Mitchell Izower, Jayson Marwaha, Lily Payvandi

Group 2

Lab 6

10/18/2020

Error Analysis

**Parser V1 Lab 5**

FalsePositive FEVER 83

FalsePositive ABDOMINAL PAIN 72

FalsePositive SHORTNESS OF BREATH 60

FalsePositive DYSPNEA 49

FalsePositive SOB - SHORTNESS OF BREATH 44

FalsePositive COUGH 39

FalsePositive CHEST PAIN 38

FalsePositive CP - CHEST PAIN 33

FalsePositive DIARRHEA 33

FalsePositive CONFUSION 21

FalseNegative FLU-LIKE SYMPTOMS 5

FalseNegative BACK PAIN 4

FalseNegative HYPOXEMIA 4

FalseNegative STOMACH ACHE 4

FalseNegative ALTERED MENTAL STATUS 3

FalseNegative WEAKNESS 3

FalseNegative ABD PAIN 2

FalseNegative ABDOMINAL MASS/DISTENTION 2

FalseNegative ABNORMAL LABS 2

FalseNegative HYPOXIA 2

**Parser V2 Lab 6**

FalsePositive ABD PAIN 105

FalsePositive FEVER 83

FalsePositive STOMACH ACHE 77

FalsePositive ABDOMINAL PAIN 72

FalsePositive SHORTNESS OF BREATH 60

FalsePositive DYSPNEA 49

FalsePositive SOB - SHORTNESS OF BREATH 44

FalsePositive COUGH 39

FalsePositive AMS - ALTERED MENTAL STATUS 38

FalsePositive CHEST PAIN 38

FalseNegative BACK PAIN 4

FalseNegative WEAKNESS 3

FalseNegative ABDOMINAL MASS/DISTENTION 2

FalseNegative ABNORMAL LABS 2

FalseNegative LETHARGY 2

FalseNegative ABCESS 1

FalseNegative ABN CT SCAN 1

FalseNegative ABNL LABS 1

FalseNegative ABNORMAL COMPUTED TOMOGRAPHY 1

FalseNegative B LEG PAIN 1

**Rationale**

The methodology that was employed to perform this lab was as follows. Raw data from the supplied CSV file was uploaded into OpenRefine, which was then clustered using NLP. Where NLP failed to cluster, manually clustering was employed. The goal of this was to produce 3 classes of clusters. Clusters marked with a “0” would be clinically designated by our team as screening negative, clusters marked with a “1” would be clinically designated by our team as screening positive, clusters marked with a “2” would be clinically designated by our team as being indeterminate. We then pulled the terms that those clusters were composed of to produce a filter that our data set was passed through, and applied the appropriate screening. For indeterminate (“2”) terms, we decided to screen them negative.

**Performance Optimization**

Our approach to designing this tool was to optimize sensitivity. Since this was designed to be a screening tool, it should theoretically prioritize sensitivity over other markers of performance since it is less harmful to overcapture positive cases (i.e., it is less harmful to have more false positives than false negatives). False negatives are more harmful in this scenario because a patient who is truly positive but, for example, placed in a “screened negative” triage tent then may contaminate other truly negative patients. The way we prioritized sensitivity in building our tool was by broadening our definition of positive chief complaints as much as possible to include as many synonyms and misspellings of various respiratory and gastrointestinal complaints as possible, with the intention to screen positive anyone who had any sort of respiratory or GI complaint. Additionally, we erred on the side of caution for chief complaints like “fever” and “abdominal pain” that contained several true negatives in the dataset by labelling them as positive as well. While this again may have resulted in a high number of false positives, it is the safer option and more appropriately fills the role of the screening tool in casting a wide net.