# Group 2 Project Submission

## File Manifest

The attached Group2FinalDeliverable.zip file contains the deliverable contents for Group 2’s submission to the Kaggle competition “Real Or Not? NLP with Disaster Tweets” (<https://www.kaggle.com/c/nlp-getting-started> ).

|  |  |
| --- | --- |
| File/Folder Name | Description |
| 1\_DataVisualizationModelEvaluation.ipynb | This Jupyter Notebook contains the scripts used to compare the various analytic models. It also includes several visualizations used during the comparison. |
| 2\_CrossVectorizationSample.ipynb | This was the groups first attempt at creating a model using Binary Vectorization through the CountVectorizer library. |
| 3\_SpacyLSVMGloveVectorization.ipynb | This was the final model the group generated using the GloVe Vectors contained in the Spacy library along with a Support Vector Machine Model |
| 4\_Twitter API.ipynb | This notebook contains a script that will pull a 100 tweets based off a key word into a csv file. This was used for random testing of the model when complete.  *Note: To use this library, a developer account needs to be requested from twitter. This can be completed at the following link:*  <https://developer.twitter.com/en/docs/twitter-api/getting-started/guide#:~:text=To%20make%20any%20request%20to,all%20requests%20to%20the%20API> |
| app.py | This script contains the code for the flask application that was tested by connecting from a web page (the web page code is found in the web UI folder)  To run this application:   1. Open an anaconda prompt window 2. Navigate to the folder on your machine where the application file (e.g. app.py) is located. If using Windows, for example, cd documents/folder 3. Type set FLASK\_APP=app.py and hit enter. 4. Type flask run and hit enter. You should get a response that tells you that the application is running |
| GloveVec.py | This is the reusable library the group created for storing the code used to create the model. This is reused in the flask application code. |
| Svm\_model.pkl | This is the compiled SVM model that was generated in the 3\_SpacyLSVMGloveVectorization notebook. This is re-used in the flask application code. |
| **Folder Name** | **Folder Description** |
| Data | This folder contains the train and test data stored in csv format that was provided as part of the competition |
| flaskUI | This folder contains a sample html page and jQuery library. This page will call the flask application to score sample text entered onto the page. |