Cleanlab Studio Testimonial and Feedback written by Steven Gawthorpe.

1. Have you used Cleanlab StudioStudio; if not, why not? What did you like/not like?

I've used Cleanlab Studio and loved the ease and expedient delivery of results. It helps when I need a quick no-code solution to check label quality. Since I don't have the support of a large data science team, it helps offset a lot of QC labor.

I also loved the automatic pre-formatting of data, addressing null values, automatic detection of label and UID columns as well as data formatting.

I also used the CLI in addition to the drag and drop. Again, super easy to use. With the pip install and copying the API key, I ran the CLI first try in less than 5 minutes. Not sure I can think of anything faster to be honest.

2. Have you edited any labels in the interface; if not, why not? What did you like/not like?

I used the editor on several occasions. I like the hotkey setup for accepting and rejecting results. For data-centric ML researchers, such as myself, it's always useful to examine the data in fine detail, which the interface allows you to do. I don't recommend any necessary changes here.

3. Have you exported your improved dataset and used it? What was the outcome?

I've exported datasets multiple times: some with labeling errors completely excluded and some with labeling errors manually or autocorrected depending on the need. I really like the option to use an agnostic model to check for labeling issues as an additional layer of scrutiny over my work (Loved your paper, "Model-Agnostic Label Quality Scoring to Detect Real-World Label Errors"). Below are some examples where I've used Cleanlab Studio for dataset exports.

1. I've used Cleanlab Studio to bootstrap and wrangle quick public datasets, from places such as HuggingFace, into a strong performing first model development.

2. I've used Cleanlab Studioto QC labeled examples given to me from annotators where I am NOT in control of the annotation policy or methodological guidelines. In two instances, I've flagged examples using Cleanlab Studio where the annotators have admitted to misclassification. This is extremely important to me when I can't be a subject-matter expert on all data sources, and I need Cleanlab Studio for this.

3. Lastly, I use weak supervision quite often and firmly believe Cleanlab Studio is an antidote to many of the weaknesses that come with the uncertainty and laborious tasks of optimizing LFs.

Additional thoughts on Issue Summary?

The "issue summary section" is very helpful to communicate with clients or other researchers. It provides a breakdown of all the label quality issues one would expect.