

# Author Identification Project

Kahlil Khan / kkhan37 / CS-6460 Educational Technology / Summer 2018

## Description of files

- **open.ipynb**: Jupyter notebook. Trains the author identification model. Status: Complete.
- **origin.py**: Python. The external tool. Provides interface to connect Canvas. Status: Draft interface works command line, not yet with Canvas (requires JSON formatting)
- **origin-test.ipynb**: Jupyter notebook. Demonstration surrogate for Canvas LMS.
- **lti.ipynb**: Jupyter notebook. Alpha-level attempt to connect to Canvas via its LTI version 1 and 2 interfaces.
- **environment.yml**: Text. Listing of development dependencies. See installation instructions below.

## Description of data

- **data/C50/C50all/**: 5051 articles from 50 authors. This is a combination of the training and test data from Reuter\_50\_50 dataset
- **data/5-300-lstm128-128-20-50-model.h5**: The trained Keras model. Trained by open.ipynb and used in origin.py.

## Installation instructions

1. Install and run mongo database: [macos](#) or [linux](#).
2. Install anaconda from <https://www.anaconda.com>
3. `conda env create -f environment.yml`
4. `source activate tensorflow_p36`
5. `jupyter lab`
6. Open **origin-test.ipynb**