



Concept Drift Detection in Clickstream Classifiers Using Lightweight Statistical Sketches

Andre Andersen^{1*}, Mai Nikolic², Arjun Assefa³

¹Department of Computer Science, Danube Center for Computer Science, Greece

²Department of Computer Science, Patagonia Research Institute of Computer Science, South Africa

³Department of Computer Science, Elbe Research Institute of Computer Science, Guyana

Article Info

Volume 1, Issue 1, 2026

doi: [10.62715/ijdwdm.2026.v1.i1.pp27-30](https://doi.org/10.62715/ijdwdm.2026.v1.i1.pp27-30)

Received: 30 March 2026

Accepted: 17 May 2026

Published: 09 July 2026

Abstract

Classifiers that score clickstream sessions for purchase intent, churn or bot activity degrade when the behaviour they were trained on changes, and the change is hard to observe directly because true labels arrive late or never. Detectors that track the error rate therefore react slowly, while detectors that compare full feature distributions are costly at clickstream volumes and awkward to run on partitioned streams. This communication argues that the probabilistic sketches already used for stream analytics are a practical middle course. Quantile sketches summarise numeric features and classifier scores, frequency sketches summarise categorical features such as page or referrer, and cardinality sketches track how many distinct values are in use; all occupy bounded memory, are updated in constant time per event and can be merged across partitions. The note describes how each kind of sketch yields a drift statistic by comparison with a reference window, sets out a monitoring loop that separates cheap unlabelled warnings from label-based confirmation, and discusses what the approach cannot see, in particular a change in the relation between features and label that leaves the feature distributions untouched. No experimental results are presented; the aim is to state the design and its limits clearly enough to be tested.

Keywords: *Concept drift, Clickstream, Data stream sketches, Quantile sketch, Count-min sketch, Model monitoring*

© 2026 Andre Andersen et al. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

A clickstream classifier assigns a label to a session or an event as it happens: likely to buy, likely to abandon, human or automated. The data are non-stationary by nature. Site redesigns move links, campaigns bring new audiences, seasons alter what people look for, and crawlers change their signatures. [Widmer and Kubat \(1996\)](#) described the resulting loss of validity of a learned model as concept drift, and the literature since has produced many methods for detecting and adapting to it ([Gama et al., 2014](#); [Lu et al., 2019](#)).

Clickstream settings place two particular demands on a detector. Labels are delayed, since a purchase

* Corresponding author: Andre Andersen, Department of Computer Science, Danube Center for Computer Science, Greece. E-mail: andre.andersen@danube.edu

© 2026 Andre Andersen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

may follow a session by days and a bot may never be confirmed, and event rates are high, with traffic split across many stream partitions. This note considers how far compact statistical sketches, which stream processors already maintain for analytics, can serve as the detector's input.

2. Why Standard Detectors Fit Poorly

The classical detectors monitor the model's errors. The drift detection method of [Gama et al. \(2004\)](#) tracks the running error rate and signals when it rises significantly above its recorded minimum. ADWIN keeps a window of recent values and shrinks it whenever two sub-windows differ in mean by more than a bound derived from the Hoeffding inequality ([Bifet and Gavalda, 2007](#)). The cumulative sum test of [Page \(1954\)](#) remains a common choice for the same purpose. All three need a stream of correct-or-incorrect outcomes, and in a clickstream that stream runs days behind the traffic it describes.

The alternative is to watch the inputs. [Kifer et al. \(2004\)](#) framed change detection as a two-sample test between a reference window and a current window of unlabelled data. Applied literally, this means retaining raw windows for every feature, which is expensive at high volume, and it sits uneasily with partitioned processing, where no single worker sees the whole window. What is wanted is a summary that is small, updatable one event at a time and combinable across workers.

3. Sketches as Drift Signals

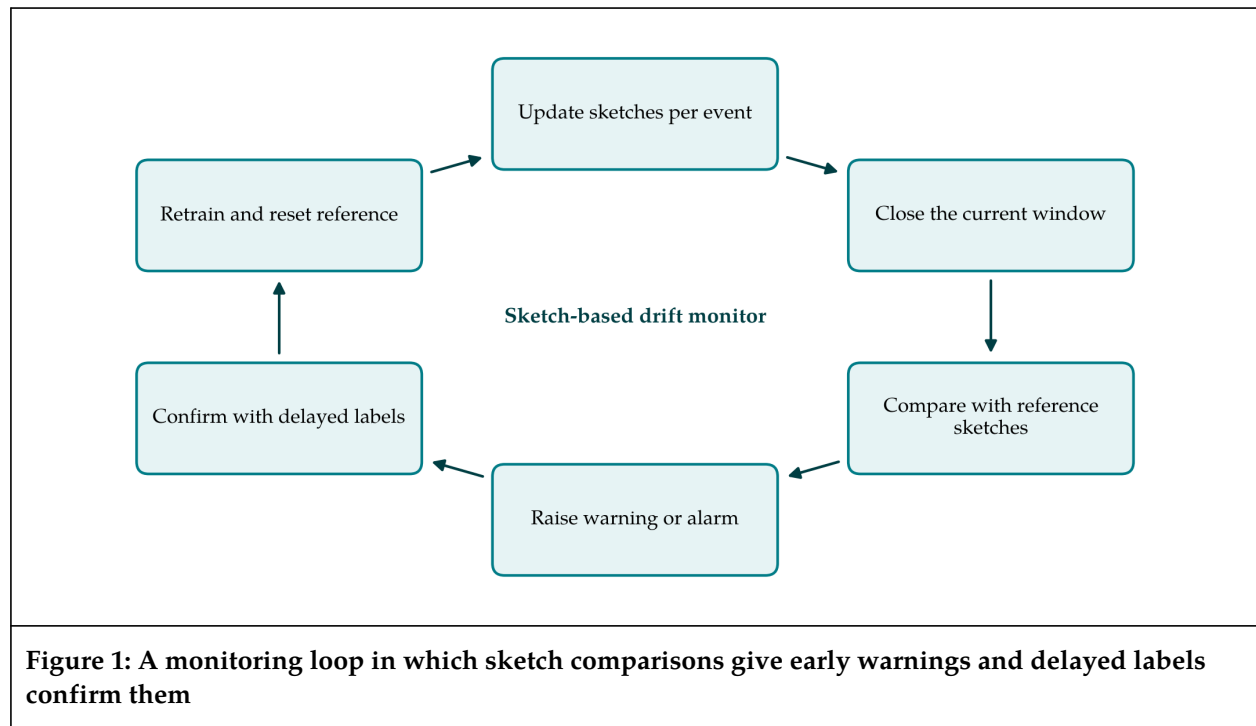
A sketch is a compact, approximate summary of a stream with a stated error guarantee. The property that matters most here is mergeability: summaries built separately over two parts of a stream can be combined into a summary of the whole with no loss of guarantee ([Agarwal et al., 2013](#)). A reference sketch can then be assembled from many partitions and many past intervals, and a current sketch from the most recent ones. Table 1 lists the three families of sketch that suit clickstream features and the drift statistic each supports.

Table 1: Sketch families suited to clickstream features and the drift statistic each supports			
Sketch family	Summarises	Drift statistic	Caveat
Quantile sketch (Greenwald and Khanna, 2001 ; Karnin et al., 2016)	Numeric features; classifier scores	Largest gap between reference and current distribution functions	Sketch error must be well below the alarm threshold
Count-min sketch (Cormode and Muthukrishnan, 2005)	Frequencies of pages, referrers, agents	Divergence between estimated frequency vectors of popular values	Overestimates rare values; weak in the tail
HyperLogLog (Flajolet et al., 2007)	Number of distinct values	Relative change in distinct count per window	Sees new values, not shifts among old ones
Counters and moments	Event rate; mean and variance	Standardised difference of means	Blind to changes of shape
Per-class score sketches	Scores, split by predicted label	Shift in predicted class proportions	Reflects the model as well as the data

For a numeric feature, such as dwell time or pages per session, a quantile sketch answers rank queries within a small additive error. Evaluating the reference and current sketches on a common grid of values gives two approximate distribution functions, and their largest difference is an estimate of the Kolmogorov-Smirnov statistic. The same construction applied to the classifier's output score is especially useful, since the score compresses all features into the one direction the model cares about. For a categorical feature, a count-min sketch returns frequency estimates that never fall below the truth and exceed it by a bounded amount, which is adequate for the popular values that dominate a divergence measure. A cardinality sketch adds a different signal: a sudden rise in the number of distinct user agents or landing pages often marks a crawler or a campaign before any distribution has visibly moved.

4. A Monitoring Loop

These statistics can be organised into a loop of six steps, shown in Figure 1. Each stream worker updates its sketches as events pass. At the end of a tumbling window the workers' sketches are merged and compared with a reference built from a period in which the model was known to perform acceptably. A statistic that crosses a lower threshold raises a warning; one that crosses a higher threshold, or stays at warning level for several consecutive windows, raises an alarm. The two-level scheme follows the practice of the error-based detectors.



An alarm from unlabelled data is a prompt and not a verdict. Feature distributions shift for reasons that leave accuracy intact, and retraining at every alarm would waste effort and could destabilise the model. The loop therefore includes confirmation: when labels for the flagged period mature, an error-based test such as ADWIN is run over them, and only a confirmed alarm triggers retraining and replacement of the reference sketches. Unconfirmed alarms are logged and used to adjust the thresholds. Because many features are tested in every window, some correction for multiple comparisons is needed, and thresholds should account for the sketches' own approximation error so that it is not mistaken for change.

5. Limits of the Approach

Three limits should be stated plainly. Sketches of individual features capture marginal distributions. A change in the dependence between features, with each marginal unchanged, passes unseen unless the classifier score happens to register it. More fundamentally, monitoring inputs detects a change in the distribution of the features and cannot detect a change in the conditional distribution of the label given the features, which the survey literature calls real concept drift (Gama et al., 2014). If customers with the same browsing pattern begin to buy less, only labels will show it. Sketch-based monitoring shortens the time to notice the first kind of change and does nothing for the second.

The third limit is practical. Window length, reference period and thresholds interact, and clickstreams have strong daily and weekly cycles that a naive comparison will report as drift every morning. Comparing a window with the same hour of earlier weeks is the obvious remedy, at the cost of keeping more reference sketches. How well any particular configuration separates harmful change from routine variation is an empirical question that this note does not answer. It should be examined on labelled clickstream data with changes of known timing before the approach is relied on.

6. Conclusion

Quantile, frequency and cardinality sketches give clickstream systems an inexpensive, mergeable view of how inputs and model scores are distributed, and comparing current sketches with a reference yields drift statistics long before labels arrive. Used as an early-warning layer in front of label-based confirmation, they fit the constraints of high-volume, partitioned streams. They do not replace labelled evaluation, since the form of drift that matters most to a classifier is the one unlabelled data cannot reveal.

Conflict of interest

To be declared by the author(s) before publication.

Funding

To be declared by the author(s) before publication.

References

- Agarwal, P.K., Cormode, G., Huang, Z., Phillips, J.M., Wei, Z. and Yi, K. (2013). [Mergeable summaries](#). *ACM Transactions on Database Systems*, 38(4): Article 26. <https://doi.org/10.1145/2500128>
- Bifet, A. and Gavaldà, R. (2007). [Learning from time-changing data with adaptive windowing](#). *Proceedings of the 2007 SIAM International Conference on Data Mining*, 443-448. <https://doi.org/10.1137/1.9781611972771.42>
- Cormode, G. and Muthukrishnan, S. (2005). [An improved data stream summary: The count-min sketch and its applications](#). *Journal of Algorithms*, 55(1): 58-75. <https://doi.org/10.1016/j.jalgor.2003.12.001>
- Flajolet, P., Fusy, É., Gandouet, O. and Meunier, F. (2007). [HyperLogLog: The analysis of a near-optimal cardinality estimation algorithm](#). *Proceedings of the 2007 Conference on Analysis of Algorithms (AofA 07), Discrete Mathematics and Theoretical Computer Science Proceedings*, 137-156.
- Gama, J., Medas, P., Castillo, G. and Rodrigues, P. (2004). [Learning with drift detection](#). *Advances in Artificial Intelligence (SBIA 2004), Lecture Notes in Computer Science 3171*, 286-295. https://doi.org/10.1007/978-3-540-28645-5_29
- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A. (2014). [A survey on concept drift adaptation](#). *ACM Computing Surveys*, 46(4): Article 44. <https://doi.org/10.1145/2523813>
- Greenwald, M. and Khanna, S. (2001). [Space-efficient online computation of quantile summaries](#). *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, 58-66. <https://doi.org/10.1145/375663.375670>
- Karnin, Z., Lang, K. and Liberty, E. (2016). [Optimal quantile approximation in streams](#). *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science*, 71-78.
- Kifer, D., Ben-David, S. and Gehrke, J. (2004). [Detecting change in data streams](#). *Proceedings of the 30th International Conference on Very Large Data Bases*, 180-191.
- Lu, J., Liu, A., Dong, F., Gu, F., Gama, J. and Zhang, G. (2019). [Learning under concept drift: A review](#). *IEEE Transactions on Knowledge and Data Engineering*, 31(12): 2346-2363. <https://doi.org/10.1109/TKDE.2018.2876857>
- Page, E.S. (1954). [Continuous inspection schemes](#). *Biometrika*, 41(1/2): 100-115. <https://doi.org/10.1093/biomet/41.1-2.100>
- Widmer, G. and Kubat, M. (1996). [Learning in the presence of concept drift and hidden contexts](#). *Machine Learning*, 23(1): 69-101. <https://doi.org/10.1007/BF00116900>

Cite this article as: Andersen A., Nikolic M. and Assefa A. (2026). [Concept Drift Detection in Clickstream Classifiers Using Lightweight Statistical Sketches](#). *International Journal of Data Warehousing and Data Mining*. 1(1), 27-30. doi: 10.62715/ijdwdm.2026.v1.i1.pp27-30