Predicting the severity of an accident

Introduction

Globally, road traffic accidents are an important public health and safety concern. To work on combating fatalities and injuries on Seattle roads, the Seattle Department of Transportation (SDOT) has introduced the Vision Zero Programme [1]. The programme aims to end traffic deaths and serious injuries on Seattle's city streets by 2030. As part of this programme, one of the key approaches that is being taken to meet the Vision Zero objective is a data driven approach [2]. Historical data on vehicle accidents and their severity as well as information relating to each incident will be analysed and used to implement smart street design, targeted enforcement, develop traffic safety education programmes and provide guidelines on engineering changes required and traffic regulations updates [1].

Using Seattle's historical collisions-dataset as provided in the coursera applied data science capstone [3], this report will be used to provide findings on the main key influencing variables that can be utilised to determine the severity of an accident. A model that can predict the severity of a collision will be identified, trained and evaluated and key insights provided. The output of such a model would be of interest to the Seattle department of transportation, residents of Seattle that are interested in understanding the predictability of the severity of an accident as well as organisations and stakeholders that are vested in the success of the vision zero programme.

The dataset

The dataset that will used to develop the model was provided by Coursera. The data is a record of Seattle City's collisions from 2004 to May 2020. It includes all collisions provided by SPD and recorded by Traffic Records of the SDOT traffic management division and is updated weekly [4].

The dataset will be utilised to identify key variables that can be used to identify the severity of an accident, which is the target variable. To identify these variables, the data will be analysed for trends, certain patterns, skewed information and correlations. This data will be applied to a supervised learning algorithm. Various supervised learning models will be evaluated to identify the one that provides the best results. A sample view of the data is provided Appendix A. Attribute information is provided in Table 1 and Table 2:

Table 1: Attribute Information

Attribute	Data type, length	Description
OBJECTID	ObjectID	ESRI unique identifier
SHAPE	Geometry	ESRI geometry field
INCKEY	Long	A unique key for the incident
COLDETKEY	Long	Secondary key for the incident
ADDRTYPE	Text, 12	Collision address type: • **Alley** • **Block** • **Intersection**
INTKEY	Double	Key that corresponds to the intersection associated with a collision

LOCATION	Text, 255	Description of the general location of the collision
EXCEPTRSNCODE	Text, 10	
EXCEPTRSNDESC	Text, 300	
SEVERITYCODE	Text, 100	A code that corresponds to the severity of the collision: • **3**—fatality • **2b**—serious injury • **2**—injury • **1**—prop damage • **0**—unknown
SEVERITYDESC	Text	A detailed description of the severity of the collision
COLLISIONTYPE	Text, 300	Collision type
PERSONCOUNT	Double	The total number of people involved in the collision
PEDCOUNT	Double	The number of pedestrians involved in the collision. This is entered by the state.
PEDCYLCOUNT	Double	The number of bicycles involved in the collision. This is entered by the state.
VEHCOUNT	Double	The number of vehicles involved in the collision. This is entered by the state.
INJURIES	Double	The number of total injuries in the collision. This is entered by the state.
SERIOUSINJURIES	Double	The number of serious injuries in the collision. This is entered by the state.
FATALITIES	Double	The number of fatalities in the collision. This is entered by the state.
INCDATE	Date	The date of the incident.
INCDTTM	Text, 30	The date and time of the incident.
JUNCTIONTYPE	Text, 300	Category of junction at which collision took place
SDOT_COLCODE	Text, 10	A code given to the collision by SDOT.
SDOT_COLDESC	Text, 300	A description of the collision corresponding to the collision code.
INATTENTIONIND	Text, 1	Whether or not collision was due to inattention. (Y/N)
UNDERINFL	Text, 10	Whether or not a driver involved was under the influence of drugs or alcohol.

WEATHER	Text, 300	A description of the weather conditions during the time of the collision.
ROADCOND	Text, 300	The condition of the road during the collision.
LIGHTCOND	Text, 300	The light conditions during the collision.
PEDROWNOTGRNT	Text, 1	Whether or not the pedestrian right of way was not granted. (Y/N)
SDOTCOLNUM	Text, 10	A number given to the collision by SDOT.
SPEEDING	Text, 1	Whether or not speeding was a factor in the collision. (Y/N)
ST_COLCODE	Text, 10	A code provided by the state that describes the collision. For more information about these codes, please see the State Collision Code Dictionary.
ST_COLDESC	Text, 300	A description that corresponds to the state's coding designation.
SEGLANEKEY	Long	A key for the lane segment in which the collision occurred.
CROSSWALKKEY	Long	A key for the crosswalk at which the collision occurred.
HITPARKEDCAR	Text, 1	Whether or not the collision involved hitting a parked car. (Y/N)

Table 2: State Collision Code Dictionary

Code	Description
0	Vehicle Going Straight Hits Pedestrian
1	Vehicle Turning Right Hits Pedestrian
2	Vehicle Turning Left Hits Pedestrian
3	Vehicle Backing Hits Pedestrian
4	Vehicle Hits Pedestrian - All Other Actions
5	Vehicle Hits Pedestrian - Actions Not Stated
10	Entering At Angle
11	From Same Direction -Both Going Straight-Both Moving- Sideswipe
12	From Same Direction -Both Going Straight-One Stopped- Sideswipe
13	From Same Direction - Both Going Straight - Both Moving - Rear End

References

[1] Seattle.gov, “Vision Zero,” Seattle.gov, February 2020. [Online]. Available: https://www.seattle.gov/visionzero. [Accessed 03 September 2020].

[2] vision zero network, “How does Vision Zero differ from the tradtional traffic safety approach in U.S Communities,” [Online]. Available: http://visionzeronetwork.org/wp-content/uploads/2016/03/VZN-Case-Study-1-What-makes-VZ-different.pdf. [Accessed 03 September 2020].

[3] Cousera, “Applied Data Science Capstone - Example Dataset,” [Online]. Available: https://www.coursera.org/learn/applied-data-science-capstone/supplement/Nh5uS/downloading-example-dataset. [Accessed 3 September 2020].

[4] Cousera, “Applied Data Science Capstone Example Dataset,” [Online]. Available: https://www.coursera.org/learn/applied-data-science-capstone/supplement/Nh5uS/downloading-example-dataset. [Accessed 03 September 2020].

Appendix A

Figure 1: Sample of collision Dataset part 1

SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	LOCATION	EXCEPTRSNCODE	EXCEPTRSNDESC	SEVERITYCODE	SEVERITYDESC
2	-122.323	47.70314	1	1307	1307	3502005	Matched	Intersection	37475	5TH AVE NE AND NE 103RD ST			2	Injury Collision
1	-122.347	47.64717	2	52200	52200	2607959	Matched	Block		AURORA BR BETWEEN RAYE ST AND BRIDGE WAY N		1	Property Damage Only Collision	
1	-122.335	47.60787	3	26700	26700	1482393	Matched	Block		4TH AVE BETWEEN SENECA ST AND UNIVERSITY ST		1	Property Damage Only Collision	

Figure 2: Sample of collision Dataset part 2

COLLISIONTYPE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	INCDATE	INCDTTM	JUNCTIONTYPE	SDOT_COLCODE	SDOT_COLDESC
Angles	2	0	0	2	2013/03/27 00:00:00+00	3/27/2013 2:54:00 PM	At Intersection (intersection related)	11	MOTOR VEHICLE STRUCK MOTOR VEHICLE, FRONT END AT ANGLE
Sideswipe	2	0	0	2	2006/12/20 00:00:00+00	12/20/2006 6:55:00 PM	Mid-Block (not related to intersection)	16	MOTOR VEHICLE STRUCK MOTOR VEHICLE, LEFT SIDE SIDESWIPE
Parked Car	4	0	0	3	2004/11/18 00:00:00+00	11/18/2004 10:20:00 AM	Mid-Block (not related to intersection)	14	MOTOR VEHICLE STRUCK MOTOR VEHICLE, REAR END

Figure 3: Sample of collision Dataset part 3

INATTENTIONIND	UNDERINFL	WEATHER	ROADCOND	LIGHTCOND	PEDROWNOTGRNT	SDOTCOLNUM	SPEEDING	ST_COLCODE	ST_COLDESC	SEGLANEKEY	CROSSWALKKEY	HITPARKEDCAR
	N	Overcast	Wet	Daylight				10	Entering at angle	0	0	N
	0	Raining	Wet	Dark - Street Lights On		6354039		11	From same direction - both going straight - both moving - sideswipe	0	0	N
	0	Overcast	Dry	Daylight		4323031		32	One parked--one moving	0	0	N