**ACCIDENT SEVIRITY PREDECTION REPORT**

1. **INTRODUCTION**

Traffic accidents are one of the leading causes of death and injuries, According to International Transport Forum, 37 133 persons lost their lives in traffic crashes in the United States in 2017.

Many factors affects the traffic accidents, some related to the road, the car, or the person.

The scope of this project is to predict the severity of an accident through supervised machine learning algorithms using historical dataset, prediction of severity of the road based on certain features or condition such as “Road Condition, Weather Condition, Light Conditions, others”.

The questions that are answered by this project

* What is the type of severity of a collision?
* What is the possibility of a person getting into a car accident and how severe it would be?

The importance of such project:

1. You can notify people to avoid the factors that have high probability of accident.
2. Emergency support can be prepared in case of an accident occurs in specific roads and can send the prober support based on the expected severity.
3. Special preparation can take place from government (Ex. Bad weather, Holidays , .. ).
4. While continues development and enhancement in this model will help in Enhancing roads, notify people, advice…

Audience:

* Government entities: Traffic Department
* Emergency department

1. **DATA**

The Dataset is provided by SPD and recorded by Traffic Records (suggested by the course material), it includes all types of collisions within this Timeframe: 2004 to Present. Collisions will display at the intersection or mid-block of a segment.

The dataset can be downloaded from the below link

<https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>.

The metadata for the dataset can be downloaded from the below link

<https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>

**The following is initial information about the dataset.**

* It contains of 194,673 rows and 38 attributes.
* Challenges
* Some missing values for specific columns, some limited rows of specific categories.
* For the complete description of the attributes and metadata, kindly follow the above link for the metadata.
* *SEVERITYCODE* is the label; it is biased label for code 1; it contains of 136,485 for severity 1 and 58188 for severity 2.

**Features that can be extracted from the data:**

* *ADDRTYPE*: Address Type

Block 126926

Intersection 65070

Alley 751 few number for this category.

* *COLLISIONTYPE* is category of 10 types
* Parked Car 47987
* Angles 34674
* Rear Ended 34090
* Other 23703
* Sideswipe 18609
* Left Turn 13703
* Pedestrian 6608
* Cycles 5415
* Right Turn 2956
* Head On 2024
* INCDATE The date of the incident.
* INCDTTM The date and time of the incident.
* JUNCTIONTYPE Category of junction at which collision took place

Mid-Block (not related to intersection) 89800

At Intersection (intersection related) 62810

Mid-Block (but intersection related) 22790

Driveway Junction 10671

At Intersection (but not related to intersection) 2098

Ramp Junction 166

Unknown

* WEATHER A description of the weather conditions during the time of the collision.
* ROADCOND The condition of the road during the collision.
* LIGHTCOND The light conditions during the collision.
* SPEEDING Whether or not speeding was a factor in the collision.

**Attributes will be investigated more to check the impact and might be involved in feature engineering.**

* PERSONCOUNT The total number of people involved in the collision
* PEDCOUNT The number of pedestrians involved in the collision.
* PEDCYLCOUNT The number of bicycles involved in the collision.
* VEHCOUNT The number of vehicles involved in the collision.
* INJURIES The number of total injuries in the collision.
* SERIOUSINJURIES The number of serious injuries in the collision.
* FATALITIES The number of fatalities in the collision.

**Attributes have some issues:** 9

* *INATTENTIONIND* (Y/N) It has only Y with 29805 but the rest are null, no values for N.
* *UNDERINFL* no enough information, there is no description for the values [1,0].

*for the complete list of attributes and description, check the url*

<https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>.