SUPPLEMENTARY FILES

A TEXT MINING FRAMEWORK FOR THE CLASSIFICATION AND PRIORITIZATION OF DISASTER-RELATED TWEETS FOR DISASTER RESPONSE

Parameters Used in the Classification Models

Simply, the classification model building step uses default parameters embedded by the functions used in RWeka. The following images show the parameters per classification model.

NB Parameters:

```
> tweets_df_data_NBmodel
Naive Bayes Classifier

                 Class
Attribute      Irrelevant   Relevant
                 (0.51)     (0.49)
=====================================
favorited
  FALSE          1996.0     1912.0
  [total]        1996.0     1912.0

favoriteCount
  mean           24.1865     7.5823
  std. dev.     274.7439    57.4714
  weight sum        1995        1911
  precision      47.3525    47.3525

ifReply
  mean            0.0887     0.0262
  std. dev.       0.2843     0.1667
  weight sum        1995        1911
  precision            1           1

userType
  #N/A           1909.0     1592.0
  CEL               2.0        1.0
  FEA               1.0       13.0
  GOV               3.0       12.0
  JOU               1.0        2.0
  LAY              82.0      230.0
  NEWS              2.0       43.0
  NGO               1.0        6.0
  OTH               2.0        4.0
  SCH               2.0       18.0
  [total]        2005.0     1921.0
```

```
truncated
  FALSE          1882.0     1713.0
  TRUE            115.0      200.0
  [total]        1997.0     1913.0

retweetCount
  mean           2548.5721  1137.1794
  std. dev.     15559.9517  2768.9205
  weight sum        1995        1911
  precision     384.6283    384.6283

isRetweet
  FALSE           639.0      826.0
  TRUE           1358.0     1087.0
  [total]        1997.0     1913.0

retweeted
  FALSE          1996.0     1912.0
  [total]        1996.0     1912.0

ifMention
  mean            0.1078     0.0251
  std. dev.       0.3101     0.1667
  weight sum        1995        1911
  precision            1           1

containsURL
  mean            0.4316     0.3124
  std. dev.       0.4953     0.4635
  weight sum        1995        1911
  precision            1           1
```

```
bagyong
  mean            0.0195     0.0288
  std. dev.       0.1667     0.1703
  weight sum        1995        1911
  precision            1           1

hundred
  mean            0.0764     0.1151
  std. dev.       0.6858     0.668
  weight sum        1995        1911
  precision         1.25        1.25

tropical
  mean                 0      0.044
  std. dev.       0.1667      0.205
  weight sum        1995        1911
  precision            1           1

safe
  mean             0.002     0.1156
  std. dev.       0.1667     0.3214
  weight sum        1995        1911
  precision            1           1

people
  mean            0.0175      0.057
  std. dev.       0.1667     0.2493
  weight sum        1995        1911
  precision            1           1

walangpasok
  mean            0.0025     0.0492
  std. dev.       0.1667     0.2163
  weight sum        1995        1911
  precision            1           1
```

```
typhoon
  mean            0.0035     0.0754
  std. dev.       0.1667     0.2718
  weight sum        1995        1911
  precision            1           1

haha
  mean            0.0356     0.0042
  std. dev.       0.1853     0.1667
  weight sum        1995        1911
  precision            1           1

city
  mean            0.0055     0.0555
  std. dev.       0.1667     0.2549
  weight sum        1995        1911
  precision            1           1

level
  mean             0.003     0.0392
  std. dev.       0.1667     0.1942
  weight sum        1995        1911
  precision            1           1

tomorrow
  mean             0.001     0.0419
  std. dev.       0.1667     0.2003
  weight sum        1995        1911
  precision            1           1

wala
  mean            0.0206     0.0387
  std. dev.       0.1765     0.1929
  weight sum        1995        1911
  precision            1           1
```

```
help
  mean              0.01     0.1999
  std. dev.       0.1667     0.4719
  weight sum        1995        1911
  precision            1           1

lang
  mean             0.017     0.0256
  std. dev.       0.1667     0.1708
  weight sum        1995        1911
  precision            1           1

rescue
  mean                 0     0.0748
  std. dev.       0.1667      0.295
  weight sum        1995        1911
  precision            1           1

stay
  mean             0.003     0.0853
  std. dev.       0.1667     0.2992
  weight sum        1995        1911
  precision            1           1

lakas
  mean             0.003     0.0612
  std. dev.       0.1667     0.2441
  weight sum        1995        1911
  precision            1           1

hangin
  mean             0.002     0.0827
  std. dev.       0.1667      0.281
  weight sum        1995        1911
  precision            1           1
```

```
pls
  mean             0.006     0.0455
  std. dev.       0.1667     0.2134
  weight sum        1995        1911
  precision            1           1

need
  mean              0.01     0.0926
  std. dev.       0.1667     0.3398
  weight sum        1995        1911
  precision            1           1

rescueph
  mean             0.001     0.1502
  std. dev.       0.1667     0.3573
  weight sum        1995        1911
  precision            1           1

jessica
  mean            0.1454           0
  std. dev.       0.3553     0.1667
  weight sum        1995        1911
  precision            1           1

jessicasoho
  mean            0.0576           0
  std. dev.       0.2394     0.1667
  weight sum        1995        1911
  precision            1           1

spirit
  mean            0.0516           0
  std. dev.       0.2429     0.1667
  weight sum        1995        1911
  precision            1           1

awareness
  mean                 0     0.0476
  std. dev.       0.1667      0.213
  weight sum        1995        1911
  precision            1           1
```

```
> tweets_df_data_SVMmodel
SMO

Kernel used:
  Linear Kernel: K(x,y) = <x,y>

Classifier for classes: Irrelevant, Relevant

BinarySMO

Machine linear: showing attribute weights, not support vectors.

        -2.8045 * (normalized) favoriteCount
 +      -0.6588 * (normalized) ifReply
 +      -1.6323 * (normalized) userType=#N/A
 +      -1      * (normalized) userType=CEL
 +       0.3668 * (normalized) userType=FEA
 +       0.3676 * (normalized) userType=JOU
 +       0.3674 * (normalized) userType=LAY
 +       1.2537 * (normalized) userType=NEWS
 +       0.3595 * (normalized) userType=NGO
 +      -0.0826 * (normalized) userType=OTH
 +      -0.2072 * (normalized) truncated=TRUE
 +      -0.055  * (normalized) retweetCount
 +      -1.3421 * (normalized) isRetweet=TRUE
 +      -0.6845 * (normalized) ifMention
 +      -0.6584 * (normalized) containsURL
 +      -0.9437 * (normalized) bagyong
 +       0.6606 * (normalized) hundred
 +       1.9996 * (normalized) tropical
 +       3.5    * (normalized) safe
 +       0.008  * (normalized) people
 +       1.9986 * (normalized) walangpasok
 +       5.999  * (normalized) typhoon
 +      -0.9953 * (normalized) haha
 +       3.7483 * (normalized) city
 +       0.7293 * (normalized) level
 +       1.2702 * (normalized) tomorrow
 +       0.2507 * (normalized) wala
 +       6.0004 * (normalized) help
 +       1.3145 * (normalized) lang
 +       3.9985 * (normalized) rescue
 +       1.5    * (normalized) stay
 +       1.3151 * (normalized) lakas
 +       3      * (normalized) hangin
 +       3.3696 * (normalized) pls
 +       1.9996 * (normalized) need
 +       2.6585 * (normalized) rescueph
 +      -3.3693 * (normalized) jessica
 +      -0.6579 * (normalized) jessicasoho
 +      -0.5    * (normalized) spirit
 +       2.658  * (normalized) awareness
 +       1.9752

Number of kernel evaluations: 2766390 (76.256% cached)
```

Here, the SMO function is one method that trains SVM models. It is perceived as a more efficient method or faster training method.

RF Parameters:

```
> tweets_df_data_RFmodel
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
```

K stands for the number of attributes to randomly investigate, 0 is equivalent to int($\log_2$(#predictors)+1)). M is the number of instances per leaf where the default is 1. V sets the minimum numeric class variance proportion of train variance for split with default value of 0.001. S sets the random seed.