"ENHANCING TRAFFIC ACCIDENT CLASSIFICATION USING MACHINE LEARNING"

PROPOSAL

Prepared by :

Ahmed Mohamed Salem

Project Description

This project is based on a countrywide car accident dataset covering 49 states in the USA.
The accident data was collected from February 2016 to March 2023, using multiple APIs that provide streaming traffic incident data.
These APIs broadcast traffic data captured by various entities, including federal and state departments of transportation, law enforcement agencies,
traffic cameras, and traffic sensors within the road networks. The dataset currently contains approximately 7.7 million accident records.

ID	Source	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	...	Roundabout	Station	Stop	Traffic_Calming	Traffic_Signal	Turning_Loop	Sunrise_Sunset	Civil_Twiligh
A-1	Source2	3	2016-02-08 05:46:00	2016-02-08 11:00:00	39.865147	-84.058723	NaN	NaN	0.010	...	False	False	False	False	False	False	Night	Nigh
A-2	Source2	2	2016-02-08 06:07:59	2016-02-08 06:37:59	39.928059	-82.831184	NaN	NaN	0.010	...	False	False	False	False	False	False	Night	Nigh
A-3	Source2	2	2016-02-08 06:49:27	2016-02-08 07:19:27	39.063148	-84.032608	NaN	NaN	0.010	...	False	False	False	False	True	False	Night	Nigh
A-4	Source2	3	2016-02-08 07:23:34	2016-02-08 07:53:34	39.747753	-84.205582	NaN	NaN	0.010	...	False	False	False	False	False	False	Night	Da
A-5	Source2	2	2016-02-08 07:39:07	2016-02-08 08:09:07	39.627781	-84.188354	NaN	NaN	0.010	...	False	False	False	False	True	False	Day	Da
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
A-7777757	Source1	2	2019-08-23 18:03:25	2019-08-23 18:32:01	34.002480	-117.379360	33.99888	-117.37094	0.543	...	False	False	False	False	False	False	Day	Da
A-7777758	Source1	2	2019-08-23 19:11:30	2019-08-23 19:38:23	32.766960	-117.148060	32.76555	-117.15363	0.338	...	False	False	False	False	False	False	Day	Da
A-7777759	Source1	2	2019-08-23 19:00:21	2019-08-23 19:28:49	33.775450	-117.847790	33.77740	-117.85727	0.561	...	False	False	False	False	False	False	Day	Da
A-7777760	Source1	2	2019-08-23 19:00:21	2019-08-23 19:29:42	33.992460	-118.403020	33.98311	-118.39565	0.772	...	False	False	False	False	False	False	Day	Da
A-7777761	Source1	2	2019-08-23 18:52:06	2019-08-23 19:21:31	34.133930	-117.230920	34.13736	-117.23934	0.537	...	False	False	False	False	False	False	Day	Da

s × 46 columns

Data Information:

- Number of Rows: 7,728,394 rows
- Number of Columns: 46 columns

Missing Data:

"End_Lat" column has a missing value percentage of approximately 44.03%.
"End_Lng" column has a missing value percentage of approximately 44.03%.
"Wind_Chill(F)" column has a missing value percentage of approximately 25.87%.
"Precipitation(in)" column has a missing value percentage of approximately 28.51%.
These percentages indicate the presence of some missing data in the dataset, which may require additional processing to ensure accurate analyses. This dataset can be used to enhance traffic accident management and improve emergency response.

GOALS & OBJECTIVE

OBJECTIVES 1

The primary goal of this project is to classify traffic accidents into four severity levels, using machine learning techniques

OBJECTIVES 2

Factor Analysis: Understand the factors contributing to accidents and their impact on accident severity.

OBJECTIVES 3

Predictive System Development: Develop models capable of predicting the impact of various factors on accidents.

THANK YOU, AND WE LOOK FORWARD TO WORKING WITH YOU.