

Leakage in Data Mining: Formulation, Detection, and Avoidance (optional)

If you want an example in more depth of how data scientists are exploring ways to detect and avoid data leakage, this technical article proposes one approach: a two-stage process based on "legitimacy tags".

If you're just interested in getting a little more background on the problem along with interesting examples, Sections 1 and 2 (Introduction and Related Work) are also useful to read on their own.

Kaufman, S., Rosset, S., & Perlich, C. (2011). [Leakage in data mining](#)<sup>↗</sup>. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*. doi:10.1145/2020408.2020496