Definition of Done Dataset

Max Bogaers

December 7, 2021

1   Introduction

One of the deliverables within this project is a dataset for other teams to use for the continuation of this project. This document describes what we want the dataset to be, how it will be delivered and when it is considered done.

2   What should the dataset be

Because this project is about the prediction of human satisfaction within the Brains for Buildings project, the data within the dataset should relate to human satisfaction. The dataset should also contain at least one a categorical value relating to human satisfaction. Preferably these values should be self collected, but with given time constrictions the possibility of mocking this data should also be looked into. In the event that we are mocking the human satisfaction values they should be based on scientific studies. and it would be nice if we are able to use more than one study to determine human satisfaction. The dataset should also be big enough to be used within a predictor model.

3   how will the dataset be delivered

The dataset will be delivered in a .CSV format for use with python machine learning models

4   when do we consider the dataset to be ”done”

The dataset is considered done when the following conditions are met:

1. The dataset should have data relating to the indoor and/or outdoor climate of classroom or office.

2. The dataset contains at least 1 column relating to user satisfaction (preferably with real data).

3. The dataset should be big enough to be used by a machine learning model. A good ballpark guess is to have at least 10x the records of the features used, but more is better.