|  |  |
| --- | --- |
| Setting up a ML codebase | ├── README.md <- You are here  │  ├── config <- Directory for yaml configuration files for model training, scoring, etc  │ ├── logging/ <- Configuration of python loggers  │  ├── data <- Folder that contains data used or generated. Only the external/ and sample/ subdirectories are tracked by git.  │ ├── archive/ <- Place to put archive data is no longer usabled. Not synced with git.  │ ├── external/ <- External data sources, will be synced with git  │ ├── sample/ <- Sample data used for code development and testing, will be synced with git  │  ├── docs <- A default place for project’s documents.  │  ├── figures <- Generated graphics and figures to be used in reporting.  │  ├── models <- Trained model objects (TMOs), model predictions, and/or model summaries  │ ├── archive <- No longer current models. This directory is included in the .gitignore and is not tracked by git  │  ├── notebooks  │ ├── develop <- Current notebooks being used in development.  │ ├── deliver <- Notebooks shared with others.  │ ├── archive <- Develop notebooks no longer being used.  │ ├── template.ipynb <- Template notebook for analysis with useful imports and helper functions.  │  ├── src <- Source data for the project  │ ├── archive/ <- No longer current scripts.  │ ├── helpers/ <- Helper scripts used in main src files  │ ├── sql/ <- SQL source code  │ ├── ingest\_data.py <- Script for ingesting data from different sources  │ ├── generate\_features.py <- Script for cleaning and transforming data and generating features used for use in training and scoring.  │ ├── train\_model.py <- Script for training machine learning model(s)  │ ├── score\_model.py <- Script for scoring new predictions using a trained model.  │ ├── postprocess.py <- Script for post processing predictions and model results  │ ├── evaluate\_model.py <- Script for evaluating model performance  │  ├── test <- Files necessary for running model tests  │ ├── true <- Directory containing sources of truth for what results produced in each test should look like  │ ├── test <- Directory where artifacts and results of tests are saved to be compared to the sources of truth. Only .gitkeep in this directory should be synced to Git  │ ├── test.py <- Runs the tests defined in test\_config.yml and then compares the produced artifacts/results with those defined as expected in the true/ directory  │ ├── test\_config.yml <- Configures the set of tests for comparing artifacts and results. Currently does not include unit testing or other traditional software testing  │  ├── run.py <- Simplifies the execution of one or more of the src scripts  ├── requirements.txt <- Python package dependencies |