THE PDS DISCIPLINE NAMESPACE FOR MACHINE LEARNING PRODUCTS. Kiri L. Wagstaff[1], Michael McAuley[1], and Minh Le[2]; [1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA (kiri.l.wagstaff@jpl.nasa.gov), [2]Stripe.

Introduction: The Planetary Data System (PDS) provides several metadata namespaces that enable the description of specialized data products (e.g., those with image, cartographic, or spectral properties) by defining new terms as extensions to the PDS4 Information Model [1]. Machine learning models are increasingly being used to analyze planetary mission data and generate products of additional value. We have designed a PDS Machine Learning Analysis (ML) discipline namespace (also known as a Local Data Dictionary, LDD) that provides relevant classes and attributes needed for machine learning products.

Machine Learning Discipline Namespace: There are many kinds of machine learning products, including per-item classifications (of images, spectra, time series, etc.), per-item real-valued predictions (e.g., temperature, thermal inertia, atmospheric dust content), image captions, image segmentations, and more. The ML namespace enables data providers to capture the information needed to understand the model that generated the product. The choice of keywords included in the ML namespace was inspired by other efforts aimed at documenting machine learning models, such as model cards [2] and model provenance [3].

Figure 1 shows the conceptual structure of the ML namespace. The primary class enables the description of the machine learning model that generated the product (`Trained_Machine_Learning_Model`). Required components include the model name, version, algorithm employed (learning style, type, name, and DOI), and training data set. Each data set is described by a version and the number of items in the data set. Optional components, enclosed in square brackets, include a validation set, test set, and test set performance results. Each performance result is described by the name of the measure and the numeric score. Full details and source for the ML LDD are available at https://github.com/pds-data-dictionaries/ldd-ml.

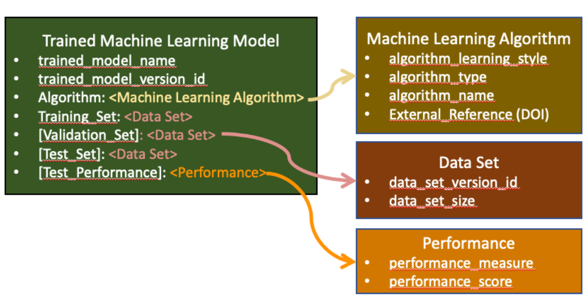

Figure 1: Machine Learning namespace concepts

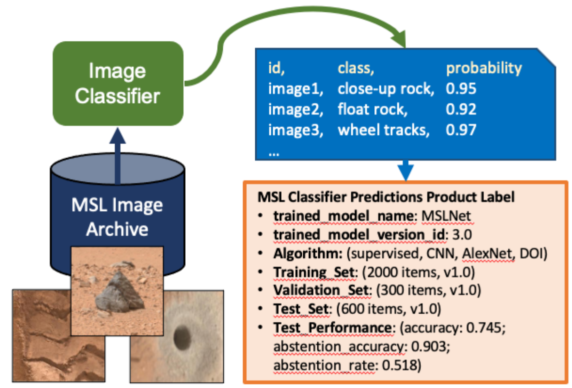

Figure 2: Example of using the ML namespace to describe a model for Mars image classifier products

Example for Mars Image Classification: We developed a PDS4 label file to describe the output of a model trained to classify content in images collected by the Mars Science Laboratory (MSL) rover [4]. The classifier is applied to the image archive and generates a product that contains, for each image, the predicted class and its posterior probability. An excerpt from the product's PDS4 label that uses terms from the ML namespace is shown in the light orange box (Figure 2). More details can be included as supplemental products, such as a complete list of the image products that are contained in the training, validation, and test sets.

Conclusion: Machine learning applied to planetary mission data generates new data products of value. The PDS Machine Learning namespace provides relevant keywords and enables these products to have the management and traceability of all PDS products.

References: [1] PDS4 Information Model Specification, https://pds.nasa.gov/datastandards/documents/im/current/index_1H00.html [2] M. Mitchell et al. (2019), "Model Cards for Model Reporting," In *Proc. of the ACM Conf. on Fairness, Accountability, and Transparency.* [3] Y. Gil et al. (2013), "PROV Model Primer," W3C Provenance Working Group Draft, http://www.w3.org/TR/prov-primer/ [4] K. Wagstaff et al. (2021), "Mars Image Content Classification: Three Years of NASA Deployment and Recent Advances," In *Proc. of the 33rd Conf. on Innovative Applications of AI.*

Acknowledgments: We thank the PDS Cartography and Imaging Sciences Node for funding and the PDS Engineering Node for supporting this work. This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.