7/8/2020

Road Accidents in Canada Explored (Project RACE)

Project Proposal

Tolu Fatoki & Syed Tasrif Ahmed

Team Member And Tasks

	Team Member	Task
1.	Tolu Fatoki	• Data Analysis and Exploration with python • Build machine learning model • Testing • Project Documentation
2.	Syed Tasrif Ahmed	• Web development • Build interactive visuals with chart.js • Build Machine learning model • Testing • Documentation

Introduction

Transportation is the process of moving from one location to another for various reasons, from business, personal to leisure. Road transportation is one of the most popular and reliable means of transportation. Road transport has become an important part of our everyday life, we commute via road to go to work, school, and leisure. Roads also play an important role in the economic development of any country and serve as a backbone for any strong economy and as such, it is very important to keep transporting on these roads safe and secure for everyday users. No wonder the Canadian Council of Motor Transport Administration (CCMTA) in its 2016 publication developed the Road Safety Strategy(RSS) 2025 with the aim to make Canadian roads one of the safest in the world (CCTMA, 2016). One of the strategies they are going to employ to achieve this is “improving the safety of vehicles and road infrastructure”. Having better and improved road infrastructure is key to ensuring road safety and reducing road accidents and in this work, we will be exploring how road configurations have contributed to the number of road accidents and how collision hotspots can be predicted based on available and help the government make right decisions in improving infrastructure in our roads.

In the spring of 2018, a now very popular road accident occurred near Armley, Saskatchewan claiming 16 lives and 13 fatal injuries (The Canadian Encyclopedia, 2018). The accident occurred at the intersection of Saskatchewan Highways 35 and 335 (Figure 1) between a semi-trailer truck and a coach bus carrying players of a junior hockey team from Humboldt, Saskatchewan. The main cause for this accident has been generally agreed to be the semi-truck driver’s failure

to yield at a flashing stop sign at the intersection. According to this report by (CBC News, 2018) a similar accident occurred on this same intersection about 2 decades before this particular Humboldt broncos incident which prompted the installation of the stop sign to prevent future occurrence of any form accidents at the intersection. It is reported that accidents at this intersection are even lower when compared to other intersections in the Saskatchewan and even in Canada as a whole as this intersection does not even feature on the list of the top 20 deadliest intersections in Canada according to this report (Markham Mitsubishi, n.d.).

The intersection on any road is just one of the many types of road configuration where accidents occur and just one of the many other reasons why accidents happen around the world. However, our work will focus on accidents that happen as a result of various road configurations as we view this cause of accidents as a major contributing factor to road accidents. Our dataset shows that road configuration contributes about 30% of the total road accidents that occur in Canada and we believe that tackling this particular issue will contribute immensely to reducing road accidents and help the government in reaching its vision of tending towards zero road accidents.

Figure 1: Aerial view of Highway 35 and 335 intersection, Armley, Saskatchewan

Source: CBC News, 2018

We have carefully titled our **project RACE** to emphasize the fact that driving is not a race against time or competition with other drivers or a Hollywood movie scene but a serious deal that requires close attention at all times and following and obeying the different instructions that have been marked out on our roads and that involves actual people's life.

Problem Statement

The challenge of road accidents has been an ever-present one all over the world and especially in Canada. According to WHO's 2018 report about 1.35 million people die each year from road accidents all over the world and this number is always increasing as many governments are not taking the required steps to reduce and curb the present challenges of road accidents. The Canadian government is committed to ensuring the safety of lives and properties in its road and has made different efforts to ensure that the number of accidents on our roads is reduced significantly. This commitment can be seen in the gradual reduction of accidents and injuries sustained over the years on our roads. We hope that with this research, we can contribute to the continued efforts of the government to reduce the number of deaths on our roads.

Several factors are always responsible for different road accidents and different categories of these factors can be identified from issues like the weather condition, road configuration, age of the driver, mental state of the driver, road surface, traffic control, time of the day, the period of the year, type and age of the vehicle, usage of recommended safety equipment. As all these factors need to be researched to get a holistic view of reducing the problem of road accidents, our main focus for this project will be on the accidents caused by road configuration which we observed has received little or no attention as no mention of any analysis done on it was available on the government of Canada website and no resources were found on the internet as regarding this crucial composition of our daily commuting. Road configuration as the name implies refers to the configuration of the road, that is how the road is built for commuting. Examples of road configuration which we will explore based on our dataset include intersection road(like the Humboldt case-study), railroad level crossing, bridges, overpass, viaduct, tunnel, underpass, ramp e.t.c.

How a road is constructed might seem like a negligible detail when it comes to road accidents but like in the case study presented above our project will highlight how road configuration has had a huge effect on the number of road accidents and how having a good understanding of collision hotspots around the country and what measures can be put in place to reduce collision in those types of roads.

Solution Overview

In this project, we propose a machine learning model that can use the ever-changing dataset of road accidents in Canada and predict the potential hotspots where we have frequent accidents and with this model, we believe government agencies can enforce laws and infrastructures in these different hotspots that will help all road users to reduce collision and possible death. In addition to this, our model will consider two other factors which are the time of the year(month/season) and weather conditions, with this new information and the road condition so that we can provide a more accurate understanding of how the period of the year, weather and the road configuration can lead to an accident, to further inform the concerned officers of changes that might be required at different road types with changing seasons. A model like this one could have been handy with the case study presented above because it will have provided prompt information on the dangers of an intersection and further investigation could have

helped to make some changes to rules governing motorists at these possible hotspots and probably save lives.

Our proposed solution (shown in the flowchart below)will provide a web application that will contain the working predictive model and also an exploratory analysis of the dataset. We will provide a detailed analysis of the data to give users of the website a good understanding of the various causes of accidents, how they affect the total number, and the different distributions over the years.

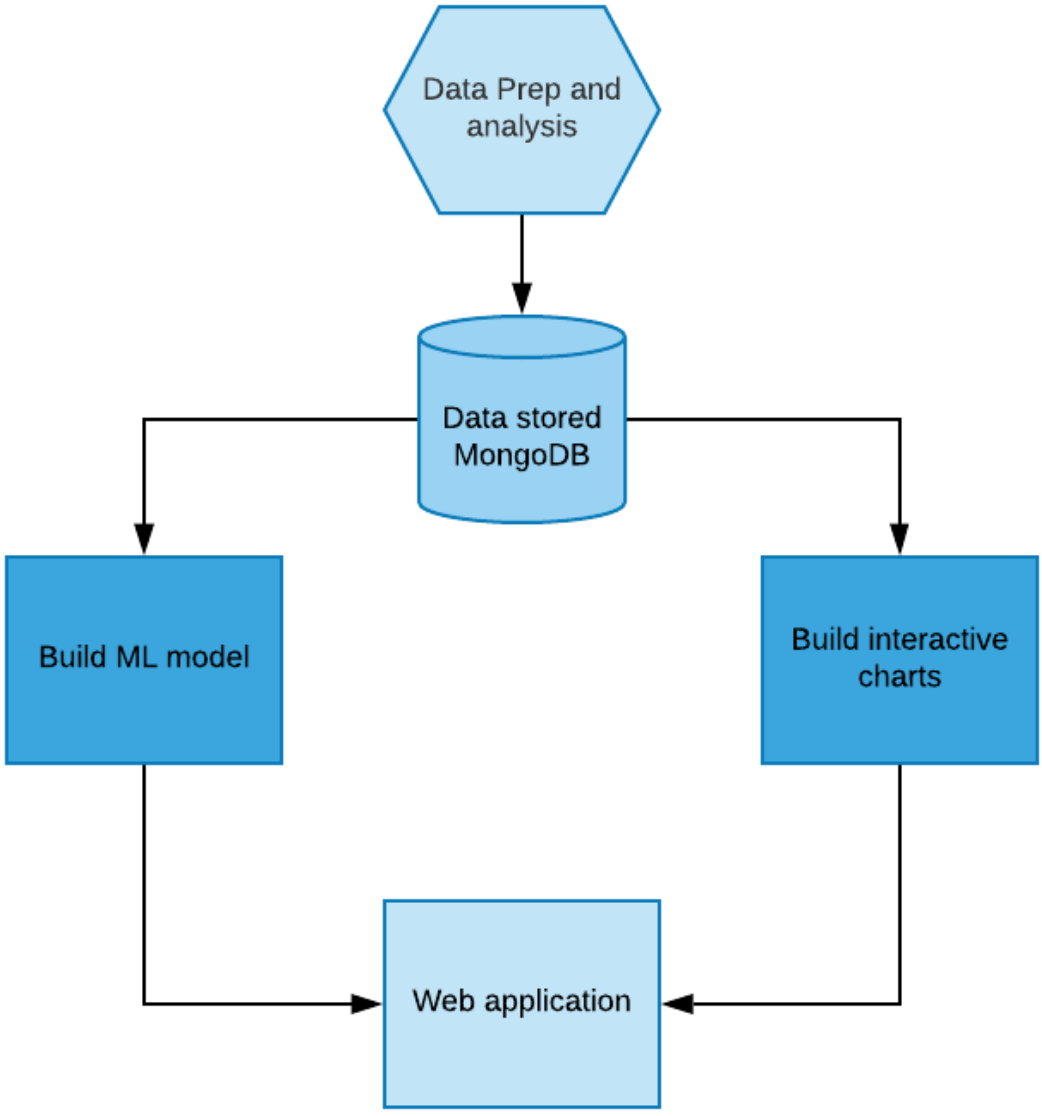

Figure 2: A flow chart showing our proposed solution

We are certain that this model will be useful for Transport Canada and various policy makers and stakeholders in the transportation industry to develop interventions aimed at reducing road accidents in Canada.

Data

The dataset we are considering is provided by the government of Canada through the National Collision Database (NCDB) and can be found here. The data dating back as far as 1999 contains all police-reported motor vehicle collisions on public roads in Canada The originally retrieved data is stored in CSV format and it contains about 6.5 million columns and 23 rows(Figure 4). A detailed description of the columns can be seen in the data dictionary provided with the dataset. The data appears to be in a structured format but a closer look shows that we have a lot of null and empty values in different columns with different irregularities, this implies that a lot of analysis and transformation are required to get the data in our desired form. We have 20 categorical columns that can be used in this analysis to categorize the dataset and 3 numerical columns which will provide the required figures for the different categories in this analysis.

We plan to store this dataset using MongoDB where we can easily extract the data for both descriptive and predictive analysis.

```
nRow, nCol = accident.shape
head = accident.head(5)
print(head)
print(f'There is a total of {nRow} rows and {nCol} columns')
```

```
   c_year c_mnth c_wday c_hour  c_sev c_vehs c_conf c_rcfg c_wthr c_rsur c_raln c_traf v_id v_type v_year p_id p_sex p_age p_psn p_isev p_safe p_user  c_case
0    1999      1      1     20      2     02     34     UU      1      5      3     03   01     06   1990   01     M    41    11      1     UU      1     752
1    1999      1      1     20      2     02     34     UU      1      5      3     03   02     01   1987   01     M    19    11      1     UU      1     752
2    1999      1      1     20      2     02     34     UU      1      5      3     03   02     01   1987   02     F    20    13      2     02      2     752
3    1999      1      1     08      2     01     01     UU      5      3      6     18   01     01   1986   01     M    46    11      1     UU      1     753
4    1999      1      1     08      2     01     01     UU      5      3      6     18   99     NN   NNNN   01     M    05    99      2     UU      3     753
There is a total of 6772563 rows and 23 columns
```

Figure 3: Dataset sample and the total number of columns and rows

```
accident.describe()
```

	c_year	c_sev	c_case
count	6.772563e+06	6.772563e+06	6.772563e+06
mean	2.007342e+03	1.983390e+00	1.273550e+06
std	5.453422e+00	1.278037e-01	7.423937e+05
min	1.999000e+03	1.000000e+00	1.000000e+00
25%	2.003000e+03	2.000000e+00	6.263120e+05
50%	2.007000e+03	2.000000e+00	1.271499e+06
75%	2.012000e+03	2.000000e+00	1.914983e+06
max	2.017000e+03	2.000000e+00	2.570235e+06

Figure 4: Description of numerical data

Tools

The tools employed for this project will include but not limited to the following;

- Python for data analysis. Python comes as a very powerful tool for data analysis and data science as it contains various data science and machine learning libraries. We will use python as our scripting language to analyze, clean and transform the data, access the stored data from the database and ultimately build our machine learning model
- Chart.js to build interactive visualization for the web application
- MongoDB to store data. The dataset will be stored in MongoDB to provide easy access and storage for our data for our web application
- HTML, CSS, and JavaScript for our web application. The web application will be developed using these tools

Timeline

To achieve this project we will be employing a waterfall model in our approach to executing the project. As we have a scheduled timeline for achieving most of our goals, we will engage the iterative approach to ensure that each module of our project is delivered successfully. The typical data analytics lifecycle approach will also be used in carrying this work. The table and Gantt chart below shows our timeline for achieving this project.

	Task	Activities	Expected Date
1.	Understand business need/Discovery	Determine the problem, construct problem as an analytic problem, determine resources needed for execution	July 1st, 2020
2.	Data transformation, preparation and analysis	Explore the data and transform it into a usable form by cleaning the data. Summarize the data and create basic visuals	July 9th,2020
3.	Modeling	Determine the variables required and the best model is selected to execute the project	July 20th,2020
4.	Build web application	Web application development	August 1st, 2020
5.	Testing	Test every module to ascertain the functionality	August 8th, 2020
6.	Communicate Results	Communicate the different findings using presentations	August 13th, 2020
7.	Documentation	Provide detailed business documentation of all the steps taken to achieve the project, the outcome, conclusions, future possibilities and any challenges encountered	August 14th, 2020

Table 1: Timeline

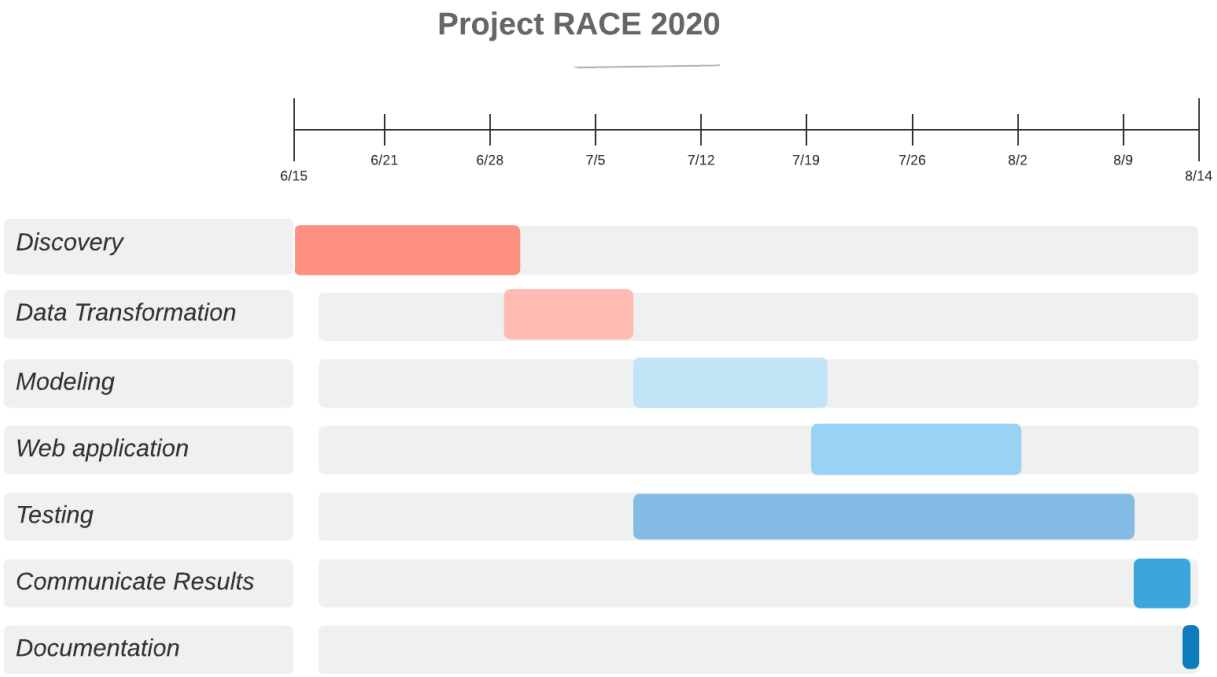

Figure 5: Gantt Chart of timeline

Expected Outcomes

Our work aims to ensure that the identified problem of continued road accidents on Canadian roads will be reduced significantly by the different findings that we discover in this project.

Upon the completion of this project, we will have a fully functioning website web application(a simple template shown in the figure below). This website will contain various interactive visualization of our descriptive analysis of the dataset from NCDB from the year 1999 to 2017. More importantly, it will contain the descriptive model that changes with the changing dataset and gives adequate information about the collision hotspots around roads in Canada and help official intensify and focus efforts on significant solutions that can be implemented around these spots. We believe that the results from the exploratory analysis and machine learning model will be the very first of its kind as no visible or detailed research has been carried out in understanding and showing the effects of the road configurations on the number of accidents that occur on Canadian roads. The research will prompt the concerned authorities to take a closer look at these collision hotspots and help them formulate policies that will assist road users.

The hosted website will available for multiple users and it can be a source of information for anyone looking for information on the road accidents in Canada as we will have a detailed analysis of the dataset on the website with interactive visuals. As a stretch goal, we are looking to share our findings and outcomes with Transport Canada and see how they can implement some of our findings and maybe grant us opportunities to further develop our model to consider other factors.

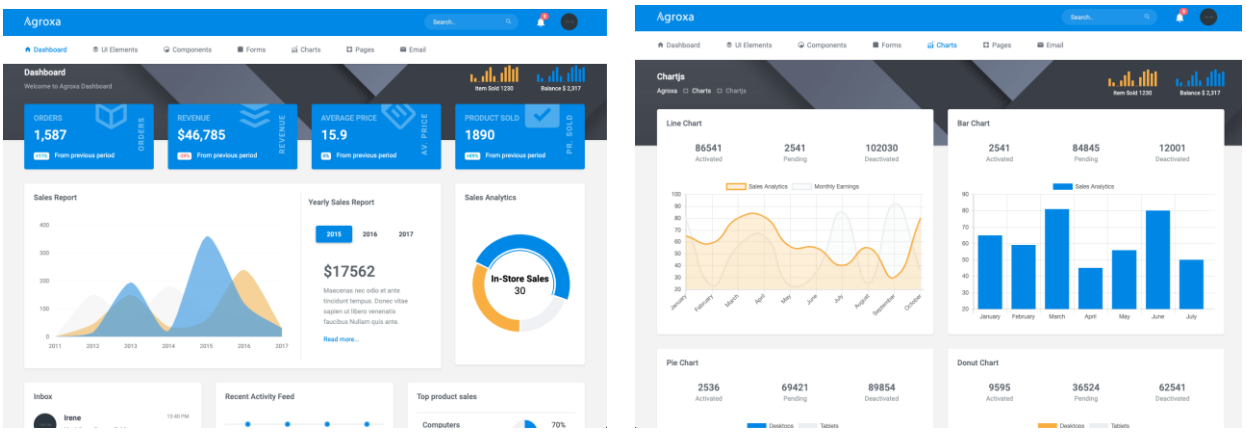

Figure 6: Sample web application

References

1. CCTMA, 2016. [online] Available at: <https://roadsafetystrategy.ca/en/strategy> [Accessed 4 July 2020].
2. Thecanadianencyclopedia.ca. 2018. *Humboldt Broncos Bus Crash | The Canadian Encyclopedia*. [online] Available at: <https://www.thecanadianencyclopedia.ca/en/article/humboldt-broncos-bus-crash> [Accessed 4 July 2020].
3. CBC News, 2018. *Scene Of Broncos Crash Haunted Sask. Truckers For Decades | CBC News*. [online] Available at: <https://www.cbc.ca/news/canada/saskatoon/scene-of-broncos-crash-haunted-sask-truckers-for-decades-1.4612253> [Accessed 5 July 2020].
4. Markhammitsubishi.ca. n.d. *Markham Mitsubishi | Most Dangerous Intersections In Canada*. [online] Available at: <https://www.markhammitsubishi.ca/en/news/view/most-dangerous-intersections-in-canada/62936> [Accessed 5 July 2020].