Northeastern University

WEEK 5

ALY 6040 20305 - Data Mining Applications

Winter 2024 CPS Quarter - Second Half

Assignment 5: Austin Bike Crash Severity App

Submitted By

Zeeshan Ahmad Ansari

Submitted To

Justin Grosz

Executive Summary

The city of Austin, known for its vibrant outdoor culture, has faced rising concerns from the cycling community regarding safety on the roads. Cyclists report feeling unprotected against motor vehicles, leading to increased incidents and accidents. In response to this, a detailed examination of cyclists' incident data was conducted. The core objective was to analyze the validity of these concerns and identify patterns that could inform preventive measures. A Streamlit application, leveraging a Random Forest Classifier enhanced with SMOTE for balanced learning from the data, was developed to predict and visualize the risk factors associated with cycling incidents in Austin.

Application Insights

The application processes input variables such as day of the week, Speed, Surface condition, Person wearing Helmet, time of the day, and other cycling conditions to predict incident severity. It employs a Random Forest Classifier with SMOTE, achieving an accuracy of approximately 67.34%. This predictive model is pivotal for understanding high-risk factors and times, providing both immediate insights for cyclists and long-term data for city planning and policy adjustments.

Key Features

Predictive Analysis: Users can input parameters such as day of the week, helmet usage, speed category, and more to receive a prediction of incident severity.

Risk Visualization: The application provides visualizations representing the factors contributing to incident risk, helping users understand the underlying causes.

Recommendations: Based on the analysis, the application suggests preventive measures to minimize risks for cyclists.

Addressing Cyclists' Concerns

The application directly addresses the cyclists' complaints by providing evidence-based insights into the factors leading to incidents. By analyzing the compiled data, it was confirmed that specific conditions and times significantly increase the risk for cyclists. For instance, high-traffic areas, lack of helmet usage, and certain times of the day were identified as high-risk factors.

How the Application Helps

Awareness: Cyclists can make informed decisions about their routes and riding times, potentially avoiding high-risk situations.

Policy Implications: The city can use the insights to implement targeted safety measures, such as enhanced traffic regulations, infrastructure improvements, and awareness campaigns.

Community Engagement: The interactive nature of the application encourages community participation, allowing cyclists to contribute data and share experiences.

What does the Product do?

The core functionality of the product is to predict the risk level of cycling at a given time based on historical incident data. The prediction model categorizes risk into three levels: low, medium, and high. It uses a Random Forest Classifier, enhanced with SMOTE to address class imbalance, ensuring more accurate and reliable predictions.

Visualization and Risk Determination

A crucial aspect of the application is its ability to visualize data, making it easier for users to comprehend how different factors contribute to the risk level. The analysis revealed that incidents peak during rush hours and in areas with inadequate cycling infrastructure. These visualizations play a vital role in advocating for safer cycling conditions in Austin.

Conclusion

The Streamlit application developed for analyzing cyclists' safety in Austin provides a dynamic platform for understanding and mitigating biking risks. It not only confirms the community's complaints but also offers actionable insights for both cyclists and city planners. By integrating data analysis with user-friendly visualizations, the application fosters a data-driven approach to enhancing cycling safety in urban environments.

APPENDIX

Screenshot of the application.