# Frequency Analysis of the Injury Reports

Abdullah Alsharef

3/20/2020

## Contents

# 1 Introduction

The goal of the this document is to prepare the Injury reports for ML/NLP analysis. This document shows the frequency of the targeted variables or attributes.

Figure 1 shows an overview of suggested process.

We are attempting to test the concept of predicting the attributes of injury reports from the injury's narratives (The textual unstructured description). This is a supervised learning problem. The attributes of the injury report are:

- Event
- Source
- Amputation
- Body Part
- Contractor
- Hospitalization
- Nature

The injury narrative's is the predictor (x variable) and the attributes (listed above) are the (Y variable).

In this folder, there are separate sub folders with the name of the attributes. Each sub folder has a CSV file. Each CSV have 3 columns (ID Number - Narratives - The labeled attribute). The number of reports in each folder is different because we removed all unlabeled data.

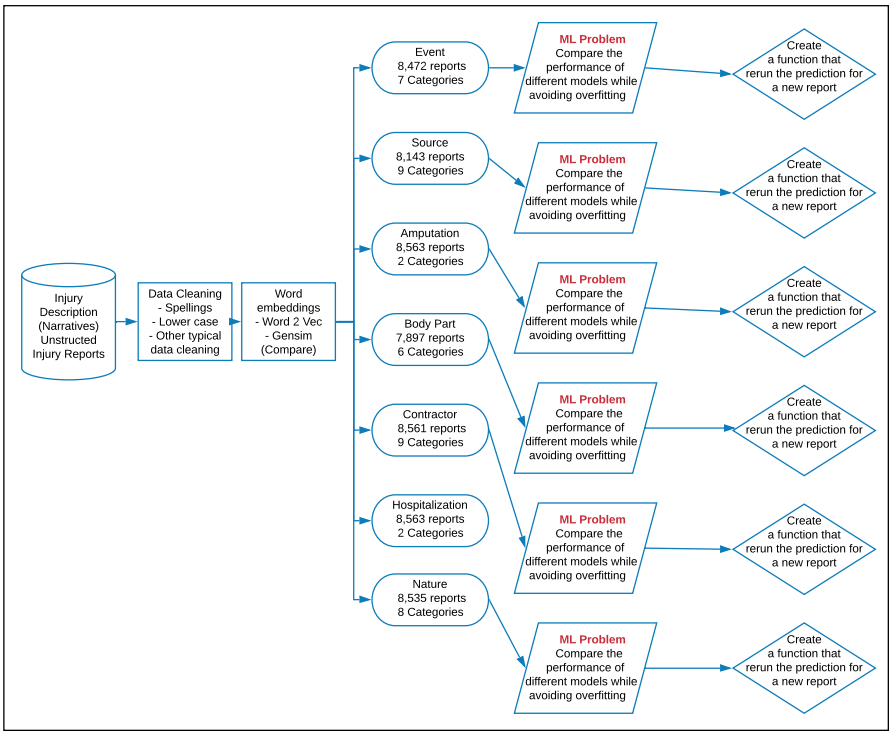


Figure 1: NLP/ML Overview

# 2 Frequency Analysis

## 2.1 Event

The number of reports (after excludeing unlabeled reports) is **8,472** reports. The number of categories for *Event* is **7**.

| Event | n | Event_p |
|---|---|---|
| falls, slips, trips | 3422 | 40.3918791 |
| contact with objects and equipment | 3388 | 39.9905571 |
| exposure to harmful substances or environments | 851 | 10.0448536 |
| transportation incidents | 546 | 6.4447592 |
| fires and explosions | 138 | 1.6288952 |
| overexertion and bodily reaction | 75 | 0.8852691 |
| violence and other injuries by persons or animals | 52 | 0.6137866 |

## 2.2 Source

The number of reports (after excludeing unlabeled reports) is **8,143** reports. The number of categories for Source is **9**.

| Source | n | Source_p |
|---|---|---|
| structures and surfaces | 2199 | 27.004789 |
| parts and materials | 1516 | 18.617217 |
| tools, instruments, and equipment | 1479 | 18.162839 |
| machinery | 1224 | 15.031315 |
| vehicles | 724 | 8.891072 |
| other sources | 462 | 5.673585 |
| containers, furniture and fixtures | 204 | 2.505219 |
| chemicals and chemical products | 176 | 2.161366 |
| persons, plants, animals, and minerals | 159 | 1.952597 |

## 2.3 Amputation

The number of reports (after excludeing unlabeled reports) is **8,563** reports. This is similar to sentiment analysis in which the Y variable (the response) is either "Yes" or "No".

| Amputation_Y_N | n | Amputation_p |
|---|---|---|
| N | 7265 | 84.84176 |
| Y | 1298 | 15.15824 |

## 2.4 Body Part

The number of reports (after excludeing unlabeled reports) is **7,897** reports. The number of categories for Source is **6**.

| Part_v2 | n | Part_v2_p |
|---|---|---|
| upper extremities | 2492 | 31.556287 |
| lower extremities | 1761 | 22.299607 |
| trunk | 1219 | 15.436242 |
| multiple body parts | 1145 | 14.499177 |
| head and neck | 801 | 10.143092 |
| body systems | 479 | 6.065594 |

## 2.5 Contractor

The number of reports (after excludeing unlabeled reports) is **8,561** reports. The number of categories for Source is **9**.

| Contractor | n | Contractor_p |
|---|---|---|
| Foundation, Structure, and Building Exterior Contractors | 2105 | 24.588249 |
| Building Equipment Contractors | 1625 | 18.981427 |
| Nonresidential Building Construction | 1266 | 14.787992 |
| Utility System Construction | 878 | 10.255811 |
| Other Specialty Trade Contractors | 752 | 8.784021 |
| Highway, Street, and Bridge Construction | 703 | 8.211657 |
| Building Finishing Contractors | 611 | 7.137017 |
| Residential Building Construction | 364 | 4.251840 |
| Other Heavy and Civil Engineering Construction | 257 | 3.001986 |

## 2.6 Hospitalization

The number of reports (after excludeing unlabeled reports) is **8,563** reports. This is similar to sentiment analysis in which the Y variable is either "Yes" or "No".

| Hospitalized_Y_N | n | Hospitalized_p |
|---|---|---|
| Y | 7652 | 89.36121 |
| N | 911 | 10.63879 |

## 2.7 Nature

The number of reports (after excludeing unlabeled reports) is **8,535** reports. The number of categories for Source is **8**.

| Nature | n | Nature_p |
|---|---|---|
| Traumatic injuries to bones, nerves, spinal cord | 3117 | 36.520211 |
| Open wounds | 2166 | 25.377856 |
| Other traumatic injuries and disorders | 1662 | 19.472759 |
| Burns and corrosions | 533 | 6.244874 |
| Intracranial injuries | 325 | 3.807850 |
| Multiple traumatic injuries and disorders | 318 | 3.725835 |
| Effects of environmental conditions | 288 | 3.374341 |
| Traumatic injuries to muscles, tendons, ligaments, joints, etc. | 126 | 1.476274 |