Supplementary Materials

Table 1: Symptom category co-occurrence matrix.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | Thinking, concentration, and decisions | Interest and Motivation | Physical: sleep | Physical: fatigue | Physical: appetite | Physical: movement | Substance use | PHQ-9-suicidal tendencies | PHQ-9-emotional: depressed and sadness | PHQ-9-emotional: worthlessness and worry | Physical: pains | Social | Mental health-related issues |
| Thinking, concentration, and decisions | 1022 | 16 | 19 | 10 | 4 | 1 | 3 | 7 | 15 | 24 | 4 | 11 | 6 |
| Interest and Motivation | 16 | 370 | 12 | 10 | 4 | 24 | 0 | 2 | 2 | 9 | 2 | 8 | 1 |
| Physical: sleep | 19 | 12 | 872 | 10 | 14 | 4 | 2 | 1 | 9 | 10 | 6 | 4 | 5 |
| Physical: fatigue | 10 | 10 | 10 | 725 | 3 | 5 | 1 | 6 | 32 | 9 | 3 | 12 | 5 |
| Physical: appetite | 4 | 4 | 14 | 3 | 695 | 6 | 7 | 2 | 8 | 5 | 2 | 6 | 1 |
| Physical: movement | 1 | 24 | 4 | 5 | 6 | 104 | 0 | 2 | 3 | 0 | 2 | 1 | 1 |
| Substance use | 3 | 0 | 2 | 1 | 7 | 0 | 185 | 0 | 4 | 0 | 0 | 0 | 2 |
| PHQ-9-suicidal tendencies | 7 | 2 | 1 | 6 | 2 | 2 | 0 | 294 | 11 | 12 | 1 | 3 | 0 |
| PHQ-9-emotional: depressed and sadness | 15 | 2 | 9 | 32 | 8 | 3 | 4 | 11 | 1612 | 32 | 12 | 14 | 20 |
| PHQ-9-emotional: worthlessness and worry | 24 | 9 | 10 | 9 | 5 | 0 | 0 | 12 | 32 | 1896 | 17 | 22 | 22 |
| Physical: pains | 4 | 2 | 6 | 3 | 2 | 2 | 0 | 1 | 12 | 17 | 377 | 1 | 5 |
| Social | 11 | 8 | 4 | 12 | 6 | 1 | 0 | 3 | 14 | 22 | 1 | 3555 | 17 |
| Mental health-related issues | 6 | 1 | 5 | 5 | 1 | 1 | 2 | 0 | 20 | 22 | 5 | 17 | 716 |

Table 1 shows the symptoms and the number of occurrences of being tagged together in an individual Tweet with other symptoms for all depression symptom annotated Tweets in data set 2.

Table 2: Summary of classifier configurations.

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier  | Binary Classifier (2 labels) | Multi-class classifier (6 labels) | Configuration |
| Naïve Bayes | Test size = 0.2 | Test size = 0.2 |   |
| Random Forest | Estimators = 600Criterion = Entropy | Estimators = 600Criterion = Entropy | Embedding layer word2vecTest size = 0.2 |
| Convolutional Neural Network (CNN) | ConvID (Filters = 100, Kernel size = 3)MaxPoolID (Pool size=2, strides=2)FlattenDropout = 0.5Dense = 10 unitsDropout = 0.52 units output (sigmoid)Loss = Binary cross entropy | ConvID (Filters = 100, Kernel size = 3)MaxPoolID (Pool size=2, strides=2)FlattenDropout = 0.5Dense = 10 unitsDropout = 0.56 units output (softmax)Loss = Categorical cross entropy | Batch 1 input layer: Embedding layer word2vecBatch 2 input layers: Embedding layer word2vec + user behavior features + linguistic featuresTest size = 0.2Epochs = 20Number of runs = 3Batch size = 128Optimizer = Adam |
| Bidirectional Long-Short Term Memory (LSTM) | Bidirectional LSTM units = 128Bidirectional LSTM units = 1282 units output (sigmoid)Loss = Binary crossentropy | Bidirectional LSTM units = 128Bidirectional LSTM units = 1286 units output (softmax)Loss = Categorical crossentropy |
| Bidirectional Long-Short Term Memory (LSTM w/ Dropout) | Bidirectional LSTM units = 128Dropout = 0.4Bidirectional LSTM units = 128Dropout = 0.42 units output (sigmoid)Loss = Binary cross entropy | Bidirectional LSTM units = 128Dropout = 0.4Bidirectional LSTM units = 128Dropout = 0.46 units output (softmax)Loss = Categorical cross entropy |
| Gated Recurrent Unit (GRU) | Bidirectional GRU units = 128Dropout = 0.5Bidirectional GRU units = 128Dropout = 0.52 units output (sigmoid)Loss = Binary cross entropy | Bidirectional GRU units = 128Dropout = 0.5Bidirectional GRU units = 128Dropout = 0.56 units output (softmax)Loss = Categorical cross entropy |

****

**Figure 1.** (a) Final model stage 1 loss and accuracy metrics. As the training run progresses, the validation accuracy is not improving, which means that 20 epochs is enough number of runs to get the highest validation accuracy. The training accuracy will improve, but the validation accuracy will not. We get the validation accuracy for this run at 0.9064 at epoch 5, which has the lowest validation loss. (b) Final model stage 2 loss and accuracy metrics. Same as with the stage 1 model, as the training run progresses, the validation accuracy is not improving, and we get the validation accuracy for this run at 0.8344 at epoch 5, with the lowest validation loss.