Chapter5

Conclusion

5.1 Summary

Through the analysis of historical data of traffic accidents in Malaysia and the construction of machine learning models, this study reveals the main factors affecting the occurrence of accidents as well as high-risk groups and periods. The results of the analysis show that motorcycle riders are the group with the highest risk of accidents, accounting for the majority of fatalities, which is related to the motorcycle's lack of protection and driving behavior. The accident rate is significantly higher at night and during peak hours, which further verifies the important role of time characteristics in traffic accidents. The positive correlation between the number of motor vehicle registrations and the risk of accidents suggests that increased traffic may be a potential driver of the increased risk of accidents. While the model has limited predictive performance (about 9.7% accuracy), feature importance analysis provides policymakers with key risk insights, such as enhancing the management of motorcycle riders and optimizing traffic flow during peak hours.

5.2 The significance of the findings

The results of this study have significant practical significance and application value: First, the analysis reveals that motorcycles and walkers as high-risk groups need special attention, and provides a scientific basis for the selection of target groups for traffic safety education; Second, the association of time characteristics (such as night and peak hours) with accident risk suggests that traffic regulation and infrastructure optimization at these times should be strengthened, such as the deployment of more lighting equipment and smart traffic lights; Finally, the continuous increase in the number of motor vehicle registrations suggests that the government needs to plan safety management measures, such as the design of separated lanes and the reduction of lane intersections, in parallel with the development of the transport system. In addition, although the performance of the accident prediction framework constructed in this study is limited, it can provide data-driven support for traffic policy optimization through feature importance analysis.

5.3 Future works

Compared with previous studies, the results of this study support the observation of Mohd Khairul Amri et al. (2017) on the high death rate of night accidents in Malaysia. However, there is still room for improvement in the prediction accuracy of the random forest model in this study, which may be caused by insufficient data features or unbalanced categories. In addition, the study did not include environmental characteristics such as weather and road construction, which may limit the predictive power of the model.

Future work directions include:

1. Expand the data range to include environmental characteristics (such as weather and road conditions) and driving behaviors (such as speed and drunk driving records);

2. Explore deep learning methods (such as LSTM or graph neural networks) to capture the spatiotemporal dependence and complex patterns of traffic accidents;

3. Optimize feature engineering, further refine time and location features, and improve model prediction accuracy;

4. Carry out regionalized studies, focusing on the detailed characteristics and treatment measures of accident black spots. By combining more comprehensive data with advanced modeling techniques, accident prediction capabilities can be further improved to provide stronger support for accurate implementation of traffic safety policies.