

Here is the checklist for the author developing the artificial intelligence-machine learning software and for the reviewer reviewing it:

Data Management Checklist

Data Management Requirements Capture

- Did you capture the objectives to be satisfied at the level of each individual data point?
- Are the data management requirements clear and well-defined?

Data Collection

- Is the data collection process properly documented?
- Was the data collected from a reliable source?
- Are there any missing or incomplete data points?

Data Labelling (only for supervised learning)

- Were the data labels accurately assigned?
- Are the labels consistent throughout the dataset?
- Are there any ambiguous or unclear labels?

Data Preparation (pre-processing, data transformation, feature engineering)

- Did you perform necessary data cleaning and preprocessing steps?
- Was data transformed to meet the requirements of the ML algorithm?
- Were relevant features engineered for the ML model?

Identification of Data Sets

- Are there distinct datasets for training, validation, and testing purposes?
- Is each dataset properly identified and documented?
- Are the datasets consistent in terms of data quality and format?

Dataset Verification (accuracy, completeness, representativeness)

- Did you verify the accuracy of the datasets?
- Are the datasets complete with no missing or incomplete data points?
- Are the datasets representative of the target population or problem domain?

Independence Requirements between Data Sets

- Are there any requirements for independence between the datasets?
- Were these requirements met during data collection and processing?

Identification and Elimination of Unwanted Bias

- Did you identify potential biases in the datasets?
- Were steps taken to eliminate or mitigate these biases?

I hope this helps! Let me know if you need further assistance.