Rail accident prediction

Cian Monnin

Abstract – Capstone project proposal

1 Background and Motivation

In light of the catastrophic derailment in East Palastine, Ohio, I became interested in rail infrastructure in the USA and Canada. This led to me want model railway accidents in Canada using data provided by the Canadian Government. Canada has over 46,000 km of railroads. The rail freight industry in Canada generates approx $9.5 billion per annum. Canadian National Railway (CN) and Canadian Pacific Railway (CPR) are the two largest operators accounting for 50% and 35% of the revenue per year, respectively. CN crosses Canada and links customers in the U.S. and Mexico while CPR operates 22,500 km in six provinces and 13 States.[1] This is a massive industry in Canada and anyway to improve the safety should be examined. Figure 1 shows that while the trend of accidents over the past ten years remains linear if the last two years were excluded and increasing trend would be observed. The second image shows that while accidents remain roughly flat the number of accidents per million miles has nearly doubled from ten years ago. Thankfully fatalities have been showing a downward trend over the last ten years.

2 Methods and Approach

I plan to use simple logistic regression to produced an interpretable logit model. Then I plan to use more deep learning techniques and compare this to the logit model with the goal of increasing accuracy. Initial features for consideration include:

- season
- time of day
- rail traffic
- rail crossing
- different segments of rail
- track owner

I expect to find a lot more features as data exploration begins.

I plan to get data provided by Transportation Safety Board of Canada and Transport Canada. Initial data-sets will include but not limited to:

- Rail transportation safety investigations and reports
- Rail transportation occurrences in 2021
- Grade Crossings Inventory
- National Railway Network - NRWN - GeoBase Series

[1] https://tc.canada.ca/en/corporate-services/policies/rail-transportation

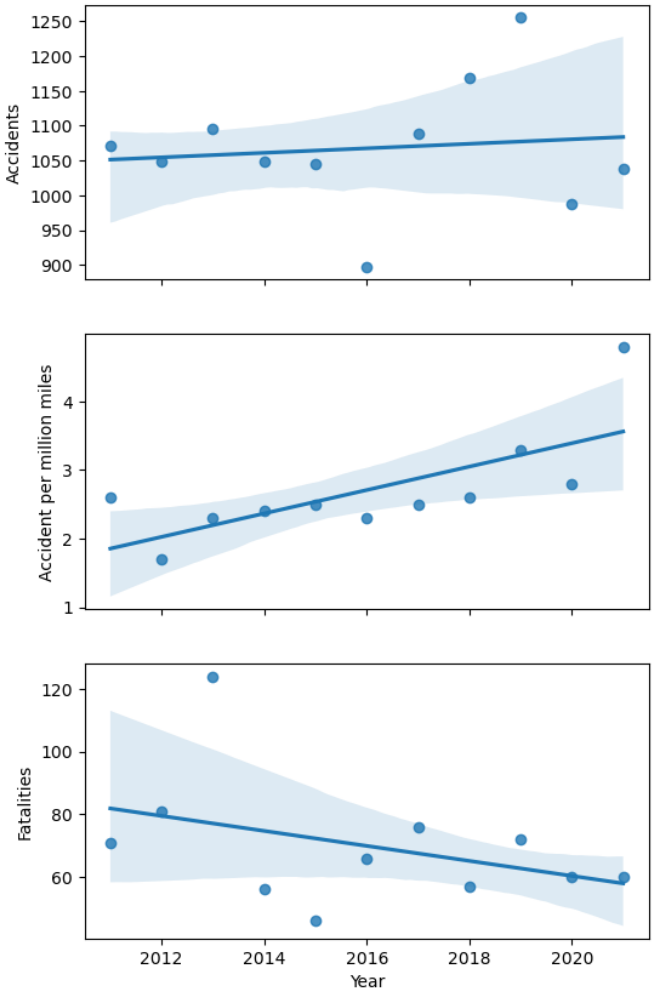

Figure 1: Changes in accidents in the last 10 years

3 Goals

- Create models both traditional logit models and deep learning models for comparison
- Potentially an SQL database to practice accessing and utilizing SQL
- A dashboard for reporting findings and displaying interesting trends or statistics discovered