

Files required for capstone project submission

- Project Proposal
- Project Report
- Development code:
 1. Exploratory Analysis and Preprocessing Notebook
Notebook for Exploring, visualizing and cleaning the data
 2. Initial model Notebook
Notebook for creating, training and testing initial model
 3. Hyperparameter Notebook
Notebook for creating pipeline, tuning hyperparameters, choosing the best model and evaluating it
 4. Testing the model with custom input
Notebook for loading the model and giving the user the ability to test the model with custom input (comment). For testing custom comments, add your comment to `input_comment` list and run last two sections. Predictions will be printed as pandas dataframe with with predicted toxicity classes. If all values are 0 - input comment is predicted as not toxic.

Software

python==3.7.3

Install requirements with `pip install -r requirements.txt`

Used libraries:

- pandas==0.24.2
- scikit-learn==0.20.3
- Keras==2.2.4
- tensorflow==2.0.0

Data

Data was provided by Kaggle competition: <https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data>

Please extract data to `data/` folder