

PREDICTION OF ACCIDENT SEVERITY



Sai Dinesh Kumar Reddy Addula



INTRODUCTION



- Road accidents are rising globally, leading to millions of deaths annually.

Key contributors to severe injuries include - Poor road conditions

- Various vehicle types
- Human behavior (e.g., distracted driving, speeding)
- Speed limits
- These factors interact in complex, non-linear ways. This complexity makes it difficult to create effective road safety models.



DATASET OVERVIEW

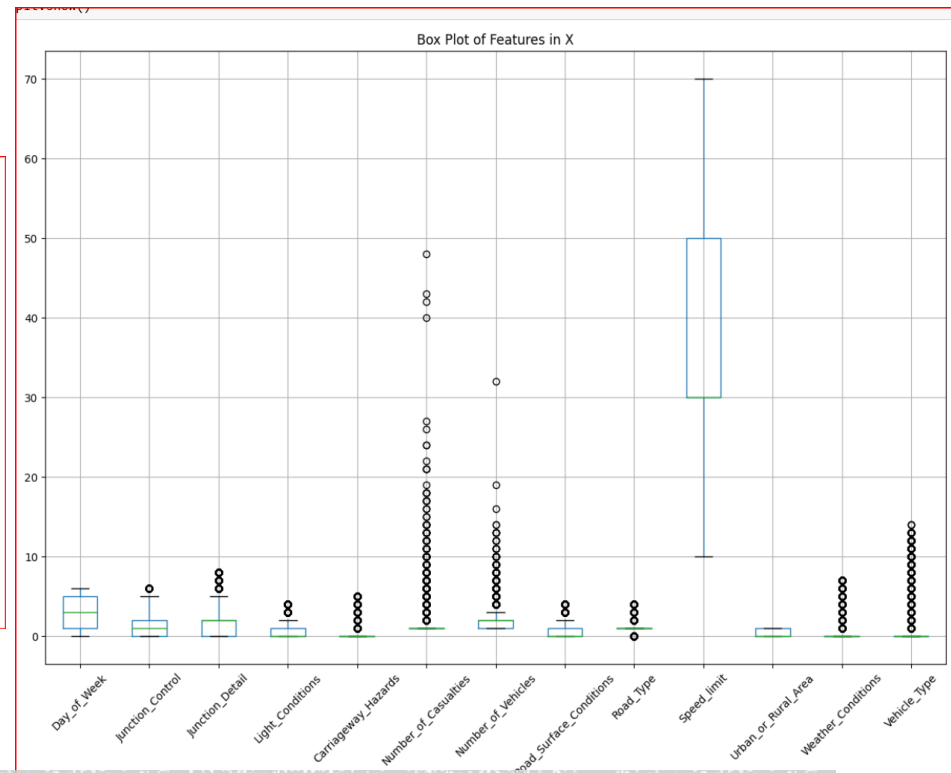
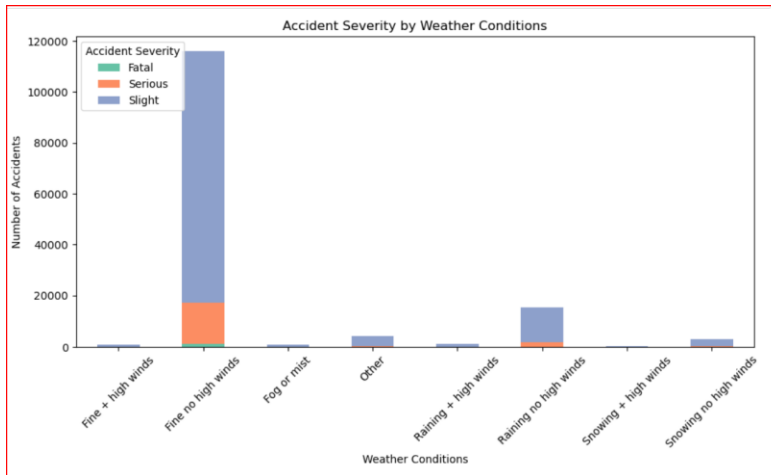
- • Source: Kaggle – US Road Accidents (2021-2022)
- • Total Records: 307,973
- • **Key Features:**
 - - Accident Severity, Weather Conditions, Road Type, Speed Limit, Light Conditions, Vehicle Type
- • **Challenges:**
 - - Class imbalance (severe vs. minor accidents)
 - - Missing values in key attributes
 - - High dimensionality requiring feature selection



EXISTING METHODOLOGIES

- **Method 1 : Road Accident Severity Prediction — A Comparative Analysis of Machine Learning Algorithms**
- **Method 2 : The random forest algorithm for statistical learning**
- **Method 3 : Neural Network**

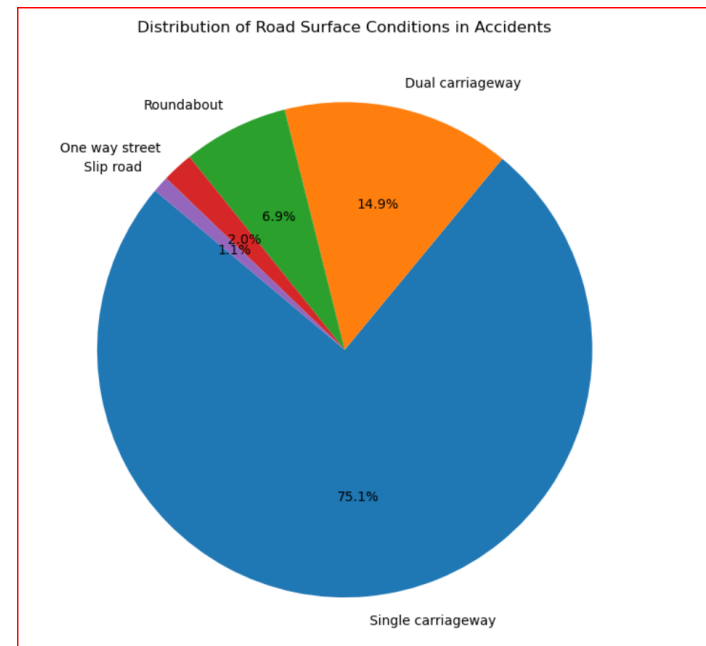
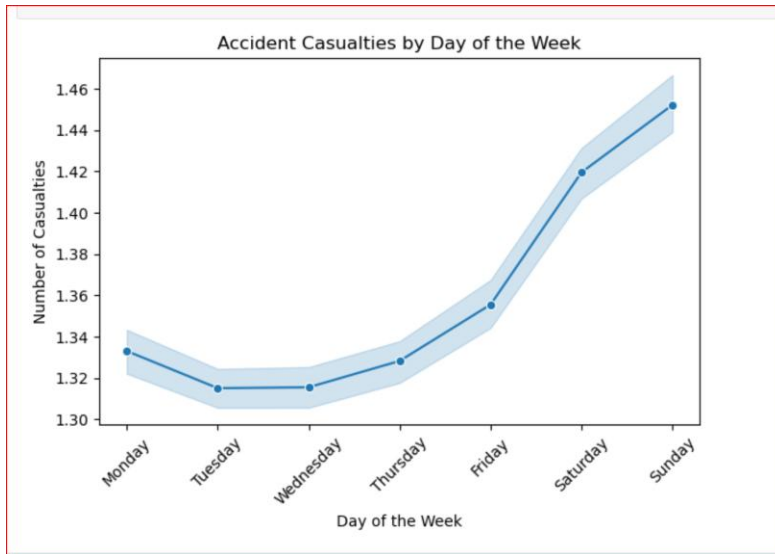




EXPLORATORY DATA ANALYSIS

- As we observe, there are a lot of outliers, and we plotted this using boxplots (via IQR)
- The Accident severity looks slight and little bit serious on days with fine with no winds

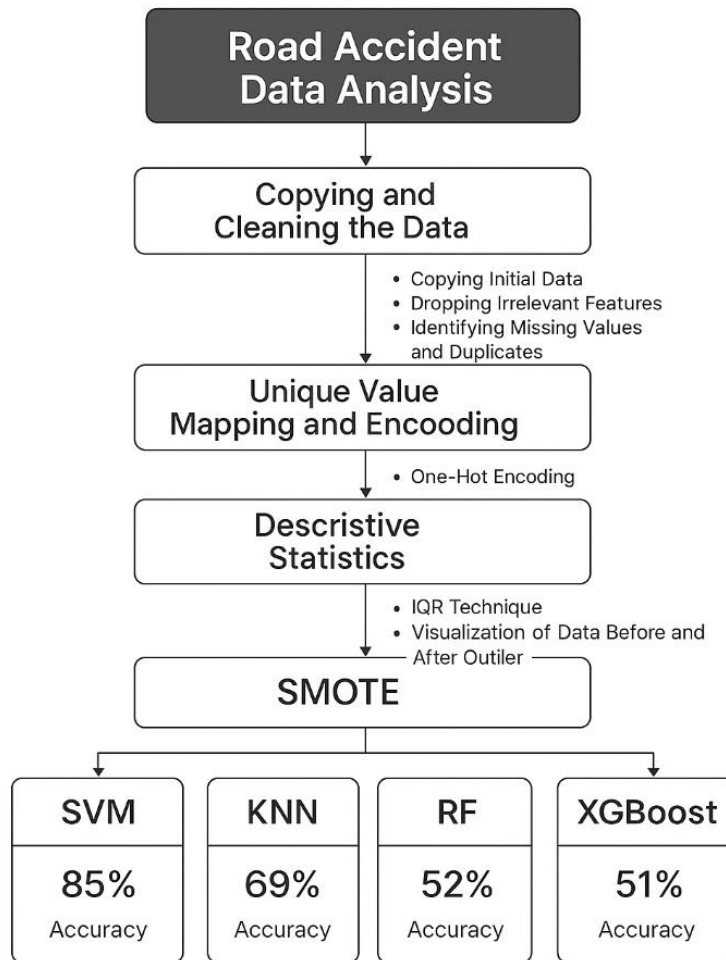




EXPLORATORY DATA ANALYSIS

- We see a rise in Accident casualties by the time of the weekend
- Single carriageway seems to be the major reason to contribute to accidents





PROPOSED METHODOLOGY

1.Cleaned and preprocessed road accident data by handling missing values and removing duplicates.

2.Applied One-Hot Encoding, and IQR for transformation and outlier treatment.

3.Conducted Exploratory Data Analysis using visual tools and correlation analysis.

4.Trained models (SVM, KNN, RF, XGBoost) with techniques like SMOTE and normalization.

5.Achieved best accuracy of 85% with SVM using SMOTE.



	precision	recall	f1-score	support
0	0.00	0.00	0.00	2734
1	0.85	1.00	0.92	16766
2	0.00	0.00	0.00	5
3	0.00	0.00	0.00	188
accuracy			0.85	19693
macro avg	0.21	0.25	0.23	19693
weighted avg	0.72	0.85	0.78	19693

SVM RESULTS

- Kernel = 'linear, C= 100, gamma= 0.1




```
Accuracy: 0.6876047326461179
Precision: 0.7587084677242417
Recall: 0.6876047326461179
F1 Score: 0.71742057546879
Confusion Matrix:
[[ 849 1876    0    9]
 [ 4039 12692    0   35]
 [    1     4    0    0]
 [   56   132    0   0]]
Classification report
```

	precision	recall	f1-score	support
0	0.17	0.31	0.22	2734
1	0.86	0.76	0.81	16766
2	0.00	0.00	0.00	5
3	0.00	0.00	0.00	188
accuracy			0.69	19693
macro avg	0.26	0.27	0.26	19693
weighted avg	0.76	0.69	0.72	19693

KNN RESULTS

- Weights = distance, algorithm = 'ball_tree', metric = 'euclidean'



	precision	recall	f1-score	support
0	0.15	0.05	0.08	2734
1	0.89	0.59	0.71	16766
2	0.00	0.00	0.00	5
3	0.03	0.50	0.05	188
accuracy			0.52	19693
macro avg	0.27	0.29	0.21	19693
weighted avg	0.78	0.52	0.62	19693


```

[[ 146 1161  700  727]
 [ 838 9919 3203 2806]
 [   1    2    0    2]
 [  12   39   43   94]]

```

RF RESULTS



	precision	recall	f1-score	support
0	0.14	0.04	0.07	2734
1	0.89	0.59	0.71	16766
2	0.00	0.00	0.00	5
3	0.02	0.51	0.05	188
accuracy			0.51	19693
macro avg	0.27	0.29	0.21	19693
weighted avg	0.78	0.51	0.62	19693
[[118 1161 700 755]				
[686 9920 3204 2956]				
[1 2 0 2]				
[11 39 43 95]]				

XGBOOST RESULTS



RESULTS

- Outliers were successfully detected and removed using the **IQR method** (Box Plot), improving data quality.
- Key accident factors identified include **weather conditions, road type, and lighting**.
- **SMOTE** effectively balanced the dataset, addressing class imbalance issues.
- Various machine learning models were trained and tested on the refined dataset.
- **Support Vector Machine (SVM)** delivered the best performance with an **accuracy of 85%**.



REFERENCES

<https://www.sciencedirect.com/science/article/pii/S2590198223000611>

<https://ieeexplore.ieee.org/document/9693055>

<https://towardsdatascience.com/what-is-eda-yes-another-post-on-eda-d8b5c06269a9>

<https://rpubs.com/Magguire/913785>

<https://www.kaggle.com/code/faressayah/support-vector-machine-pca-tutorial-for-beginner>

<http://prody.csb.pitt.edu/manual/reference/dynamics/pca.html>

<https://www.kaggle.com/code/shubhisaxena12/eda-and-pca-analysis>

<https://journals.sagepub.com/doi/10.1177/1536867X20909688>



CONCLUSION



- Our project aims to leverage machine learning for accident severity prediction.



- Addressing challenges in data imbalance, feature selection, and model optimization.



- Results can contribute to enhanced road safety policies and accident prevention.



- Future work includes incorporating real-time accident data sources.



QUESTIONS?



Thank you for your
attention!



We welcome your
questions and feedback.

