Artifact for the ESEC/FSE'22 Paper: "23 Shades of Self-Admitted Technical Debt: An Empirical Study on Machine Learning Software":

This artifact contains the code, data, and additional information for reproducing the results found in the ESEC/FSE 2022 paper entitled "23 Shades of Self-Admitted Technical Debt: An Empirical Study on Machine Learning Software". This study analyzes the occurrence of self-admitted technical debt (SATD) in a dataset consisting of 2,641 open-source machine learning repositories. The artifact contains the Boa scripts ran to acquire the comment data, as well as the Python scripts which were used to filter the dataset into 68,820 SATD comments. During the dataset creation, a sample was taken for two authors to independently label before settling disagreements in discussion of a moderator. The authors' labels and the agreed upon labels are included within the artifact. Instructions for reproducing our dataset can be found in our GitHub repository: https://github.com/DavidMOBrien/23Shades.

Paper Title: 23 Shades of Self-Admitted Technical Debt: An Empirical Study on Machine Learning Software
Authors: David OBrien, Sumon Biswas, Sayem Mohammad Imtiaz, Rabe Abdalkareem, Emad Shihab, and Hridesh Rajan

A. Badges:

We are applying for all three badges **Evaluated – Functional, Evaluated – Reusable,** and **Available**. We defend our application for these badges in the STATUS file found on the GitHub repository.

B. Installation and Requirements:

The included artifact contains both the data and scripts. Because the data gathered is so large, the output for our Boa scripts is supplied and is named "1-output.txt". The Python scripts can be run with a default installation of Python 3, no additional packages are needed to be installed. We supply the expected output of each script in our artifact for reviewers to compare with.

C. Artifact Access:

The artifact can be accessed at https://github.com/DavidMOBrien/23Shades, where a Zenodo link is supplied to the larger datasets that could not be uploaded to GitHub due to file size constraints. The README contains the necessary instructions to reproduce our results.