

Data Sheets and Model Cards

Ashish Mahabal
aam at astro.caltech.edu
Caltech Astronomy
Center for Data Driven Discovery, Caltech
IVOA, KD, 2021-11-03 (Remote)

Datasheets for Datasets (Gebru et al. 1803.09010)

Dataset creators:

careful reflection on the process of creating,
distributing, and
maintaining a dataset,
including any underlying assumptions,
potential risks or harms, and
implications of use.

Dataset consumers:

have the information needed to make informed decisions about using the dataset

Other's touched ...

policy makers,

consumer advocates,

individuals whose data is included in those datasets, and

those who may be impacted by models trained or evaluated on those datasets.

Leads to transparency

(Origin: social arenas)

Questions sheets should answer

Instances

Stats

Gaps

Relationships

Splits

Links

...

Provenance

Distribution

Maintenance

Impacts

Model Cards for Model Reporting (Mitchell et al. 1810.03993)

Many of the same topics as for data sheets:

Creators

Consumers

Biases, exhaustiveness, applicability, transparency, interpretability, trust, ...

Model Cards

Model Details.

Person or organization developing model

Model date

Model version

Model type

Intended Use.

Factors.

Metrics.

Evaluation Data.

Training Data.

Quantitative Analyses

Ethical Considerations

Caveats and Recommendations

Astronomy Specific Use-cases

Different surveys and their data releases (their compatibility)

Bands

Aperture

Area

Depth

Number of epochs

Nature of gaps

ZTF science groups

Galactic

Extra-Galactic

Solar System

Parameter space covered, their overlap

AGNs, SNe, TDEs

LSST/Vera Rubin Observatory ..

Different Science Groups

Varied requirements

Compatibility

Collaborations

(e.g. Plasticc challenge)

Models developed for one set ...

Transfer learning - when is it futile?

IVOA can take a lead in encouraging groups to maintain standardised information so that data fusion becomes simpler