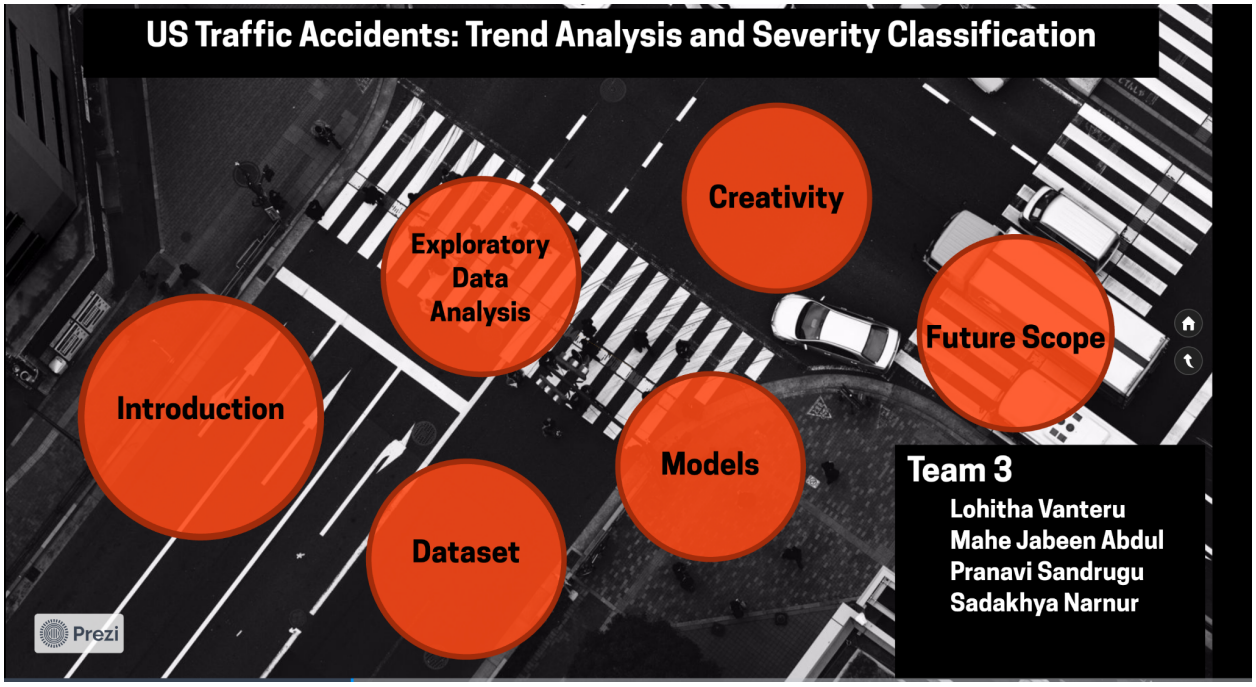

US Traffic Accidents: Trend Analysis and Severity Classification
Creativity
Exploratory Data Analysis
Future Scope
Introduction
Models
Dataset
Team 3
Lohitha Vanteru
Mahe Jabeen Abdul
Pranavi Sandrugu
Sadakhya Narnur
Prezi


Introduction
Road traffic accidents (RTAs) can have serious consequences, resulting in injuries, fatalities, and significant property damage.
Statistics per year according to NHTSA:
• There are over 37,000 fatalities.
• 2.35 million injuries or disabilities.
• economic cost of these accidents is around $230.6 billion per year.
On a global scale, accidents are ranked as the 9th leading cause of death.
Motivation
Goals
Community Contribution
Prezi

Motivation : Making America's Roads Safer

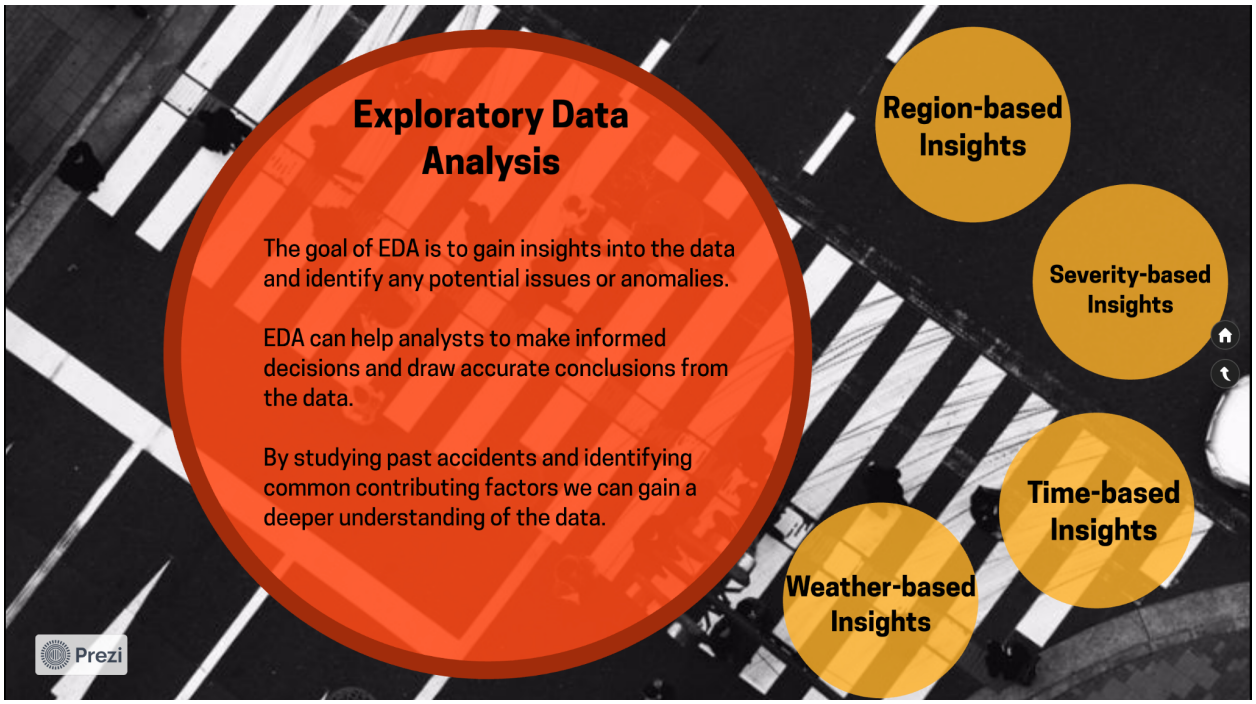
- Traffic accidents have a significant economic and societal impact on the United States, costing billions of dollars annually.
- Advanced data analytics techniques can be used to analyze vast amounts of data on USA accidents and uncover patterns and trends.
- Conducting an in-depth analysis of road accidents is a crucial step toward making our roads safer and reducing the devastating impact of accidents on individuals, families, and society.

Goals : Putting Accidents in the Rearview

- To analyze historical accident data to identify patterns and trends in accident occurrence, contributing factors, and potential solutions.
- To develop a predictive model that accurately identifies accident-prone areas and helps reduce the frequency and severity of accidents in the USA.
- To identify the most significant predictors of accidents in the USA, including road conditions, vehicle types, and weather patterns.
- To explore the use of advanced technologies, such as machine learning and artificial intelligence, to improve the accuracy and effectiveness of accident prediction models.
- To develop recommendations and guidelines for policymakers, transportation agencies, and other stakeholders to help prevent accidents and reduce their impact on public safety.

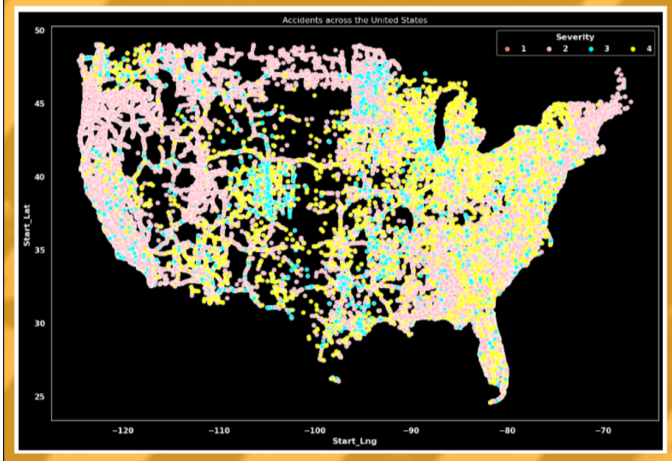
Community Contribution : Why Our Work Matters?

- **By sharing our findings with the community, we hope to raise awareness to prevent accidents**
  - *Improved road signage and traffic signals*
  - *Increased law enforcement presence in high-risk areas*
  - *Public education campaigns to promote safe driving practices*
- **By open sourcing our model, we aim to empower communities to improve road safety**
  - *Real-time accident alerts and risk assessment to educate drivers in accident-prone areas*
  - *Optimized route planning, avoiding high-risk areas and reducing the likelihood of accidents*

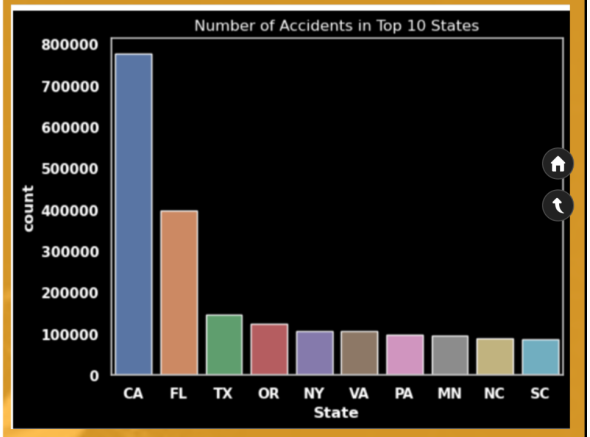


Mapping the Mayhem

A region-based analysis of accident hotspots



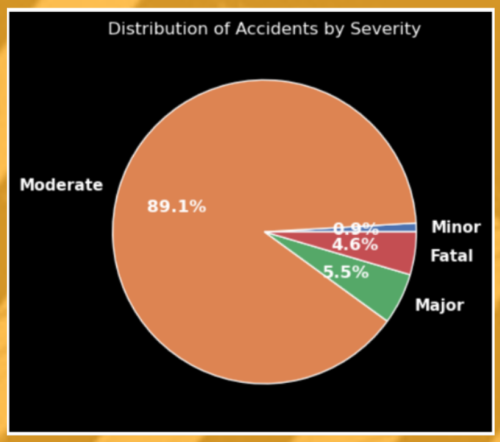
Top 10 states with the highest numbers of accidents



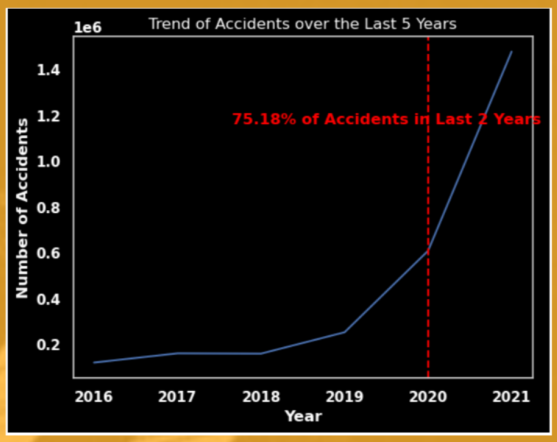
Prezi

From Fender-Benders to Total Wrecks

A Severity Spectrum of Accidents



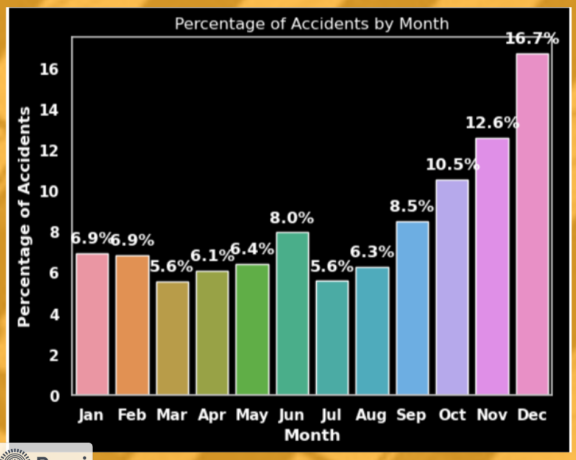
Uncovering the Timing of Traffic Troubles



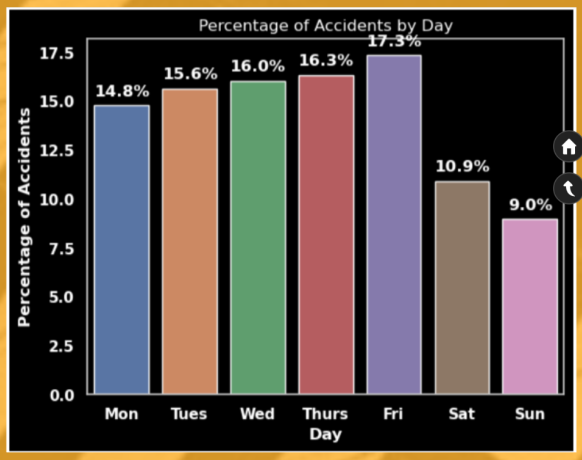
Prezi

Uncovering the Timing of Traffic Troubles

Percentage of accidents by Month

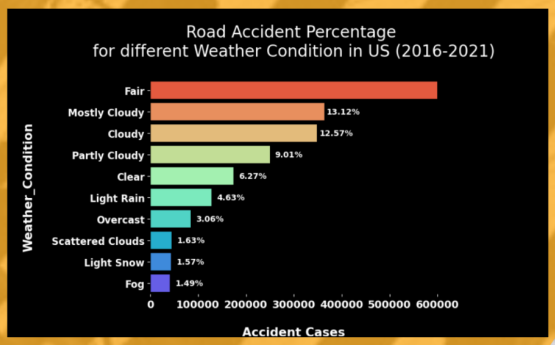


Percentage of Accidents by Day



Prezi

A look at how weather conditions impact road safety



Prezi

Dataset Description

US accidents dataset (Feb 2016 to Dec 2021):

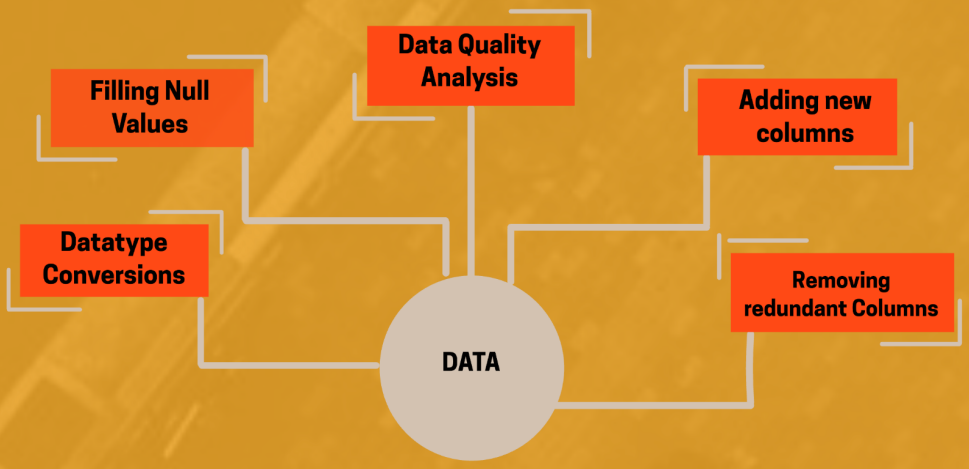
- This dataset covers 49 states of the United States.
- It includes information on 2.8 million traffic accidents and has 47 features.

US Census Demographic dataset:

- It covers all 52 states of America including DC and Puerto Rico.

Data Preparation

Data Preparation : Getting Our Data in Shape



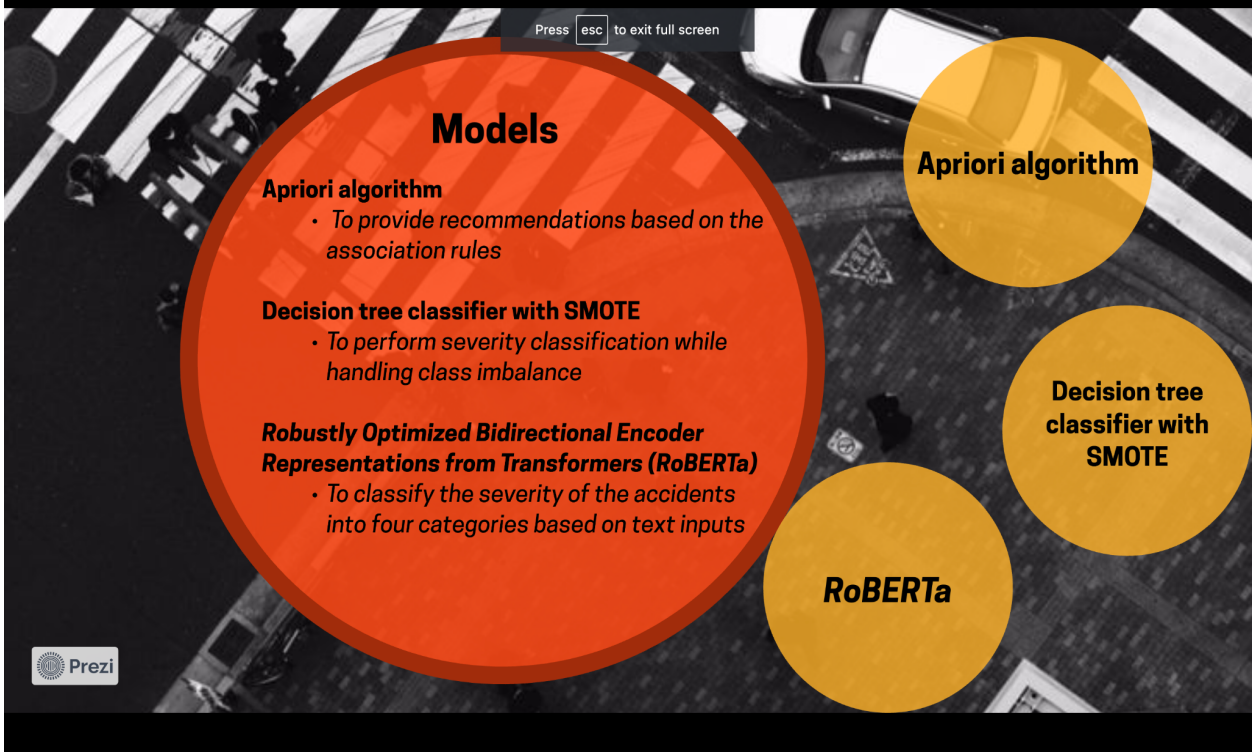
Data Preparation : Getting Our Data in Shape

```
acc_data['Duration(min)'] = (acc_data['End_Time'] - acc_data['Start_Time']).astype('timedelta64[m]')
acc_data['Start_Hour'] = (acc_data['Start_Time']).dt.hour
acc_data['Day'] = (acc_data['Start_Time']).dt.weekday
acc_data['Month'] = acc_data['Start_Time'].dt.month
acc_data['Year'] = acc_data['Start_Time'].dt.year
# Weekend column where 1 = weekend, 0 = weekday
#acc_data['Weekend'] = acc_data['Day'] <= 5
acc_data['Weekend'] = acc_data['Day'].map({0: 0, 1: 0, 2: 0, 3: 0, 4: 0, 5: 1, 6: 1})
```

- 2.4% of Temperature are NaN --> fill with mean temp
- 19.3% of Precipitation are NaN --> fill with 0, meaning no rain
- 2.5% of Humidity are NaN --> fill with mean
- 2.0% of Pressure are NaN --> fill with mean
- 2% of visibility are NaN --> Fill with mean
- 16% of Wind Speed are NaN --> fill with mean
- 2% of Weather Condition are NaN --> Drop these rows
- 0.1% of Sunrise_Sunset are NaN --> Drop these rows
- 0.004% of city are Nan --> Drop these rows

```
[ ] data['Temperature(F)'] = data['Temperature(F)'].fillna(data['Temperature(F)'].mean())
    data['Precipitation(in)'] = data['Precipitation(in)'].fillna(0)
    data['Humidity(%)'] = data['Humidity(%)'].fillna(data['Humidity(%)'].mean())
    data['Pressure(in)'] = data['Pressure(in)'].fillna(data['Pressure(in)'].mean())
    data['Visibility(mi)'] = data['Visibility(mi)'].fillna(data['Visibility(mi)'].mean())
    data['Wind_Speed(mph)'] = data['Wind_Speed(mph)'].fillna(data['Wind_Speed(mph)'].mean())
    data = data[data['Weather_Condition'].notna()]
    data = data[data['Sunrise_Sunset'].notna()]
    data = data[data['City'].notna()]
```

```
# convert the Start_Time & End_Time Variable into Datetime Feature
acc_data.Start_Time = pd.to_datetime(acc_data.Start_Time)
acc_data.End_Time = pd.to_datetime(acc_data.End_Time)
```

Apriori algorithm

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
112	(No delay or block)	(Day, Severity_2, Clear)	0.898313	0.338422	0.304684	0.339173	1.002219	0.000675	1.001136	0.021772
88	(No delay or block)	(Day, Severity_2, 05)	0.898313	0.338045	0.309207	0.344209	1.018233	0.005537	1.009399	0.176094
110	(Day)	(Severity_2, No delay or block, Clear)	0.852983	0.349543	0.304684	0.357198	1.021901	0.006530	1.011909	0.145773
111	(Severity_2)	(Day, No delay or block, Clear)	0.846197	0.346433	0.304684	0.360062	1.039342	0.011533	1.021298	0.246114
86	(Day)	(Severity_2, No delay or block, 05)	0.852983	0.346810	0.309207	0.362501	1.045245	0.013385	1.024614	0.294432
...	...	...	...	...	...	...	...	...	...	...
13	(Clear)	(No delay or block)	0.459335	0.898313	0.419847	0.914034	1.017500	0.007221	1.182869	0.031811
77	(Day, Severity_2, 05)	(No delay or block)	0.338045	0.898313	0.309207	0.914692	1.018233	0.005537	1.191997	0.027051
21	(Day, 05)	(No delay or block)	0.376025	0.898313	0.345773	0.919549	1.023640	0.007985	1.263959	0.037011
33	(Severity_2, 05)	(No delay or block)	0.376402	0.898313	0.346810	0.921382	1.025680	0.008683	1.293432	0.040150
3	(05)	(No delay or block)	0.432005	0.898313	0.400999	0.928229	1.033302	0.012924	1.416817	0.056741

124 rows × 10 columns

On sorting the rules based on confidence:

- Severity 2 has very little impact on blockages and delays.
- Day and month had a greater confidence with Severity 4
- Accidents in Day time of May are less likely to cause delays or blocks.

Prezi

Decision tree classifier with SMOTE

	precision	recall	f1-score	support
0	0.82	0.94	0.88	44864
1	0.79	0.77	0.78	44799
2	0.62	0.62	0.62	45575
3	0.62	0.55	0.58	45106
accuracy			0.72	180344
macro avg	0.71	0.72	0.71	180344
weighted avg	0.71	0.72	0.71	180344

- Decision tree classifier despite balanced classes done with SMOTE gave a low accuracy of 71%
- Model couldn't learn how to classify based on the patterns in the data like bumps, crossing, signals, etc. which are the main features in road traffic data.

Prezi

Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa)

	precision	recall	f1-score	support
0	0.73	0.80	0.76	10
1	1.00	0.70	0.82	10
2	0.75	0.90	0.82	10
3	1.00	1.00	1.00	10
accuracy			0.85	40
macro avg	0.87	0.85	0.85	40
weighted avg	0.87	0.85	0.85	40

- Our proposed BERT model for solving this problem showed a greater accuracy 85% with few training data and computational resources.
- Fatal accidents are perfectly classified based on F1 Score.

```
output type:  <class 'transformers.modeling_outputs.SequenceClassifierOutput'>
tensor([[    0, 43725,  1703,  ...,     1,     1,     1],
        [    0,  3750,   382,  ...,     1,     1,     1],
        [    0, 43725,  1703,  ...,     1,     1,     1],
        ...,
        [    0, 43725,  1703,  ...,     1,     1,     1],
        [    0, 11428,  7878,  ...,     1,     1,     1],
        [    0,  3750, 12016,  ...,     1,     1,     1]], device='cuda:0')
output type:  <class 'transformers.modeling_outputs.SequenceClassifierOutput'>
Epoch: 100%|          | 3/3 [12:57<00:00, 259.28s/it]Train loss: 0.3934102550866132
```

A loss of 0.3 is observed that could account to the less epochs it is trained for. Hence it is expected to perform more better having trained it for longer with powerful resources

Prezi

Creativity

- Classifying accident severity has been a challenge in research for a long time, and traditional ML algorithms have not shown good performance when considering all important features.
- RoBERTa, a classification model using BERT and transfer learning, is proposed as a solution, typically used for NLP problems.
- Processing accident features as contextual sentences using RoBERTa allows for more relevance to all details and better learning of nuances, even with limited training data

Prezi

Future Scope

- Conduct similar analyses in different countries and examine driver and vehicle characteristics to gain insights into psychological factors influencing driving behavior.
- Integrate study findings into a real-time accident risk prediction model.
- Implement the BERT classification model as an accident help bot, providing contextual and conversational insights into accident severity and preparation for handling the situation.

Any Questions?

Prezi

THANK YOU

Prezi