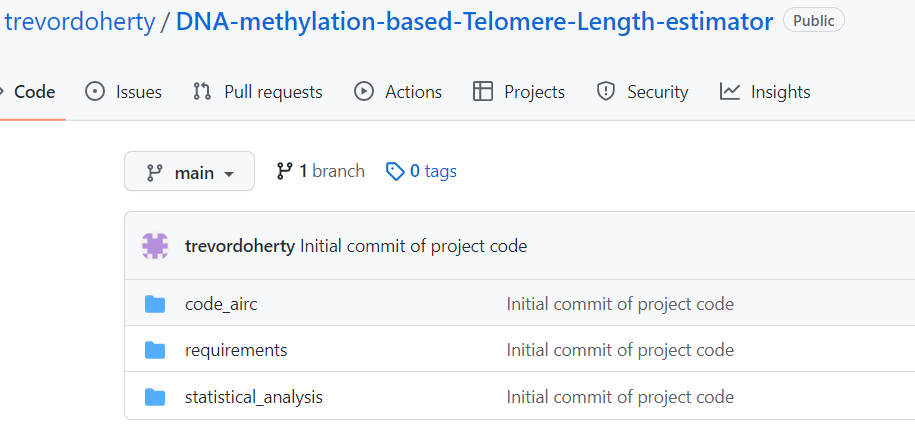
**Readme file for Telomere Length Estimator Project**

First, create a virtual environment using the supplied requirements.txt file. This file contains all the Python packages necessary to implement the models.

To run the models on Windows, you can:

1. Download and Install Anaconda. Make sure to allow Anaconda to be added to your path environment variable (this should be an option as part of the installation process).
2. Next, set up a virtual environment using python 3.9.9.
3. Go to command prompt.
4. Type conda create -n name python=3.9.9 (where “name” is the name you want to give the virtual env)
5. Activate the virtual environment with the command: activate name (where “name” is the name of the virtual environment)
6. Create a folder containing the subfolders *code\_airc*, *requirements and statistical analysis*.



1. Within the *code\_airc* folder, there are 2 subfolders:

Graphical user interface, text, application, email

Description automatically generated

1. The folder *nested\_cv\_analysis* contains the scripts for running the nested cross-validation models which utilise the Dunedin data set. *main4.py* is the executable script which calls the other supporting scripts. *get\_data4.py* contains the functionality to read in the data files, applying preprocessing, cleaning and structuring steps. *models4\_gridsearch.py* contains the code for applying the nested CV analysis with hyperparameter tuning and model selection. *feature\_selection4.py* contains the functions for applying the feature selection and transformation methods. Note that *launch\_en\_timings.sh* is the shell script for running models on the local cluster.

Graphical user interface, text, application, email

Description automatically generated

1. The folder *final\_estimators* contains scripts for running the models which were developed using the discovery data set (Dunedin) and evaluated on the 2 independent data sets (EXTEND and TWIN). *main.py* is the executable script which calls the other supporting scripts. *get\_data.py* contains the functionality to read in the data files, applying preprocessing, cleaning and structuring steps. *feature\_selection.py* contains the functions for applying the feature selection and transformation methods. *test\_sets\_evaluation.py* contains the functions and code for developing models on the discovery data and evaluating these models on the 2 independent test data sets. The *utils.py* script contains supporting utility functions.

Graphical user interface, text, application, email

Description automatically generated

1. The *statistical\_analysis* folder contains the following files:

Graphical user interface, text, application, email

Description automatically generated

The Jupyter notebook *tuning\_and\_feat\_sel.ipynb* contains functions for the creation of graphs for the paper and visualisation of results. The plots showing the performance of feature selection approaches in the nested CV analysis and the number of features remaining in models at each stage of feature reduction are shown in this file.

The notebook *final\_statistical\_analysis\_fs\_paper.ipynb* contains the various statistical analysis including correlation analysis between actual/estimated TL and a range of subject clinical and phenotypic data such as blood cell concentrations etc. It also contains the association analyses between actual/estimated TL and age controlling for a range of confounding variables. Additionally, the correlation plots between age and TL, actual/estimated TL and monocyte counts etc. are generated in this file.

1. The *requirements* folder contains the requirements.txt file for loading the necessary python libraries to run models.
2. Navigate to the *requirements* folder within the terminal and enter the command: pip install -r requirements.txt

(This will load all necessary Python packages within the activated virtual environment)

1. It should now be possible to run Jupyter notebooks and python scripts within this virtual environment.
2. For any Jupyter notebooks, navigate to the folder containing the relevant notebook and type: jupyter notebook - this should activate Jupyter notebook in a browser window.