**EXAMINING THAILAND’S RECENT MOTORCYCLE HELMET COMPLIANCE CAMPAIGNS**

JEFFREY ORCUTT

University of Wisconsin – Green Bay

DS 785: Capstone

04/08/2025

**Abstract**

More than 1.7 million motorcyclists required hospital treatment from 2018 to 2023, according to data from the Thai Government’s Injury Surveillance System. The research sought to assess the effectiveness of helmet usage promotion and enforcement efforts to reduce accident-related hospitalizations in Thailand through random forest models. Partial dependencies from the models were analyzed to provide insights into the demographic and environmental correlational factors. Age proved to be the most important factor within the models. For the youngest riders, the predicted probability of wearing a helmet was relatively low, then rose sharply through the early-to-mid 20s to reach a peak, a predicted probability of helmet wearing of approximately 0.20, in the late 20s or early 30s. After this point, the curve gradually declines, settling for a more moderate probability for individuals in their 50s and 60s, though it remains higher than for the youngest riders.

Keywords: Thailand, motorcycles, helmet, random forest, machine learning

**Contents**

[**Examining Thailand’s Recent Motorcycle Helmet Campaigns' Efficacy** 7](#_Toc196555945)

[Background 7](#_Toc196555946)

[Project Objectives 8](#_Toc196555947)

[Significance of the Project 8](#_Toc196555948)

[**Literature Review** 9](#_Toc196555949)

[Risk Factors and Behavioral Patterns Contributing to Road Traffic Accidents 10](#_Toc196555950)

[Helmet Legislation and Community Engagement Interventions 14](#_Toc196555951)

[Helmet Use Disparities and Behavioral Influences on Road Safety 16](#_Toc196555952)

[Summary 19](#_Toc196555953)

[**Data Science Application** 20](#_Toc196555954)

[Data Overview 21](#_Toc196555955)

[Data Cleaning and Processing 22](#_Toc196555956)

[Exploratory Data Analysis 23](#_Toc196555957)

[***Overall Exploratory Data Analysis*** 23](#_Toc196555958)

[***2023 Increase Exploration*** 25](#_Toc196555959)

[***Motorcycle-Specific Exploratory Analysis*** 27](#_Toc196555960)

[***Helmets and Negative Outcomes Analysis*** 30](#_Toc196555961)

[Feature Engineering 30](#_Toc196555962)

[Model Development, Deployment, and Evaluation 31](#_Toc196555963)

[***Sample Selection, Preprocessing, and Feature Selection*** 32](#_Toc196555964)

[***Modeling Pipeline*** 32](#_Toc196555965)

[***Overall Dataset Modeling Results*** 34](#_Toc196555966)

[***2023 Modeling Results*** 36](#_Toc196555967)

[Summary 37](#_Toc196555968)

[**Discussion of Analysis** 38](#_Toc196555969)

[Interpreting Random Forest Models 38](#_Toc196555970)

[2018 – 2023 Results 39](#_Toc196555971)

[***Age*** 40](#_Toc196555972)

[***Location*** 42](#_Toc196555973)

[***Alcohol*** 43](#_Toc196555974)

[***Sex*** 45](#_Toc196555975)

[***Time of Day*** 46](#_Toc196555976)

[2023 Results 48](#_Toc196555977)

[***Age*** 49](#_Toc196555978)

[***Location*** 50](#_Toc196555979)

[***Alcohol*** 51](#_Toc196555980)

[***Sex*** 53](#_Toc196555981)

[***Time of Day*** 53](#_Toc196555982)

[Conclusion 54](#_Toc196555983)

[**Discussion** 55](#_Toc196555984)

[Summary of Findings 55](#_Toc196555985)

[**References** 59](#_Toc196555986)

[**Appendix** 63](#_Toc196555987)

**List of Figures**

[**Figure 1** Distribution of Hospitalizations by Vehicle Type (2018-2023) 22](#_Toc195708701)

[**Figure 2** Histogram of Injuries by Year (2018-2023) 23](#_Toc195708702)

[**Figure 3** 2019-2023 provincial increases compared to increases in 2022-2023. 24](#_Toc195708703)

[**Figure 4** Distribution of Motorcycle Hospitalizations by Age and Gender 25](#_Toc195708704)

[**Figure 5** Hospitalizations by Time and Helmet Status 26](#_Toc195708705)

[**Figure 6** Hospitalizations by Time and Alcohol Usage 27](#_Toc195708706)

[**Figure 7** Motorcyclist Head Injury Outcomes 28](#_Toc195708707)

[**Figure 8** Partial Dependence Plot for Age 38](#_Toc195708708)

[**Figure 9** Partial Dependence Plot of Provinces 40](#_Toc195708709)

[**Figure 10** Partial Dependence of Alcohol Status 42](#_Toc195708710)

[**Figure 11** Partial Dependence Plot of Sex 44](#_Toc195708711)

[**Figure 12** Partial Dependence Plot of Time Category 45](#_Toc195708712)

[**Figure 13** 2023 Partial Dependence Plot of Age 47](#_Toc195708713)

[**Figure 14** 2023 Partial Dependence Plot of Provinces 49](#_Toc195708714)

[**Figure 15** 2023 Partial Dependence Plot of Alcohol Use 50](#_Toc195708715)

**List of Tables**

[**Table 1** Random Forest Model Hyperparameter Table 31](#_Toc195769276)

[**Table 2** Comparison of Models Performance 33](#_Toc195769277)

[**Table 3** Model Feature Importance 37](#_Toc195769278)

[**Table 4** 2023 Model Feature Importance 46](#_Toc195769279)

# **Examining Thailand’s Recent Motorcycle Helmet Campaigns' Efficacy**

Thailand faced substantial public health and economic strain from motorcycle accidents, which resulted in high mortality rates and severe injuries. The World Health Organization (WHO) identified Thailand as one of the leading nations for traffic fatalities, where motorcycles cause more than 70% of road-related deaths (Thai Road Safety Coalition, 2024; World Health Organization, 2025). Although helmets serve as an effective prevention tool for traumatic brain injuries, their usage does not match legal expectations because enforcement obstacles and socioeconomic and behavioral barriers persist. The research sought to assess the overall effectiveness of helmet usage promotion combined with enforcement efforts to reduce accident-related hospitalizations in Thailand through an analysis of data and provide any insights from recent data.

## Background

The ubiquitous use of motorcycles for transportation in Thailand has resulted in rising traffic injuries and deaths (Nishi et al., 2018). There are more than 22,700,000 registered motorcycles in Thailand (Injury Prevention Division, Thailand Center for Disease Control, 2024b). Despite the Thai government's implementation of safety measures, such as helmet requirements and road safety programs, enforcing these measures remains insufficient. The Global Status Report on Road Safety reveals that although Thailand implemented helmet requirements for motorcycle drivers in 1994 and their passengers in 2006, their enforcement remains weak, especially in non-urban regions (World Health Organization, 2025). Passengers in rural areas frequently avoid wearing helmets because they lack awareness about helmet laws, face difficulties accessing quality helmets, and helmet law enforcement was often sparse or ineffective (Suriyawongpaisa et al., 2013).

The research examined the overall effectiveness of helmet-encouraging practices related to the number of accident hospitalizations. Stakeholders and policymakers could use the insights created to craft specific measures to boost adherence, which will lower both death rates and medical expenses resulting from motorcycle accidents.

## Project Objectives

The research used data science to study how helmet usage promotion and enforcement efforts correlate with accident-related hospitalizations in Thailand. The project studied hospitalization data sets to explore helmet usage adherence patterns. These data sets comprise hospitalization records for individuals involved in vehicle accidents requiring hospitalization. This study employed data science techniques, including extensive extract, transform, and load processes (ETL Processes), exploratory data analysis, and classification algorithms, to measure helmet law effectiveness and highlight compliance determinants. The research aimed to deliver evidence-based insights to policymakers and health organizations to boost helmet adoption rates and minimize motorcycle-related injuries.

## Significance of the Project

Achieving project objectives yielded important information about helmet compliance patterns, accident severity levels, and enforcement measure effectiveness. The findings could benefit evidence-based policy design across policymaking bodies, law enforcement agencies, and public health departments.

Key benefits may include societal and economic impacts such as:

* Reduced Road Traffic Fatalities**:** The application of data-driven policy recommendations could improve helmet compliance rates and decrease the number of preventable deaths.
* Lower Healthcare Costs: When severe motorcycle injuries decrease, the financial strain on hospitals and government healthcare programs should lessen.
* Improved Road Safety Policies**:** The study's findings may inform better education and enforcement resource distribution.
* Economic Impact**:** As motorcycle accidents and hospitalizations decrease, dependable workforces could benefit motorcycle transport-dependent businesses like food delivery services, motorcycle taxis, and ride-hailing platforms. Thailand's universal healthcare system covers medical expenses for most citizens, so decreasing severe motorcycle injuries may ease public healthcare financial stress and enable funds to be redirected to more essential medical requirements.

# **Literature Review**

Research on motorcycle safety and helmet use has become an essential aspect of traffic accident prevention studies, especially in areas experiencing high numbers of motorcycle-related deaths. This literature review examines existing research on motorcycle accidents and helmet use, structured into three key thematic areas: risk factors and behavioral patterns contributing to road traffic accidents, helmet legislation, community engagement interventions, helmet use disparities, and behavioral influences on road safety. These thematic areas provided a framework for this study to uncover insights from the completed analysis. The review structures this chapter to deliver a high-level analysis of motorcycle safety issues, which examines accident causes and evaluates injury reduction interventions.

The studies reviewed here encompass multiple thematic areas because they address a wide-reaching subject matter. Each study in this review has been allocated to its main topical area to ensure clarity and coherence. This approach organizes findings systematically, recognizing the interrelated effects of risk factors, policy interventions, and behavioral patterns on road safety results. The following sections analyze the main thematic categories through key studies, revealing trends and challenges and offering recommendations for enhancing motorcycle safety with evidence-based approaches.

## Risk Factors and Behavioral Patterns Contributing to Road Traffic Accidents

Motorcycle accidents in Thailand create substantial public health problems and disproportionately impact vulnerable groups such as children and food delivery drivers. The study by Chiangkhong et al. (2024) analyzed motorcycle-related fatalities among children under 15 years in Thailand and discovered that boys aged 10–14 years had disproportionately high death rates, which represented 64% of child motorcycle fatalities between 2015 and 2020. The link between motorcycle collisions and helmet non-use indicates a significant association with accidents intensified by prevalent alcohol consumption. Hossain and Iamtrakul's (2007) analysis showed that head injuries made up 39.18% of injury patterns, while helmet use during crashes occurred in just 9.2% of cases. Stephan et al. (2011) discovered that young males aged between 15 and 19 encounter the highest risk of transport-related injuries, which frequently result from alcohol use combined with a lack of helmet use. Every study emphasizes that helmet use can prevent deaths on the road.

Environmental conditions and systemic infrastructure deficiencies exacerbate accident severity, increasing the likelihood of fatal outcomes. Mahikul et al. (2022) researched what caused severe bus accidents and concluded that driver fatigue amplifies fatal accident risk by 3.40 times, whereas drunk driving increases fatality risk by 4.79 times. The lack of frontage roads and road slopes played a significant role in increasing accident severity. Baral and Kanitpong (2015) found that motorcyclists faced elevated danger during night-to-morning hours on weekends and holidays like Songkran and New Year, resulting in severe accidents. The primary risk factors for significant accidents involved collisions with larger vehicles along with alcohol consumption and accidents occurring at curve locations and junctions. A study by Sae-Tae et al. (2020) confirmed these results after examining 37,659 cases. They determined that helmetless male and elderly motorcycle riders experienced significantly higher risks of severe injury and death. Older car users who fail to wear seat belts experience higher mortality rates. Analysis through logistic regression demonstrated a direct correlation between not wearing helmets and avoiding seatbelts with increased fatality rates. Chiangkhong et al. (2024) also found that school-aged motorcyclists faced the highest accident risks between 3 p.m. and 8 p.m., with accident rates increasing significantly during semester breaks.

Kasantikul (2001) conducted an extensive study of motorcycle accidents in Bangkok, Thailand, interviewing more than 700 motorcyclists and 2300 motorcyclists to establish their knowledge and behaviors. Kasantikul's research found that rider error stands as the leading primary cause, and alcohol involvement preceded about 40% of these accidents. Accidents occurred frequently due to ineffective traffic strategies like improper following distances and excessive speeding. The study found that other vehicle drivers contributed to motorcycle accidents through dangerous turning maneuvers across motorcyclists' paths or by neglecting traffic control rules. Accidents initiated by mechanical failures seldom occur but are mostly linked to insufficient maintenance instead of built-in motorcycle defects.

Kasantikul concentrated on the protective performance of helmets and their usage regulation in the study. During the accidents, the study observed that about 66% of motorcyclists wore helmets, substantially decreasing during nighttime, while around 30% of passengers used protective helmets. Riders confessed they wore helmets primarily in places where they expected police enforcement. The research highlighted the importance of implementing compulsory helmet regulations alongside regular safety inspections of available helmets to maintain safety standards. Kasantikul supported public education campaigns to raise awareness about the protective advantages of helmets and additional safety equipment like eye protection, which lowered accident involvement rates.

The research determined environmental and roadway factors as primary influences on motorcycle accident rates. The number of nighttime crashes rose because of roads that received insufficient maintenance and inadequate traffic control measures. The issues described did not directly result in accidents but were important contributing factors in more serious injuries. The research suggested that motorcycle driver education, road infrastructure enhancements, and better lighting conditions could mitigate these risks. Through comparative analysis, Kasantikul determined essential risk factors for motorcycle accidents and effective preventive measures by studying riders who experienced accidents versus those who safely traveled the same routes. The study advanced motorcycle safety knowledge through comprehensive methods, incorporating on-scene investigations, exposure data analysis, and interviews.

Kumphong et al. (2018) analyzed helmet usage factors among motorcyclists in Khon Kaen City, a major northern city, through logistic regression analysis of CCTV footage at arterial road intersections. Helmet usage statistics revealed that only 67% of motorcyclists wore helmets, as determined by various influential factors. Helmet wearing was 2.7 times more common among riders than passengers, and adults wore helmets 2.8 times more frequently than children. Motorcycle engine size affected helmet usage as riders on bikes with engine displacements greater than 125 cc wore helmets 1.9 times more frequently than those on smaller bikes. Motorcycle riders demonstrated maximum helmet usage during morning and afternoon hours because visibility and law enforcement presence were higher than in evening hours, which showed reduced compliance with helmet use.

Observations at intersections with police booths resulted in a 2.2-fold increase in helmet-wearing amongst riders. Riders who ran red lights wore helmets at 0.6 times the rate of law-abiding riders, suggesting risk-taking behavior leads to helmet law violations. While gender did not significantly impact helmet usage patterns, research demonstrated that location-based enforcement measures were crucial to helmet use. The study emphasized the necessity of strict enforcement measures at intersections during evening hours and implementing surveillance technology, including CCTV systems, for compliance assurance. Public awareness campaigns focused on high-risk groups like child passengers and red-light runners emerged as vital strategies for better helmet use and reduced motorcycle-related deaths in Thailand.

Prakobkarn et al. (2024) investigated how motorcycle delivery drivers in Bangkok faced a 35.1% accident rate among surveyed participants. The research showed that accident risks rose due to improper driving methods like improper or incorrect helmet use, excessive delivery quotas, and other unsafe behaviors, which multiplied risks 1.754 times. Research findings demonstrated substantial positive connections among knowledge levels, safety attitudes, and safety practices, highlighting education and behavioral change as essential strategies for preventing accidents.

The research highlights how road traffic injuries stem from multiple factors that require precise interventions. Research studies repeatedly identify systemic failures, which consist of insufficient traffic safety education in schools, inconsistent enforcement of helmet laws, and inadequate protection for high-risk individuals like children, food delivery drivers, and older motorcyclists. Research evidence backs targeted measures that include stricter helmet regulation enforcement, public education programs, infrastructure changes to add motorcycle-specific lanes, and improved road signage. These studies build a solid foundation for policy development to decrease preventable road traffic deaths and injuries in Thailand despite their methodological limitations, including secondary data reliance and self-reported behaviors.

## Helmet Legislation and Community Engagement Interventions

Investigating Thailand's helmet legislation and enforcement effectiveness highlights both positive outcomes and existing limitations of such approaches. Ichikawa et al. (2003) analyzed Thailand's helmet law by reviewing hospital statistics, which revealed a fivefold rise in helmet-wearing-injured motorcyclists and a 41.4% decline in head injuries after implementing the law. Helmet use diminished injury severity but failed to lower mortality rates because enforcement weaknesses at night and incorrect helmet use persisted. Research suggests that additional steps like public information campaigns and enhanced helmet quality requirements must accompany helmet laws to achieve better protection from motorcycle accidents. Additionally, Ichikawa et al. (2003) recognized systemic barriers such as fragmented trauma data and the narrow range of hospital research that could impact the broader understanding of policy implementations. Research supports combining helmet regulations with quality standards, public awareness campaigns, and strict enforcement to enhance road safety interventions.

Alternative methods to improve helmet use emerge from community-driven approaches for helmet compliance. An analysis by Ratanavaraha and Jomnonkwao (2013) showed that helmet-wearing rates rose by 13.23% in Nakhonpathom Province after a year-long initiative of educational programs and public meetings enabled community-based decision-making. Helmet usage exhibits noticeable differences across demographic groups, with younger riders and female motorcyclists showing lower compliance rates. The study reveals that road safety improves when pairing enforcement measures with targeted community interventions for specific high-risk groups. Research reveals that participatory approaches yield benefits but recognize demographic resistance and sustainability concerns as fundamental obstacles requiring ongoing community participation and policy support. This study promotes discussions about integrated road safety programs incorporating educational activities, enforcement strategies, and community empowerment initiatives.

Educational interventions demonstrate how modifying behavior enhances motorcycle safety. Swaddiwudhipong et al. (1998) evaluated a community-based motorcycle rider education program in rural Thailand. They found that intervention area participants achieved higher valid license ownership (69.7% vs. 46.5%) and helmet usage rates (46% vs. 20.5%) than those in control areas. The educational program achieved reduced yearly injury rates, demonstrating the effectiveness of structured educational initiatives in enhancing safety behaviors of at-risk communities. However, like Ichikawa et al., the study revealed that educational interventions did not lead to a significant decrease in fatality rates, which indicates that education itself cannot fully address motorcycle-related dangers. As with other studies, there are continuing risks associated with alcohol consumption, whether by the motorcyclist or other drivers, alongside low rates of helmet use, demonstrating the need to combine educational efforts with stronger legal enforcement measures and public awareness activities.

Lohitakul, in 2008, used funding from the World Bank to focus on 111 villages and schools in northeastern Thailand to educate children and provide appropriate motorcycle helmets. One of the main takeaways from the study initially is that at least 96% of parents thought that it was either not very important or not important at all for their children to wear helmets while on a motorcycle.

These studies further illustrate the need for a holistic approach to motorcycle safety in Thailand. While helmet legislation (Ichikawa et al., 2003) provides a necessary legal framework, community engagement (Ratanavaraha & Jomnonkwao, 2013) and educational programs (Lohitakul, 2009; Swaddiwudhipong et al., 1998) play essential roles in fostering behavioral change. A multi-layered intervention strategy integrating sustained legislative enforcement targeted public awareness campaigns, and localized community involvement is essential to improving motorcycle safety and reducing injuries and fatalities.

## Helmet Use Disparities and Behavioral Influences on Road Safety

Thailand's motorcycle helmet regulations have effectively reduced fatalities, but substantial differences in helmet usage remain across the country. Nishi et al. (2018) stated that motorcycles drive 73% of traffic deaths in Thailand, and helmet use is uneven throughout the nation, with a 44% average rate recorded in 2010. Helmet compliance rates show significant geographic and demographic contrasts because Bangkok achieves rates above 80% while rural areas such as Khon Kaen face much lower adherence, especially among teenage populations. The research demonstrates that achieving a 90% rate of helmet use among motorcyclists could lead to a 23% reduction in motorcycle-related fatalities and emphasizes the necessity of law enforcement and awareness campaigns. Progress is being deterred by behavioral resistance alongside insufficient law enforcement and inadequate research into targeted interventions. Nishi et al. (2018) recommended a community-specific mixed intervention strategy that supports international road safety objectives during the Decade of Action for Road Safety (2011–2020). They emphasize the need for coordinated action among researchers, law enforcement, and policymakers to create sustainable improvements in helmet compliance.

Examining regional disparities, Suriyawongpaisa et al. (2013) evaluated helmet use variations across Thai provinces through national roadside surveys and injury sentinel surveillance data collected in 2010. The data showed significant disparities in helmet use across regions, where the national average stood at 43.7%, while 53.3% of drivers complied compared to only 19.3% of passengers. Bangkok had the highest helmet use reported at 81.8%. At the same time, other regions exhibited significantly lower compliance, with helmet use in the least-compliant province at approximately 3%, resulting in a highest-to-lowest ratio of 28.5.

The statistical analysis revealed that helmet usage increased alongside population density and traffic violation conviction rates, which showcased law enforcement's impact on helmet compliance. The research uncovered inconsistencies between roadside survey findings and injury surveillance data, which presents obstacles to developing reliable policies. Suriyawongpaisa et al. (2013) call for enhanced police resource distribution and focused educational campaigns to tackle disparities in rural and low-income areas. The research presents essential information on how structural and behavioral obstacles influence helmet usage in Thailand, even though cross-sectional data prevents clear causality establishment.

Beyond helmet usage, understanding driver behavior and financial support for road safety interventions is essential. Jomnonkwao et al. (2021) explored willingness to pay (WTP) for accident risk reduction among personal car drivers in Thailand, employing the Theory of Planned Behavior (TPB) as a framework. Surveying 1,650 drivers, Jomnonkwao et al. found that perceived behavioral control exerted the most substantial influence on WTP, followed by attitudes and subjective norms, with all relationships statistically significant (p < 0.01). The Value of Statistical Life (VSL) is one way to measure how much, on average, a person is willing to pay to reduce a given risk by a set amount. Policymakers also use VSL as an unbiased metric for allocating resources to minimize risk in an economical fashion. Jomnonkwao et al. found that drivers would be willing to spend an average of 23 baht, or approximately $0.67, per person for every 50 km trip to achieve a 50% decrease in fatality or injury risk, which equates to a VSL of around $844,163. This relatively low value indicates that road safety policies would benefit from increased public investment if psychological and behavioral dimensions were incorporated.

The study demonstrates that road safety policies need more significant public investment to succeed when psychological and behavioral factors become part of the planning process. Research limitations include ignoring sociodemographic elements that influence WTP and concentrating solely on adult drivers, which excludes younger or unlicensed populations. The study by Jomnonkwao et al. (2021) proposed targeted awareness programs and infrastructure enhancement projects derived from behavioral research to improve traffic safety policy formulation and resource management in high-risk areas.

Jiwattanakulpaisarn et al. (2013) conducted interviews with motorcyclists in Thailand in 2009 to ascertain their knowledge of the helmet laws in Thailand. Only 60% of drivers and 28% of passengers reported always wearing their helmets. The low prevalence of passengers not wearing helmets could be caused by a lack of knowledge of the law and a widespread belief that the police do not enforce the law (Jiwattanakulpaisarn et al., 2013). Practical outreach to educate the public, along with a concerted enforcement campaign, is necessary to change the public perception.

The collective findings from these studies demonstrate the intricate relationship between regulatory enforcement and financial factors alongside human behavior concerning road safety practices. While Nishi et al. (2018), Suriyawongpaisa et al. (2013), and Jomnonkwao et al. (2021) highlighted the urgent need for more vigorous helmet law enforcement and fair resource distribution in their 2013 study. Research from Jomnonkwao et al. (2021) advances the conversation by illustrating how behavioral economics influences road safety investment decisions. A comprehensive, evidence-based strategy integrating legislative enforcement, demographic-specific interventions, and behavioral research is essential to reducing traffic fatalities and improving road safety outcomes in Thailand.

## Summary

Studies show that both helmet use and pre-ride alcohol consumption, along with demographic details like age and gender, directly influence motorcycle accident outcomes and death rates in Thailand. Studies by Chiangkhong et al. (2024), Hossain and Iamtrakul (2007), and Sae-Tae et al. (2020) confirm that strengthening helmet laws and improving helmet quality standards are crucial in reducing fatal head injuries. However, inconsistent and insufficient enforcement of existing helmet laws and insufficient school-based traffic safety education exacerbate risks, particularly in rural areas where compliance remains low (Chiangkhong et al., 2024). Additionally, road safety challenges extend beyond motorcycles, as other transportation systems, including cars and buses, also experience high incidences of drunk driving and reckless behaviors (Baral & Kanitpong, 2015; Mahikul et al., 2022). Dangerous road conditions and adherence to traffic regulations, such as the lack of frontage roads, motorcycle traffic flowing in reverse direction, and hazardous intersections, further contribute to accident severity, necessitating infrastructure improvements alongside behavioral interventions (Baral & Kanitpong, 2015; Mahikul et al., 2022).

Addressing these pervasive issues requires a comprehensive, multi-faceted strategy that integrates legislative enforcement, public education initiatives, and targeted safety interventions for high-risk populations, particularly young male motorcyclists and commercial delivery riders (Prakobkarn et al., 2024; Stephan et al., 2011). Only through a coordinated approach that combines policy reforms, community engagement, and sustained enforcement efforts can Thailand achieve meaningful reductions in road traffic fatalities and improve long-term road safety outcomes.

# **Data Science Application**

The primary goal of the project was to study helmet usage patterns by comparing the performances of different predictive models and developing insights into at-risk groups, locations, and times. The analysis of six years of collected data led to the development and deployment of three distinct models, which enabled an examination of long-term trends and variations in helmet usage during motorcycle accidents that resulted in hospitalization. Additionally, a separate model was constructed using only 2023 data to analyze current trends and changes in behavior. Ensemble techniques were chosen because Random Forest classifiers demonstrate powerful performance, excelling at processing diverse data types while providing robust feature importance analysis (Biau et al., 2008; Schonlau & Zou, 2020). This method was preferred over logistic regression and gradient boosting, as previous research has indicated that Random Forest offered an optimal combination of interpretability and predictive ability for the dataset (Fernandez-Delgado et al., 2014). Although advanced methods such as deep learning and support vector machines were considered, they were not utilized due to the excessive computational power and complexity required by the dataset size and the significant challenges in deriving actionable insights from deep learning models (Cortes & Vapnik, 1995; Louppe et al., 2013).

The decision to segment the analysis into aggregated and recent data models was both strategic and data-driven. The aggregated models provided insights into overarching patterns across extended timelines, while the dedicated 2023 model precisely captured recent developments that might have been obscured if combined with older data. This dual analytical approach enabled a deeper examination and more precise temporal insights into the evolution of helmet usage and its influencing factors. Using nested cross-validation and thorough hyperparameter optimization, reliable models were established that produced well-supported conclusions and practical recommendations for public safety.

This project’s models followed a pipeline style of analysis with the following steps:

1. Extract, Transform, Load
2. Exploratory Data Analysis
3. Imputation and Feature Engineering
4. Model Development, Deployment, and Evaluation

## Data Overview

This dataset originates from Thailand's Department of Disease Control Injury Surveillance System, accessible through Thailand's Open Government Data Center repository. The dataset contains documented injury reports from vehicle accident victims from 2018 to 2023, where treatment was sought at sentinel hospitals. These sentinel hospitals are located throughout Thailand and serve as a broad sample base that reflects the population’s injuries and illnesses, where they collect and forward pertinent information to the Injury Surveillance System (Santikarn & Santijiarakul, 2002). The demographic information fields within the dataset consist of age, province, and gender. The dataset contains specific details about accidents that capture accident date and time, vehicle type, and the person's role in the vehicle. Thailand uncovered multiple risk factors, including helmet, seat belt usage, cell phone operation, and inebriating drug or alcohol consumption. The dataset recorded 2,283,220 injuries that required medical treatment from sentinel hospitals across Thailand.

## Data Cleaning and Processing

Structurally, column names were first addressed from year to year, where they were not standardized and had different cases and spacing within the column names. Next, three pairs of date and time columns—integrating both the Gregorian and Thai calendar systems—were parsed and consolidated into three columns using NumPy's standard datetime object. With a lack of a unique patient identifier or more information about each patient, duplicated rows within the dataset may be unique records. For example, if two 23-year-old men sustain similar injuries in the same vehicle accident, their records in the dataset would appear the same. Nonetheless, such rows were removed from the dataset, totaling 5,915 records (0.26% of all documents), to preserve the integrity of the analysis.

Minimal data needed to be imputed: blanks and any data that did not match the most recent Information Surveillance System's Data Dictionary and the data dictionary's acceptable value list resulted in a transformation of that value into a standard unknown value in each column (Injury Prevention Division, Thailand Center for Disease Control, 2024). Before choosing this method, imputation methods were considered, and no clear commonalities between missing values could be determined as they occur in all provinces and across different columns. It is unclear whether the deviations from the acceptable field values are due to variations in local procedure, transcription errors, or various human factors. This “unknown value” imputation approach had three exceptions: age, gender, and province—if a row was missing an age or province, or the row’s gender was not male or female, the row was removed from the dataset to ensure the actionability of results. An additional filter was applied later to the models, ensuring that ages were non-negative and did not exceed 100 years.

Since this study focuses on motorcycle helmet usage, several columns were removed, representing information such as three columns narrowing down locations to neighborhoods, manual text entry columns because most entries are blank, and reported generic activity. Since this study focuses on motorcycle helmet usage, several columns were removed, representing information such as three columns narrowing down locations to neighborhoods, manual text entry columns because most entries are blank, and a generic activity, such as "transportation," that was reported.

## Exploratory Data Analysis

### ***Overall Exploratory Data Analysis***

During the six years, motorcycle incidents accounted for 86.18% of all injuries, totaling 1,839,982 occurrences from 2,266,903. Motorcycle-related injuries represent an essential safety and health matter because of their dominant share in overall injury statistics, as shown in the plot below. Motorcycles face significant transportation risks because of their vulnerability to traffic conditions, minimal protective measures, and unsafe riding habits. According to the literature analysis, community-based education, engagement funding, and policy modifications for motorcycle safety are crucial for reducing rider injuries.

**Figure 1**Distribution of Hospitalizations by Vehicle Type (2018-2023)

A graph of medical personnel

AI-generated content may be incorrect.

Figure 1 shows a rank-ordered count of motorcycle accident hospitalizations by vehicle type, highlighting that motorcycle hospitalizations occur at a rate of more than four to one versus every other vehicle type. From 2022 to 2023, vehicle accident injuries surged dramatically, as shown in Figure 2 below, breaking away from previous annual patterns. The apparent increase prompts an investigation into possible root causes, including traffic pattern changes, driver education improvements, behavior shifts, regulatory adjustments, or other external elements affecting road safety. A detailed analysis of the causes behind this increase prompted an exclusive modeling of the 2023 data.

**Figure 2**Histogram of Injuries by Year (2018-2023)

A graph of injury injuries

AI-generated content may be incorrect.

### ***2023 Increase Exploration***

The exact cause for the nearly doubling of accidents from 2022 to 2023 is unclear – the number of sentinel hospitals could have increased in every province, resulting in a potential skewing of the dataset. An area that may be worth studying is the rebound of tourism post-COVID-19. According to Thailand's Ministry of Tourism and Sports, in 2023, tourism rebounded to 70% of 2019, reaching 28.15 million visitors (Thailand Government Public Relations Department, 2024). Figure 3 below compares the increase in injuries resulting from motorcycle accidents from 2019 to 2023 and from 2022 to 2023. The comparison from 2019 was to show the consistency in increases despite COVID-19’s effect on tourism. Furthermore, every province showed increased motorcycle accident hospitalizations from 2022 to 2023. From 2019 to 2023, the data reflects some provinces where the number of accidents decreased, further complicating the analysis of this dataset because it discounts the possibility that the number of sentinel hospitals has changed.

The maps contained in Figure 3, which plot the difference in injuries from 2019-2023 on the left and 2022-2023 on the right, show some interesting trends. Provinces that serve as tourist hot spots, like Chon Buri and Surat Thani, experienced the most significant increase in motorcycle injuries between 2022 and 2023, with over 32,554 additional injuries. Bangkok registered only 385 extra motorcycle injuries throughout the same timeframe. Motorcycles are a practical but dangerous transportation solution for tourists in regions with limited or expensive public transportation alternatives. Popular tourist areas are not the only provinces with a marked increase in motorcycle accidents. The primarily industrial nature of Si Sa Ket, Rayong, and Songkhla suggests that multiple factors, besides tourism, contribute to increasing motorcycle accident hospitalizations.

**Figure 3**2019-2023 provincial increases compared to increases in 2022-2023.

A map of thailand with a number of different colored areas

AI-generated content may be incorrect.A map of thailand with a number of states

AI-generated content may be incorrect.

### ***Motorcycle-Specific Exploratory Analysis***

The literature review showed that previous research has documented motorcycles as a widely favored transportation method, and this trend appears in the demographic data of the dataset. The data shows that male injuries total 1,129,407 while female injuries total 710,575, and both genders exhibit the same mean age of 33.1 years, together with a median age of 28.0 years. This demonstrates that both men and women within a comparable age range face similar risks due to motorcycle usage, illustrating the popularity of motorcycles among young adults.

**Figure 4**Distribution of Motorcycle Hospitalizations by Age and Gender

A graph of a motorcycle accident

AI-generated content may be incorrect.

Figure 4 shows the distribution of injury counts by age and gender and shows a significant right-skewed distribution, demonstrating that injuries primarily occur among people aged 13 to 25. The data shows that younger populations, especially teenagers and young adults, experience injuries at higher rates, possibly because they rely on motorcycles for cheap and affordable transportation. As the literature shows, they are more likely to engage in riskier activities and have limited driving experience while facing more dangerous environments. Identifying this pattern provides essential information for developing models to inform prevention initiatives and educational programs to help decrease injury rates among this high-risk group.

**Figure 5**Hospitalizations by Time and Helmet Status

A graph of a number of people

AI-generated content may be incorrect.

**Figure 6**Hospitalizations by Time and Alcohol Usage

A graph of a bar chart

AI-generated content may be incorrect.

Figures 5 and 6 explore the relationships between accidents and helmet and alcohol usage. Figure 5 shows a relationship between nighttime riding and a lower proportion of helmet usage, indicating the possibility of a weaker enforcement presence, fatigue, or heat-related behavioral resistance. Figure 6 displays the increase in alcohol use and injuries resulting from motorcycle use. The marked increase in alcohol use during the evening, night, and into the early morning hours, coupled with the information from Figure 6, indicated that risk-taking extends beyond just not wearing a helmet and that the two may be interlinked.

### ***Helmets and Negative Outcomes Analysis***

Of the 1,835,588 injuries resulting from motorcycle accidents, 1,329,369 were reported not wearing a helmet during their accident, 242,438 were reported to have an unknown status, and 263,779 complied with the helmet law. Those who choose not to wear a helmet are 7.3% more likely to suffer a head injury than those who comply with the law. Figure 7 below shows that wearing a helmet significantly reduces the likelihood of a head injury.

**Figure 7**Motorcyclist Head Injury Outcomes

A graph of injury injuries

AI-generated content may be incorrect.

## Feature Engineering

Feature engineering occurred in several columns in this dataset. For example, some files did not have a head\_injury column but had a series of br1 through br6 columns representing numerical body regions. If the body region in any of those columns was one, the value in the column indicated a head injury occurred and head\_injury was updated accordingly. A column was added indicating whether a person was deceased upon arrival or became deceased throughout treatment, which was inferred from the treatment and release status codes from the hospital. Similarly, the five risk columns for alcohol, drug, seatbelt, helmet, and cell phone usage were converted into categorical variables, mapping blanks and out-of-definition values to unknown values. Lastly, a categorical time variable representing time in six four-hour periods was added to the dataset.

For convenience, the province column was translated from Thai into English. The vehicle descriptions were translated into English and consolidated into fewer vehicle types for easy analysis during the exploratory phases. During the modeling phase, one-hot encoding was employed to encode categorical variables, dropping the base case.

Critically, during the modeling phase, 243,353 rows where helmet status was unknown were dropped, constituting about 13.2% of the dataset. The unknown values were analyzed to understand if there were patterns to their status, but they are distributed throughout the dataset with no discernible pattern. Therefore, the analysis did not include records containing unknown values to maximize the discernment between those wearing helmets and those injured without them.

## Model Development, Deployment, and Evaluation

The model development process in this project was comprehensive and multifaceted, incorporating robust data preprocessing, feature engineering, hyperparameter tuning, deployment strategies, and rigorous evaluation for each modeling approach. Four models were created that address the entire dataset: three random forests using different optimization techniques and a logistic regression model. Two additional random forest models were created that look at 2023 data only, with one model’s hyperparameters optimized by a random grid search and the other utilizing Bayesian searching. Models that lack interpretability, such as Support Vector Machines and Neural Networks, did not align with the goals of this project and were not used.

### ***Sample Selection, Preprocessing, and Feature Selection***

The dataset is large and required employing a stratified sampling technique, supplying 20% of the dataset for cross-validation while maintaining the distribution of values for helmet\_used in the training and test sets (He & Garcia, 2009). The raw dataset is initially transformed by creating a binary target variable ("helmet\_used") based on helmet status. The helmet status as a target variable causes a severe imbalance within the dataset. Put another way, over the six years, 81.73% of hospitalized motorcyclists did not wear a helmet at the time of the accident after considering the filtering of the dataset described in the previous sections. This is also known as the “no-information rate” (NIR) and is one metric against which the models can be measured. Stratified sampling was chosen to maintain target variable proportion throughout model selection and evaluation to mirror accurate class distribution.

Predictors were carefully selected and preprocessed using scaling for numerical features and one-hot encoding for categorical variables. The predictors are chosen using SelectKBest from age, sex, time\_category, province, drug\_impairment\_status, alcohol\_status, and cellphone\_status. This ensures that the best models can choose the features that best describe their relationship with helmet usage (Breiman, 2001).

### ***Modeling Pipeline***

For the logistic regression model, a pipeline integrating automatic feature selection via ANOVA F-tests alongside logistic regression, with hyperparameters tuned through GridSearchCV and nested cross-validation ensuring unbiased performance estimates via ROC AUC scores. Similarly, the Random Forest (RF) models employ pipelines that combine data preprocessing with extensive hyperparameter searches—grid, random, and Bayesian searches are implemented—with nested 5-fold inner and outer cross-validation to first tune the hyperparameters and then assess model performance rigorously. The grid search tests every combination of hyperparameters and is computationally expensive for large parameter spaces. Grid searches engage in fine-tuning by systematically searching the parameter space for estimators count, tree depth, and splitting criteria (Snoek et al., 2012). The random grid search tests 20 randomly created combinations of hyperparameters, sometimes producing models that perform well by random chance. Bayesian optimization works differently, though, by identifying and tuning impactful hyperparameters adaptively. Table 1 below displays the hyperparameters tested. Each model was scored using the Receiver Operating Characteristic – Area Under Curve (ROC AUC) metric to determine the best hyperparameter combination and the best model overall.

**Table 1**   
Random Forest Model Hyperparameter Table

|  |  |  |
| --- | --- | --- |
| Hyperparameter Name | Values | Description |
| classifier\_\_n\_estimators | 100, 200, 300, 500 | Number of trees in Random Forest |
| classifier\_\_max\_depth | None, 5, 10, 15, 20 | Maximum depth of each tree |
| classifier\_\_min\_samples\_split | 2, 3, 5, 10 | Minimum n of samples required to split a node |
| classifier\_\_min\_samples\_leaf | 1, 2, 4 | Minimum n of samples required at a leaf node |
| classifier\_\_max\_features | sqrt, log2, None | Number of features to consider for each split |

ROC AUC was chosen as the scoring measurement because it provides information on how well each model could differentiate between the classes of the target variable, for instance, predicting if someone used a helmet. Simply put, a higher value means better discrimination. This is especially important for this imbalanced dataset, where the no-information rate exceeds 80%.

After the best combination of features and hyperparameters are selected, the package joblib is employed to write each model to disk for easy scalability in deployment and reuse. Loading saved models offers fast evaluation capabilities that enable potential real-time implementation. Full dataset metrics like ROC AUC and accuracy are computed for all models, and model interpretability is enhanced by examining feature coefficients and importance. The pipeline approach enables the creation of dependable predictive models and their practical implementation in production systems with ongoing performance tracking.

### ***Overall Dataset Modeling Results***

Four distinct predictive models were built using the process described above and then assessed using several key performance measures to evaluate the effectiveness of the designed modeling pipeline to predict the target variable across the entire dataset. Table 3 summarizes these models, comparing their ability to discriminate between helmet-wearing and non-helmet-wearing individuals (using ROC AUC), their overall accuracy, and the computational cost required to train the models. This evaluation framework allows a quantitative evaluation of which model most accurately predicts helmet use but also considers practical aspects such as processing time and resource efficiency.

By comparing methods ranging from Logistic Regression to various Random Forest approaches, including those optimized via Randomized Search, Bayesian Optimization, and exhaustive Grid Search, one can assess the trade-offs between incremental performance improvements and the computational demands of each technique. Although all models exceed the no-information rate, subtle differences in ROC AUC and accuracy indicate that the Random Forest methods are particularly effective. The detailed assessment provided in Table 3 thus offers a comprehensive perspective on model performance, enabling the selection of a model that balances predictive accuracy with practical efficiency for real-world applications.

**Table 2**Comparison of Models Performance

|  |  |  |  |
| --- | --- | --- | --- |
| Model | ROC AUC | Accuracy | Run Time (approx.) |
| Logistic Regression | 0.6819 | 0.8195 | 3 minutes |
| Randomized Search RF | 0.7285 | 0.8216 | 6 hours 34 minutes |
| Bayesian Optimized RF | 0.7166 | 0.8211 | 2 hours 39 minutes |
| Grid Search RF | 0.7167 | 0.8210 | 4 days, 5 hours |

Table 3 compares three different predictive methods based on three performance measures: ROC AUC, Accuracy, and modeling run time. As previously discussed, the Receiver Operating Characteristic – Area Under Curve (ROC AUC) provides information on how well each model could differentiate between the classes of the target variable, for instance, predicting if someone used a helmet - a higher value means better discrimination. "Accuracy" shows the percentage of correct predictions. Run time is the time the model took to process from start to finish, including feature selection, hyperparameter optimization, and final fitting, including evaluation.

In plain language, the Logistic Regression method correctly predicts about 81.9% of the cases and has a moderate ability to distinguish between classes (ROC AUC ≈ 0.68). The Random Forest model tuned via Randomized Search does a slightly better job (ROC AUC ≈ 0.73) while improving accuracy by 0.2%, and the Random Forest model tuned using Bayesian Optimization has the best discrimination (ROC AUC ≈ 0.72) with about the same accuracy. While the exhaustive Grid Search has a similar discrimination and accuracy to the Bayesian Optimized model, it was incredibly resource-intensive to model, making it more expensive for slightly less performance. While all methods are roughly 82% accurate, the Random Forest approaches, particularly those optimized with Random Grid Searches and Bayesian techniques, are better at correctly identifying helmet usage. This is reinforced when considering the processing time needed to fit the model. All models of the entire dataset provided slightly better accuracy than the dataset’s NIR (81.73%).

### ***2023 Modeling Results***

The study examining injured motorcyclists from 2023 showed that the optimized Random Forest models exceeded the dataset's No-Information Rate (NIR) benchmark of 82.39%. The models show valuable predictive capabilities beyond random guessing because they deliver improvements above the strong baseline accuracy. The Bayesian Optimized Random Forest model displayed its competency in outcome prediction for injured motorcyclists with an ROC AUC score of 0.7274 and an accuracy percentage of 82.62%, showing a slight yet valuable improvement over the No-Information Rate.

The Random Grid Search Optimized Random Forest model expanded its predictive power, resulting in an ROC AUC of 0.7378 and an accuracy figure of 82.66%. The random grid search hyperparameter tuning strategy proved effective in identifying essential model parameters for this data subset, which enabled the classifier to fully utilize underlying data patterns. These findings demonstrate that advanced optimization methods enhance Random Forest model performance measurably when they are used with public health and safety data because minor improvements in ROC AUC and accuracy produce more dependable insights.

## Summary

After creating an extensible and thorough modeling pipeline, four different models were built to capture how the features in the dataset describe helmet-wearing behaviors among injured motorcyclists across the entire period from 2018 to 2023. All models performed better than the no-information rate (81.73%), which represents the accuracy achieved by naïvely predicting the most frequently occurring class of the target variable. The overall best performing model—the Random Grid Search Optimized Random Forest—was selected due to its balanced computation time, an accuracy of 82.16%, and an ROC AUC of 0.7285. This model correctly predicted nearly 800 additional riders compared to its counterparts with only an acceptable increase in computing time.

Focusing on the 2023 subset of injured motorcyclists, two targeted Random Forest models were developed, each demonstrating performance beyond the subset's no-information rate of 82.39%. The Bayesian Optimized 2023 Random Forest achieved an ROC AUC of 0.7274 and an accuracy of 82.62%. Again, the Random Grid Search Optimized 2023 Random Forest proved to have improved performance, with an ROC AUC of 0.7378 and an accuracy of 82.66%. These results illustrate how refined hyperparameter tuning strategies can yield incremental gains, particularly in time-specific contexts. Together, the overall modeling results and the focused analysis on the 2023 subset underscore the strength of the developed pipeline, providing actionable insights into helmet use while validating the effectiveness of sophisticated optimization techniques in improving model performance.

# **Discussion of Analysis**

A total of six models, with four covering the entire period and two focused on injuries in 2023, were created and evaluated against each other during the modeling process, resulting in models whose accuracy all outperformed the NIR for their datasets. This chapter discusses the best-performing model's specific results and actionable insights, focusing on key topics: age, location, alcohol use, sex, and time of day. These focus areas allow a framework to present actionable insights that allow an updated understanding of conditions and motorcyclists' choices surrounding their injuries. Additionally, the model exploring 2023's increases will be explored using the same framework, highlighting differences in significant features.

## Interpreting Random Forest Models

Random Forest models were used in this project to allow for actionable insights to be derived from the model. Interpreting Random Forests is a two-step process. First, the importance of a feature was extracted, typically expressed as the percentage contribution a feature makes, on average, in reducing misclassification errors in the target variable. This important value represents how much information the value provides the model in discerning the target value's classes (Friedman, 2001).

Second, the marginal impact of a feature on the probability of the target variable can be determined by partial dependency analysis. In this analysis, all other features in the model are held to representative values, such as their mean or median values, and the resulting change in the probability was caused by changing the value of the feature of interest. Often, these marginal impacts are presented in a Partial Dependence Plot that shows the range of values the feature may have on the x-axis, and the y-axis shows the marginal impact the feature's values have on the probability of the target variable (Friedman, 2001).

## 2018 – 2023 Results

In the best-performing model overall, the Random Grid Optimized Random Forest model, the top five most impactful factors on whether a person wore a helmet are age (46.02%), riding in Bangkok (6.03%), not consuming alcohol (6.02%), riding in Nakhon Ratchasima (2.80%), and being female (2.30%). Table 4 below indicates the ten most important features for the model. Nine out of ten of the most impactful features fall within the framework previously described.

**Table 3**Model Feature Importance

|  |  |  |
| --- | --- | --- |
| Feature | Description | Importance (%) |
|  |  |  |
| num\_\_age | Age | 46.02 |
| cat\_\_prov\_Bangkok | Injured in Bangkok | 6.03 |
| cat\_\_alcohol\_status\_No Alcohol | No Alcohol Consumption | 6.03 |
| cat\_\_prov\_Nakhon Ratchasima | Injured in Nakhon Ratchasima | 2.80 |
| cat\_\_sex\_2 | Female | 2.30 |
| cat\_\_time\_category\_Morning (06:00-09:59) | Morning (06:00-09:59) | 2.22 |
| cat\_\_prov\_Phuket | Injured in Phuket | 2.06 |
| cat\_\_time\_category\_Evening (18:00-21:59) | Evening (18:00-21:59) | 1.41 |
| cat\_\_time\_category\_Night (22:00-01:59) | Night (22:00-01:59) | 1.31 |
| cat\_\_cellphone\_status\_No Cellphone | Not using Cellphone | 1.27 |

### ***Age***

The age feature, shown in Table 4, emerged as the primary contributor in the best model, having a massive 46.02% feature importance for helmet use. The significant importance score suggests that age represents multiple dimensions, such as rider behavior patterns, informed risk evaluation methods, safety norms, educational aspects, and regulatory adherence. Younger individuals may be more prone to risk-taking or place less emphasis on safety measures. In contrast, older or formally educated riders could show heightened caution or familiarity with helmet mandates. Nevertheless, it was not entirely clear how age influences different population segments without further demographic information about riders.

**Figure 8**Partial Dependence Plot for Age

A graph of a growth

AI-generated content may be incorrect.

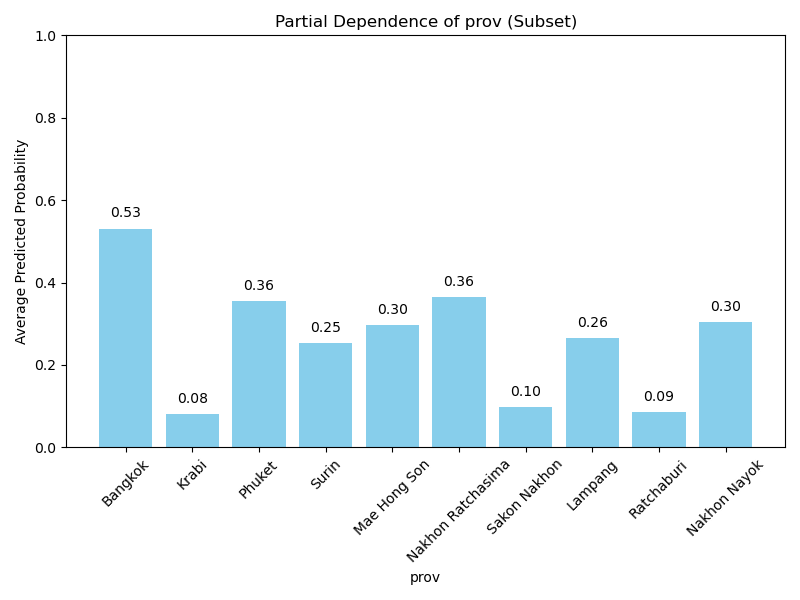
A closer inspection via the Partial Dependence Plot (PDP) in Figure 8 above for the age feature reveals a non-linear relationship with helmet-wearing probability. For the youngest riders, the model's predicted probability of wearing a helmet was relatively low, then rose sharply through the early-to-mid 20s to reach a peak (above 0.20) in the late 20s or early 30s. After this point, the curve gradually declines, settling for a more moderate probability for individuals in their 50s and 60s, though it remains higher than for the youngest riders. This pattern likely reflects changing attitudes and behaviors over time: younger riders may either be unaware of legal requirements or more inclined to take risks, those in their late 20s and early 30s could adopt helmets as they mature or respond to targeted interventions, and older individuals may ride less frequently or rely on different risk assessments. Additionally, families could face difficulties in locating and purchasing appropriate helmets for their younger children.

Combinations of overall feature significance and local partial dependence analysis suggest improvements for safety measures targeted at specific age brackets. Legal enforcement, together with educational outreach, emerges as the key method to increase helmet usage among young people. The use of protective gear needs to be emphasized through reinforcement strategies among older demographic groups. By examining how age affects helmet usage quality across all age groups and specific age categories, policymakers and public health officials can develop intervention strategies that ensure consistent helmet usage.

### ***Location***

The partial dependence plot highlights how much motorcycle riders’ helmet-wearing behavior varies across different provinces in Thailand. Bangkok stands out with a notably higher predicted probability of helmet use—around 0.53—suggesting that riders from the capital region, on average, are more inclined to wear helmets. In contrast, provinces like Krabi (0.08) and Lampang (0.09) show this dataset’s lowest average predicted probabilities. These findings underscore considerable regional variation and may reflect differences in enforcement intensity, road safety culture, or urbanization levels that encourage or discourage the use of protective gear.

**Figure 9**Partial Dependence Plot of Provinces



From a feature-importance standpoint, the table further confirms that province-level variables strongly influence helmet use decisions. Bangkok's indicator ranks highest among provinces at 6.03%, while Nakhon Ratchasima (2.80%) and Phuket (2.06%) also emerge as key differentiators in the model's ability to predict whether an individual will wear a helmet. Although these values may appear modest compared to the absolute top factors (e.g., age), they still signal that the specific location plays a meaningful role in shaping rider behavior. The existence or absence of specific provincial indicators can quantifiably influence the way Random Forest algorithms partition and refine their predictions.

The analysis reveals that each province's social factors, policy frameworks, and infrastructure heavily influence helmet usage patterns. Bangkok exhibits a relatively elevated partial dependence value, possibly due to more decisive law enforcement actions, denser traffic, which induces cautious driving, and comprehensive safety education programs. Regions with low probability scores could see improvements from targeted campaigns and supplemental actions like boosting police presence or organizing safety workshops. Regional insights allow policymakers to customize road safety strategies for different areas to enhance helmet promotion outcomes.

### ***Alcohol***

The model considers alcohol influence to be a significant factor. Looking at the feature importance results underscores this point: The variable "No Alcohol" stands as one of the model's top predictors with an importance score of 6.03%. The model makes a clear distinction between riders who consistently state they abstain from alcohol use and those who do not, which indicates that non-drinking attitudes and behaviors are associated with helmet usage patterns. On the other hand, "Unknown" shows a much lower importance (0.92%), indicating that missing or ambiguous data around alcohol consumption contributes far less to the model's predictive power. Nonetheless, its partial dependence probability of 0.12 suggests that incomplete knowledge of a rider's alcohol status may also coincide with lower helmet use. However, the model relies far more on explicit declarations of abstinence.

**Figure 10**Partial Dependence of Alcohol Status

A bar chart with text and numbers

AI-generated content may be incorrect.

In this partial dependence plot, figure 9, for the categorical variable representing alcohol use, the category "No Alcohol" clearly stands out with the highest predicted probability of helmet use at around 0.21. In contrast, individuals who have either "Unknown" alcohol status (0.12) or who are explicitly marked as "Alcohol" (0.11) both exhibit lower probabilities of wearing helmets. The disparity demonstrates that riders who report no alcohol consumption tend to adopt safer practices than those who drink alcohol or whose alcohol status is unclear. The data shows alignment between these probability figures and existing safety literature that demonstrates a link between alcohol intake and unclear reporting patterns with risk-taking behavior.

The collected findings stress the real-world value of documenting precise alcohol status to direct safety measures effectively. Motorists who deny alcohol consumption show a higher propensity to employ helmets, which suggests shared tendencies toward avoiding risks. For policymakers and safety advocates, reinforcing the dual benefits of avoiding alcohol before riding and consistently wearing a helmet may prove especially compelling for public health campaigns, given that both behaviors are strongly associated with safer outcomes in traffic accidents.

### ***Sex***

The partial dependence plot in Figure 11 for sex indicates that both males and females (labeled as "Male" and "Female" on the x-axis) share virtually the same predicted probability of helmet use, coming in at approximately 0.18. At first glance, the model, on average, does not distinguish significantly between men and women when it comes to wearing helmets. Despite these similar partial dependence values, the model may still use "sex" in targeted splits for specific subgroups, primarily if it interacts with other features like age or location. In other words, while the global average shows minimal difference, there might be localized pockets where being female, represented in the data as "cat\_\_sex\_2," meaningfully contributes to the model's predictive accuracy.

**Figure 11**Partial Dependence Plot of Sex

A graph with blue squares

AI-generated content may be incorrect.

Indeed, examining feature importance reveals that being female carries an importance score of 2.3%, which places it behind top predictors but still indicates a nontrivial role in reducing classification error for helmet usage. This seemingly small contribution could become critical when combined with other attributes such as age and provincial location. Thus, the overall takeaway was that "sex" alone may not drastically shift the model's prediction on a broad scale, reflected by similar partial dependence values. Still, it remains relevant enough for more nuanced distinctions, especially in interactive effects.

### ***Time of Day***

The model showed three different time categories in the top ten list of most important features, indicating that the time of day significantly impacts the model's ability to differentiate between helmet and non-helmet usage. The model assigns higher importance to Morning times (2.22%) compared to evening (1.41%) and night (1.31%) because it finds more distinguishing features in riders' behavior during early daytime hours. The importance scores for time-of-day features may appear minor when compared to prime predictors like age. However, they demonstrate that the Random Forest model can use specific daily periods to identify helmet use patterns. Morning commutes require adherence to traffic regulations, which result in regular safety practices, unlike nighttime and evening hours, where riders are more likely to be recreational or tired. Implementing public campaigns and law enforcement efforts timed to coincide with periods of high and low helmet compliance might lead to more consistent helmet use throughout the day.

**Figure 12**Partial Dependence Plot of Time Category

A graph with blue squares

AI-generated content may be incorrect.

The partial dependence plot in Figure 12 reveals that helmet-wearing probability shows substantial variation across different times of the day. The Morning (06:00–09:59) category shows the highest predicted probability of about 0.22, which exceeds Late Morning (10:00–13:59) at 0.20 and afternoon (14:00–17:59) at 0.19. In contrast, the Night category from 22:00- 01:59 registers the least helmet-wearing probability of 0.15, while evening (18:00–21:59) follows closely at 0.16, and Early Morning (02:00–05:59) reaches 0.17. The differences in helmet use between daytime hours and nighttime or late-evening hours may result from daytime conditions that promote helmet use through commuting traffic patterns and visible police activity, while nighttime conditions, such as rider fatigue and lessened enforcement, create riskier behavior patterns.

## 2023 Results

Like the overall model, the best-performing model was a random grid search model. The top five most influential features in the model are age (42.21%), not consuming alcohol (4.08%), driving in Phuket province (3.34%), being female (3.23%), and driving in Nakhon Ratchasima province (3.19%). The remaining features are listed by importance in Table 4.

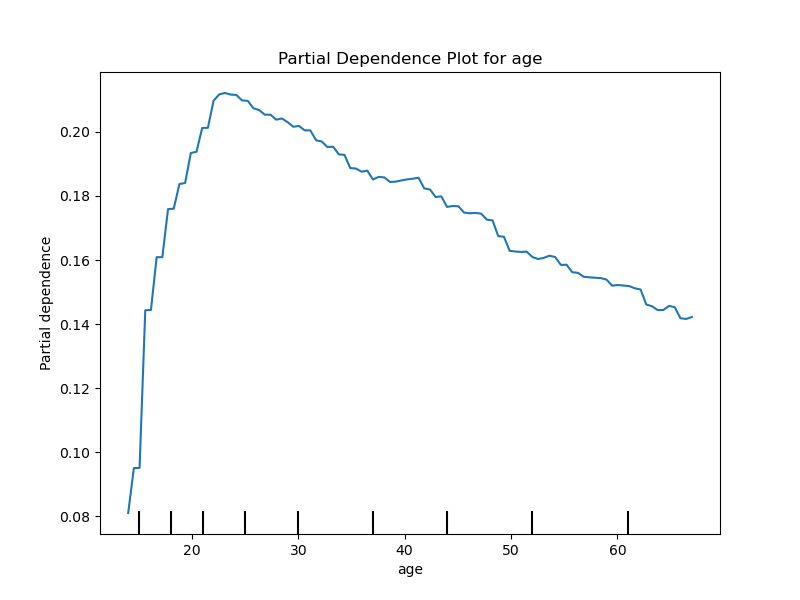
**Table 4**2023 Model Feature Importance

|  |  |  |
| --- | --- | --- |
| Feature | Description | Importance (%) |
|  |  |  |
| num\_\_age | Age | 42.21 |
| cat\_\_alcohol\_status\_No Alcohol | No Alcohol Consumption | 4.08 |
| cat\_\_prov\_Phuket | Injured in Phuket | 3.39 |
| cat\_\_sex\_2 | Female | 3.23 |
| cat\_\_prov\_Nakhon Ratchasima | Injured in Nakhon Ratchasima | 3.19 |
| cat\_\_prov\_Chon Buri | Injured in Chon Buri | 2.45 |
| cat\_\_prov\_Bangkok | Injured in Bangkok | 2.20 |
| cat\_\_time\_category\_Morning (06:00-09:59) | Morning (06:00-09:59) | 2.01 |
| cat\_\_prov\_Rayong | Injured in Rayong | 1.66 |
| cat\_\_time\_category\_Evening (18:00–21:59) | Evening (18:00-21:59) | 1.50 |

### ***Age***

When comparing the 2023 partial dependence plot from Figure 13 for age against the multi-year (2018–2023) findings in Figure 8, the overall shape of the curve remains broadly consistent: The predicted probability begins as low for young riders but quickly rises throughout their early 20s until it reaches a peak between late 20s and early 30s before showing a steady decline for older age groups. However, the peak was notably lower than the peak probability for the overall dataset. Analysis shows younger populations continue to display more risk-taking tendencies and less understanding of legal requirements, while young adult riders demonstrate the highest likelihood of helmet use. The 2023 data reveal variations in probability magnitudes, which suggest recent changes in enforcement policies, public awareness efforts, or social behavior trends have slightly modified how age affects helmet-wearing compliance.

**Figure 13**2023 Partial Dependence Plot of Age



The 2023 results validate age as a leading factor for helmet usage at 42.2%, which matches previous findings where age explained over 46% of a model's prediction strength. The data from 2023 shows that unique demographic or contextual elements of that year influenced the exact peak and slope of the partial dependence curve, especially in the late 20s to early 30s age range. The existing foundation for age-specific intervention strategies, including awareness programs for younger riders and reinforcement efforts for older groups, is valid. The slight changes in the 2023 partial dependence plot demonstrate the necessity for continuous evaluation and model enhancement to adapt helmet safety strategies according to shifting rider behaviors and local environmental changes.

### ***Location***

An analysis of the new partial dependence plot (PDP), Figure 14 below, for provinces alongside earlier research results shows alignments and variations in the location's impact on helmet-wearing patterns. In the 2023-specific subset analysis, Bangkok shows a predicted probability of approximately 0.41, while its earlier probability estimate was 0.53 over an extended timeframe. The difference between the datasets could be attributed to variations in enforcement practices and rider demographics, as well as external influences that arose after the initial collection of data. The provinces of Phuket, with a probability of 0.35, and Nakhon Ratchasima, with a probability of 0.34, demonstrate high probabilities, which show how urban development levels alongside local regulations and safety campaigns significantly affect helmet usage rates.

**Figure 14**2023 Partial Dependence Plot of Provinces

A graph with blue bars and black text

AI-generated content may be incorrect.

From a feature-importance perspective, the overall analysis remains intact: Helmet behavior shows that geographical location influences helmet usage levels, as Bangkok consistently shows higher compliance than other provinces. The dataset from 2023 shows a distinct change in probability rankings, which suggests local or temporal factors such as policing strategies and cultural changes may have a more specific influence on current behavior. The comparison between previous partial dependence values and the recent 2023 snapshot reveals ongoing provincial-level helmet-use disparities and their evolution, emphasizing the need for ongoing monitoring and refinement of region-specific safety measures.

### ***Alcohol***

Comparing the partial dependence plot in Figure 15 for alcohol status with the previously discussed findings reveals both continuity and nuance in how alcohol consumption, or lack thereof, relates to helmet use. While "No Alcohol" still shows the highest predicted probability of helmet-wearing, now at 0.18 instead of around 0.21, "Unknown" (0.13) and "Alcohol" (0.11) remain lower, indicating that abstinence correlates with safer riding practices. This modest shift in probabilities may reflect changes in enforcement, reporting habits, or broader social attitudes toward drinking and road safety over time. Nevertheless, the overarching trend remains that individuals who categorically report abstaining from alcohol appear more inclined to wear helmets compared to those who either consume alcohol or have ambiguous reporting.

**Figure 15**2023 Partial Dependence Plot of Alcohol Use

A bar chart with text and numbers

AI-generated content may be incorrect.

The model’s feature importance list emphasizes how the modeling approach underscores the significance of a clear "No Alcohol" designation. The pattern maintains that rider-reported abstinence remains a powerful predictor in the model despite any small changes in its importance score or exact probability. The "Unknown" status has a minor role in model splits but indicates that missing information could relate to decreased helmet usage according to its partial dependence value. For policymakers and public health advocates, the 2023 subset thus echoes earlier recommendations: Two critical approaches for public health include accurate reporting and promoting the safety benefits of riding without alcohol consumption. When authorities adjust their intervention messages to match existing rider behaviors, they can more effectively respond to new trends that impact helmet compliance.

### ***Sex***

The analysis from 2023 demonstrates that male and female helmet use probabilities hover around 0.17, according to partial dependence results. In the current specific analysis, the observed probabilities remained just a bit below the broader context's 0.18 value. Earlier research findings align with these results, showing that model evaluations give equal weight to both genders for population-level assessments.

The variable "sex" demonstrated modest growth in its feature importance ranking, with a 3.23% score showing its value, particularly when analyzed alongside features such as age or province. When used for detailed subgroup analysis and interaction studies, the model reveals different outcomes between women and men. Basic data analysis displays minimal sex-based differences in helmet-wearing probabilities, yet a detailed investigation shows sex played a small but substantial role in certain situations.

### ***Time of Day***

The 2023 partial dependence plot for the time category mirrors the overall pattern observed in the full‐period analysis: Helmet‐wearing probability reaches its highest level during the morning (06:00–09:59) time slot but declines during both the evening (18:00–21:59) and night (22:00–01:59) periods. In this subset, morning maintains the highest predicted probability at 0.20, late morning follows closely with 0.18, and afternoon is third at 0.17. The lowest observed probabilities occur during night and evening, approximately 0.14, while early morning (02:00–05:59) exhibits a likelihood of 0.16. While the absolute probabilities are somewhat reduced compared to the full‐dataset PDP, which saw morning reach 0.22 and night reach 0.15, the order of time‐of‐day categories and their ability to differentiate remain stable.  
The feature importance rankings reinforce this consistency: The full-period model demonstrates morning as having the most significant weight of 2.01% compared to the evening category at 1.50%. At the same time, the 2023 subset maintains the same ranking order of marginal effects. Despite slight variations in helmet-use probabilities across different periods, the highest and lowest helmet compliance times do not change. The relevance of interventions for morning commuters and nighttime riders has stayed consistent between 2023 and the 2018–2023 timeframe.

## Conclusion

Random Forest models tuned with Random Grid Search showed superior performance over simple benchmarks consistently across the 2018–2023 dataset and its 2023 subset while demonstrating strong predictive capability for helmet usage. Age was the primary predictor across all analyses because it accounted for nearly half of the model’s explanatory power while showing a clear non-linear connection with helmet‑wearing probability. Location, alcohol status, sex, and time of day each contributed additional, meaningful signals: Urban areas such as Bangkok alongside tourist destinations like Phuket exhibited higher adherence rates while non-drinkers reported more frequent helmet use and gender effects prominently emerged when combined with other characteristics; morning trips proved safest whereas evening and nighttime journeys presented increased danger.

The strong correlation between feature-importance rankings and partial dependence patterns demonstrates the consistent reliability of these results over time. The 2023 subset revealed slight changes in absolute probabilities, including reduced peak probability in Bangkok and a diminished morning effect, while preserving relative risk and compliance factors sequence.

# **Discussion**

## Summary of Findings

This chapter synthesized and translated the modeling results into concrete policy, enforcement, and public health practice recommendations. Building on the identified key drivers of helmet use—age, provincial context, alcohol abstinence, gender, and time of day—this section explored how these insights can inform targeted interventions, resource allocation, and legislative enhancements. By examining global feature importance and localized behavior patterns, actionable strategies are illustrated for stakeholders to improve helmet compliance, reduce injury severity, and ultimately enhance road safety outcomes across Thailand.

## Conclusions by Project Objectives

### ***Project Objective One***

First, this project looked at overall trends and found that above 80% of all vehicle hospitalizations are from motorcycle accidents, and that motorcycle accident hospitalizations have been on an upward trajectory over the six years, totaling more than 1.8 million. Youths and young adults aged 13 to 25 are most frequently injured. Approximately 72.4% of all injured motorcyclists were reported not to wear a helmet.

### ***Project Objective Two***

Secondly, identifying and highlighting the key determinants of helmet‐wearing compliance allowed the analysis to move beyond aggregate rates and uncover the specific drivers of safe behavior. In the overall Random Forest model, age emerged as the dominant predictor, capturing 46.02% of the total feature importance—a measure of its relative contribution to impurity reduction across the forest. Partial dependence analysis showed that younger riders had the lowest predicted probabilities of helmet use, while those in their late twenties to early thirties peaked at over 0.20. Provincial context also proved influential: the Bangkok indicator contributed 6.03% and Nakhon Ratchasima 2.80% of the model’s total importance, reflecting substantial regional variation. Similarly, riders reporting no alcohol consumption accounted for 6.03% of importance versus 0.92% for those with unknown status, underscoring the behavioral link between sobriety and safety. Additional factors—being female (2.30%), riding in the morning (2.22%), or not using a cellphone (1.27%)—added further nuance about when and where helmet use was most and least likely.

### ***Project Objective Three***

Lastly, stakeholders can allocate limited enforcement, education, and subsidy resources far more strategically by pinpointing these determinants. For example, targeted outreach to younger riders—both the least compliant and the most influenced by age-related risk-taking- can be paired with enhanced helmet law enforcement in rural provinces where compliance lags. Likewise, leveraging the high compliance in Bangkok and Phuket suggests that policymakers could adapt best-practice campaigns for lower-compliance regions. Timing interventions around high-risk periods, such as late evening and night when predicted helmet use dips below 0.15, and coupling them with alcohol-awareness initiatives addresses two major risk factors simultaneously. This way, a data-driven focus on compliance determinants transforms raw model outputs into concrete, high-impact actions that maximize safety returns across Thailand’s diverse riding population.

## Implications for Thailand

The core drivers of helmet use have remained unchanged, even though enforcement intensity, public campaigns, and social behaviors have experienced changes over time.

These findings provide a definite path for developing specific interventions:

* Educational programs and enforcement measures designed for different age groups to improve younger riders' compliance while supporting older commuters' established safety habits.
* Regional enforcement and outreach should be increased in provinces with lower helmet use compliance while maintaining effective programs in areas with high compliance levels.
* Public awareness campaigns that unify messages about sobriety with helmet use benefit from the established connection between avoiding alcohol and engaging in safe behavior.
* Time‑of‑day enforcement campaigns utilize morning checkpoints alongside nighttime patrols to address compliance shortfalls.

By continuously updating the models with fresh data and monitoring shifting partial dependence curves, policymakers can adapt these strategies in real time, maximizing the impact of limited resources and steadily improving helmet‑wearing rates across Thailand.

## Limitations of Results and Suggestions for Future Research

Although this analysis was strong in numerous areas, its limitations prevent findings from being broadly applied. Data collection focused solely on hospitalized motorcycle accident cases from 2018 to 2023, which could mask the frequency of milder incidents and exaggerate hospitalization-related risk elements. The study lacked access to essential contextual variables, including real-time law-enforcement intensity and precise road-condition metrics. This could have led to omitted-variable bias affecting modeled relationships. The Random Forest analysis revealed intricate relationships, yet functioned strictly as a correlational tool that cannot provide definite causal conclusions regarding helmet-use mandates or alcohol-reduction campaigns. The failure to collect demographic information such as income level, educational background, and age and sex data restricted the ability to customize interventions for specific subpopulations with distinct risk factors. Researchers should use wider data sources and more experimental methodologies to verify these findings across various environments.

## Conclusion

This study harnessed advanced data‐science methodologies to uncover the key factors driving helmet-wearing behavior among injured motorcyclists in Thailand. By processing six years of hospital records through rigorous extraction, transformation, exploratory analysis, and by deploying optimized Random Forest classifiers, it quantified the relative influence of demographic, behavioral, and contextual variables, most notably age, provincial location, alcohol abstinence, and time of day. Partial dependence analyses further illuminated non-linear and interaction effects, revealing, for example, a pronounced peak in compliance during the late 20s to early 30s and markedly higher helmet use in urban centers like Bangkok.

Targeted interventions, which include intensified enforcement in non-compliant provinces, rider education tailored to different age groups, sobriety campaigns combined with helmet checks, and road-safety initiatives specific to certain times of day, should be considered instead of one-size-fits-all solutions. These findings help to direct resources and improve laws to achieve higher helmet wearing rates and decrease head injury cases from motorcycle accidents in Thailand.

# **References**

Baral, S., & Kanitpong, K. (2015). Factors Affecting the Severity of Motorcycles Accidents and Casualties in Thailand by Using Probit and Logit Model. *Journal of the Eastern Asia Society for Transportation Studies*, *11*, 2175–2188. https://doi.org/10.11175/easts.11.2175

Biau, G., Devroye, L., & Lugosi, G. (2008). Consistency of Random Forests and Other Averaging Classiﬁers. *Journal of Machine Learning Research*, *9*. https://jmlr.csail.mit.edu/papers/v9/

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(5–32). https://doi.org/10.1023/A:1010933404324

Chiangkhong, A., Surakarn, A., Kleebbua, C., & Pounglek, W. (2024). A Profiles of Motorcycle Accident Mortality and Risk Behavior in Thai Children. *International Journal of Geoinformatics*, *20*(3), 54–63. https://doi.org/10.52939/ijg.v20i3.3133

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we Need Hundreds of Classiﬁers to Solve Real World Classiﬁcation Problems? *Journal of Machine Learning Research*, *15*. https://jmlr.csail.mit.edu/papers/volume15/delgado14a/delgado14a.pdf

Friedman, J. H. (2001). 1999 Reitz Lecture Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hossain, M., & Iamtrakul, P. (2007). Medical Investigation of Motorcycle Accidents in Thailand. *Proceedings of the Eastern Asia Society for Transportation Studies*, *2007*, 360–360. https://doi.org/10.11175/eastpro.2007.0.360.0

Ichikawa, M., Chadbunchachai, W., & Marui, E. (2003). Effect of the helmet act for motorcyclists in Thailand. *Accident Analysis & Prevention*, *35*(2), 183–189. https://doi.org/10.1016/S0001-4575(01)00102-6

Injury Prevention Division, Thailand Center for Disease Control. (2024a). *คู่มือการลงรหัสแบบบันทึกข้อมูลการเฝ้าระวังการบาดเจ็บ และการบันทึกลงคอมพิวเตอร์ (IS Data Dictionary)*.

Injury Prevention Division, Thailand Center for Disease Control. (2024b). *สถานการณ์ อุบัติเหตุทางถนน ประเทศไทย กองป้องกันการบาดเจ็บ กรมควบคมโรค (Situation of Road Injury Accidents)*. https://ddc.moph.go.th/uploads/publish/1587620240712091713.pdf

Jiwattanakulpaisarn, P., Kanitpong, K., Ponboon, S., Boontob, N., Aniwattakulchai, P., & Samranjit, S. (2013). Does law enforcement awareness affect motorcycle helmet use? Evidence from urban cities in Thailand. *Global Health Promotion*, *20*(3), 14–24. https://doi.org/10.1177/1757975913499030

Jomnonkwao, S., Wisutwattanasak, P., & Ratanavaraha, V. (2021). Factors influencing willingness to pay for accident risk reduction among personal car drivers in Thailand. *PLoS ONE*, *16*(11), 1–17. https://doi.org/10.1371/journal.pone.0260666

Kasantikul, V. (2001, September). *Motorcycle accident causation and identification of countermeasures in Thailand: Volume I: Bangkok study.* Chulalongkorn University. https://smarter-usa.org/wp-content/uploads/2017/06/2.-2001-vol-i-motorcycle-accident-research-in-thailand-bangkok.pdf

Kumphong, J., Satiennam, T., & Satiennam, W. (2018). The determinants of motorcyclists helmet use: Urban arterial road in Khon Kaen City, Thailand. *Journal of Safety Research*, *67*, 93–97. https://doi.org/10.1016/j.jsr.2018.09.011

Lohitakul, N. (2009, October 16). Community Youth Helmet Use Project in Thailand. *International Trafic Safety Data and Analysis Group*. 4th IRTAD Conference. https://www.itf-oecd.org/sites/default/files/docs/p11\_lohitakul.pdf

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems*, *26*. https://papers.nips.cc/paper\_files/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html

Mahikul, W., Aiyasuwan, O., Thanartthanaboon, P., Chancharoen, W., Achararit, P., Sirisombat, T., & Singkham, P. (2022). Factors affecting bus accident severity in Thailand: A multinomial logit model. *PLoS ONE*, *17*(11), 1–17. https://doi.org/10.1371/journal.pone.0277318

Nishi, A., Singkham, P., Takasaki, Y., Ichikawa, M., Chadbunchachai, W., Shibuya, K., & Tanasugarn, C. (2018). Motorcycle helmet use to reduce road traffic deaths in Thailand. *Bulletin of the World Health Organization*, *96*(8), 514-514A. https://doi.org/10.2471/BLT.18.215509

Prakobkarn, P., Luangwilai, T., Prempree, P., & Kunno, J. (2024). Prevalence of motorcycle accidents among food delivery drivers and its relation to knowledge, attitudes, and practices in urban areas in Bangkok, Thailand. *PLoS ONE*, *19*(5), 1–15. https://doi.org/10.1371/journal.pone.0303310

Ratanavaraha, V., & Jomnonkwao, S. (2013). Community participation and behavioral changes of helmet use in Thailand. *Transport Policy*, *25*, 111–118. https://doi.org/10.1016/j.tranpol.2012.11.002

Sae-Tae, N., Lim, A., & Dureh, N. (2020). Determinants of severe injury and mortality from road traffic accidents among motorcycle and car users in Southern Thailand. *International Journal of Injury Control & Safety Promotion*, *27*(3), 286–292. https://doi.org/10.1080/17457300.2020.1774616

Santikarn, C., & Santijiarakul, S. (2002). *From Provincial to National: The Development of Thailand Injury Surveillance*. https://www.cdc.gov/nchs/ppt/ice/cuernavaca/santikarn\_thailand.ppt

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3–29. https://doi.org/10.1177/1536867X20909688

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, *25*. https://papers.nips.cc/paper\_files/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

Stephan, K., Kelly, M., Mcclure, R., Seubsman, S., Yiengprugsawan, V., Bain, C., & Sleigh, A. (2011). Distribution of transport injury and related risk behaviours in a large national cohort of Thai adults. *Accident Analysis & Prevention*, *43*(3), 1062–1067. https://doi.org/10.1016/j.aap.2010.12.011

Suriyawongpaisa, P., Thakkinstian, A., Rangpueng, A., Jiwattanakulpaisarn, P., & Techakamolsuk, P. (2013). Disparity in motorcycle helmet use in Thailand. *International Journal for Equity in Health*, *12*(1), 74–79. https://doi.org/10.1186/1475-9276-12-74

Swaddiwudhipong, W., Boonmak, C., Nguntra, P., & Mahasakpan, P. (1998). Effect of motorcycle rider education on changes in risk behaviours and motorcycle-related injuries in rural Thailand. *Tropical Medicine & International Health*, *3*(10), 767–770. https://doi.org/10.1046/j.1365-3156.1998.00301.x

Thai Road Safety Coalition. (2024, December 31). *Thai Road Safety Coaltion*. Accident Database. thairsc.com

Thailand Government Public Relations Department. (2024, September 8). *Tourist Arrivals in Thailand Increase Significantly*. https://thailand.prd.go.th/en/content/category/detail/id/48/iid/314311

Watcharapong Rueankham & Narongsak Nuson. (2019). Motorcycle Accidents in Thailand: Epidemiology Perspective. *Journal of Health Sciences and Wellness*, *23*(1), Article 1.

World Health Organization. (2025, February 26). *Road safety*. Thailand - Road Safety. https://www.who.int/thailand/our-work/road-safety

# **Appendix**

Code Located: [Google Drive Link](https://drive.google.com/drive/folders/1jY72JEVqrr4X0OywF5R-PLYK7Ln5XWh8?usp=sharing)