

Determining Car Accident Severity with ML



INTRODUCTION
DATA
METHODOLOGY
RESULTS
DISCUSSION



Introduction

- **Question:** How to determine accident severity (1 or 2) based on different variables in the dataset such as road conditions, collision type, weather, incident time, and vehicle conditions?
 - **Audience:** Traffic Police and Marketing Campaigns
 - **Goal:** Interpret & Analyze real-time data to help devise better mitigation strategy to reduce traffic collisions and subsequently save
-



Data

- Collision data available from Seattle Department of Transport
 - Contains 37 attributes and more than 19000+ cases reported
 - Timeframe covered: 2004 to 2020
 - Attributes include Location, Address Type, Collision Type, Incident Time, Road ,Lighting , weather conditions, etc.
-



METHODOLOGY

- **Data Cleaning:**
 - **Excluded:** geographic information, drop non-numeric values, empty columns and nominal data which didn't seem to impact the severity
 - **Included:** Address Type, Weather, Road Conditions, Collision Type, Light Conditions, Day of the Week
 - **Data Pre-processing:** Feature scaling
 - **Data Preparation:**
 - One hot-encoding translation of variables to analyze binary impact and help in classification
-



TECHNIQUES

Deployed supervised Machine Learning using Sci-kit learning and Pandas in Python

K-Nearest Neighbors

Decision Tree

Support Vector Machine

Logistic Regression



RESULTS

- Used 80:20 ratio for training : testing to train the model comprehensively and increase its robustness for fitting of the analysis
- Cross Validation done using k folds CV for K Nearest Neighbor to ensure that the overfitting or underfitting of the model doesn't occur
- Based on accuracy evaluation on test size on all data, logistic regression seems the *fastest algorithm with 70% + accuracy*

Algorithm	Jaccard Index	F1-score	LogLoss
K Nearest Neighbour	0.73	0.70	NA
Decision Tree	0.75	0.69	NA
Support Vector Machine	0.75	0.69	NA
Logistic Regression	0.75	0.69	0.49



DISCUSSION

- 70%+ accuracy achieved by ML techniques in predicting severity of an accident in the future
 - Conditions/Parameters that minimize the risk of an accident
 - Dry roads
 - Good lighting
 - Sunny weather
 - Scope for further Improvement
 - Training on more data could allow for improving data model precision and reduce sensitivity further
-



CONCLUSION

- Machine Learning Algorithms can help determine the severity of the accident in real-time
 - 75% accuracy with supervised learning based predictive model based on logistic regression
 - Allows better public safety
 - Opportunity to improve analysis using strong, more efficient data analysis software and exploring the option of bringing geospatial data
-



**Thank you for your
attention!!**

