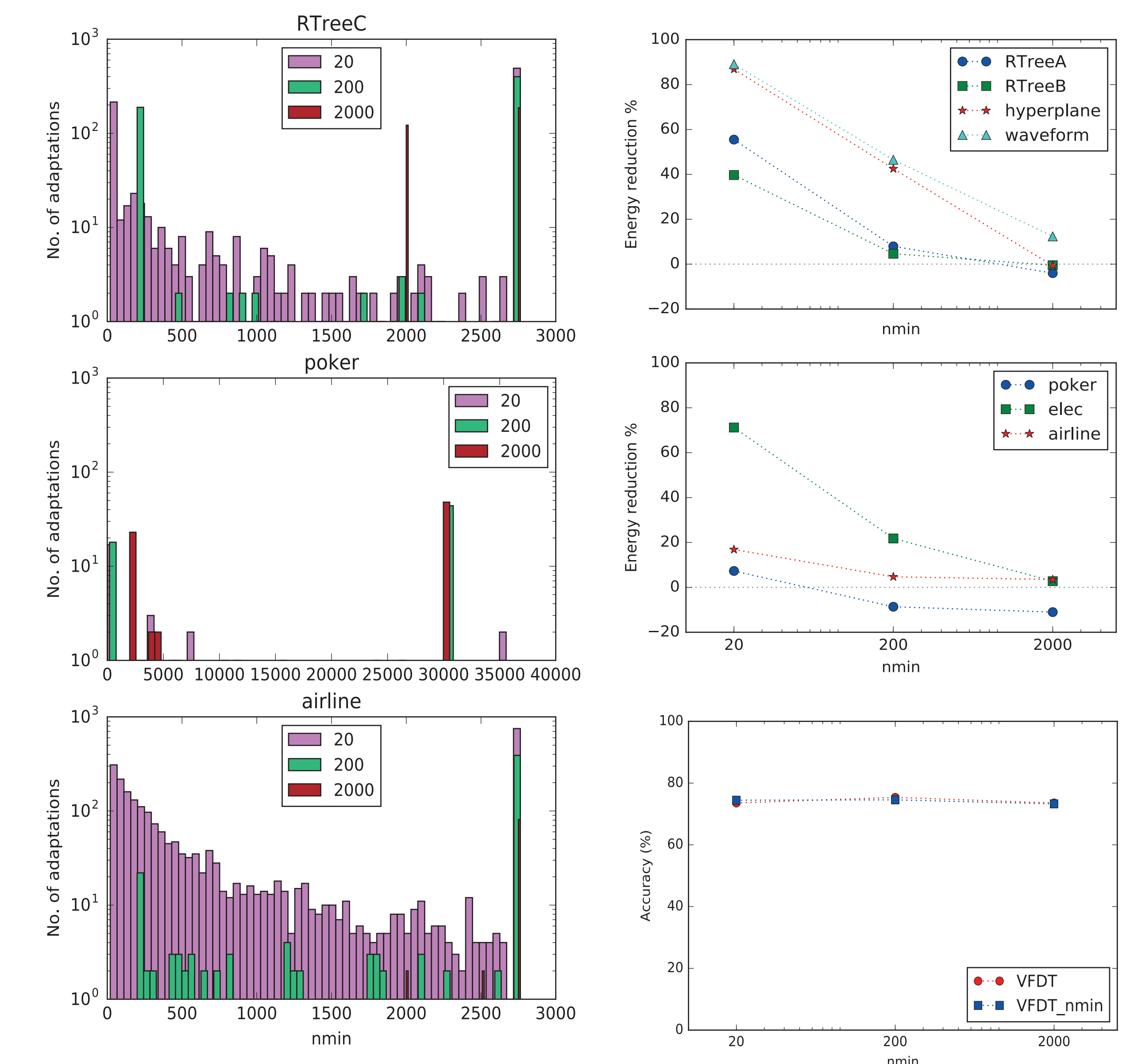


## EXPERIMENT

- Comparison: AVFDT vs VFDT
- 7 datasets: 4 artificial, 3 real
- 3 values of  $nmin$ : 20, 200, 2000

## RESULTS

- AVFDT consumes 23 % less energy in average
- On average, AVFDT achieves <1% less accuracy
- AVFDT obtains a maximum 89% energy decrease



## References

[1] P. Domingos and G. Hulten. "Mining high-speed data streams". In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2000, pp. 71–80.

[2] A. Shehabi, S. Smith, D. Sartor, R. Brown, M. Herrlin, J. Koomey, E. Masanet, N. Horner, I. Azevedo, and W. Lintner. United States data center energy usage report. Tech. rep. Lawrence Berkeley National Laboratory, Berkeley, California, 2016.

• Icons: The noun project. <https://thenounproject.com>