Comparative Analysis of Machine Learning Models for Predicting Risky Driving Behaviors from a Naturalistic Experiment

8. Road safety and resilience

Stella Roussou[1] , *Virginia Petraki*[1], *George Yannis*[2]

[1] Ph.D. Candidate, Research Associate, Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece

[2] Professor, Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece

Background

Road safety is still a critical public health concern, and human behavior has been identified as the main cause of road crashes. However, vehicle-specific attributes also play a crucial role in influencing driving behavior. The Horizon 2020 i-DREAMS project was designed to study these interactions through a large-scale naturalistic driving experiment, aiming to determine how vehicle characteristics impact risky driving behaviors. This study extends existing research on the topic by comparing different machine learning models-CatBoost, Random Forest, and XGBoost-on the predictive capabilities for estimating the role of vehicle attributes in shaping driver risk profiles.

In this experiment, the data was collected from 48 drivers and 4,922 trips in Greece. Vehicle features were considered, including vehicle age, engine capacity, fuel type, and gearbox type, among others, along with driving event data. The analysis of the association between vehicle attributes and drivers' behavior may allow policymakers to implement targeted interventions toward improving road safety.

Machine learning models provide a robust framework for detecting risky driving behaviors by identifying patterns that conventional statistical methods might overlook. Boosting-based algorithms, such as CatBoost and XGBoost, have gained significant attention in predictive analytics lately because of their efficiency in handling categorical variables. Random Forest, while widely known for its interpretability, has demonstrated limitations in handling imbalanced datasets. This study, seeks to examine the performances of these models on predicting critical driving events with an emphasis on their practical implementation for road safety analysis.

Methods

Data collection included in-vehicle monitoring systems that recorded driving events such as harsh acceleration and deceleration. Additional vehicle characteristics, acquired from the participants questionnaire, enriched the dataset for analysis of its influence on driver risk. This research work utilized machine learning methods for the classification of risky and non-risky events, using different algorithms for comparison in their performance.

Initially, CatBoost was chosen because it has the highest capability for processing categorical features powerfully. Then, Random Forest and XGBoost models were developed to make the results more robust. Each model was tuned with a grid search using 5-fold cross-validation, targeting its best possible performance. For addressing class imbalance, the SMOTE technique was applied to ensure that the models could generalize across a range of driving event frequencies.

Feature selection was done based on the most relevant predictors of risky driving behavior. Each model was used to conduct feature importance analysis in order to identify which vehicle attributes had the greatest influence. Model performance metrics included accuracy, precision, recall, and F1-score. The confusion matrix was analyzed to assess misclassification rates, identifying potential weaknesses in each approach. The comparative analysis was designed to determine the optimal balance between predictive accuracy and interpretability, offering insights into the practical implementation of these models for road safety analysis.

Results

The feature importance analysis revealed that vehicle age, engine capacity, and intervention phase were the strongest predictors of risky driving events across all models. Although the CatBoost performed high on precision at 72% for acceleration and 60% for deceleration, XGBoost yielded a high recall of 75% in acceleration and 63% for deceleration, thus remaining more effective to detect risky driving behaviors. However, Random Forest, though highly interpretable, performed low, with an AUC of 68% and 58%, respectively, in the acceleration and deceleration of the participants.

Cross-validation confirmed that CatBoost maintained stable accuracy across different driving conditions. XGBoost showed some variability due to its sensitivity to outliers but demonstrated superior detection of high-risk behaviors. Random Forest, though robust in handling balanced datasets, struggled with rare risky event classification.

Studies have showed that vehicle age and engine capacity were always influential features for all models, which further confirmed their influence on unsafe driving. The analysis also emphasized how effective phased interventions were in shaping driver behavior, as model outputs demonstrated. Drivers who received post-trip feedback and gamification had fewer risky events, underlining the potential of data-driven interventions in improving road safety.

Beyond classification accuracy, the interpretability assessments using SHAP values confirmed that external interventions and vehicle-related parameters played a significant role in shaping model predictions. This study showed that machine learning techniques can be used to gain further insight into driver risk profiles that will inform future safety initiatives.

Discussion and Conclusion

This research highlights that the addition of vehicle-specific data is integral to any predictive modeling when the driving behavior analysis is concerned. Indeed, aging vehicles and high-performance cars become the target for interventions since their findings suggest these are more susceptible to having a risky event. Machine learning models provide rich insight into the creation of data-driven safety measures which complement traditional methods of analyzing road safety.

The superiority of XGBoost in classifying risky behavior was because it had higher recall among the tested models. On the other hand, CatBoost provided a very balanced approach where it kept its accuracy high along with efficient categorical feature handling. Random Forest is interpretable, but it proved limited when working with imbalanced data. It brings out both strengths and trade-offs for every model that can guide future applications in intelligent transportation systems.

Further research should enlarge the dataset to encompass a wider variety of vehicle types and road conditions. Consideration of real-time traffic conditions, weather, and even driver psychological factor integration could be done to bring more accuracy to the model predictions. Besides, deep learning approaches such as using LSTMs or transformers could be explored to model sequential driving behaviors more precisely.

It finds its practical application in the potential ability of such models to be embedded into ADAS for real-time evaluation of risks. These inputs can be employed by insurance companies for making customized risk profiles, encouraging responsible driving. Policymakers may also use these results in constructing regulatory policies that reduce crashes based on vehicle-specific safety interventions.