This file describes all the submission files as part of our project.

**Python Notebooks:**

1. csc2515-project-stackoverflow.ipynb: Pre-process the original [Stack Overflow 10% dataset](https://www.kaggle.com/stackoverflow/stacksample) and extract numerical features.It outputs train.csv as the machine learning training set to feed into the Models\_Training.ipynb.
2. Models\_Training.ipynb: training the different classical and modern machine learning models, including linear regression, random forest, multi layer perceptron (MLP), CNN, LSTM.
3. 1D-CNN-LSTM.ipynb: Similar to Models\_Training.ipynb, it trains the CNN-LSTM model.

**Data:**

1. Train.csv: contains the pre-processed dataset which we then use to train our machine learning models. The file can be found here: https://drive.google.com/file/d/1MS\_GfLCWG21EfmcbvZWsH-k5jAbh7Rq-/view?usp=sharing