**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…
5. **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.
* *ADDRTYPE*: Address Type

Address Type has fewer number of type Alley

Block 126926

Intersection 65070

Alley 751 few number for this category.

* *COLLISIONTYPE* is category of 10 types
* *JUNCTIONTYPE*  has number for this some categories.
* *INATTENTIONIND* (Y/N) Whether or not collision was due to inattention, this attribute has issue, 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].
* *WEATHER* description of the weather conditions during the time of the collision, has fewer number for this some categories like (Sleet/Hail/Freezing Rain, Blowing Sand, Dirt-Crosswind).
* *Some other attributes:ROADCOND , LIGHTCOND, PEDROWNOTGRNT, SDOTCOLNUM, SPEEDING, HITPARKEDCAR.*

*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>.