Colorado Car Accident Severity Prediction

By

Joanna Johnson

Applied Data Science Capstone

Course 9 of IBM Data Science Professional Certificate

Table of Contents

[Abstract 2](#_Toc53562659)

[Introduction / business problem 3](#_Toc53562660)

[Data 4](#_Toc53562661)

[Methodology 7](#_Toc53562662)

[Results 8](#_Toc53562663)

[acknowledgement 9](#_Toc53562664)

# Abstract

Explain in 3 paragraphs or less

# Introduction / business problem

Here’s the scenario, you are driving to work. It’s cloudy and windy, the roads are wet and slushy from last night’s snow. Traffic is slow as expected but it’s not too bad (you already left home earlier anticipating slow traffic), until you get on the highway. Long lines of cars barely moving, police cars and ambulances passing by. You turn on the radio, there’s an accident happening about 10 miles ahead of you. Now your navigation system shows the highway segments ahead of you all jammed and marked in dark red. You are wondering if you should get off the highway at next exit to take an alternative route (longer way to work, it usually takes more time, but probably faster than the jammed highway now). But what if the accident ahead of you is not that bad after all and will be cleared soon?

Wouldn’t it be great if there’s something in place that could warn you, given the weather, road conditions and other information about the severity of the accident, so you can make an informed decision?

That’s exactly what I’m going to work on in this project. It will help all drivers, as well as companies who provide navigation services to make better informed decisions in this situation. Also redirect traffic from the scene of the accident to minimize its impact to the traffic.

# Data

The data I’m going to use for this project is called Colorado-Accidents, a subset of US-Accidents dataset, which covers Colorado traffic events data from February 2016 to June 2020. There are **49,731** accident records in this dataset. It’s in csv format. Below is the table of attributes.

| **#** | **Attribute** | **Description** | **Nullable** |
| --- | --- | --- | --- |
| 1 | Rush\_Hour | Indicates if the accident happens in rush hour (1-yes, 0-no). It’s derived from Start\_Time. Rush hours are 6:00-10:00 and 15:00-19:00. | No |
| 2 | Duration\_in\_Hour | Shows how long (in hours) the impact on traffic caused by the accident. It’s calculated as (End\_Time – Start\_Time) | No |
| 3 | Interstate | Shows if the accident happens on an interstate highway (1-yes, 0-no). It’s derived from Street attribute. | No |
| 4 | ID | This is a unique identifier of the accident record. | No |
| 5 | Source | Indicates source of the accident report (i.e. the API which reported the accident.). | No |
| 6 | TMC | A traffic accident may have a [Traffic Message Channel (TMC)](https://wiki.openstreetmap.org/wiki/TMC/Event_Code_List) code which provides more detailed description of the event. | Yes |
| 7 | Severity | Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay). | No |
| 8 | Start\_Time | Shows start time of the accident in local time zone. | No |
| 9 | End\_Time | Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed. | No |
| 10 | Start\_Lat | Shows latitude in GPS coordinate of the start point. | No |
| 11 | Start\_Lng | Shows longitude in GPS coordinate of the start point. | No |
| 12 | End\_Lat | Shows latitude in GPS coordinate of the end point. | Yes |
| 13 | End\_Lng | Shows longitude in GPS coordinate of the end point. | Yes |
| 14 | Distance(mi) | The length of the road extent affected by the accident. | No |
| 15 | Description | Shows natural language description of the accident. | No |
| 16 | Number | Shows the street number in address field. | Yes |
| 17 | Street | Shows the street name in address field. | Yes |
| 18 | Side | Shows the relative side of the street (Right/Left) in address field. | Yes |
| 19 | City | Shows the city in address field. | Yes |
| 20 | County | Shows the county in address field. | Yes |
| 21 | State | Shows the state in address field. | Yes |
| 22 | Zipcode | Shows the zipcode in address field. | Yes |
| 23 | Country | Shows the country in address field. | Yes |
| 24 | Timezone | Shows timezone based on the location of the accident (eastern, central, etc.). | Yes |
| 25 | Airport\_Code | Denotes an airport-based weather station which is the closest one to location of the accident. | Yes |
| 26 | Weather\_Timestamp | Shows the time-stamp of weather observation record (in local time). | Yes |
| 27 | Temperature(F) | Shows the temperature (in Fahrenheit). | Yes |
| 28 | Wind\_Chill(F) | Shows the wind chill (in Fahrenheit). | Yes |
| 29 | Humidity(%) | Shows the humidity (in percentage). | Yes |
| 30 | Pressure(in) | Shows the air pressure (in inches). | Yes |
| 31 | Visibility(mi) | Shows visibility (in miles). | Yes |
| 32 | Wind\_Direction | Shows wind direction. | Yes |
| 33 | Wind\_Speed(mph) | Shows wind speed (in miles per hour). | Yes |
| 34 | Precipitation(in) | Shows precipitation amount in inches, if there is any. | Yes |
| 35 | Weather\_Condition | Shows the weather condition (rain, snow, thunderstorm, fog, etc.) | Yes |
| 36 | Amenity | A [POI](https://wiki.openstreetmap.org/wiki/Points_of_interest) annotation which indicates presence of [amenity](https://wiki.openstreetmap.org/wiki/Key:amenity) in a nearby location. | No |
| 37 | Bump | A POI annotation which indicates presence of speed bump or hump in a nearby location. | No |
| 38 | Crossing | A POI annotation which indicates presence of [crossing](https://wiki.openstreetmap.org/wiki/Key:crossing) in a nearby location. | No |
| 39 | Give\_Way | A POI annotation which indicates presence of [give\_way](https://wiki.openstreetmap.org/wiki/Tag:highway%3Dgive_way) in a nearby location. | No |
| 40 | Junction | A POI annotation which indicates presence of [junction](https://wiki.openstreetmap.org/wiki/Key:junction) in a nearby location. | No |
| 41 | No\_Exit | A POI annotation which indicates presence of [no\_exit](https://wiki.openstreetmap.org/wiki/Key:noexit) in a nearby location. | No |
| 42 | Railway | A POI annotation which indicates presence of [railway](https://wiki.openstreetmap.org/wiki/Key:railway) in a nearby location. | No |
| 43 | Roundabout | A POI annotation which indicates presence of [roundabout](https://wiki.openstreetmap.org/wiki/Tag:junction%3Droundabout) in a nearby location. | No |
| 44 | Station | A POI annotation which indicates presence of [station](https://wiki.openstreetmap.org/wiki/Key:station) in a nearby location. | No |
| 45 | Stop | A POI annotation which indicates presence of [stop](https://wiki.openstreetmap.org/wiki/Key:stop) in a nearby location. | No |
| 46 | Traffic\_Calming | A POI annotation which indicates presence of [traffic\_calming](https://wiki.openstreetmap.org/wiki/Key:traffic_calming) in a nearby location. | No |
| 47 | Traffic\_Signal | A POI annotation which indicates presence of [traffic\_signal](https://wiki.openstreetmap.org/wiki/Tag:highway%3Dtraffic_signals) in a nearby location. | No |
| 48 | Turning\_Loop | A POI annotation which indicates presence of [turning\_loop](https://wiki.openstreetmap.org/wiki/Tag:highway%3Dturning_loop) in a nearby location. | No |
| 49 | Sunrise\_Sunset | Shows the period of day (i.e. day or night) based on sunrise/sunset. | Yes |
| 50 | Civil\_Twilight | Shows the period of day (i.e. day or night) based on [civil twilight](https://en.wikipedia.org/wiki/Twilight#Civil_twilight). | Yes |
| 51 | Nautical\_Twilight | Shows the period of day (i.e. day or night) based on [nautical twilight](https://en.wikipedia.org/wiki/Twilight#Nautical_twilight). | Yes |
| 52 | Astronomical\_Twilight | Shows the period of day (i.e. day or night) based on [astronomical twilight](https://en.wikipedia.org/wiki/Twilight#Astronomical_twilight). | Yes |

I will use the traffic events (e.g., accident, lane closed, congestion, etc), weather (e.g., visibility, wind speed, weather condition), and time (e.g., day of week, hour of day, and period of day) to build a machine learning model for live prediction of car accidents’ severity level (1-4) in terms of impact on traffic, level 1 indicates the least impact on traffic (e.g., short delay), level 4 indicates a significant impact on traffic (e.g., long delay).

# Methodology

Explain in detail.

# Results

Will be added later

# acknowledgement

* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. [“A Countrywide Traffic Accident Dataset.”](https://arxiv.org/abs/1906.05409), arXiv preprint arXiv:1906.05409 (2019).
* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. [“Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.”](https://arxiv.org/abs/1909.09638)