Project 4: Enhancing Cyclist Safety in Austin through Data-Driven Risk Assessment

Pei-Yu Jheng

ALY6040 – Data Mining Applications, Northeastern University

Professor Justin Grosz

March 24, 2024

Problem Introduction

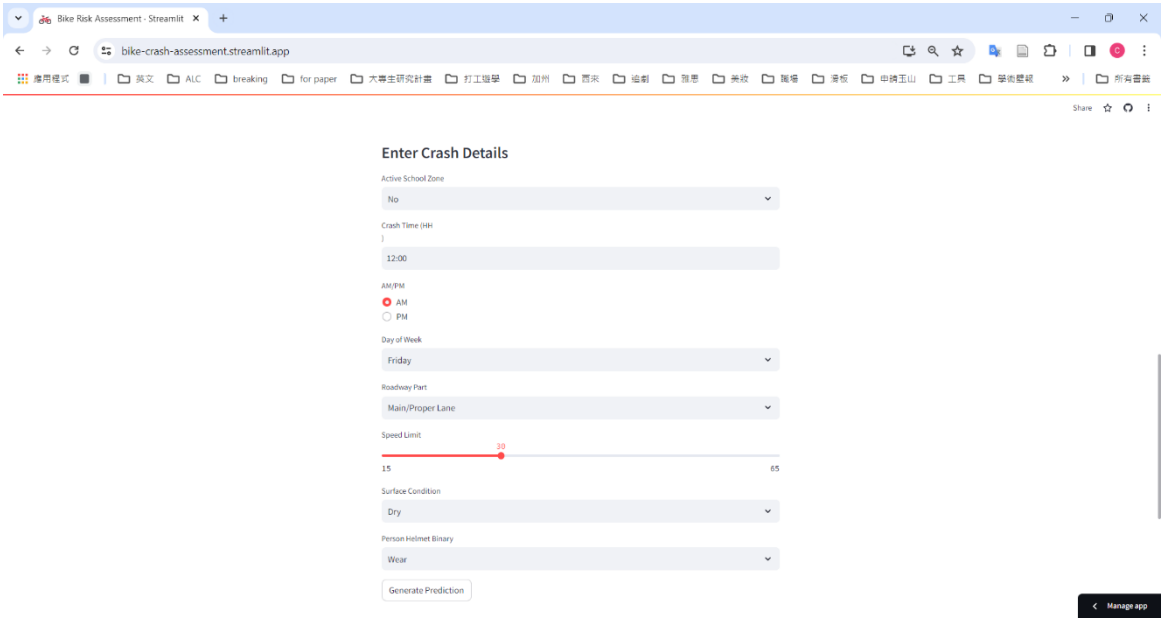
The city of Austin has been facing concerns from its cycling community regarding the lack of adequate protection from motor vehicles. To address these complaints and gain insights into the issue, the city has compiled a dataset of cyclist-related incidents. We have developed a data-driven solution to assess the risk level for cyclists in Austin based on various factors. This report summarizes the developed product, its functionalities, and how it contributes to resolving the challenges faced by Austin's biking community.

Seven columns were dropped from the DataFrame because they contain details of the event that are not relevant to the prediction task. The remaining variables, such as 'Active School Zone Flag', 'Crash Severity', 'Crash Time', 'Day of Week', 'Roadway Part', 'Speed Limit', 'Surface Condition', and 'Person Helmet', are more directly related to the prediction and are therefore retained for analysis.

Summary of the Developed Product

The link to the Bike Risk Assessment app: is https://bike-crash-assessment.streamlit.app/

Figure 1



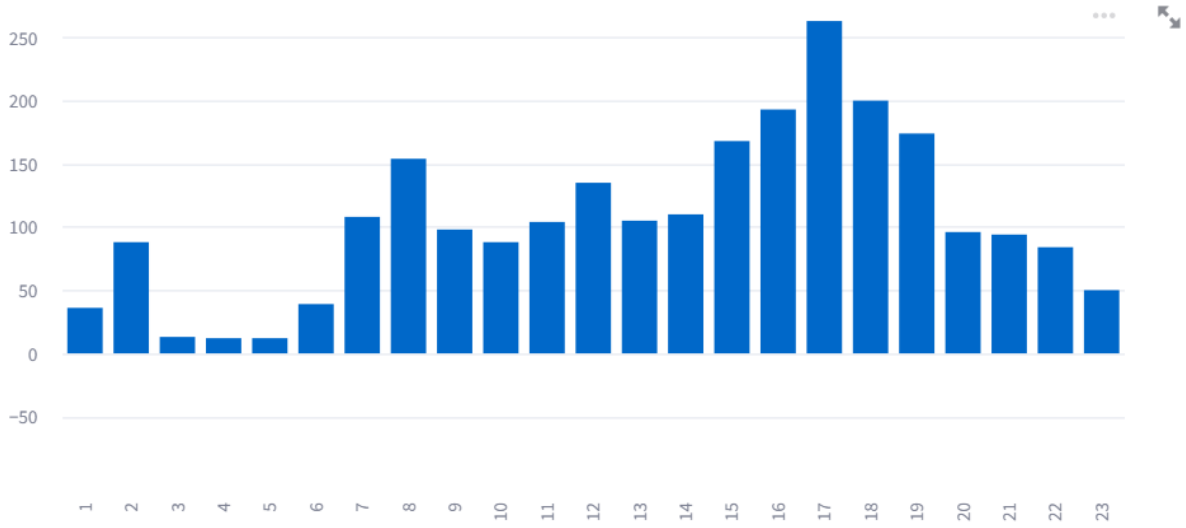
The product is a Streamlit application that analyzes the provided bike crash dataset and predicts the risk level for cyclists in Austin. The application allows users to input relevant information such as active school zone, day of the week, roadway part, speed limit, surface condition, helmet usage, and hour of the day. Based on these inputs, the application utilizes a trained decision tree model to predict the risk level (low or high) for a potential bike crash.

Product Functionalities

The application offers the following functionalities:

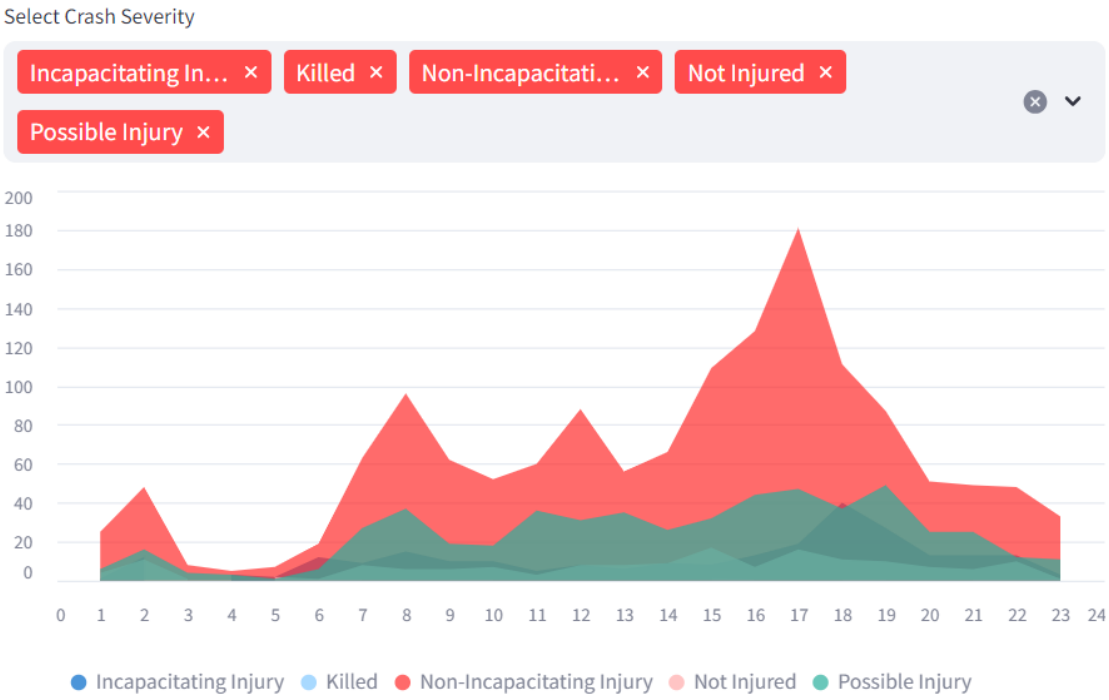
1. Crash Distribution by Hour: A bar chart visualizes the distribution of bike crashes across different hours of the day, enabling users to identify peak hours for crashes.

Figure 2 – Crash Distribution by Hour



2. Crash Severity by Hour: An area chart presents the crash severity levels (e.g., incapacitating injury, non-incapacitating injury) across different hours. Users can select specific severity levels to focus on.

Figure 3 – Crash Severity by Hour

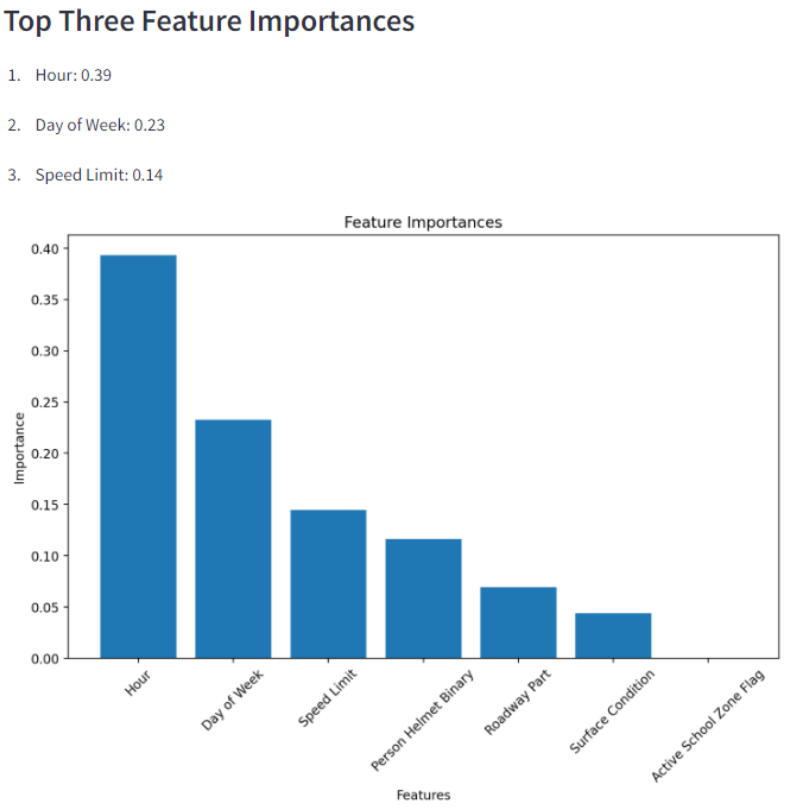


3. Risk Prediction: Users can input details such as active school zone, crash time, day of the week, roadway part, speed limit, surface condition, and helmet usage. Based on these inputs, the application predicts the risk level (low or high) for a potential bike crash.

Figure 4

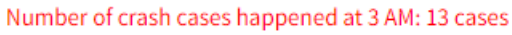
4. Feature Importances: The application displays the top three features that have the most significant impact on risk prediction, providing insights into the key factors influencing bike crashes.

Figure 5



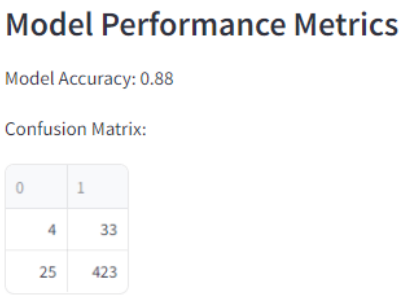
5. Crash Count: The application shows the number of crash cases that occurred at the specified crash time, highlighting the historical data for that particular time.

Figure 6



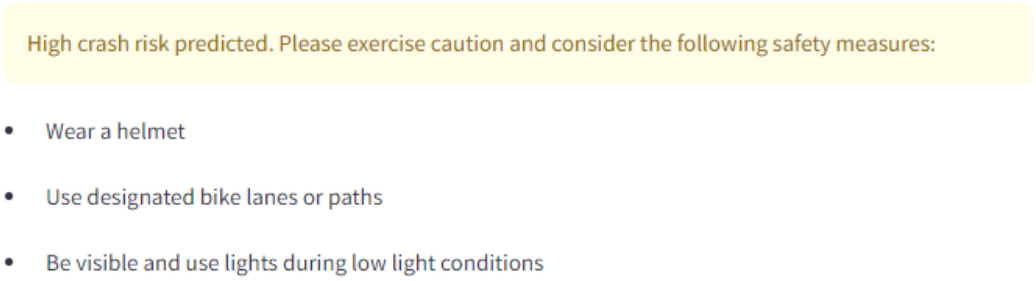
6. Model Performance Metrics: The accuracy and confusion matrix of the trained decision tree model are displayed, allowing users to assess the model's performance.

Figure 7



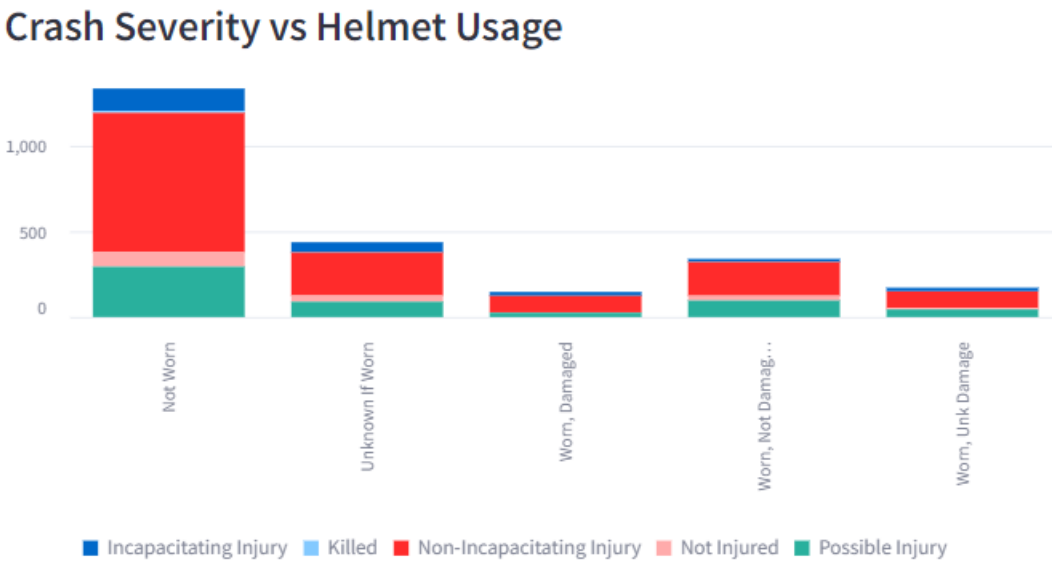
7. Safety Recommendations: Based on the predicted risk level, the application provides safety recommendations for cyclists.

Figure 8



8. The "Crash Severity vs Helmet Usage" chart: This visualization emphasizes the importance of helmet usage in mitigating the severity of injuries in the event of a bike crash.

Figure 9



Conclusion

The developed Streamlit application serves as a powerful tool for assessing the risk level for cyclists in Austin. By leveraging historical bike crash data and user inputs, the application

provides data-driven insights to help cyclists make informed decisions and prioritize their safety. The application's functionalities, visualizations, and safety recommendations contribute to addressing the concerns raised by Austin's biking community and support the city's efforts in enhancing cyclist protection.

By utilizing this application, city officials can identify high-risk areas and factors, guiding their efforts to improve cycling infrastructure and implement targeted safety measures. The insights gained from the application can also facilitate data-driven decision-making and resource allocation to create a safer environment for cyclists in Austin

Reference

Data Source:

ALY6040 Data Mining Application Course of Northeastern University,

https://northeastern.instructure.com/courses/165136/assignments/2071170