Data Stewardship DMP Exercise

ID (HTTP-ORCID)	Created	Modified	Language
https://doi.org/10.5281/zenodo.2648751	22-04-2019	22-04-2019	en

	Data Officer	Moritz Leidinger, moritz.leidinger@student.tuwien.ac.at https://orcid.org/0000-0002-4537-6648 (HTTP-ORCID)
I	**Data Characteristics**	
I.1	**Description of the data**	UCI Arrhythmia Dataset This experiment aims to apply two machine learning models to solve a classification task for two different datasets. Each dataset is evaluated with each of the machine learning models, using different parameter settings and preprocessing strategies to compare the respective results and analyze across datasets and/or machine learning models. Type: dataset Language: en Keywords: Breast Cancer, Arrhythmia
II	**Documentation and Metadata**	
II.1	**Metadata standards**	
II.2	**Documentation of data**	
II.3	**Data quality control**	• We use a readme file to describe the data quality and provide the source code of the experiment.
III	**Data availability and storage**	
III.1	**Data sharing strategy**	ID-1: https://doi.org/10.5281/zenodo.202648713 (HTTP-ORCID) UCI Repository: A publicly available dataset about Arrhythmia. Size: <10MiB Data access: open URL: https://archive.ics.uci.edu/ml/machine-learning-databases/arrhythmia/arrhythmia.data Format: text/csv
III.2	**Data storage strategy**	
IV	**Legal and ethical aspects**	
IV.1	**Legal aspects**	License: • Valid from: 2019-04-22 https://opensource.org/licenses/MIT
IV.2	**Ethical aspects**	Personal data: unknown Preservation statement: Must be preserved to keep research & development between stakeholders intact. Sensitive data: unknown Ethical issues: unknown