**Predicting Psychiatric Consultation in Inpatient Hospitals Using NLP-**

MILESTONE

Robert F. Arao

**THE PROBLEM:**

As the number of people with mental health problems increases, there appears a coinciding decrease in health-care resources. As a result, admitted inpatients may receive consultation with a psychiatrist/psychologist too late or not at all. The problem is that these patients need to be identified as early as possible and somehow flagged so that they may receive the necessary behavioral intervention as soon as possible.

**THE CLIENT**

The client are psychiatrists and psychologists and other mental health-care providers at inpatient hospitals.

**THE DATASET**

The dataset was created by extracting electronic health care records from two hospitals – University of Washington Medical Center (UWMC) and Harborview Medical Center (HMC). The data includes patients who are between the ages of 18 and 100 and admitted as inpatients between January 1, 2015 and September 20, 2017, and who were not deceased. The dataset was created using a quasi case-control study design. That is, cases were matched 1:1 using propensity score matching.

My final dataset consisted of 6,775 medical records with medical narratives (blob content) and consisted of 2,675 cases and 2,675 controls. Most of these text-based medical notes were already collated into a single blob content. That is, on average, each patient was linked to at least one blob content comprising the complete history of medical notes.

For wrangling component, I vertically merged the output excel files that was exported from SQL, deleted two missing records (controls that had empty blob content), checked for duplicates, change the date variable into datetime format, and changed the blob content from object to string.

**OTHER POTENTIAL DATASETS**

I tried using a 1:2 case:control propensity score matching, but this caused my computer to crash. So there are no other datasets I can use.

**INITIAL FINDINDINGS**

1. The propensity scores matched nicely (Figure 1). The differences before and after matching was assessed using chi square tests (Table 1).

Figure 1. Propensity score matching between cases and controls in unmatched and matched cohort

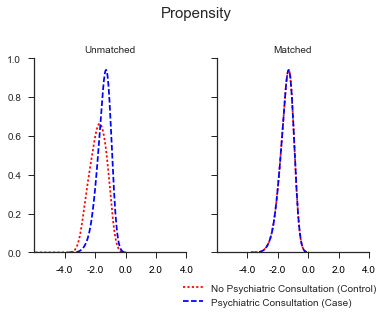


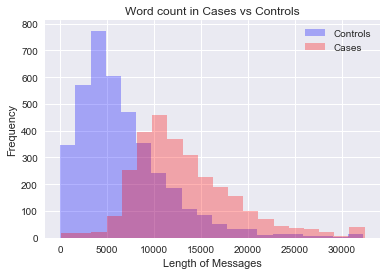
Table 1

Table 1. Comparison of patient characteristics before and after matching

|  | *Unmatched* | | |  | *Matched (1:1)* | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Characteristics* | *controls* | *cases* | *p-value* |  | *controls* | *cases* | *p-value* |
| Total | 105,763 | 2,767 | < 0.0001 |  | 2,767 | 2,767 | 1.0000 |
| UWMC | 46,451 | 451 |  |  | 451 | 451 |  |
| HMC | 59,312 | 2,316 |  |  | 2,316 | 2,316 |  |
|  |  |  |  |  |  |  |  |
| Gender |  |  |  |  |  |  |  |
| Female | 50,303 (47.6) | 1,090 (39.4) | < 0.0001 |  | 1,097 (39.6) | 1,090 (39.4) | 0.8474 |
| Male | 55,448 (52.4) | 1,677 (60.6) |  |  | 1,670 (60.4) | 1,677 (60.6) |  |
| Unknown | 12 (0.0) | 0 (0.0) |  |  | 0 (0.0) | 0 (0.0) |  |
|  |  |  |  |  |  |  |  |
| Race |  |  |  |  |  |  |  |
| White | 74,064 (70.0) | 2,006 (72.5) | < 0.0001 |  | 2,037 (73.6) | 2,006 (72.5) | 0.5360 |
| Black or African American | 13,098 (12.4) | 367 (13.3) |  |  | 372 (13.4) | 367 (13.3) |  |
| Asian | 8,441 (8.0) | 155 (5.6) |  |  | 158 (5.7) | 155 (5.6) |  |
| Native Hawaiian or Other | 1,121 (1.1) | 24 (0.9) |  |  | 14 (0.5) | 24 (0.9) |  |
| America Indian or Alaskan | 2,698 (2.6) | 87 (3.1) |  |  | 76 (2.7) | 87 (3.1) |  |
| Multiple races | 1,124 (1.1) | 63 (2.3) |  |  | 53 (1.9) | 63 (2.3) |  |
| Unknown | 5,217 (4.9) | 65 (2.3) |  |  | 57 (2.1) | 65 (2.3) |  |
|  |  |  |  |  |  |  |  |
| Primary Language Spoken |  |  |  |  |  |  |  |
| English | 94,469 (89.3) | 2,605 (94.1) | < 0.0001 |  | 2,609 (94.3) | 2,605 (94.1) | 0.6074 |
| Spanish | 4,306 (4.1) | 50 (1.8) |  |  | 54 (2.0) | 50 (1.8) |  |
| Other | 6,639 (6.3) | 105 (3.8) |  |  | 101 (3.7) | 105 (3.8) |  |
| Unknown | 349 (0.3) | 7 (0.3) |  |  | 3 (0.1) | 7 (0.3) |  |
|  |  |  |  |  |  |  |  |
| Veteran Status |  |  |  |  |  |  |  |
| Yes | 8,142 (7.7) | 230 (8.3) | < 0.0001 |  | 216 (7.8) | 230 (8.3) | 0.5358 |
| No | 90,263 (85.3) | 2,402 (86.8) |  |  | 2,429 (87.8) | 2,402 (86.8) |  |
| Unknown | 7,358 (7.0) | 135 (4.9) |  |  | 122 (4.4) | 135 (4.9) |  |
|  |  |  |  |  |  |  |  |
| Employment status |  |  |  |  |  |  |  |
| Full-time | 30,508 (28.8) | 292 (10.6) | < 0.0001 |  | 297 (10.7) | 292 (10.6) | 0.9854 |
| Part-time | 5,026 (4.8) | 78 (2.8) |  |  | 78 (2.8) | 78 (2.8) |  |
| Self-Employee | 3,720 (3.5) | 50 (1.8) |  |  | 49 (1.8) | 50 (1.8) |  |
| Not Employed | 41,143 (38.9) | 1,862 (67.3) |  |  | 1,854 (67.0) | 1,862 (67.3) |  |
| Student | 3,120 (2.9) | 36 (1.3) |  |  | 39 (1.4) | 36 (1.3) |  |
| Active Duty | 79 (0.1) | 1 (0.0) |  |  | 0 (0.0) | 1 (0.0) |  |
| Retired | 18,708 (17.7) | 362 (13.1) |  |  | 369 (13.3) | 362 (13.1) |  |
| Unknown | 3,459 (3.3) | 86 (3.1) |  |  | 81 (2.9) | 86 (3.1) |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Year of Admission |  |  |  |  |  |  |  |
| 2015 | 38,513 (36.4) | 1,028 (37.2) | 0.0020 |  | 1,045 (37.8) | 1,028 (37.2) | 0.8492 |
| 2016 | 39,898 (37.7) | 960 (34.7) |  |  | 960 (34.7) | 960 (34.7) |  |
| 2017 | 27,352 (25.9) | 779 (28.2) |  |  | 762 (27.5) | 779 (28.2) |  |
|  |  |  |  |  |  |  |  |
| Age, mean (std) | 49.2 (18.0) | 47.4 (16.4) | < 0.0001 |  | 47.6 (16.5) | 47.4 (16.4) | 0.7095 |

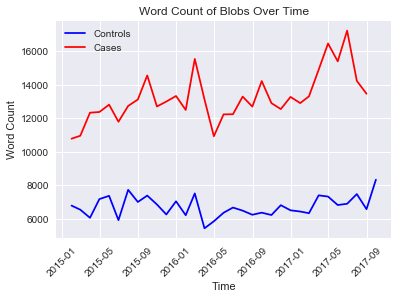
1. A histogram of the word count between cases and controls showed a unique area where cases uses more words in their blob contents (Figure 2). This supports the idea that there are certain words that may be better predictors of psychiatric consultation. The mean word count per message among cases (13,317 (95%CI: 13,111, 13,521)) was significantly higher than that among controls (6,679 (95%CI: 6,525, 6,832)).

Figure 2. Distribution of word count in cases and controls



1. There were no observable time trends among cases nor controls (Figure 3).

Figure 3. Distribution of word count in cases and controls over time



1. I created a single-word model using the multinomial Naïve Bayes classifier algorithm. After fitting the model to the training (Table 2) and test dataset (Table 3) in the single-word model, the overall accuracy of the training-fitted model was slightly higher than the test-fitted model (88.5% vs.87.6%). The suggests the model was not overfitted. In identifying single words, the overall precision was found to be 90% with a recall rate of 88%. The F1-score was also fairly high (88%) and approximates the recall rate.

Table 3 Classification report of single-word model prediction on training data

precision recall f1-score support

0 1.00 0.77 0.87 2754

1 0.76 1.00 0.86 1988

avg / total 0.90 0.87 0.87 4742

Accuracy=0.885280

Table 4. Classification report of single-word model prediction on test data

precision recall f1-score support

0 0.99 0.80 0.88 1207

1 0.77 0.99 0.87 826

avg / total 0.90 0.88 0.88 2033

Accuracy=0.876045

**NEED TO DO:**

1. Improve model using k fold cross validation
2. Variations of the model such as using n-grams or Random Forrest.
3. Pull out top 10 words
4. Predict a priori list that psychiatrists provided.