

ANALYSING CRASH DATA AND SEVERITY PREDICTION

Table of Contents

- Introduction
- Dataset Overview
- Research Questions
- Data Exploration
- Modeling Approach
- Evaluation
- Recommendations
- Next Steps
- Conclusions

INTRODUCTION

- This project involves a thorough analysis of traffic crash data to gain valuable insights and create predictive models for the severity of traffic accidents. The data used in this study was obtained from the Chicago Police Department (CPD) and covers incidents within the City of Chicago. The primary goals of this project are to contribute insights that can be used to enhance road safety and reduce the frequency and impact of traffic accidents.

DATASET OVERVIEW

The dataset used in this project contains information on a variety of attributes related to traffic crashes, such as weather conditions, road surface characteristics, vehicle types involved, and injury classifications. This comprehensive dataset provides the foundation for the exploratory data analysis and predictive modeling efforts undertaken in this project.

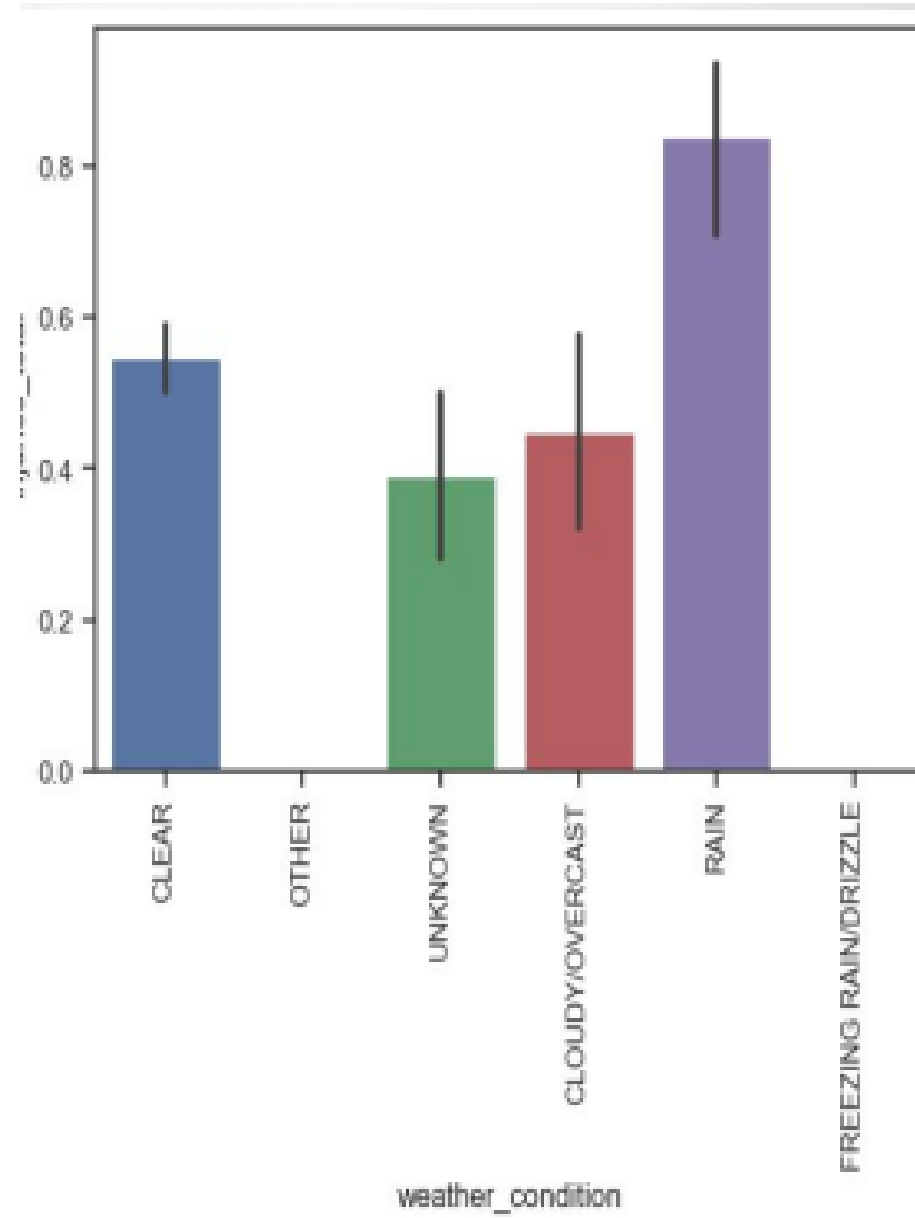
RESEARCH QUESTIONS

Our project addresses the following key research questions:

- What are the key factors that significantly influence the severity of traffic crashes, and can these factors be used to accurately predict the level of crash severity?
- How does the usage of safety equipment, such as seatbelts or airbags, impact the occurrence and outcomes of traffic crashes?
- Can the analysis identify geographic areas or "hotspots" with higher frequencies of traffic crashes, and can this information be used to recommend targeted prevention strategies?

DATA EXPLORATION

The initial exploration of the traffic crash dataset has revealed several important patterns and insights. The analysis uncovered correlations between various factors, such as weather conditions and crash severity. Temporal patterns were also observed, with certain hours and days of the week exhibiting higher frequencies of crash occurrences. Additionally, the investigation examined the influence of road surface conditions and defects on the likelihood and severity of traffic crashes.



MODEL APPROACH

- The project's modelling approach combined exploratory data analysis, data preprocessing, and the application of several machine learning techniques. To address the research questions and predict outcomes, the team utilized a variety of models, including Random Forest Classifier, Decision Tree Classifier, Logistic Regression, and K-Nearest Neighbors. This multi-pronged modelling approach allowed the researchers to explore the data from different perspectives and develop predictive capabilities for the various research objectives.

MODEL 1

- The first predictive model developed in this project was focused on forecasting the severity of traffic crashes. To achieve this, the researchers leveraged a variety of features, including the posted speed limit, prevailing weather conditions, and the age of the individuals involved in the crashes.
- This model demonstrated strong predictive performance, achieving an overall accuracy of approximately 90%. Additionally, the researchers provided detailed evaluation metrics for the model, such as precision, recall, and F1-score, for each of the different severity classes (e.g., minor, moderate, or severe injuries). These comprehensive performance metrics allow for a more nuanced understanding of the model's capabilities in accurately classifying the various levels of crash severity.
- The high accuracy and robust evaluation metrics suggest that this model can be a valuable tool for transportation authorities and policymakers to identify the key factors that contribute to the severity of traffic crashes. By understanding these influential factors, targeted interventions and safety measures can be implemented to mitigate the impact and frequency of severe crashes.

MODEL 2

- The second predictive model developed in this project focused on understanding the impact of safety equipment usage, such as seatbelt and airbag deployment, on the outcomes of traffic crashes. By incorporating data on the utilization of these safety features, the researchers aimed to gain insights into how they influence the likelihood and severity of injuries sustained in crashes.
- This model demonstrated promising results in its ability to predict the likelihood of injuries based on the usage of safety equipment. The evaluation metrics, particularly the strong recall and precision scores, indicate that the model was able to accurately identify cases where safety equipment usage was associated with reduced injury severity.
- These findings underscore the importance of safety equipment in mitigating the consequences of traffic crashes. The model's performance suggests that it could be a valuable tool for transportation authorities and policymakers to promote and encourage the use of seatbelts, airbags, and other safety features among drivers and passengers. By understanding the protective effect of these safety measures, targeted educational campaigns and enforcement efforts can be implemented to further enhance road safety and reduce the burden of traffic-related injuries.

MODEL 3

- The third model developed in this project focused on identifying geographic hotspots or areas with higher frequencies of traffic crashes. Leveraging Logistic Regression, the researchers were able to predict the likelihood of crash occurrences at specific locations. This predictive capability allowed the project team to pinpoint the areas that would benefit most from targeted safety interventions.
- By mapping the geographic hotspots and understanding the factors contributing to the elevated crash frequencies in these areas, the researchers were able to recommend prevention strategies tailored to the unique characteristics of each location. This granular, location-based approach enables transportation authorities and city planners to allocate resources and implement safety measures more efficiently, addressing the specific needs and vulnerabilities of high-risk areas.
- The insights gained from this model can inform a range of interventions, such as infrastructure improvements, traffic calming measures, enhanced enforcement, or educational campaigns. By focusing on these geographic hotspots, the project's recommendations have the potential to significantly improve road safety and reduce the overall frequency and impact of traffic crashes within the city.

Model Performance Comparison

- The research team conducted a comprehensive comparative analysis of the predictive models developed in the project. This analysis aimed to assess the strengths and limitations of each modeling approach by evaluating a range of performance metrics.
- The researchers measured and compared the accuracy, precision, recall, and F1-score of the various models to gauge their reliability and effectiveness in addressing the project's research questions. These metrics provided a multi-faceted assessment of the models' capabilities, allowing the team to understand the trade-offs and nuances in their predictive power.
- By thoroughly evaluating the models' performance across these key statistical measures, the researchers were able to gain deeper insights into the suitability and applicability of each approach. This comparative analysis enabled the identification of the model(s) that demonstrated the most robust and consistent predictive performance, as well as areas where further refinement or alternative techniques might be needed to improve the overall reliability of the forecasting capabilities.
- The detailed comparative evaluation ensures that the project's findings and recommendations are grounded in a rigorous assessment of the models' strengths and limitations. This level of analytical rigor enhances the credibility of the research and provides transportation authorities and policymakers with a comprehensive understanding of the predictive tools available to support data-driven decision-making and targeted interventions.

Conclusion

1. The predictive model developed in this project exhibited strong performance in accurately forecasting the severity of traffic accidents, with a particular emphasis on correctly identifying cases involving incapacitating injuries. The high accuracy rate achieved by the model underscores its capability to reliably distinguish between different levels of crash severity, providing valuable insights that can inform targeted interventions and resource allocation. By accurately pinpointing accidents with incapacitating injuries, the model enables transportation authorities and emergency response teams to prioritize and allocate resources more effectively. This granular understanding of crash severity can support the implementation of tailored safety measures, such as infrastructure improvements, enhanced enforcement, or targeted public awareness campaigns, to mitigate the risk and impact of most severe crashes. The model's robust predictive power in this domain suggests that it can be a reliable tool for transportation planners and decision-makers to enhance road safety and improve overall outcomes for the community.
2. Ensuring road safety in the Chicago area is an enduring challenge that demands ongoing attention and the refinement of strategies based on data-driven insights and evolving circumstances. Given the dynamic nature of transportation patterns, infrastructure changes, and shifting societal factors, maintaining safe roads and reducing the frequency and severity of crashes requires a proactive and adaptive approach. Continuous monitoring and evaluation of the effectiveness of safety interventions, as well as the identification of emerging trends or emerging risk factors, are crucial for informing the continuous improvement of safety strategies. By leveraging the insights gained from the predictive models developed in this project, as well as monitoring and analyzing real-time data on traffic crashes and safety indicators, transportation authorities and policymakers in the Chicago area can stay attuned to the evolving landscape. This will enable them to make data-driven decisions, allocate resources efficiently, and implement impactful safety measures that adapt to the changing needs and conditions of the local community. Embracing this cyclical process of data-driven evaluation, strategy refinement, and proactive implementation will be key to enhancing road safety and reducing the burden of traffic-related incidents in the Chicago metropolitan area over the long term.

Recommendations

1. **Maintain Road Conditions:** Continuously monitor and promptly address road surface defects, such as potholes, cracks, and uneven surfaces, to ensure the road infrastructure is well-maintained and safe for all users.
 2. **Optimize Road Layouts:** Leverage the insights gained from crash data analysis to identify high-risk areas and redesign or enhance the layouts of roads in those locations, to improve safety and reduce the frequency and severity of crashes.
 3. **Collaborative Enhancements:** Foster partnerships with the automotive industry and educational institutions to develop technological advancements that can enhance vehicle stability and integrate comprehensive road safety education measures into school curriculums. This approach aims to promote safe road use behaviours among the community, starting from a young age.
- By proactively addressing road conditions, implementing data-driven infrastructure improvements, and collaborating across sectors to enhance both vehicles and education, transportation authorities can take a multifaceted approach to improving road safety and reducing the burden of traffic-related incidents within the community.

Next Steps

1. **Incorporate Real-Time Data.** Leverage real-time data sources, such as connected vehicle technologies, traffic sensors, and incident reporting systems, to obtain a more accurate and up-to-date understanding of current road conditions and traffic patterns. Integrate this real-time data into the predictive models to enhance the accuracy and timeliness of the accident severity forecasts, enabling more effective and responsive interventions. Develop data aggregation and analysis platforms that can quickly process and synthesize real-time information, providing transportation authorities with actionable insights in near-real-time.
2. **Expand the Analysis:** Broaden the scope of the model evaluation and comparative analysis to include a wider range of variables, such as road surface conditions, weather data, and driver behaviour metrics (e.g., speeding, distracted driving, impaired driving). Investigate the influence of these additional factors on accident severity and incorporate them into the predictive models to gain a more comprehensive understanding of the underlying dynamics. Conduct sensitivity analyses to determine the relative importance of each variable and how they interact to influence the model's predictive power.
3. **Collaborate with Local Authorities:** Engage closely with local transportation agencies, law enforcement, and emergency response teams to ensure the recommended strategies align with their operational needs and priorities. Jointly develop implementation plans for the suggested safety interventions, such as infrastructure improvements, enforcement initiatives, and public education campaigns. Establish data-sharing agreements and feedback loops to monitor the real-world effectiveness of the implemented strategies, enabling continuous refinement and optimization. Facilitate knowledge-sharing and capacity-building workshops to empower local stakeholders in adopting and sustaining the recommended approaches.

By incorporating real-time data, expanding the analytical scope, and collaborating with local authorities, the research team can further enhance the relevance, accuracy, and impact of the predictive models and safety recommendations. This comprehensive approach will enable transportation decision-makers to make more informed, data-driven choices and implement effective, sustainable solutions to improve road safety in the community.