

datadriftR: An R package for data drift detection

Ugur Dar ¹ and Mustafa Cavus ¹

¹ Eskisehir Technical University, Department of Statistics, Eskisehir, Turkey

DOI: [10.21105/joss.09481](https://doi.org/10.21105/joss.09481)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Chris Vernon](#)  

Reviewers:

- [@fangzhou-xie](#)
- [@benyamindsmith](#)

Submitted: 17 November 2025

Published: 30 March 2026

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Deployed machine learning models frequently encounter degradation in predictive accuracy when the statistical properties of incoming data evolve over time, a condition known as data drift. This phenomenon can manifest in several forms, most notably concept drift, which occurs when the functional relationship linking predictor variables to the outcome changes, thereby undermining model reliability. Conventional drift detection strategies often rely on aggregate performance indicators or univariate distributional summaries, approaches that may overlook nuanced yet consequential shifts in the data-generating mechanism. Within the broader machine learning operations (MLOps) framework, continuous model monitoring has emerged as a critical practice for safeguarding the stability and dependability of production systems ([Biecek, 2019](#); [Mougan & Nielsen, 2023](#)). `datadriftR` is an open-source R package designed to address these challenges by providing real-time detection of data drift in univariate streaming data. The package implements a comprehensive suite of widely recognized statistical methods for monitoring distributional changes, including error-rate-based detectors that track classification performance (DDM ([Gama et al., 2004](#)), EDDM ([Baena-García et al., 2006](#))), Hoeffding-bound methods that employ adaptive windowing to detect mean shifts (HDDM-A and HDDM-W ([Frías-Blanco et al., 2015](#))), adaptive windowing for change detection (ADWIN ([Bifet & Gavaldá, 2007](#))), a sliding-window Kolmogorov–Smirnov test for distribution comparison (KSWIN ([Raab et al., 2020](#))), the cumulative-sum-based Page–Hinkley test for detecting persistent shifts ([Page, 1954](#)), histogram-based Kullback–Leibler divergence monitoring for measuring distributional divergence ([Kullback & Leibler, 1951](#)), and a functional profile comparison method for analyzing temporal patterns ([Kobylińska & others, 2023](#)).

Statement of need

Data drift detection is a fundamental challenge in deployed machine learning systems and adaptive analytics ([Kobylińska & others, 2023](#)). When the underlying data-generating process changes over time, model performance can deteriorate silently, leading to incorrect predictions and suboptimal decision-making. Early detection of such shifts enables timely interventions—such as model retraining, recalibration or triggering alerts—thereby maintaining system reliability in production environments.

The R ecosystem lacks a dedicated package for streaming drift detection despite widespread availability in Java (MOA ([Bifet et al., 2010](#))) and Python (scikit-multiflow ([Montiel et al., 2018](#))). While individual R packages address specific aspects of change-point detection or distribution testing, no existing toolkit consolidates canonical online detectors—DDM ([Gama et al., 2004](#)), EDDM ([Baena-García et al., 2006](#)), HDDM-A and HDDM-W ([Frías-Blanco et al., 2015](#)), ADWIN ([Bifet & Gavaldá, 2007](#)), KSWIN ([Raab et al., 2020](#)), Page–Hinkley ([Page, 1954](#)), and KL divergence ([Kullback & Leibler, 1951](#))—under a unified framework for incremental analysis.

`datadriftR` brings these methods together in a single R package, providing a consistent interface for updating detectors observation by observation. For convenience, the package also includes

a high-level wrapper (`'detect_drift()'`) that processes whole vectors and returns detected drift points. This makes it straightforward to incorporate drift monitoring into streaming workflows and to compare alternative detectors on the same data.

Examples of Use

To illustrate the package's unified interface, we generate a synthetic binary stream with an abrupt distributional shift at index 501 and demonstrate minimal usage for representative detectors. All examples process the same stream to enable direct comparison.

```
library(datadriftR)

set.seed(123)
# Generate pre-drift and post-drift segments
pre <- sample(c(0,1), 500, replace = TRUE, prob = c(0.7, 0.3))
post <- sample(c(0,1), 500, replace = TRUE, prob = c(0.3, 0.7))
stream <- c(pre, post)

# 1) DDM (Drift Detection Method)
ddm <- DDM$new()
for (i in seq_along(stream)) {
  ddm$add_element(stream[i])
  if (ddm$change_detected) {
    message("DDM drift detected at index ", i)
    break
  }
}

# 2) Page Hinkley
set.seed(1)
x <- c(rnorm(300, 0, 1), rnorm(200, 3, 1))
detect_drift(x, method = "page_hinkley", delta = 0.05, threshold = 50)
```

The package also includes HDDM-A, HDDM-W, ADWIN, KL-divergence, ADWIN, and ProfileDifference detectors. Each follows the same instantiate–update–check pattern. For complete examples, method comparisons, and detailed usage, see the online documentation and vignettes at <https://ugurdar.github.io/datadriftR/>, the package vignette (`vignette("datadriftR-intro")`), the README, and the individual detector documentation pages on CRAN.

References

- Baena-García, M., Campo-Ávila, J. del, Fidalgo, J., Bifet, A., Gavaldá, R., & Morales-Bueno, R. (2006). Early drift detection method. *ECML PKDD 2006 Workshop on Knowledge Discovery from Data Streams*.
- Biecek, P. (2019). *Model development process*. <https://arxiv.org/abs/1907.04461>
- Bifet, A., & Gavaldá, 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>
- Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: Massive online analysis. *Journal of Machine Learning Research*, 11, 1601–1604. <https://www.jmlr.org/papers/volume11/bifet10a/bifet10a.pdf>

- Frías-Blanco, I., Campo-Ávila, J. del, Ramos-Jiménez, G., Morales-Bueno, R., Ortiz-Díaz, E., & Caballero-Mota, Y. (2015). Online and nonparametric drift detection methods based on hoeffding's bounds. *IEEE Transactions on Knowledge and Data Engineering*, 27(3), 810–823. <https://doi.org/10.1109/TKDE.2014.2345382>
- Gama, J., Medas, P., Castillo, G., & Rodrigues, P. (2004). Learning with drift detection. *Advances in Artificial Intelligence – SBIA 2004*.
- Kobylińska, E., & others. (2023). *A survey on concept drift adaptation*. <https://arxiv.org/abs/2308.11446>
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <https://doi.org/10.1214/aoms/1177729694>
- Montiel, J., Read, J., Bifet, A., & Abdesslem, T. (2018). Scikit-multiflow: A multi-output streaming framework. *Journal of Machine Learning Research*, 19(72), 1–5. <https://jmlr.org/papers/v19/18-251.html>
- Mougan, C., & Nielsen, D. S. (2023). Monitoring model deterioration with explainable uncertainty estimation via non-parametric bootstrap. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37, 15037–15045. <https://doi.org/10.1609/aaai.v37i12.26758>
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1-2), 100–115. <https://doi.org/10.1093/biomet/41.1-2.100>
- Raab, M., Heusinger, M., & Schleif, F.-M. (2020). REDC: Regularized drift detection for MOA. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2019.11.111>