Today’s focus has been on how to construct the program’s main structure. I decided to explore the foundations for the program through the Machine Learning Life Cycle.

Define the input for the program.

Previous work has been focused on establishing a pipeline and implementing a function to track the changes. The next step will be to evaluate upon whether the current procedures are in the correct direction and to implement code for tracking the repository into the integration system.

Client working on machine learning repository.

Watches the repo for changes: GitHub

GitHub: check the last time of the updates. Use the Id to check whether there have been any changes.

Tracking code: examine the commit ids.

Create local copies of the data for comparison.

Git computes differences and stores the code.

**How to focus on tracking the data in the repository?**

Timestamp to track the data changes (simple)

Examine the changes on the single repository.

Version Control : use the existing version control on the sources.

**Optional**: Design a source control for the data.

Installed git-python to use Python to track the Git repository versions.

I have added comments for the files pipeline.py and pipeline\_example.py.

Career: PHD research project for funding and advertised PHD project roles.

Application process: interview.

Institutes with programs: Biology and Welcome Trust Program: Interview + Selection Process.

<https://wellcome.org/grant-funding/schemes/four-year-phd-programmes-studentships-basic-scientists>

Key identification: Department + Research Topics.

Topics: science and machine learning.