* See the GitHub repository for the specific variables gathered from each data source.
* I created 5 versions of each fold, each with a different selection of features (see Table 5 for a summary and the GitHub repository for a spreadsheet giving the specific features used in each subset).