

# Car Accident Severity

Devansh Goel

September 7, 2020

## Introduction

### 1.1 Background

Road accidents are a serious concern for the majority of nations around the world because accidents can cause severe injuries and fatalities. According to the World Health Organization's Global Status Report, approximately 1.25 million people deaths happened per year are because of road accident injuries, and most fatality rates were in lower income countries. Our motivation is to predict the accident severity of any road, which will play a crucial factor for traffic control authorities to take proactive precautionary measures.

### 1.2 Problem

The purpose of this project is to predict the severity of an accident by training an efficient machine learning model with the help of existing accidents data. This project is majorly focused on predicting rarer classes accurately such as Serious and Fatal.

## Data

### 1.1 Data Understanding

The data consists of 37 independent variables and 194,673 rows. The dependent variable, "SEVERITYCODE", contains numbers that correspond to different levels of severity caused by an accident from 1 to 2.

### 1.2 Extract Dataset and Convert

In its original form, this data is not fit for analysis. For one, there are many columns that we will not use for this model. Also, most of the features are of type object, when they should be numerical type.

We must use label encoding to covert the features to our desired data type.

With the new columns, we can now use this data in our analysis and ML models!

### **1.3 Balancing the dataset**

Our target variable SEVERITYCODE is only 42% balanced. In fact, severity code in class 1 is nearly three times the size of class 2.

We can fix this by down sampling the majority class.