

Assessing Data Quality

One of the biggest difficulties with ML is dealing with messy data. This is a common and reoccurring problem.

CleanLab

[CleanLab](#) is a product that attempts to use statistical methods to clean up data and labels. I need to read more about exactly how it works.

They have some tutorials on how to use their system to clean up text for processing [here](#)

Revision #2

Created 26 October 2022 07:46:44 by James

Updated 21 January 2024 14:49:27 by James