COMPSCI 361 – Assignment 6
Data Stream Mining

This is worth 5% of your final grade. Due Date: June 11, 2020, 23:59 NZT

Problem

The goal will be to try out suitable/appropriate algorithms for data streams. You are given a data stream in the following files on Canvas.

Electricity is another widely used dataset described by M. Harries [1]. This data consist of data from the Australian New South Wales Electricity Market. The dataset contains 45, 312 instances. The class label identifies the change of the price relative to a moving average of the last 24 hours.

You may use either MOA, Scikit-Multiflow package or any other suitable packages. If you intend to use other packages please do clear it with me first.

Question 1: Discuss the accuracy of the Hoeffding Tree. Please try different parameters settings and use an evaluation technique (i.e., Prequential) for this data stream. **Discuss your findings, evaluation technique, and how your parameter settings affected your results.**

Question 2: Compare the accuracy in Question 1 to using the Hoeffding Adaptive Tree classifier. **Discuss your findings, evaluation technique, and your how your parameter settings affected your results.**

What to submit?

You will need to provide short answers (approx. 1 – 3 sentences) the questions above. Please keep your report to 1 page maximum. Please clearly indicate your answers in your report. Clarity of the report is important. Please ensure you provide sufficient details to reproduce your findings. The final report has to be deposited to Canvas. Please name your report file **"Your_UPI.pdf"**.

Grading rubric

- Reproduciblity – 0.5 mark

- Appropriate Evaluation Technique – 0.5 mark
- Clarity of the discussion for Question 1 – 2 mark
- Clarity of the discussion for Question 2 – 2 marks

References

[1] M. Harries. Splice-2 comparative evaluation: Electricity pricing. Technical report, The University of South Wales, 1999.