The file `main.ipynb` can be used for data preparation, cleaning, feature-engineering, and generating the final models.

Ensure that utility files

- `data_utils.py`

- `nlp_utils.py`

- `model_utils.py`

are all in the same working directory as `main.ipynb`. File `requirements.txt` lists packages and versions required for reproducibility.

`GoogleScholar.ipynb` contains the code for extracting project outcome information from Google Scholar, which depends upon library `scholarly`. `scholarly` searches are often obstructed by Google's reCAPTCHA, resulting in IP bans, especially if searching by title rather than by author.