CAR ACCIDENT SEVERITY

CAPSTONE PROJECT REPORT

SUMEET GREWAL

1 INTRODUCTION

One of the most dangerous and complex activities we engage in on a daily basis is driving a vehicle. Increasing population density in major cities such as Seattle increases the propensity for traffic incidents and collisions. According to a dataset on collisions in the city of Seattle, there have been over 190 000 incidents since 2004. Traffic collisions have an enormous impact, most notably in terms of physical injury and in property damage.

The key audience is the City of Seattle as they would ideally undertake initiatives to reduce the number and severity of traffic collisions. From a mitigation perspective, it would be unwise to apply resources uniformly across the city in order to address a specific issue. Mitigation efforts should be concentrated where they would have the greatest impact. It would thus be constructive to identify which intersections are the most dangerous. This way, the City of Seattle can optimally allocate their resources and mitigation tactics towards addressing key factors in key intersections.

It is also possible that certain environmental factors and driver behaviours increase the severity of collisions. Some of these conditions could be specific days and times, weather and road conditions, driver inattention, and speeding.

The objective of this project is to understand which factors have the greatest impact on the severity of collisions in intersections, and identify which intersections could most benefit from mitigation tactics.

2 DATA

The selected dataset contains all collisions provided by the Seattle Police Department (SPD) from 2004 - Present. It contains over 194000 records and 37 fields which are described in detail in the associated metadata file. The objective is to develop a model to accurately predict the ‘severitycode’ label. This field currently either contains a value of ‘1’ indicating property damage only or a value of ‘2’ indicating physical injury from the collision.

There are many columns that describe the various details of the collision. The focus of this study will be on identifying the key contributors to the collision. Some of the fields that will be assessed for their impact on severity are:

ATTRIBUTE	DATA TYPE	DESCRIPTION
INCDTTM	Text	Incident Date and Time
WEATHER	Text	The weather conditions at the time of the collision
ROADCOND	Text	The conditions of the road at the time of the collision
LIGHTCOND	Text	The light conditions at the time of the collision
INATTENTIONIND	Text (Y/N)	Whether or not the collision was due to inattention by the driver
SPEEDING	Text (Y/N)	Whether or not the collision was due to speeding by the driver
PEDROWNOTGRNT	Text (Y/N)	Whether or not the pedestrian right of way was not granted by the driver
UNDERINFL	Text (Y/N)	Whether or not the driver was under the influence of drugs or alcohol

Once the data has been prepared, the model will be trained on this data to predict the severity. This will help to identify which factors are most significant in the severity of a collision. As the objective is to determine the class label of a categorical target attribute, a classification model is most appropriate. As the data is binary and we are looking to understand the impact of various features, a Logistic Regression model is likely the best choice. Other models such as Support Vector Machine and Decision Trees may also be evaluated for their accuracy.