Research Data Services

Data Literacy Series, n.01/2026

The Value of Model Cards in Machine Learning Research

Machine Learning (ML) models are mathematical representations learned from data that enable machines to make predictions, classifications, or decisions. Without proper documentation, they function as opaque black boxes. In a fast-moving research landscape, model cards help ground innovation in clarity and trust. Think of them as READMEs for models with rich metadata that make ML models easier to discover, evaluate, reproduce, and share.

OVERVIEW

The concept of Model Cards was first introduced in 2018* as a way to create consistent and structured documentation for machine learning models, supporting clearer communication, greater accountability, and more responsible use in real-world settings.

They are designed to openly communicate key details about a trained model, how it was developed, the assumptions made during training, and how its behavior and performance may vary across different cultural, demographic, or phenotypic groups.

*Seminal Paper
Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229). doi.org/10.1145/3287560.3287596

WHY USE MODEL CARDS?

To openly and transparently describe what a model is, how it was built, and how it should (and should not) be used.

To promote scientific collaboration, reproducibility, and responsible reuse, supporting faster discovery and the efficient use of shared models.

To support attribution and proper credit for shared models, similarly to other valuable research outputs (e.g., data and code).

We're looking for ML research projects on campus and would love to connect to see how we can best support your work. Interested?

rds@library.ucsb.edu

UC **SANTA BARBARA**
Library

RECOMMENDED SECTIONS

- **Model Details**
  Key information about the model, including developers, version, type, training approach, publications, licensing, and contact points.
- **Intended Use**
  Primary use cases, intended users, and scenarios to avoid.
- **Factors**
  Characteristics or conditions (demographic, environmental, or technical) that may affect performance.
- **Metrics**
  Performance measures and evaluation methods, including thresholds and variability assessment.
- **Evaluation Data**
  Datasets and preprocessing steps used for evaluation, along with the rationale for their selection.
- **Training Data**
  Details about the training data and how it can be accessed. If access is limited, include the main distributions and characteristics.
- **Quantitative Analyses**
  Evaluation results, both overall and across intersecting factors.
- **Ethical Considerations**
  Potential biases, fairness issues, and ethical implications.
- **Caveats & Recommendations**
  Limitations, safe usage guidance, and deployment best practices.

TEMPLATES & GUIDELINES

Some dedicated ML hubs offer templates and guidelines to help researchers create clear, effective model cards. Check out their websites for more information:

🤗 **huggingface_hub**

PyTorch Hub

TensorFlow Hub

www.library.ucsb.edu