AIDRIN: A Comprehensive Toolset for Automating Data Preparation for AI

aidrin.io

Kaveen Hiniduma[1*], **Jean Luca Bez**[2*], **Ravi Madduri**[3], and **Suren Byna**[1]

[1]*The Ohio State University,* [2]*Lawrence Berkeley National Laboratory,* [3]*Argonne National Laboratory*

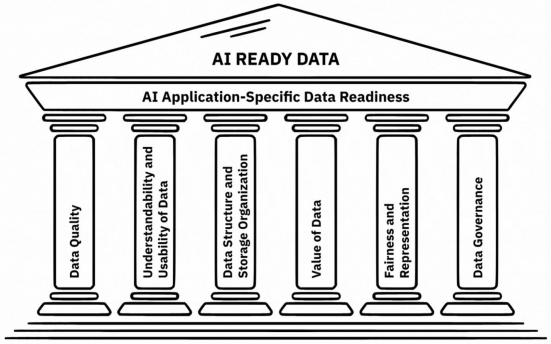
[*]*These authors contributed equally to this work*

AIDRIN aims to automate dataset readiness for AI training by assessing **data readiness** for AI across six pillars. Our framework provides real-time insights to improve data preparation efficiency for highly performant and trustworthy AI.

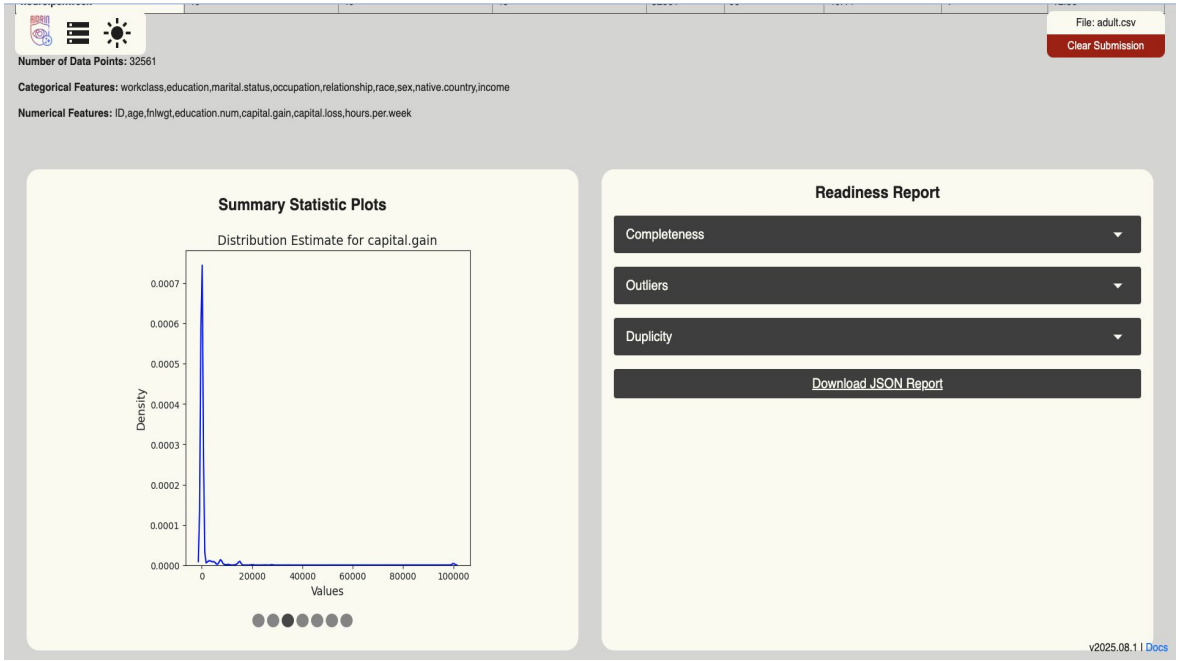
Data Readiness for AI is Critical

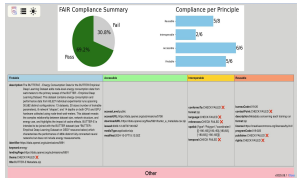
- Poor quality data negatively impacts AI model outcomes
- Manual readiness evaluations are time consuming and error-prone
- Community needs automated & reliable readiness evaluation tools

A High-Dimensional Data Readiness Framework



- **AIDRIN** assesses readiness of data across a high-dimensional metric space that are organized into **six** pillars.
- Provides data readiness reports with metric evaluation visualizations
- Allows user-defined metrics, evaluation rules, and remedies
- Inspect your data at https://**aidrin.io**





FAIR principles compliance evaluation

Applying user-defined metrics and remedies

github.com/**idtlab/AIDRIN**

aidrin.readthedocs.io

byna.1@osu.edu, **jlbez@lbl.gov**