MARYLAND STATEWIDE VEHICLE CRASHES – PREDICTING SEVERITY OF INJURY

By

Ujwala Namineni – OI46298

INTRODUCTION

- The dataset used in this project is the Maryland Statewide Vehicle Crashes dataset, which was obtained from https://opendata.maryland.gov/ This dataset contains information on vehicle crashes that occurred in Maryland state between 2015 and 2022. The dataset is composed of three sub-datasets, which are:
- Maryland Statewide Vehicle Crashes: https://opendata.maryland.gov/Public-Safety/Maryland-Statewide-Vehicle-Crashes/65du-s3qu
- Person Details: https://opendata.maryland.gov/Public-Safety/Maryland-Statewide-Vehicle-Crashes-Person-Details-/py4c-dicf
- Vehicle Details: https://opendata.maryland.gov/Public-Safety/Maryland-Statewide-Vehicle-Crashes-Vehicle-Details/mhft-5t5y

PROBLEM STATEMENT

Predicting the severity of the injury after the accident: This project aims to develop a machine-learning model that predicts the severity of injuries sustained by individuals involved in vehicle crashes in Maryland. Predicting the severity by the deployment of the model.

DATASET

- The vehicles and persons datasets are merged on ‘VEHICLE_ID’.
- The obtained dataset and crashes dataset are merged on ‘REPORT_NO’.
- After preprocessing, the shape of the dataset is (619031, 22).

PROCESS


Data Preprocessing
Exploratory Data Analysis
Modeling
Deployment

TOOLS USED

DATA PREPROCESSING

- To create a comprehensive dataset, we combined data from the three sub-datasets and selected a subset of relevant columns for our analysis.
- Cleaned and pre-processed data using Python.

EXPLORATORY DATA ANALYSIS

- Visualized data using Tableau.
- Link for my Tableau public account is Profile - ujwala.namineni1131 | Tableau Public
- The frequency of accidents is influenced by various factors such as road surface conditions, weather conditions, county, and road division type.

DASHBOARD - 1

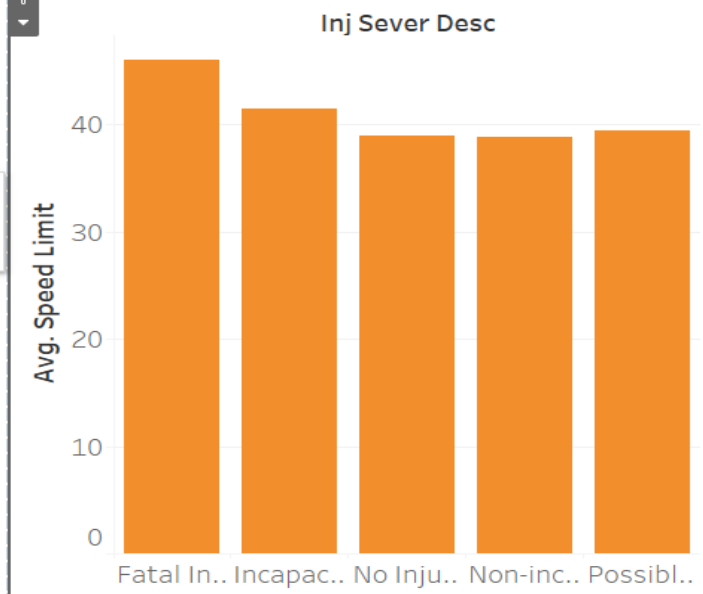
Sheet 7

Surf Cond Desc	
Dry	402,556
Ice	6,006
Mud, Dirt, Gravel	303
Other	454
Snow	6,162
Unknown	447
Wet	203,103

Sheet 2

Inj Sever Desc: **No Injury**
Avg. AGE: **41.324**

Sheet 3



Count of Inj Sever ..
2 295,170

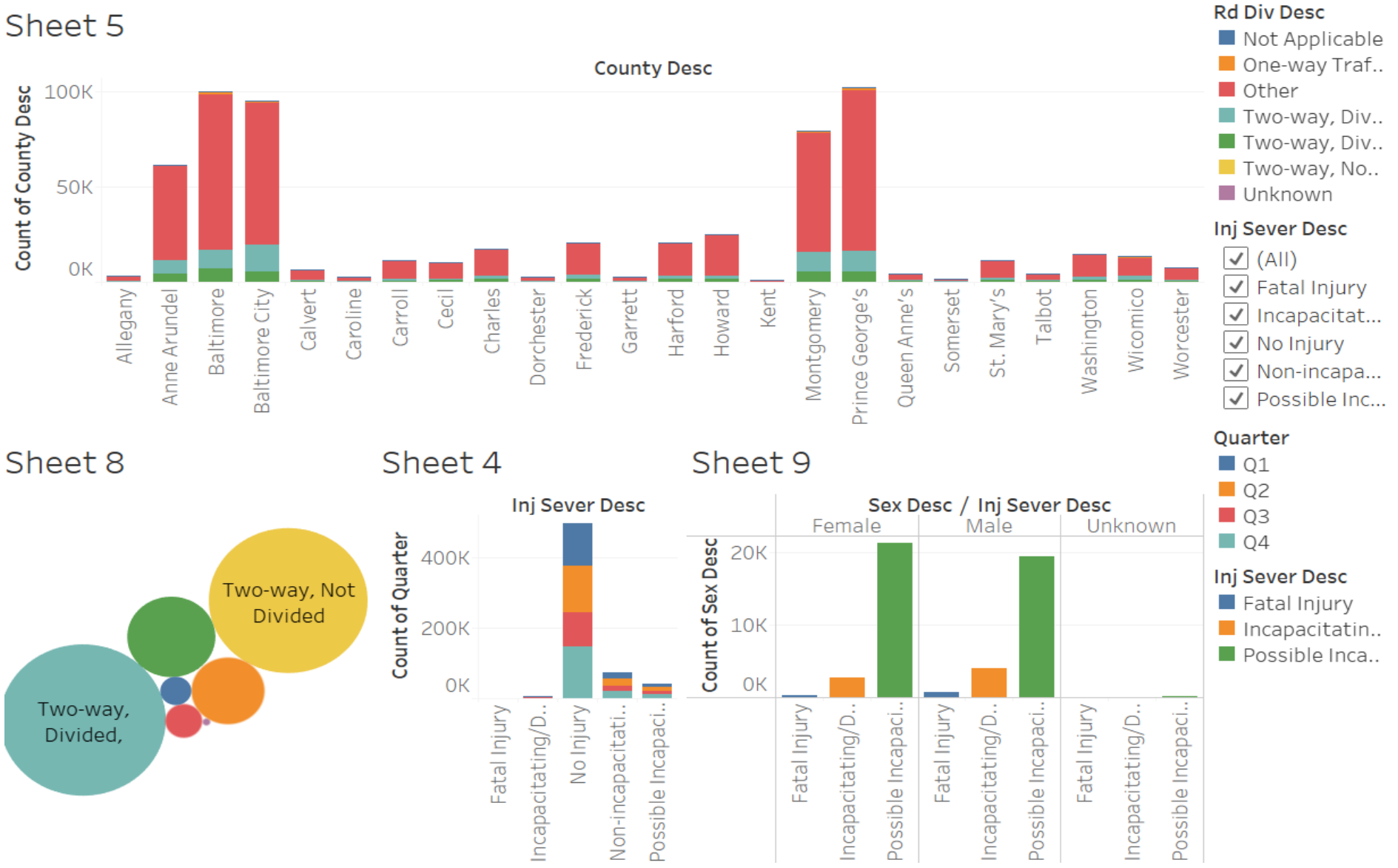
Inj Sever Desc
Fatal Injury
Incapacitating..
No Injury
Non-incapacit..
Possible Incap..

Avg. AGE
219.705

Sheet 6

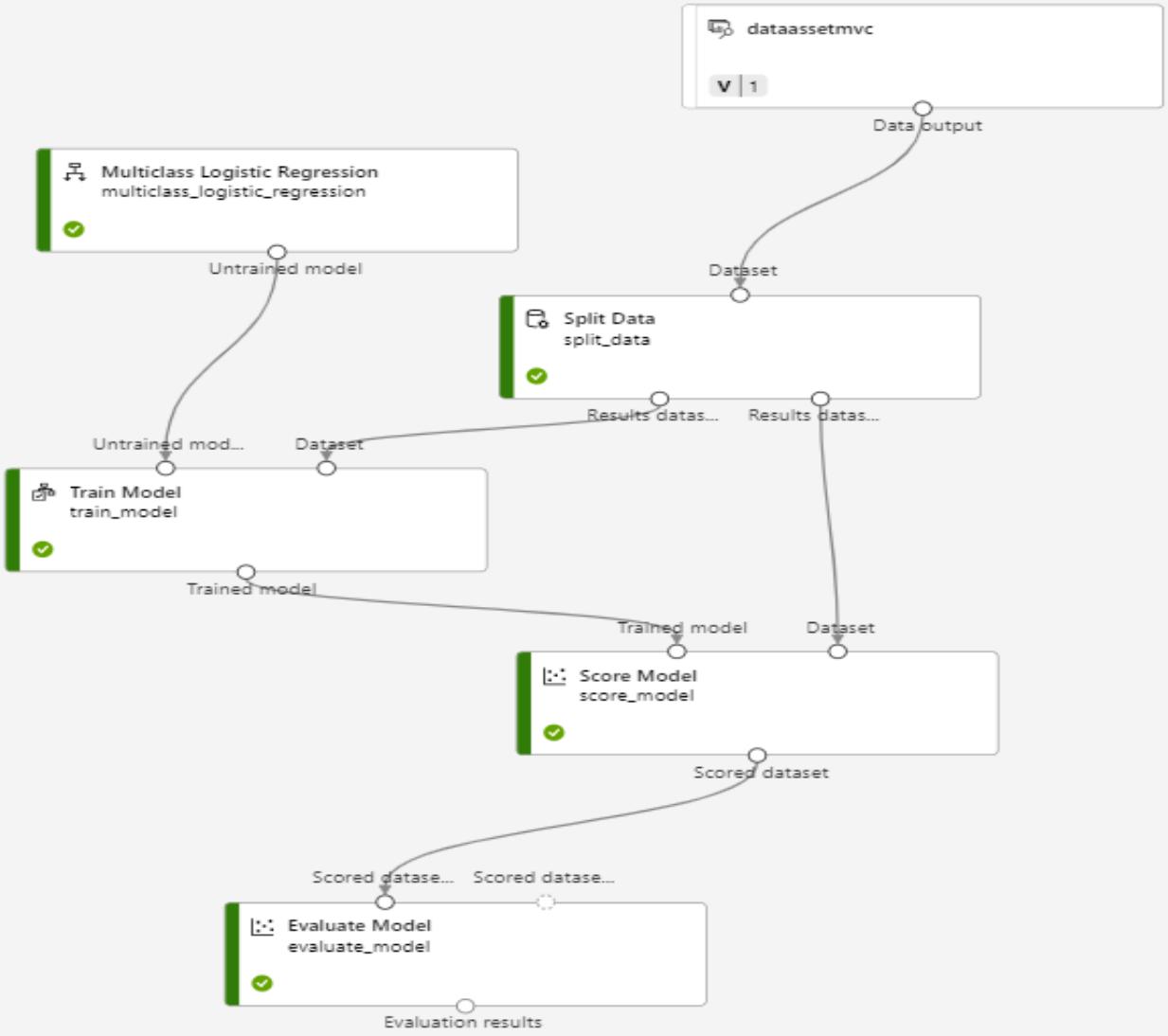
	Weather Desc									
Inj Sever Desc	Blowin..	Blowin..	Clear	Cloudy	Foggy	Raining	Severe ..	Sleet	Snow	Wintry ..
No Injury	44	966	295,170	44,858	5,256	139,830	1,517	977	5,634	3,367
Non-incapacitati..	3	108	44,985	6,838	711	18,488	191	128	719	442
Possible Incapac..	4	94	23,567	3,794	570	12,057	128	75	355	348
Incapacitating/D..		9	4,618	631	102	1,346	16	15	35	40
Fatal Injury		2	650	103	27	190	6	3	6	8

DASHBOARD - 2


Sheet 5
County Desc
Count of County Desc
100K
50K
0K
Allegany
Anne Arundel
Baltimore
Baltimore City
Calvert
Caroline
Carroll
Cecil
Charles
Dorchester
Frederick
Garrett
Harford
Howard
Kent
Montgomery
Prince George's
Queen Anne's
Somerset
St. Mary's
Talbot
Washington
Wicomico
Worcester
Rd Div Desc
Not Applicable
One-way Traf..
Other
Two-way, Div..
Two-way, Div..
Two-way, No..
Unknown
Inj Sever Desc
(All)
Fatal Injury
Incapacitat...
No Injury
Non-incapa...
Possible Inc...
Sheet 8
Two-way, Not Divided
Two-way, Divided,
Sheet 4
Inj Sever Desc
Count of Quarter
400K
200K
0K
Fatal Injury
Incapacitating/D..
No Injury
Non-incapacitati..
Possible Incapaci..
Sheet 9
Sex Desc / Inj Sever Desc
Female
Male
Unknown
Count of Sex Desc
20K
10K
0K
Fatal Injury
Incapacitating/D..
Possible Incapaci..
Quarter
Q1
Q2
Q3
Q4
Inj Sever Desc
Fatal Injury
Incapacitatin..
Possible Inca..

MODELING

- Used a Storage account in Microsoft Azure to store the CSV file of the dataset after preprocessing.
- Used Azure Machine Learning workspace and Azure ML studio to create different types of pipelines of the machine learning models.
- Developed machine learning models using Microsoft Azure, including logistic regression, decision tree, and random forest regression.
- Among the models, logistic regression performed the best for this data, achieving an accuracy of 80.25%.



Macro_Precision
0.3647783

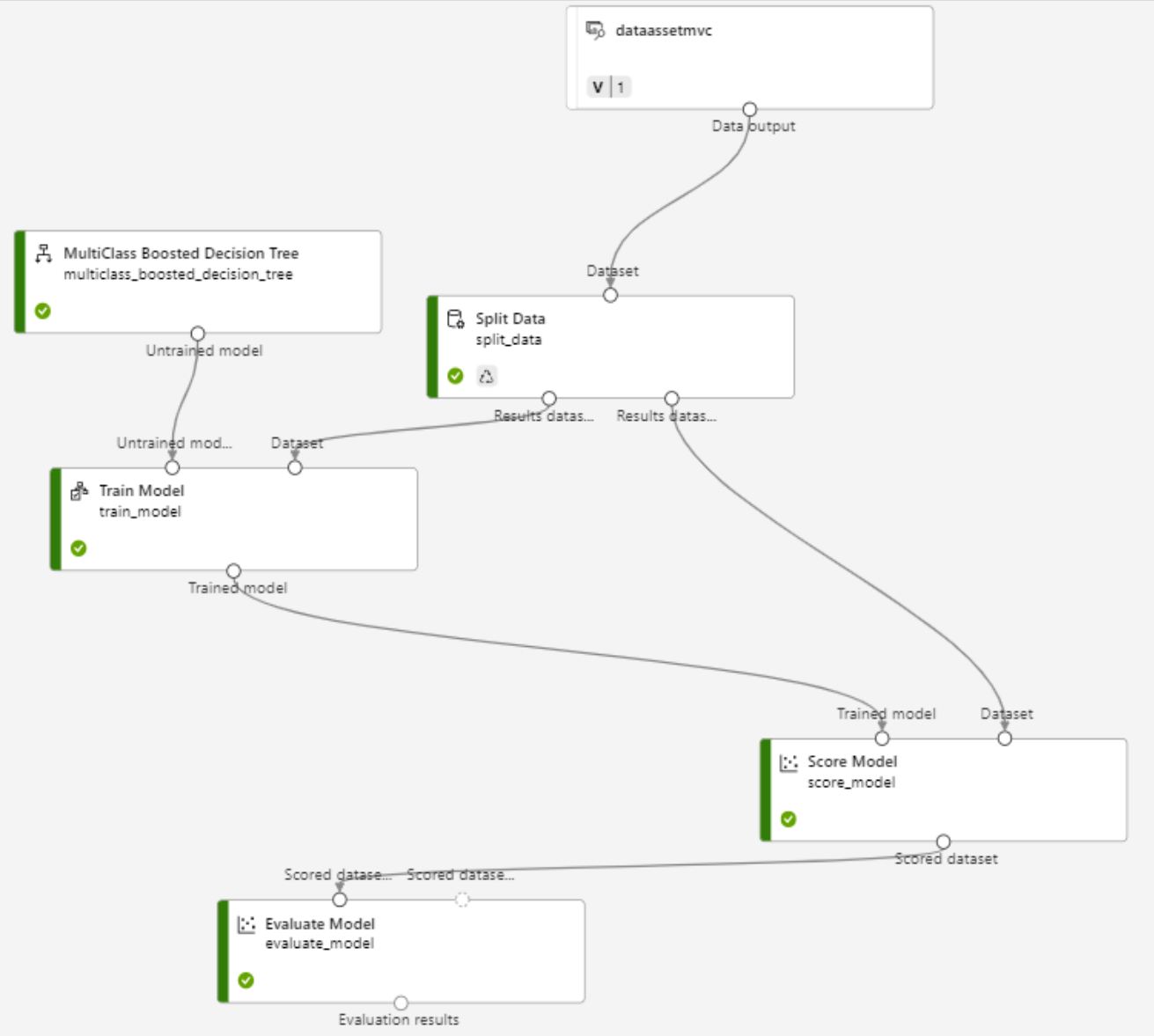
Macro_Recall
0.2315823

Micro_Precision
0.8024759

Micro_Recall
0.8024759

Overall_Accuracy
0.8024759

Logistic Regression



Macro_Precision
0.2742970

Macro_Recall
0.2138486

Micro_Precision
0.7759936

Micro_Recall
0.7759936

Overall_Accuracy
0.7759936

Decision Tree algorithm

DEPLOYMENT

- Developed a web application using Streamlit and Python.
- Implemented a web page via the 'app.py' file.
- Integrated the machine learning model into the web page.
- Defined a function to predict whether a person is injured or not.
- The page is created in localhost (http://localhost:8501/).

Injury Severity Prediction

Vehicle Year

2000

Speed Limit

30

Year_accident

2022

dob_year

2000

AGE

25

Predict Injury Severity

Predicted Injury Severity:

	0
0	No Injury

CHALLENGES

- To ensure accurate and reliable data analysis, it is crucial to validate data integrity and avoid merging duplicate entries or introducing additional rows due to improper merging practices.
- Developing web pages with Microsoft Azure web services is complex and requires technical expertise.
- Using Streamlit, Working with many categorical features can make predicting target variables complex.

THANK YOU