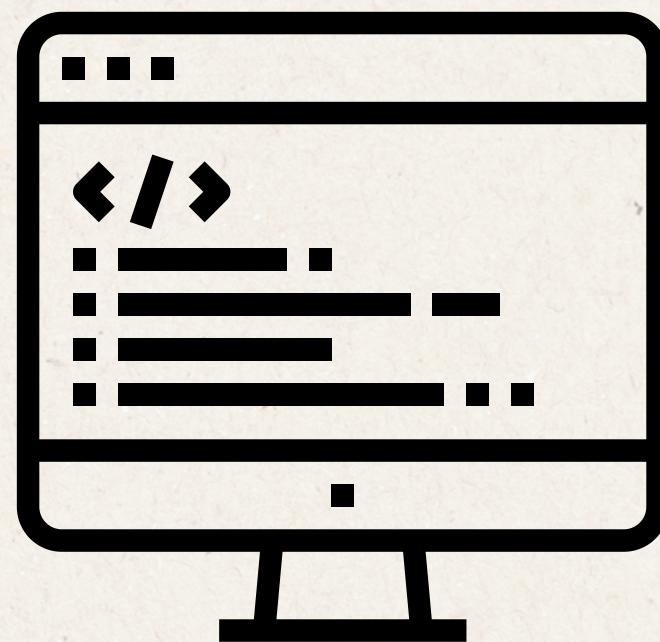




POLITECNICO  
MILANO 1863

01/09



# CAPYMOA VS RIVER

**Streaming Machine Learning**

**NAME OF PROJECT:**  
**SML Notebooks in CapyMOA**

**PRESENTED BY:**  
Sebastian Ballesteros

**PRESENTED TO:**  
Prof. Della Valle and Lorenzo Iovine

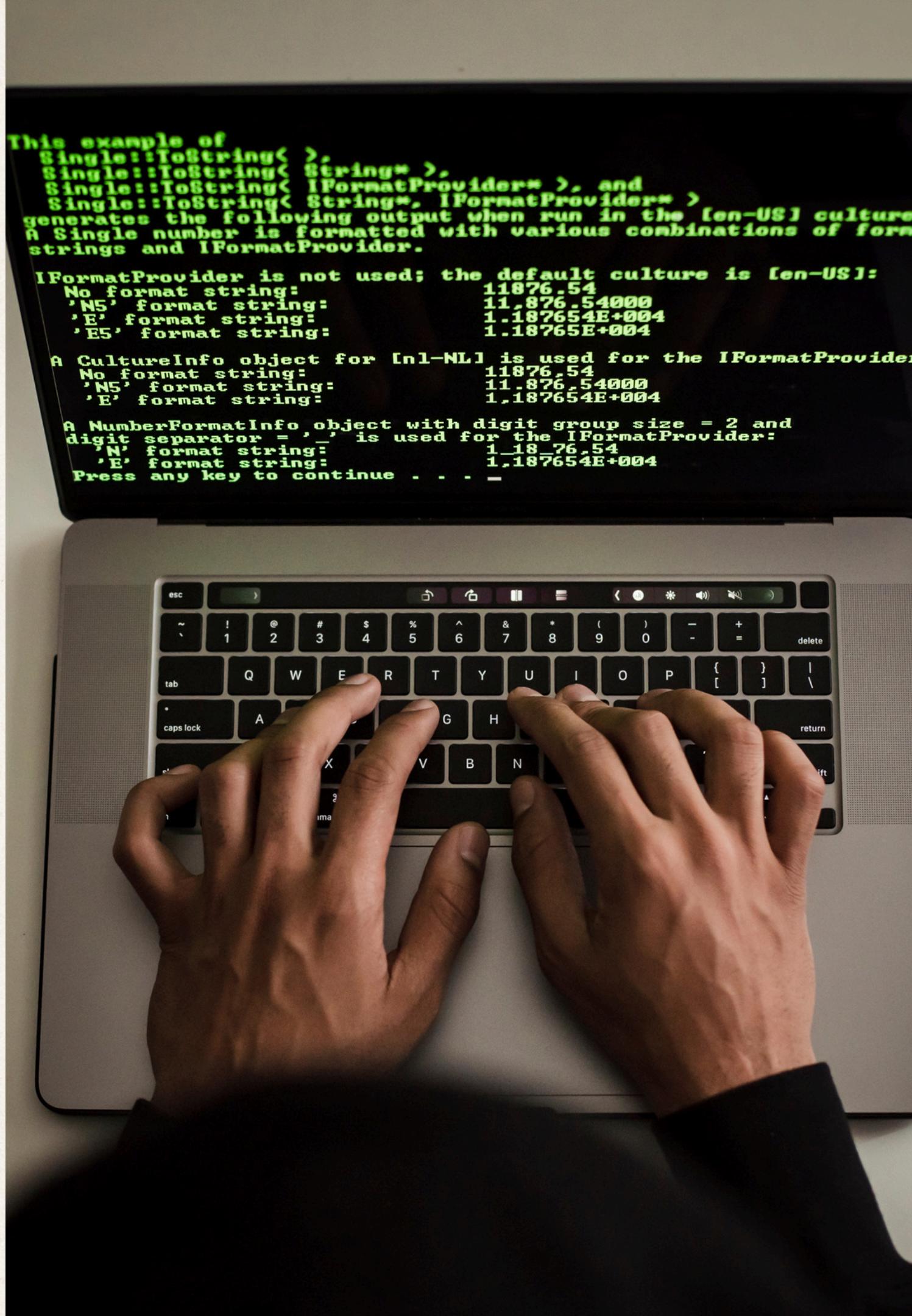
**02/09**

**01** Reproduce CapyMOA tutorial notebooks to understand how the library works.

**02** Compare the time performance of some parts of the CapyMOA code with the same features implemented in River.

**03** Analyze CPU/RAM and check whether time optimization in CapyMOA resulted in sacrifices in accuracy or computational efficiency.

**04** Identify any features present in River but absent in CapyMOA.

A photograph showing a person's hands typing on a laptop keyboard. The laptop screen is visible above the keyboard, displaying a command-line interface with text output related to number formatting. The text on the screen includes:

```
This example of
Single::ToString< >,
Single::ToString< String* >,
Single::ToString< IFormatProvider* >, and
Single::ToString< String*, IFormatProvider* >
generates the following output when run in the [en-US] culture.
A Single number is formatted with various combinations of form
strings and IFormatProvider.

IFormatProvider is not used; the default culture is [en-US]:
No format string: 11876.54
'N5' format string: 11.876.54000
'E' format string: 1.187654E+004
'E5' format string: 1.18765E+004

A CultureInfo object for [nl-NL] is used for the IFormatProvider:
No format string: 11876.54
'N5' format string: 11.876.54000
'E' format string: 1.187654E+004

A NumberFormatInfo object with digit group size = 2 and
digit separator = ',' is used for the IFormatProvider:
'N' format string: 1.18.76.54
'E' format string: 1.187654E+004

Press any key to continue . . .
```

# CapyMOA

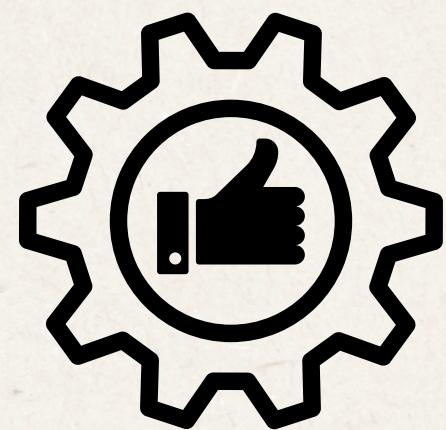
CapyMOA is a recent library for Streaming Machine Learning based on MOA

# Key findings



## Way faster

CapyMOA performed significantly faster than River in all tasks.



## Easy to use

The tutorials and the API are understandable and helpful.



## Still performant

There was not a compromise in performance due to optimization.

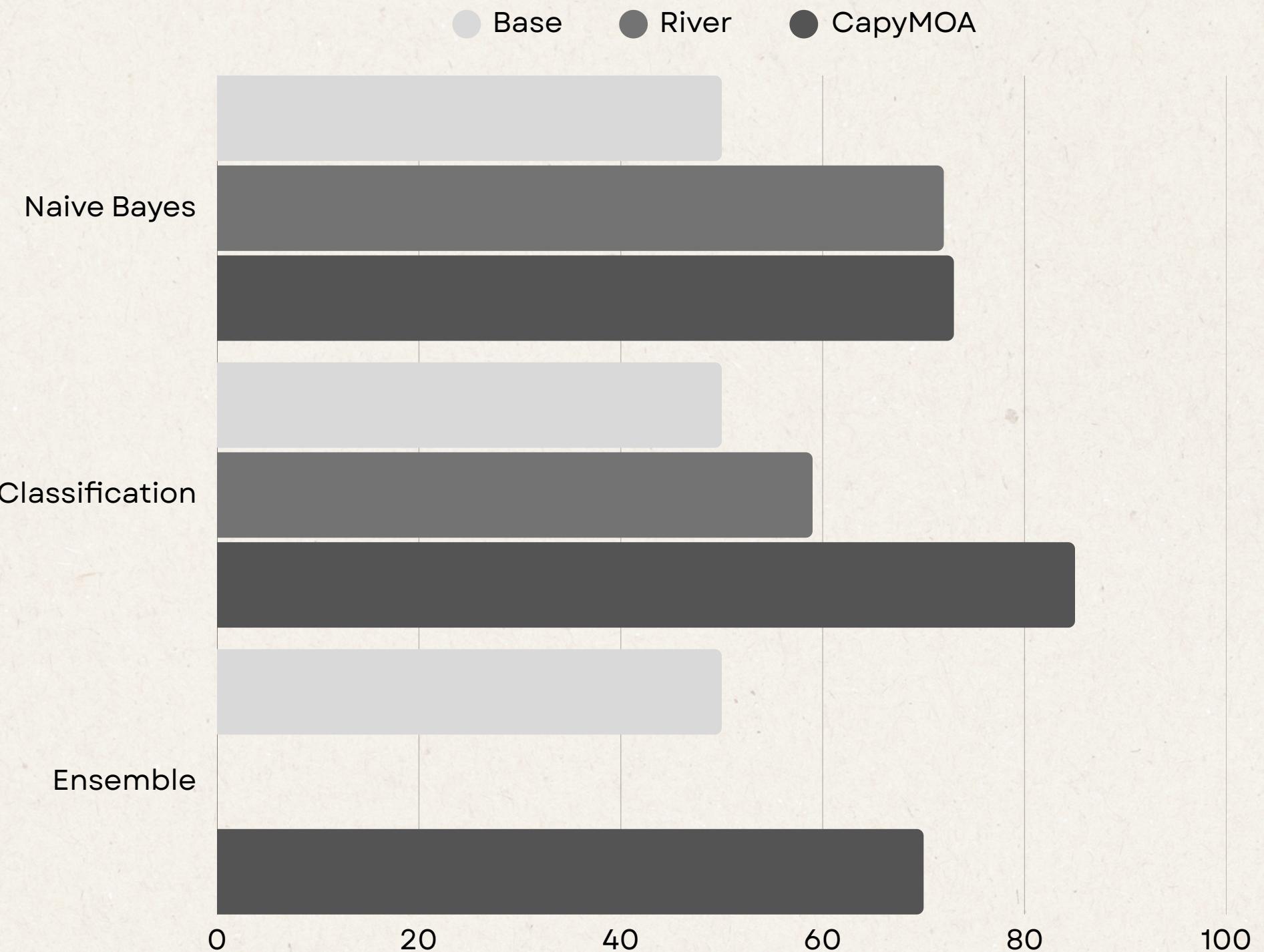
# Performance

## CapyMOA

CapyMOA performed better than River in all of the tasks (i.e. Classification and Regression). Ensemble methods were trained on a custom dataset to introduce drifts.

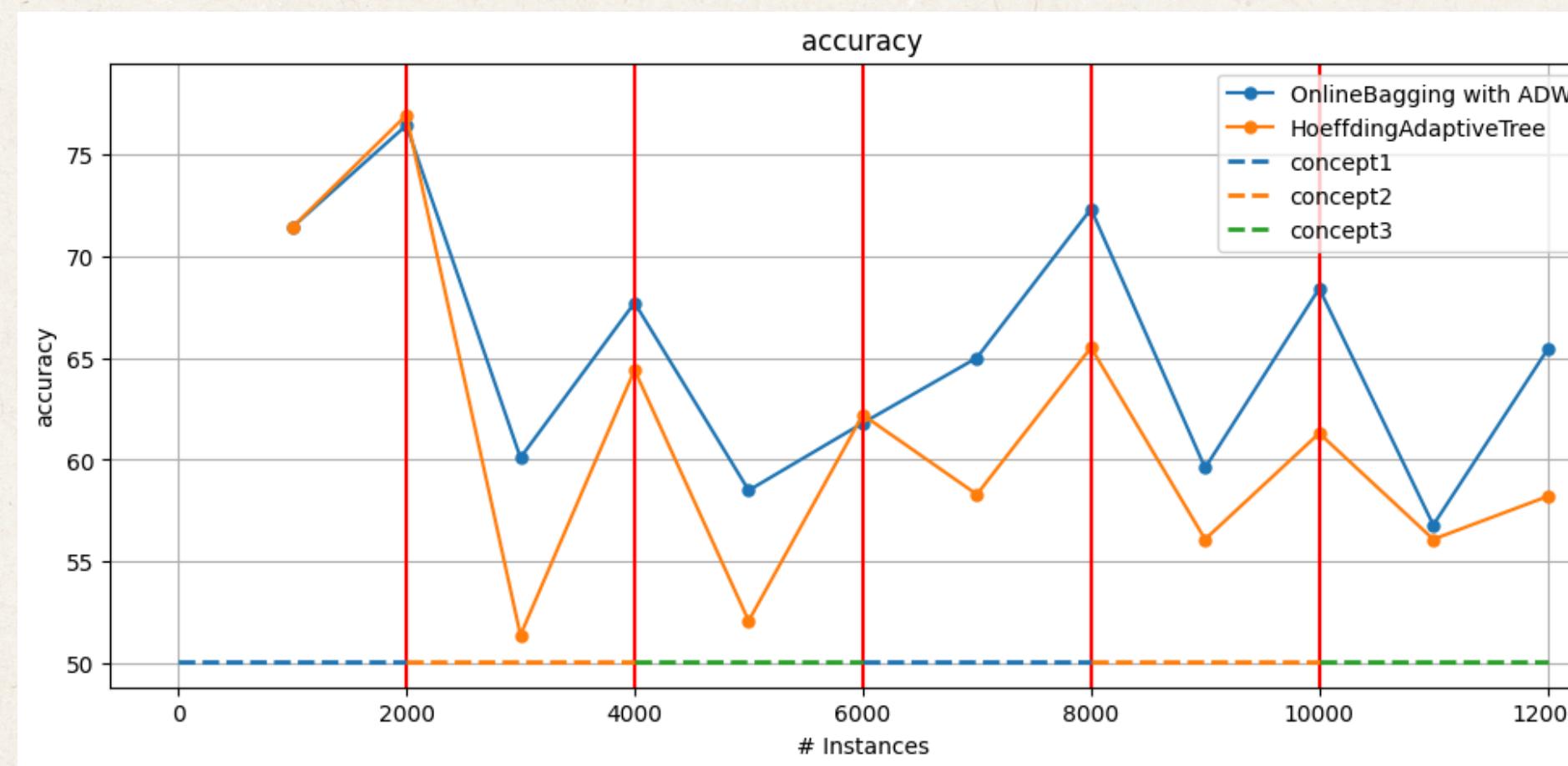
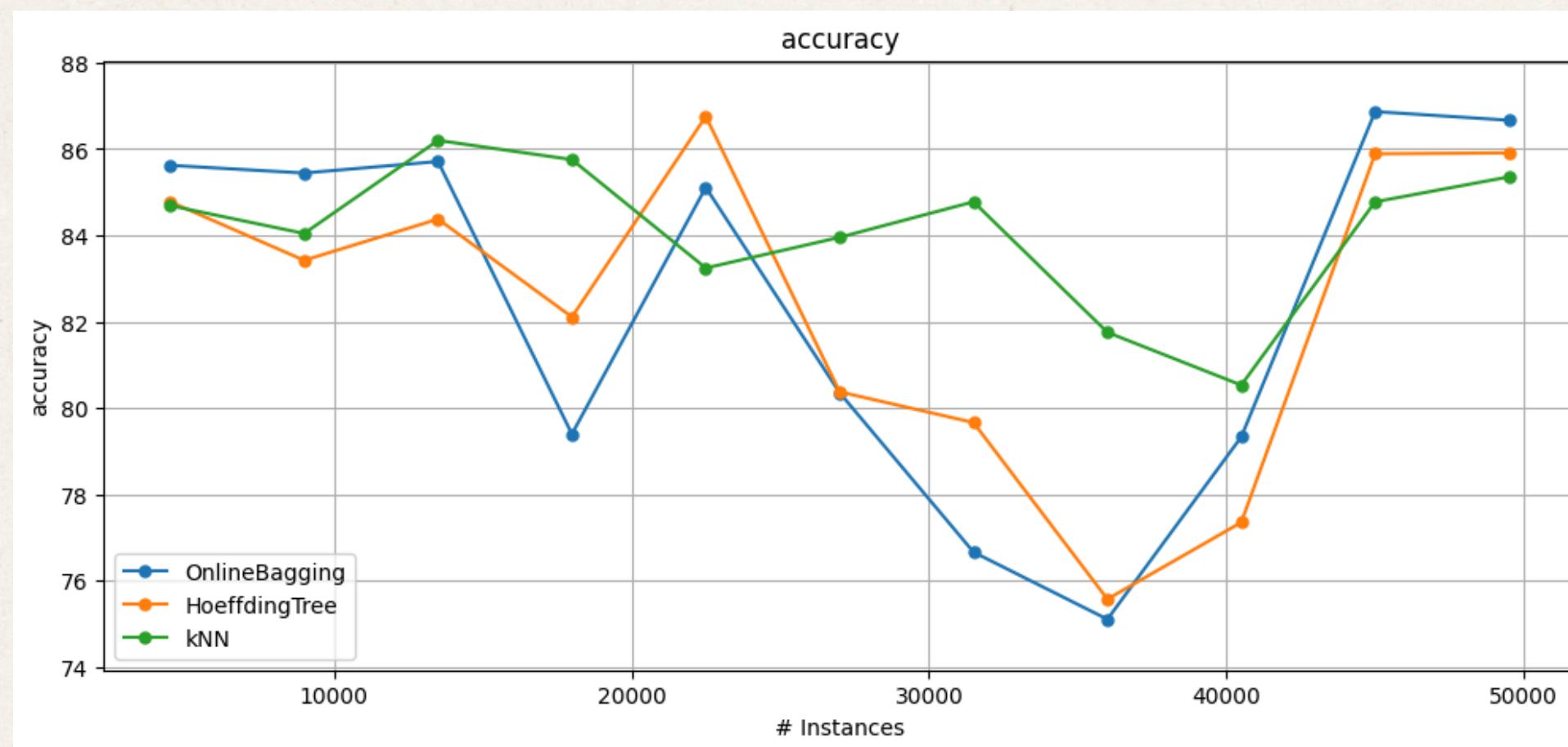
## River

For River, there is no data for ensemble methods since it took more than 3 hours to run.



## Classification

Online Bagging  
 Hoeffding Tree  
 Online KNN  
 Electricity Dataset

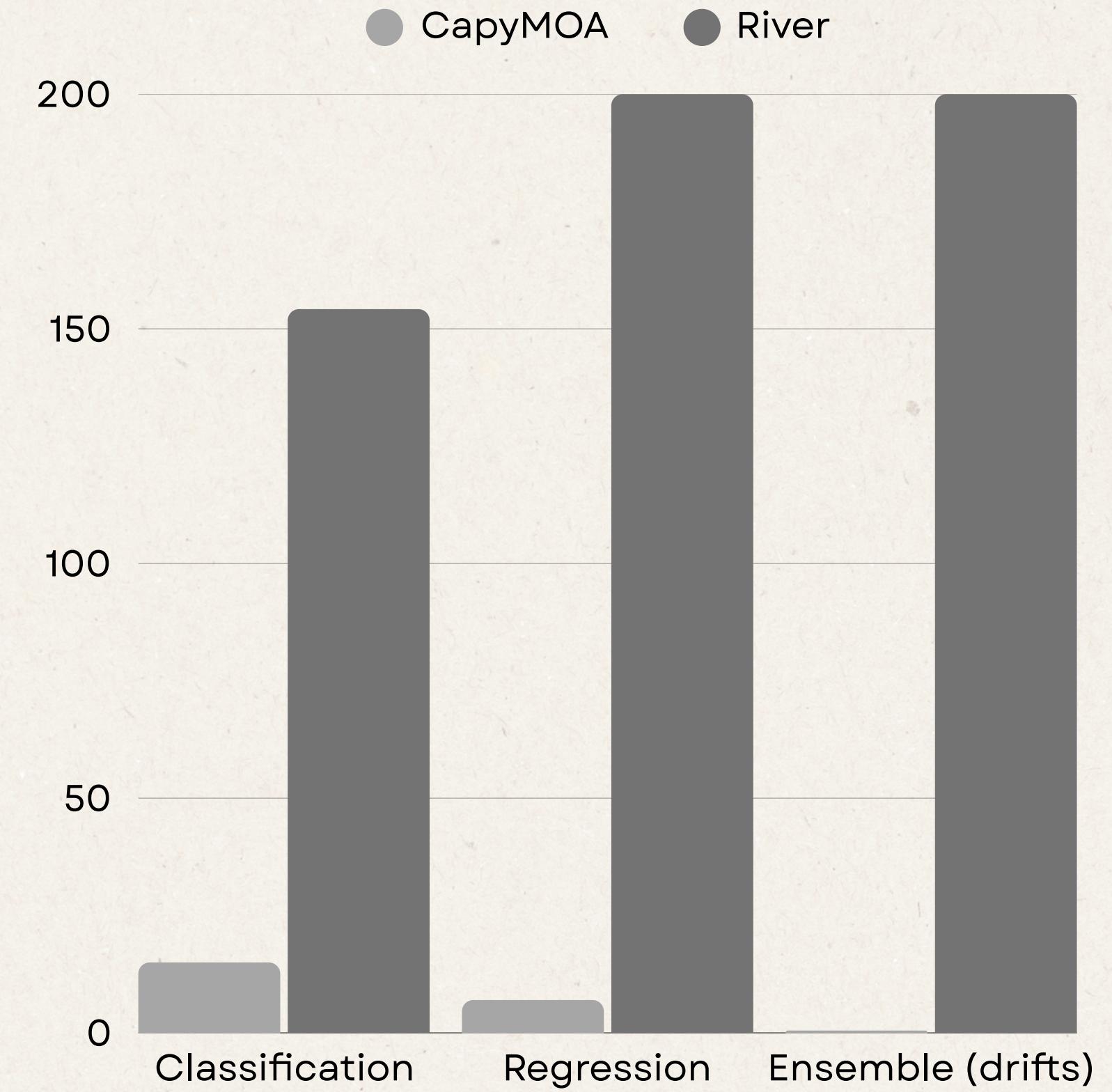
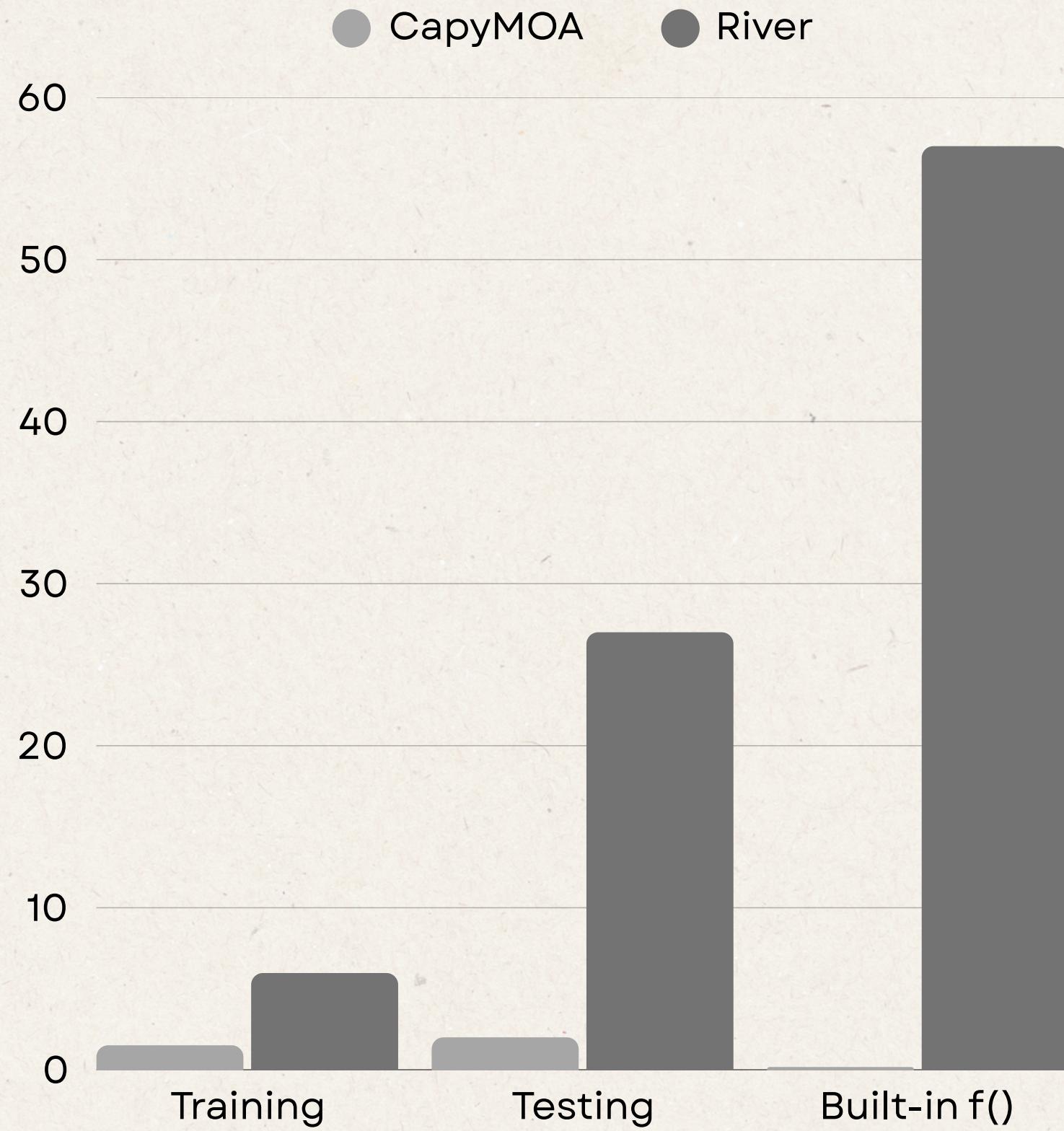


## Drift Detection with Ensemble models

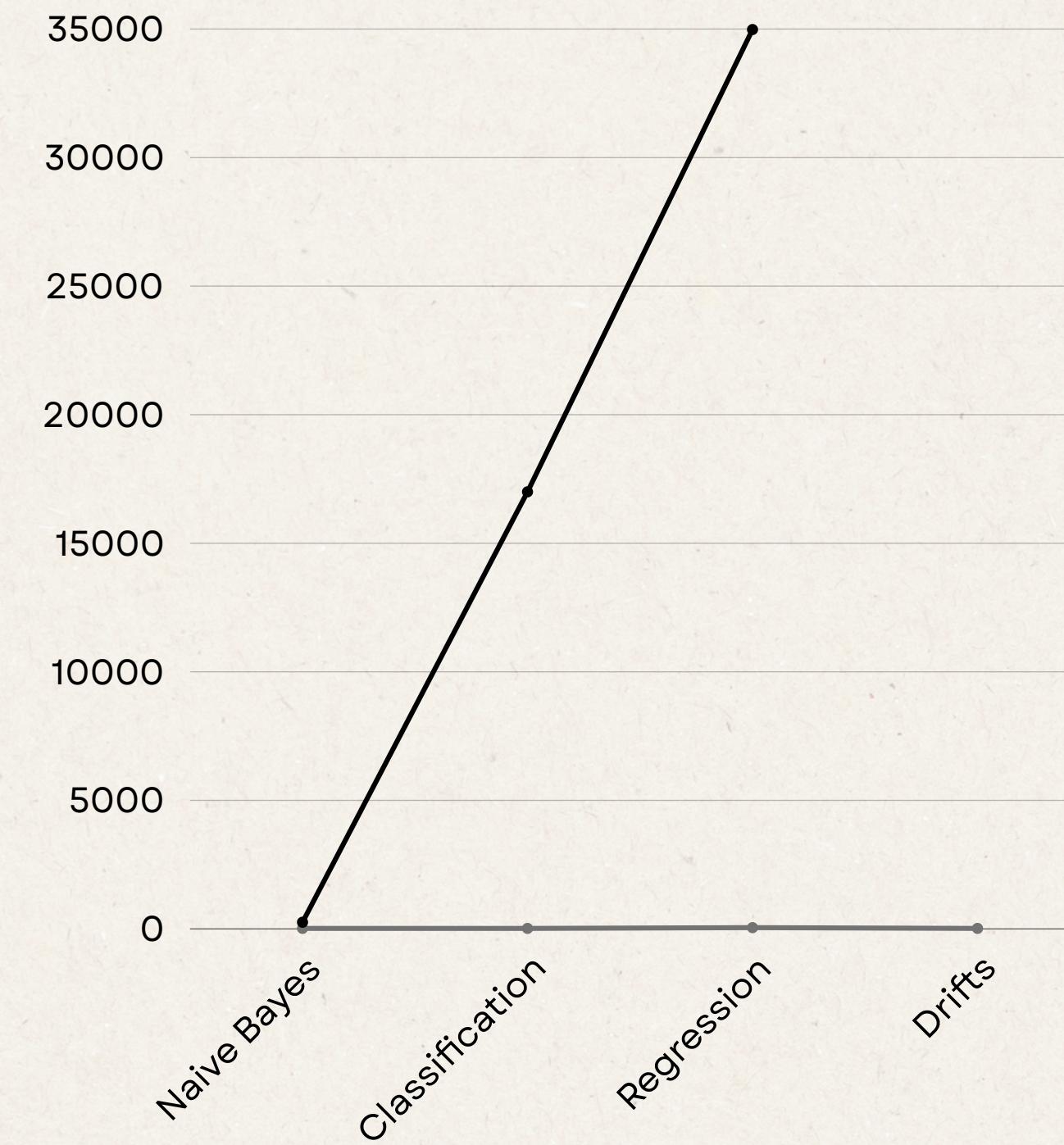
Online Bagging with ADWIN  
 Hoeffding Adaptive Tree  
 3 concept drifts  
 Custom dataset

# Time

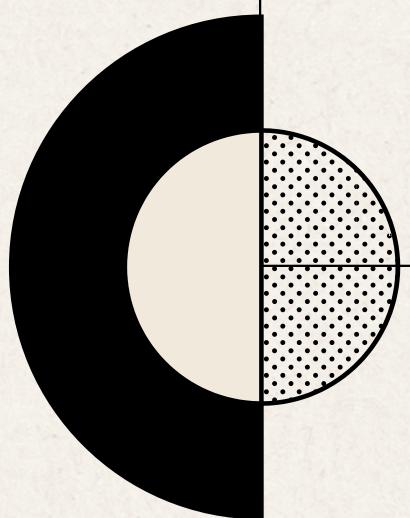
06/09



# CPU (%) and RAM (Kb)



# Feature Comparison



CapyMOA	River
Multiple model evaluation built in functionality	Composable pipelines
Continual Learning	Multi-class classification
Multivariate Drift Detectors	Time Series

# Thank you