Factors affecting perceptions in transport - A deep dive into the motorbike ban in Hanoi, Vietnam

Minh Kieu[1,*][0000−0001−7798−6195], Alexis Comber[2][0000−0002−3652−7846], Eric Wanjau[3], Kristina Bratkova[3], Hang Nguyen Thi Thuy[4], Thanh Bui Quang[5], Phe Hoang Huu[6], and Nick Malleson[3,7][0000−0002−6977−0615]

[1] Department of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand
`http://www.cee.auckland.ac.nz/`
[2] School of Geography, University of Leeds, LS2 9JT, UK `https://environment.leeds.ac.uk/geography`
[3] Leeds Institute for Data Analytics, University of Leeds, UK
[4] VNU Vietnam Japan University, Hanoi, Vietnam
[5] Faculty of Geography, VNU University of Science, Hanoi, Vietnam
[6] R&D Consultants, Hanoi, Vietnam
[7] Alan Turing Institute, NW1 2DB, UK `https://www.turing.ac.uk/` else
* Corresponding author: `minh.kieu@auckland.ac.nz`

Abstract. The dependence on motorbikes has contributed to traffic problems in Hanoi, Vietnam. Policymakers have considered a controversial ban on nonelectric motorbikes in parts of the city in an effort to reduce congestion and pollution. However, understanding of individual perceptions on critical transport policies, such as this potential ban is lacking, especially in the Global South, with implications for evidence-based policy making. This paper presents the results of some exploratory data analysis and a machine learning application using a travel survey recently conducted in Hanoi. It aims to understand how residents perceive a potential motorbike ban, their perceptions of different mobility modes, as well as their future plans for mobility if motorbikes are banned. This data-driven analysis of policy scenarios shows that awareness of the potential ban, distance to public transport, and individual transport modal choice determine the acceptability of the proposed motorbike ban and its likely success. It also shows that policymakers in Hanoi should also consider citizens' plans for future vehicle ownership, as the analysis results suggest that cars are likely to replace motorbikes if the ban is implemented.

Keywords: Individual perceptions, transport policies, Global South

1 Introduction

It is vital to understand the needs of the public before investing in infrastructure. Transport projects that were constructed without thorough public consultation are often deemed to be inadequate for meeting the public's needs (Calvo-Poyo et al. 2020). Transport policymakers need to understand citizens' perceptions of new policies and the overall transport systems to make informed decisions that will cater for their needs.

This paper aims to augment our understanding of the factors affecting individuals' perceptions of transport policies and mobility modes. Usually, this is challenging to achieve because perceptions, emotions, intentions, and desires are personal and subjective. Travel surveys are a well-understood method to interrogate these aspects that have proven useful in analysing the effects of existing policies (Lin et al. 2020; Wallner et al. 2018) and to make informed decisions to meet future transport needs (Charleux 2018). However, only a handful of studies have been conducted on the perception of road users in developing countries. Joewono and Kubota (2007) explores how users and paratransit drivers think about the safety and security of the service in Bandung, Indonesia. Das and Pandit (2013) discusses the importance of individual perception in determining the level of service of public transit services in India. Nordfjærn and Rundmo (2009) compares the perception of traffic risk between an industrialised and a developing country (Norway and Ghana). Similarly, little attention has been paid to the perceptions of travellers in motorcyclist-dominated cities such as Hanoi. Hongsranagon et al. (2011) investigates the behaviours and perceptions of motorcyclists towards the traffic safety risks. Santos et al. (2019) investigates the relationship between work overload and risk behaviours in Brazilian motorcyclists. On the same topic of traffic accident risk perceptions, Manley et al. (2020) proposes a latency-based hazard perception test for Thai drivers. Rusli et al. (2020) uses on-road observation methods to investigate risky riding behaviours among motorcyclists in Malaysia.

Perhaps one of the main challenges of perception-focused research is the lack of a large-scale survey of individual perceptions. Most of the existing studies in the literature for developing countries focus largely on risk-taking behaviour and perceptions – which are the main concerns for motorcyclists – rather than on perceptions on transport policies. To overcome this gap, we have developed a large-scale travel survey of more than 26,000 Hanoi residents that aims specifically to capture individuals' perceptions of transport policies and transportation modes. This article explores perceptions of an important future transport policy: aproposed ban on motorcycles in Hanoi. The main novelty of this work is the use of state-of-the-art data analytics on this large-scale dataset to get useful policy implications, as motorbike-dependent developing cities are not often the focus of research, especially for large-scale surveys. We employ data-driven methods in Machine Learning, in particular exploratory data analysis and classification models to answer three research questions (RQs):

1. Can we predict an individual's perception of a future transport policy?
2. What policy implications can we obtain from the data and the modelling?
3. What factors are most and least important in explaining individuals' perceptions?

The first question aims at anticipating a person's perception of an important transport policy. While our travel survey in Hanoi has collected individuals' perceptions directly through their responses, a prediction model of individuals' perception to a transport policy may allow us to apply the model to a much larger dataset than our travel survey (e.g. census data). By analysing the importance of each variable to the classification model, the second question aims to enable us to understand how different factors are affecting individuals' perceptions. Finally, question three enables us to obtain further policy implications regarding motorbike dependency in Hanoi.

The remainder of this paper includes a brief description of the transport system in Hanoi (Section 2), the adopted modelling methods (Section 3), the explanatory data analysis (Section 4), the classification modelling results (Section 5), the policy scenarios (Section 6) and finally a conclusion of study (Section 7).

2 Transport in Hanoi and related works

Hanoi, located in Southeast Asia, is the capital of Vietnam. The motorbike is the primary mode of transport in Hanoi. Since the introduction of the Doi-Moi policy, in less than 20 years, motorcycles have replaced bicycles as the main means of transport at a "remarkable pace" (Hansen 2017, 2022). The number of motorcycles in Hanoi has increased tenfold and in 2014 there were 4 million motorbikes in Hanoi alone (Hansen 2016, 2017). More than 90% of the vehicles driven in Hanoi are motorbikes and there are two and a half motorbikes per person (Van et al. 2009). Motorbike dependency is also not a Hanoi-specific problem, with 93.9% of the registered vehicles in Vietnam being motorcycles (National Traffic Safety Committee 2021). Vietnam also has the highest two- and three-wheeler ratio per capita among countries of South East Asia, with 422 vehicles per 1000 people (Kitamura et al. 2018).

The rapid increase in motorbikes has contributed to traffic-related problems in Vietnam. Many studies have shown that traffic congestion has dramatically increased in the last decades (Pham et al. 2021; Van et al. 2009; Lim 2018; Ly et al. 2020). Hanoi was the 2nd most polluted city in South East Asia in 2018 (Huu et al. 2021). Much of the increase in air pollution and emissions has been found to be associated with motorbikes (Thuy et al. 2012). This has also been evidenced by Ly et al. (2020) who found that more than 90% of the emissions of volatile organic compounds are contributed by motorbikes. Some other studies also found a significant connection between Vietnamese motorcycle dependence and increased traffic accidents (Pham et al. 2021; Tuan 2015; Truong et al. 2016; Chou et al. 2022). In addition to causing increased congestion and pollution, motorcycles have fundamentally changed the character of the city for residents. A city that used to be famous for "tranquil" streets full of bicycles is now characterised by "almost constant buzzing and honking" (Hansen 2022).

Several cities with a demographic similar to that of Hanoi have already implemented or planned a ban on motorbikes. For example, in China Guangzhou reduced the percentage of motorbikes from 31% to 7%, by improving public transport and implementing a ban on motorbikes in urban areas (Pucher et al. 2007). A similar ban was also successful in Shenzen (Dong and Liu 2017). One reason for these bans come from the fact that the number of motorbike sales in China increased more than 13 times from the period between 1991 to 2008 (Yan and Crookes 2010). From August 2015, the High Court of Karachi, Pakistan implemented a ban on operation of three-wheelers, commonly known locally as Qingqis, across the province. An analysis by Ahmed and Fatmi (2016) showed that both the number of vehicles and the number of traffic injuries were significantly reduced after the implementation of the ban. Many existing studies tend to agree that limiting motorbikes will reduce traffic congestion, accidents, and emissions (Dong and Liu 2017; Tuan 2015; Singkham 2016).

However, the dependency on motorbikes in Hanoi is not easy to deal with. They are woven into the fabric of Hanoi's society as they provide a flexible, independent and, relatively effortless form of mobility (Thuy et al. 2012; Hansen 2016, 2017). In short, motorbikes are "absolutely vital" to the lives of most people in Vietnam (Hansen 2022). In recent years, many government initiatives have encouraged the use of public transport, including purchasing over 1500 new buses, subsiding bus fares and upgrading bus facilities (Pham et al. 2021). However, the current public transport system is largely limited to buses alone and despite these investments there are only eight bus routes with a regularity of less than 10 minutes during peak hours (Bray and Holyoak 2015). Although a new rapid transit rail line opened in November 2021, most of the new metro system is still under construction without a definite completion date. Thus the adoption of public transport has been low.

The problem of low use of public transport despite investments and initiatives by governments is prevalent in many countries in South East Asia, including Indonesia and Thailand (Nguyen et al. 2018). A motorbike ban may also lead to other negative impacts. In Yangon, Myanmar, motorbikes have been banned since 2003. This decision has lead to fewer travel options, more traffic congestion and reduced employment opportunities (Gupta 2019). Increased car ownership poses even greater risks of traffic problems compared to motorbikes. Van et al. (2009) showed that if 30-40% motorbike users switched to a car, traffic conditions would worsen and resemble gridlock-like situations. Laksana (2019) quoted a popular view in Indonesia, a country with a very similar traffic system to that of Vietnam, that motorbikes can be twice as fast as cars in Jakarta, while only using 1/10 of the fuel and are generally a much more efficient use of space compared to cars. Local and international planners have also advised that a motorbike ban ban may privilege other private modes of transport, such as cars, while punishing those without cars (Kim 2017), and may lead to "mobility injustice" on Hanoi's streets (Turner 2020).

Despite these concerns, the Hanoi People's Council issued Decision No. 5953/QDUBND, a plan to ban motorbikes in specific areas of Hanoi (Huu et al. 2021). This plan suggests a stepwise limit on all nonelectric motorcycles by 2030 and increases the share of public transport modes to 65% (Hanoi People's Council 2017). This plan will lead to many mobility-related questions affecting the majority of Hanoians, such as the aforementioned questions in the Introduction. This paper aims to answer the research questions using data-driven analysis from travel survey data in Hanoi, Vietnam, in 2021.

3 Methodology

Figure 1 shows the proposed methodology. After an explanatory data analysis (Step 1) we develop a machine learning classification model to capture individuals' perceptions of a potential motorbike ban policy in Hanoi (Step 2). Step 3 uses variable importance analysis to estimate the strongest determinants of an individual's perception of the policy. Finally, Step 4 involves using the model to make a prediction about what might make people more or less favourable to the policy.

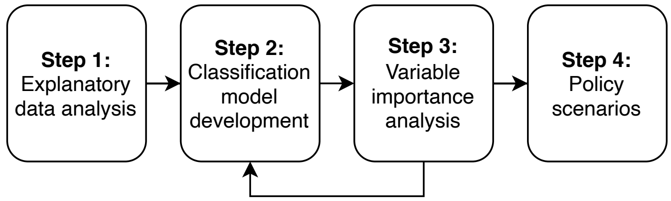


Fig. 1: Study workflow

3.1 Survey design and implementation

The original survey was designed collaboratively by all authors. After an initial prototyping period to test the survey, the project aimed to survey $10,000$ households. The survey was initially conducted by Vietnam National University (VNU), University of Science in 2020. The survey format was a face-to-face questionnaire conducted by VNU staff and students using Kobo Toolbox (`https://www.kobotoolbox.org/`) as the main data collection tool. After receiving approximately 1,500 responses the COVID-19 pandemic reached Vietnam and survey collection had to terminate; not only would it have been impossible and unethical to recruit survey staff but the pandemic disrupted travel patterns to such a degree that the responses would not have been useful beyond the course of the pandemic anyway. Once pandemic suppression was measures lifted in Hanoi and more typical travel patterns began to emerge in 2021, rather than trying to restart face-to-face collection, the survey was converted to online-only and performed by the Vietnam General Statistics Office, Ministry of Planning and Investment. The Ministry is responsible for conducting the Vietnam Census and therefore has considerable experience in telephone interviews and in obtaining largely representative samples. At the time of writing, the survey reached $N = 26,339$ respondents, considerably more than the initial $10,000$ that was hoped for. With the telephone interviews, a single family member was asked to fill out the questionnaire survey.

Although our survey is relatively large, its sample size still covers only a small proportion of the Hanoi population. Our aim is to provide the best possible representative of the population in Hanoi, both geographically and demographically. Figure 2 illustrates the distribution of households chosen for the survey versus the distribution of households in the census. Since the potential motorbike ban, if implemented, will be near Hanoi's CBD, the survey focusses on the urban districts and not the suburban areas. Our samples represent the actual population distribution in Hanoi's urban area, with the majority of people are living in the West bank of the Red River.

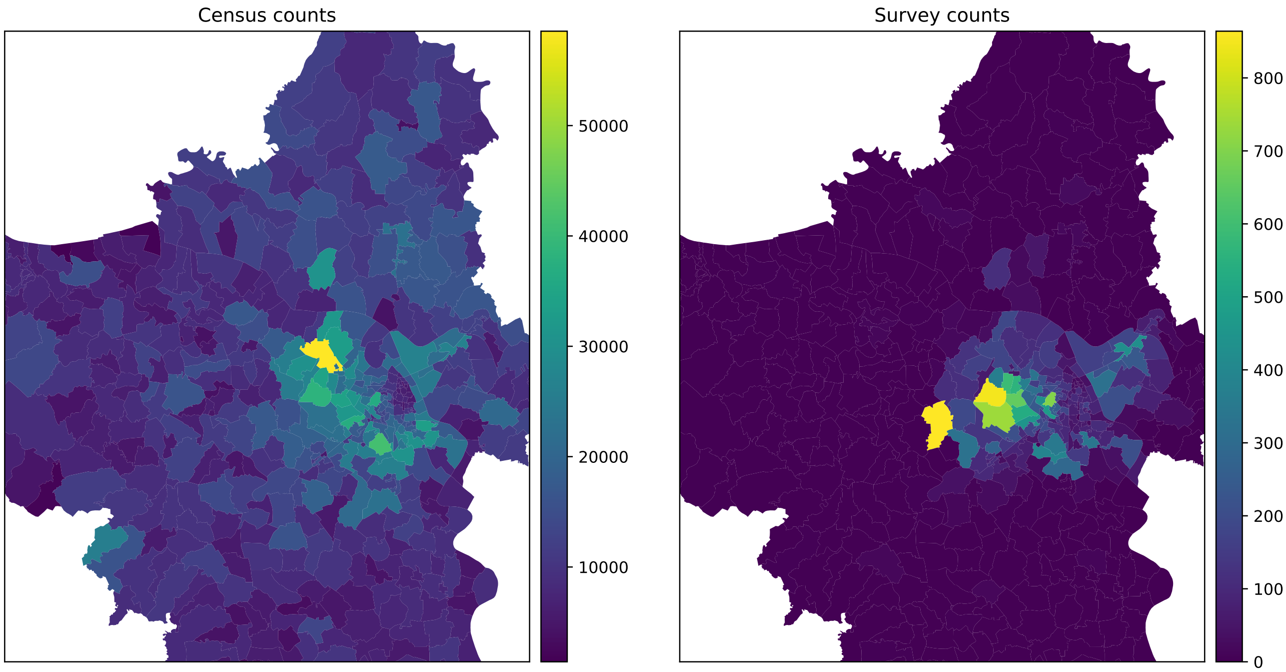


Fig. 2: Map of surveyed households, versus the distribution of households in the census

The questions in the survey were designed to reflect general demographics, travel patterns, behaviours, and especially the perceptions of people about transportation policies and transportation modes. The main survey questions are illustrated in Table 1. The 'general' questions collect the demographics and home location of the respondent. The 'living conditions' is a series of questions aimed at understanding an individual's quality of life, and access to necessities such as open space, school, and hospital. The next class of questions, 'household composition and vehicle ownership', aims to collect the family structure of the respondent and how many vehicles of each type are owned in the household. The 'primary trip' is the regular trip that the respondent often makes. The 'mode choice' section collects individuals' frequency of usage and perceptions on each mode of transport. Finally, the 'motorbike ban' section explores the individuals' awareness of the ban, their agreement/disagreement with the ban, and their plans of alternative vehicles if the motorbike ban is implemented.

Group	**Questions (selected)**
General	Age, gender, home location, occupation
Living conditions	Living duration, property type, status, home ownership, water quality, open space, noise, school access, market access, hospital access, bank access, security, leisure access
Household composition and vehicle ownership	household car ownership, household motorbike ownership, household e-bike ownership, household bike ownership
Primary trip	Origin, destination, travel purpose, mode choice, the reason for mode choice, travel time, frequency per day, frequency per week, frequency per month
Mode choice	Frequency usage of a car, motorbike, e-bike, bike, bus; future purchase; reason not to buy a certain vehicle; distance to public transport; opinion of of a certain transport mode
Motorbike ban	Awareness of the potential motorbike ban, opinion, alternative vehicle: car, e-bike, bike, taxi, bus, light rail, taxi, walk; reason for vehicle ban: convenience, cost, parking, other

Table 1: Main survey questions

3.2 Exploratory data analysis

The first step in our 4-step methodology for this paper is an exploratory data analysis (EDA). This is a comprehensive exploration on the data to discover the trends, patterns and relationships that are not readily apparent. The EDA is also important for choosing variables for the classification model in Step 2, as it sums up the distributions in the data and the relationships between variables.

3.3 Classification model development

To predict an individual's perceptions of a future transport policy, we plan to achieve the first defined objective by using a classification model in machine learning. This is because the data is complex, noisy and many of the variables are categorical, which would make it difficult for statistical models to learn from the data. We will develop the model to predict the binary variable of an individual's perception of the motorbike ban, i.e. whether a person 'agrees' or 'disagrees' with the ban. The inputs of these models are answers for all the 142 questions in the survey, as described in Table 1.

Among various classification techniques in machine learning, we adopt Extreme Gradient Boosting (XGBoost). Gradient boosting is one of the latest and most powerful methods to develop a predictive model in machine learning. The main idea behind boosting is to build a strong model from an ensemble of weak models in series, with each model fitted to the residuals of the previous models. Unlike bagging algorithms (e.g. Random Forest), which only deals with the high variance in a model, gradient boosting deals with the bias-variance trade-off, and is thus generally considered more powerful. Extreme Gradient Boosting (XGBoost) is not a single method but can be considered a class of optimised distributed gradient boosting models. Although there could be multiple types of XGBoost models, we focus on the gradient-boosted decision trees, as our classification model fits well to decision trees, and this is also the most popular XGBoost algorithm in the literature. Interested readers should refer to the original paper (Chen and Guestrin 2016) for a more detailed description of the XGBoost algorithm.

3.4 Variable importance analysis

To meet the third objective, an important modelling step is to investigate the impact of each variable through the importance of the variables. It measures how much each variable contributes to the classification models to make accurate predictions. Variable importance shows how much each variable can be used to explain the perception of the motorbike ban.

Variable importance analysis is also an important step in the development of classification models. Variables with low importance are omitted from the classification models to simplify the models. Here we look at a ranked list of variable importance using Gini-based importance index (or Gini impurity) (Breiman 1996), and then iteratively omit variables until we no longer can maintain the same accuracy. The final list $\mathcal{X}_\mathcal{I}$ forms the simplest model we can have (least number of variables) and informs us that a shorter survey can be done at future data collection without compromising models' accuracy. $\mathcal{X}_\mathcal{I}$ is also helpful for Hanoi transport authorities in understanding the factors that affect individual perceptions of transport policies.

Finally, we also consider using the trained classification model to predict whether a specific hypothetical person is in $\mathcal{X}_\mathcal{C}$. This enables us to understand how each individual's demographics and travel attributes contribute to their perceptions of transport policy. To do this, we adopt a method from cooperative game theory named Shapley value Štrumbelj and Kononenko (2014), using a Monte-Carlo approximation as described in Štrumbelj and Kononenko (2014) because there are such a large number of variables.

$$\hat{\phi}_j = \frac{1}{M} \sum_{m=1}^{M} \left(\hat{f}\left(x^m_{+j}\right) - \hat{f}\left(x^m_{-j}\right) \right) \tag{1}$$

where $\hat{f}\left(x^m_{+j}\right)$ is the outcome of the prediction of variable x, but with a random number of values of characteristics replaced by values from a random data point. The x-vector $\hat{f}\left(x^m_{-j}\right)$ is also from the same random sample. We calculate the Shapley value for each variable in $\mathcal{X}_\mathcal{C}$.

4 Explanatory data analysis: a deep dive into individual perceptions on the motorbike ban

This section takes a deep dive into the travel survey in Hanoi, Vietnam to understand the data, especially related to individual's perceptions of transport policies. We focus on survey questions that are related to the potential motorbike ban (see Table 1 for more details) as these are the most important for this paper. The understanding from this section is important for the development of the classification model (RQ 2), as well as to derive policy implications regarding the motorbike ban (RQ 3).

4.1 Age and Gender

We first look at how respondents of different age and gender think about the potential motorbike ban in Figure 3. Between the genders, females are more likely to oppose the ban. There are around 40% more female respondents within the 28 to 35 years old bracket who disagrees with the ban compared to those who agrees. We can also observe a similar pattern in other age brackets for female respondents. On the other hand, there are more males within 18 to 25 years old who are in support of the ban. We hypothesise that females may use more motorbikes than males in Hanoi. This hypothesis is then confirmed in Figure 4. Figure 4(a) shows that the number of motorbike trips are relatively similar between the two genders, but in Figure 4(b), we can see that there are in fact more males than females in our dataset, which means that each female makes more motorbike trips than each male respondent. When we look at the distribution of mode choice for the primary trip, males make more trips by car than females, whereas the motorbike is the dominant mode for females in Hanoi.

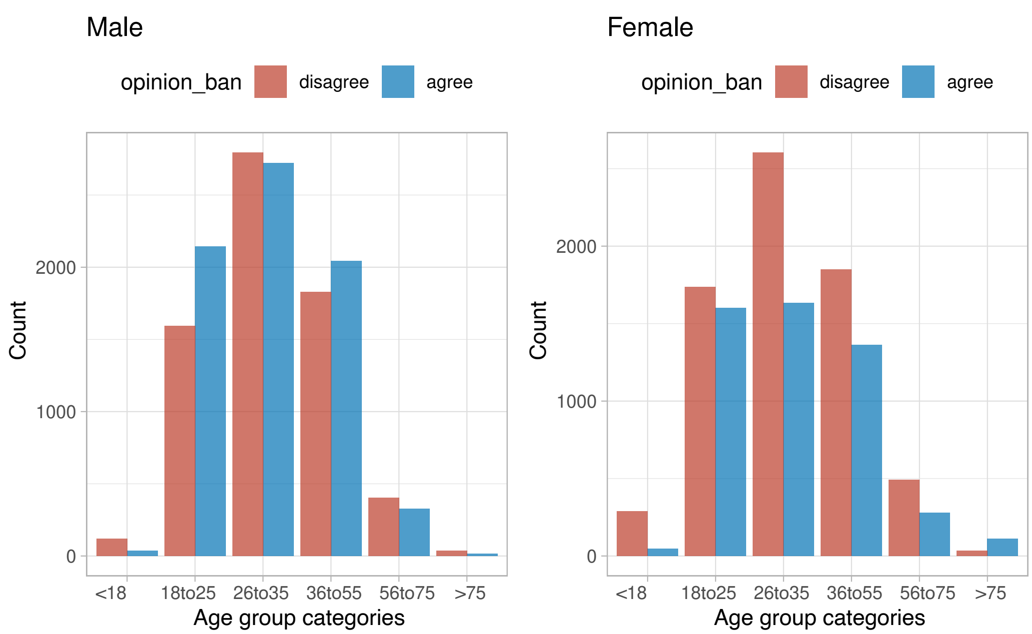
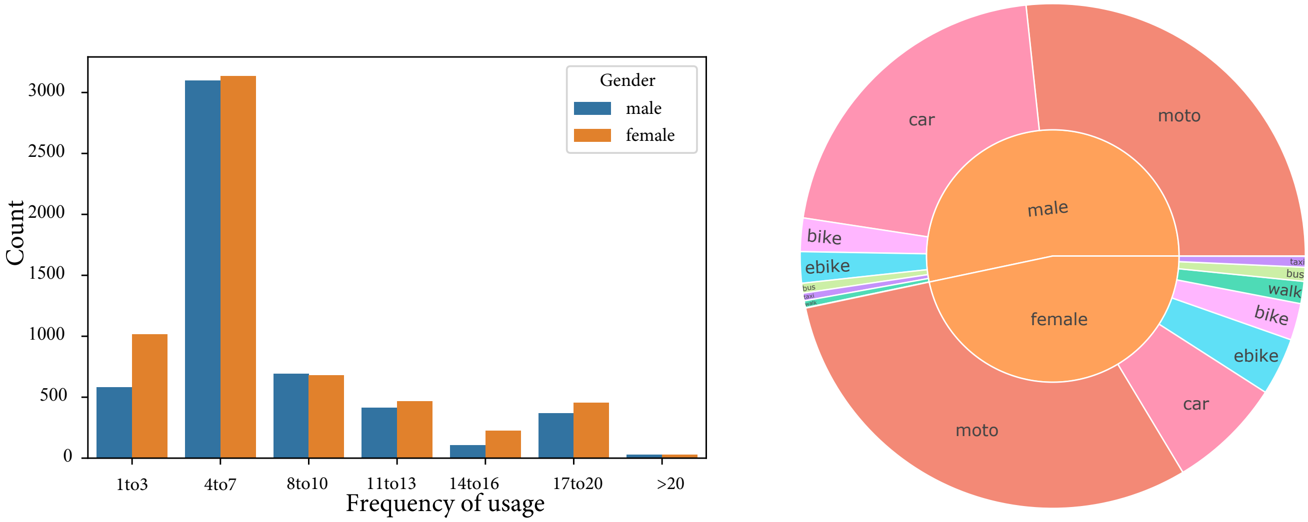


Fig. 3: The distribution of opinions on the ban in different age and gender

4.2 Vehicle ownership and mode choice

This section explores the distribution of vehicle ownership and choice of travel mode in the data, as they are directly related to individual's perceptions of the motorbike ban and the access to an alternative vehicle if the ban is implemented.

Figure 5 illustrates the overall distribution of vehicle ownership in households. The x-axis shows the number of owned vehicles in the household for 4 types of vehicles (car, motorbike, e-bike and bicycle). The y-axis shows the proportion of households having zero to more than or 5 vehicles of one of the 4 types. Figure 5 illustrates the motorbike popularity in Hanoi, where 95% of households have at least one motorbike, with the majority of them having 2 or more motorbikes. We also see that not only that Vietnam is the motorbike haven, cars are surprisingly popular, with more than 60% of households having at least one car. Motorbikes are still the only type of vehicle that households keep more than 3 of. Depending on the future development of ownership for private travel modes such as cars and motorbikes, the traffic congestion and safety in Hanoi would be very different.



(a) Gender vs frequency of travels with motorbikes

(b) Gender vs mode choice for the primary trip

Fig. 4: Gender and travel mode choice

We now explore the reasons why a travel mode is and is not chosen. Figure 6(a) shows that convenience is the main reason why someone in Hanoi chooses many of the travel modes, e.g. motorbikes, taxis and cars. Cars and motorbikes are both private vehicles, and similar to many other cities, people in the survey think that they are convenient and time-saving. ‘Cost’ is the biggest reason why taking a bus and walking are chosen in Hanoi as the main mode of transport (the fare for a bus trip in Hanoi varies between USD \$0.3 to \$0.5), along with ‘other’. While it is not known what could be the reason for the choice of ‘other’, the lack of vehicle ownership may be a potential reason why ‘other’ has been chosen as the reason why people walk or take a bus.

On the other hand, Figure 6(b) illustrates the largest concerns that people have with regards to a certain travel mode. The figure shows how each mode should be improved to attract more riders. For bikes, the main issue is speed, as around 60% of respondents chose ‘Slow’ as the main concern. It is also interesting to see that ‘unsafe’ is only associated with e-bikes and not bikes. This could be due to the fact that they are fast but also quieter than petrol motorbikes; this perception may change as e-bikes become more commonplace. Traffic jam is the main reason why a person dislikes motorcycles in Hanoi, and people also do not choose e-bikes for the same reason. However, traffic jams are not an issue for cars in Hanoi, although one would probably experience more travel delays in a car than on a motorbike. It may be that the health and temperature benefits of driving cars outweigh concerns about traffic jams. The cost is the main concern people in Hanoi have with regard to cars, which suggests that many more people will buy cars when the cost of owning a car is less of a concern to them. We explore this point further by looking at the intention of choosing an alternative travel mode if a motorbike is banned (Figure 7(a)), or purchasing a future vehicle (Figure 7(b)).

Figure 7(a) shows a “flow” of people from their current primary mode of transport on the left, to a future transport mode on the right if motorbike is banned (there is no ‘motorbike’ option on the right). The size of the bands show the number of respondents. Figure 7(a) shows that motorbikes and cars are currently the most popular modes of transport in Hanoi. The mode shares for the alternative vehicle are almost equal across all the possible modes in Hanoi, which shows that people are willing to take the currently less popular modes in Hanoi, such as taxi, buses and walking.

However, Figure 7(a) also shows that the majority of car users are going to stick with cars if motorbikes are banned. When being asked which vehicle they will buy in the future, motorbikes and cars remain the most popular options, apart from “no plan”. Figure 7 shows a concerning fact that majority of motorbike users plan to either buy a car or another motorbike in the future, which suggests that the number of private vehicles will increase. Many cars users are planning to buy another car as well. While the majority of people has no plan to buy more private vehicles, the road authorities in Hanoi should still plan to reduce the 40% of motorbike users who plan to upgrade to cars.

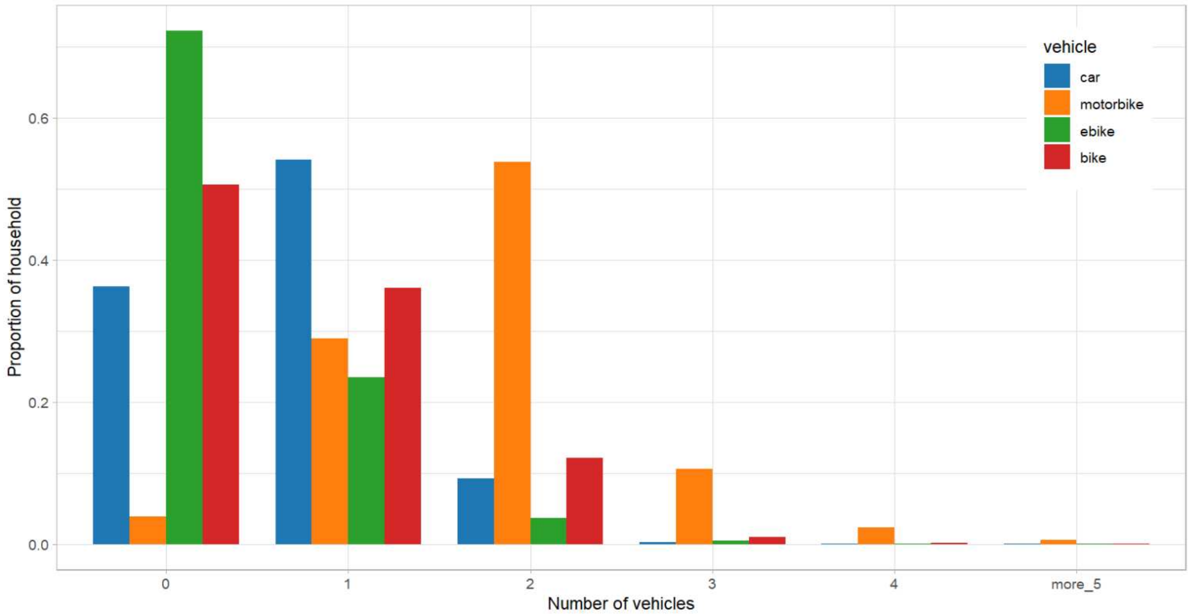


Fig. 5: Distribution of family vehicle ownership

4.3 Individual opinion on the potential motorbike ban

Figure 8 shows the proportion of individuals' awareness of the motorbike ban and whether they agree with it. The smaller the circle in Figure 8 the smaller the proportion of people who are aware of the ban ('yes'), or not ('no'), or do not care about the ban ('donotcare'). Figure 8 shows that around half of people in the survey have already heard about the ban previously, and among the rest, around 15% of them do not care about it. The larger circle in Figure 8 illustrates the proportion of people who agree or disagree with the motorbike ban. For all 3 groups, the number of people who disagree is always larger than the number of people who agree. However, the proportion of agreement is higher among people who are aware of the ban. It shows that if people hear about the ban for the first time while answering the survey, it is more likely that they will oppose it. This suggests public announcements and consultations are an effective way to educate people about proposed policies and to make them more acceptable.

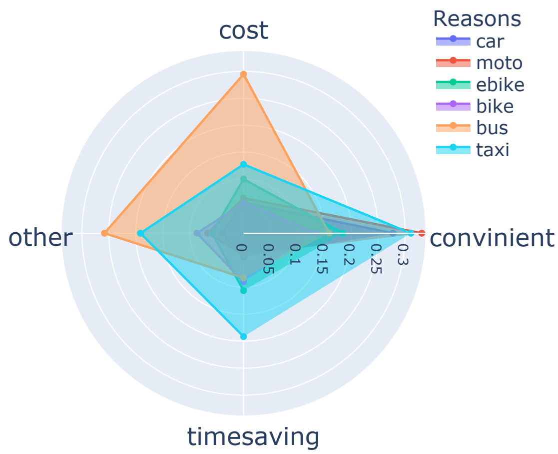
5 Classification modelling results and variable importance

As discussed in Section 6, a policy that involved restricting petrol motorbikes from some parts of Hanoi would be one possible way of reducing pollution in the city centre. However, such a policy would be extremely contentious. To better understand the underlying factors that might make people more or less amenable to a ban, we now develop a machine learning classification model to explore the relationship between an individuals' attributes and their perception of a motorbike ban policy. The classification model aims to use all of the attributes in the survey to predict whether an individual person 'agrees' or 'disagrees' with the policy. Figure 9 illustrates the process we take to develop and analyse the outputs from this model.

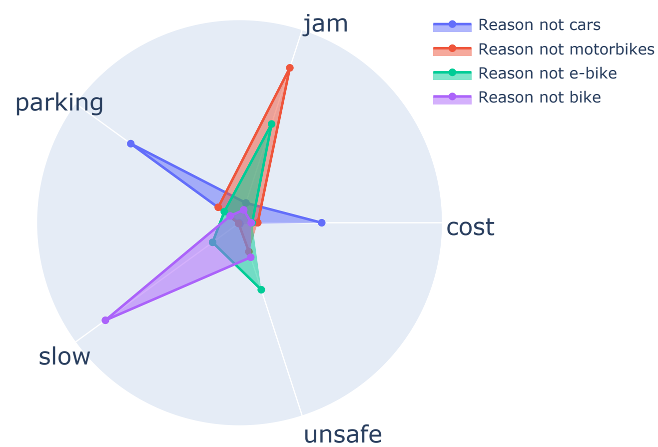
We start by choosing the classification method to develop our model. The next step (data budgeting) processes and cleans the input data for this model. The core model development then optimises the hyperparameters for this model, to make sure that it fits best to the training data. Finally, after evaluating the model on an unseen testing dataset, we analyse the importance of each variable to understand its impacts on the motorbike ban in Hanoi.

5.1 Classification Method: XGBoost

Extreme Boosted Trees (XGBoost) is a supervised classification model as the classification method in this paper. XGBoost is an ensemble model of multiple decision trees, where the ensembling is performed by boosting (weak learners are trained sequentially to minimize training error) Chen and Guestrin (2016). We adopt XGBoost in this paper because this is one of the most powerful and versatile methods in classification, which usually outperforms other popular machine learning methods, such as Random Forest, Support Vector Machines or Logistic Regression

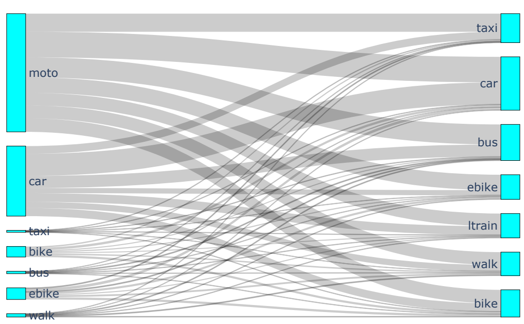


(a) Reasons for a travel mode to be chosen

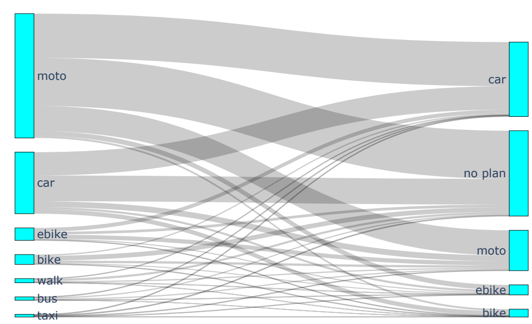


(b) Reasons for a travel mode NOT to be chosen

Fig. 6: Reasons for travel mode choice



(a) Individual plan for an alternative travel mode



(b) The intention to buy a future vehicle

Fig. 7: Analysis of alternative and future vehicles

in classification tasks Freeman et al. (2016). In this section, we describe the steps we took to develop, tune, and analyse the outputs of the XGBoost model when it is used to classify individual respondents into two groups: those who agree and disagree with the potential motorbike ban.

5.2 Data budgeting

We first split the data into two distinct sets, namely, the training set and the test set. The training set (typically larger) is used to develop and optimise the model, while the test set is used as the final arbiter to determine the efficacy of the model. The test set remains unseen to the model and is only used to evaluate the model in the last step (as seen in Figure 9). We randomly split the data into 70% for training and the rest for testing. Other data budgeting steps include:

- We use 'one hot encoding' to convert all categorical variables into dummy variables.
- We drop all 'living conditions' variables in Table 1, as many data points in these variables are missing.
- For other variables, we impute the missing values by either the median value (numerical variables), or the mode (categorical variables)
- We drop near zero variance predictors i.e., predictors with only a handful of unique values that occur with very low frequencies

Fig. 8: Individuals' awareness of the motorbike ban, and their opinion on it

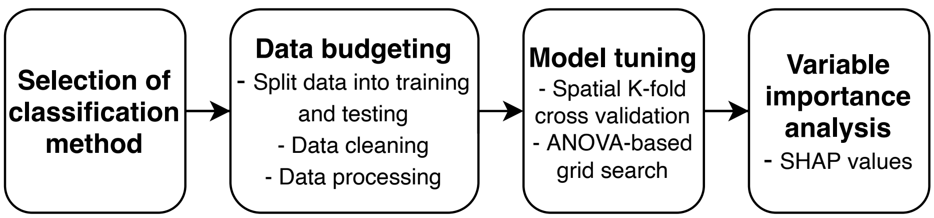


Fig. 9: The classification model development and analysing workflow

– We remove 'tram' as one of the travel option in the survey because this mode is brand new in Hanoi and most respondents filled in the survey before this mode was available.

5.3 Model tuning

Model tuning is the most important step in developing a machine learning model, where its hyperparameters are optimised for the best accuracy against the test dataset and avoid overfitting. XGBoost presents several hyperparameters related to the boosting process (e.g learning rate), decision trees used (e.g tree depth) and model complexity (e.g regularisation). If properly tuned, these hyperparameters can greatly improve the model's flexibility and accuracy. Hyperparameter values cannot be estimated directly from the data. Instead, resampled datasets created from the training set together with an appropriate parameter search method are used to determine the best possible values of the hyperparameters a priori.

K-fold cross validation is one of the most popular technique to avoid over-fitting in model tuning. The data is first divided into k subsets. In each cross-validation step, one of the k subsets is used as the testing dataset, and the other $k-1$ subsets form a training dataset. The error is then averaged over all subsets.

However, our target variable (opinion on the motorbike ban) is spatially correlated. Figure 10 shows a strong relationship between geospatial location and agreement to the motorbike ban. Here the locations of individuals were aggregated to their constituent district—as defined for the 2019 Vietnamese population and housing census—and then ordered by the percentage of respondents who agreed with the ban. It appears that people in more populated districts (larger bubbles) tend to agree with the ban more.

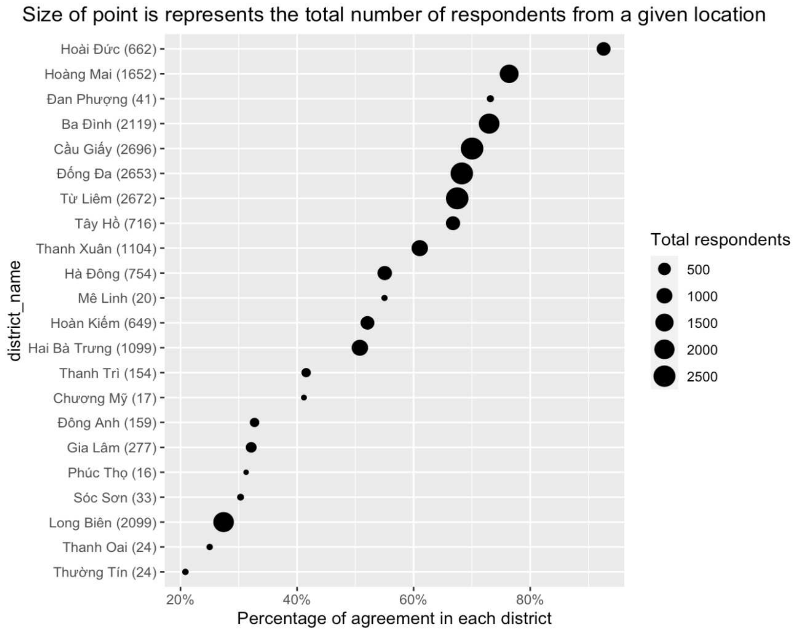
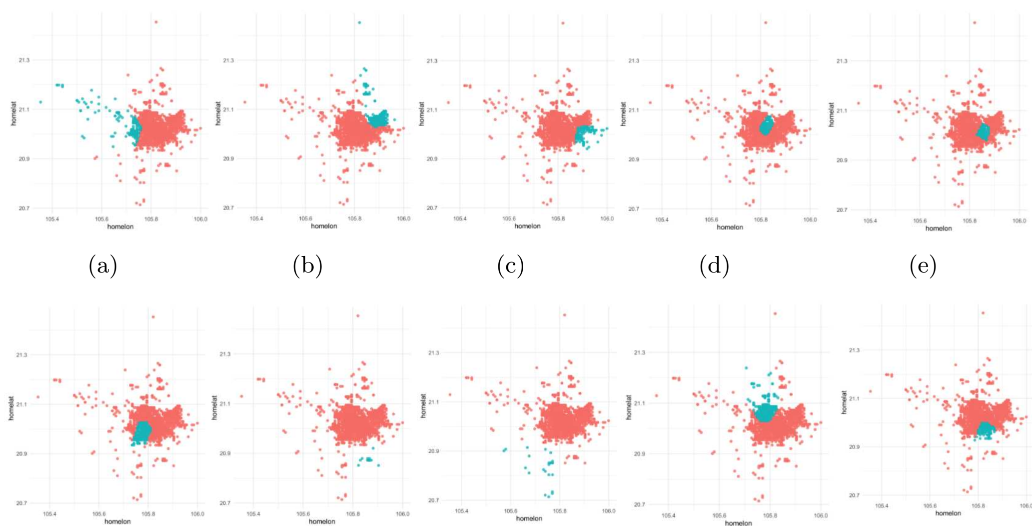


Fig. 10: Percentage of agreement at different districts in Hanoi

To address for the spatial correlation in the target variable, we adopt a spatial variation of the K-fold cross validation. Spatial cross validation splits the data into k groups of disjointed sets using k-means clustering of spatial coordinates. This ensures that we account for the presence of spatial autocorrelation in geospatial data. Figure 11 illustrates the 10 spatial clusters that are the result of our spatial cross-validation based on k mean.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

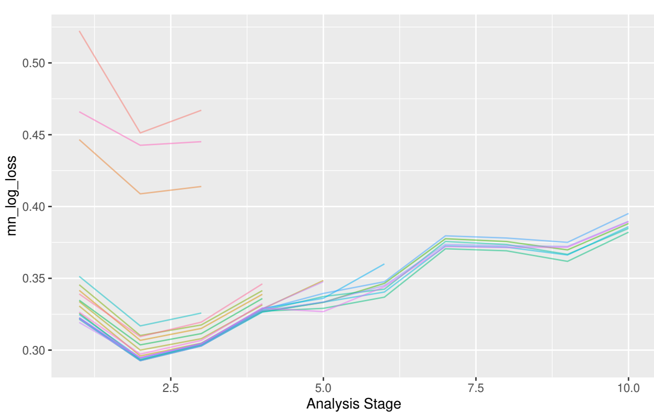
Fig. 11: Spatial cross-validation cluster using individual home location and K-means clustering

We then adopt a grid search technique to find the best hyperparameters for the XGBoost model. We begin this process by setting the learning rate, $\eta = 0.01$. η is the step size shrinkage to prevent overfitting and make the model learning process more conservative. The grid search procedure then focusses on two important variables:

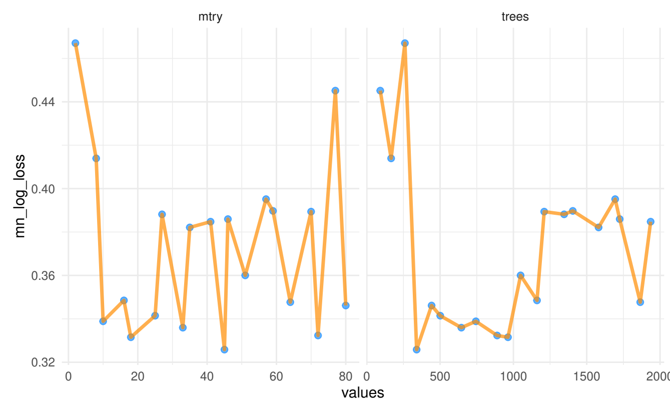
– $mtry$ represents the proportion of predictors that are sampled at each split during the creation of tree models. It accounts for the fact that some predictors were strongly related to the outcome.

– *tree* represents the total number of successive trees in the ensemble and was tuned to avoid reaching a point where the addition of more trees would just try to explain the residuals that are random, resulting in overfitting.

We adopt an ANOVA-based grid search technique, as described in Kuhn (2014), to improve the efficiency of hyperparameter optimisation compared to the classical grid search. Unlike the traditional grid search when all combination of variables are evaluated equally, the idea is to evaluate a set of metrics (e.g. accuracy) and for each iteration, we eliminate combinations of parameters that are unlikely to be the best using an ANOVA model. 20 random combinations of potential hyperparameter values were selected using a maximum entropy design (Shewry and Wynn 1987) to produce values that cover the parameter space with the smallest chance of overlapping values. We then eliminate combinations of hyperparameters that are unlikely to return the best prediction accuracy using *mean log loss* as the performance index. The next iteration is then more likely to return a better combination of parameters, as illustrated in Figure 13(a). Figure 13 (b) shows how precision varies with different combinations of hyperparameters *mtry* and *trees*. The figure shows that there is a range of values for these hyperparameters that the XGBoost model will provide the lowest *mean log loss*.



(a) Incremental elimination of tuning parameters



(b) Accuracy varies with different combinations of hyperparameters *mtry* and *trees*

Fig. 13: Results from the ANOVA-based grid search

Finally, using the best-fit model from the ANOVA-based model tuning procedure (with spatial cross-validation), we can evaluate our model performance on the test dataset. Recall that test data have not been seen to the model thus far. The performance of the final developed model on the test data is showed in Table 2.

Metric	Value
Accuracy	0.878
Recall	0.853
Specificity	0.919
PPV	0.945
F-measure	0.897
ROC-AUC	0.953

Table 2: XGBoost model performance against the test data

Table 2 shows that the developed model has high performance on all the evaluation metrics. The *Accuracy* metric shows that the model predicts accurately 87.8% of the test data. The high value of *Recall* shows that the model has 85.3% of detecting a person who would agree with the motorbike ban, while the model is also very good at avoiding false positives with *Specificity* equal to 91.9%.

5.4 Variable importance

Having attained satisfactory performance from the classification model, this section explores the importance of the independent variables in the developed XGBoost model to understand how each one affects individual perceptions of the motorbike ban. We adopt one of the recent techniques to interpret black-box models that has proven effective among others. SHAP (SHapley Additive exPlanations) (Molnar 2020). SHAP shows the importance of variables by comparing the model's prediction with and without the variable in every possible order. The y-axis in Figure 14 shows the most important variables with an aggregated SHAP important value equal to or greater than 0.1. The higher the aggregated variable importance value, the more important the variable to the XGBoost model output.

Figure 14 also comprises individual SHAP values for each data point in our test dataset. These values are shown on the x-axis of Figure 14, and are coloured with two colours for the dummy variables and a full spectrum of colours for the continuous variables (thus the patterns for the dummy variables are easier to identify compared to the continuous variables). For each respondent and each variable, positive SHAP values contribute to the approval of the motorbike ban, and vice versa.

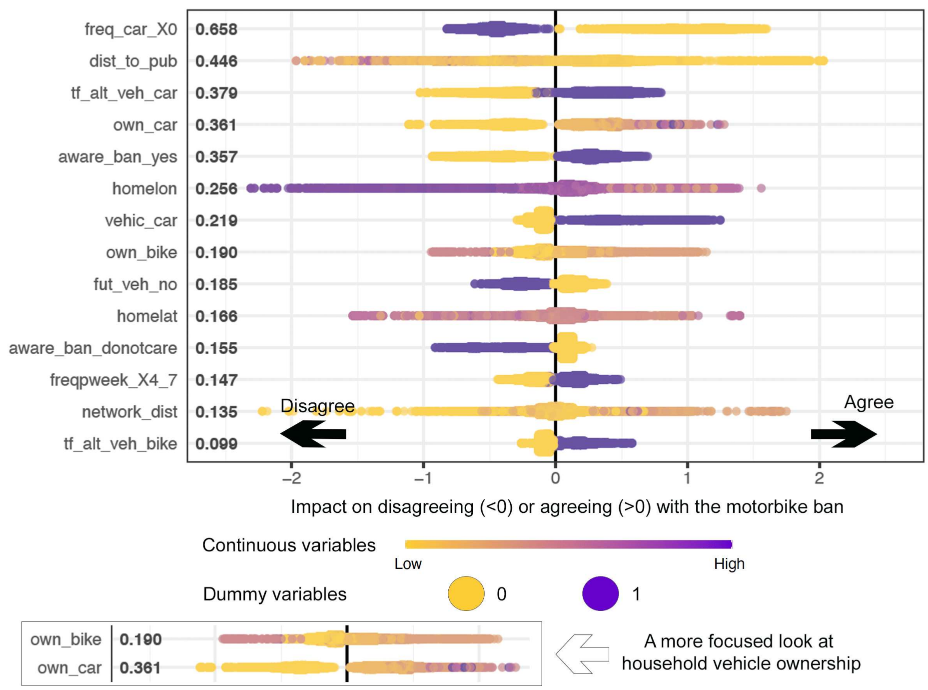


Fig. 14: Variable importance using SHAP values. x-axis: individual SHAP values, y-axis: aggregated variable importance values

Figure 14 shows a few important patterns that shed some light into individual perceptions on the motorbike ban. As a reminder, the categorical variables have all been encoded into 'dummy variables' of binary values for modelling.

- The '$freq\_car\_X0$' is a dummy variable, that has a value of 1 if the respondent does not travel by cars, and 0 otherwise. This variable has the largest impact in an individual's acceptance of the motorbike ban. Low $freq\_car\_X0$ values, coloured yellow, are associated with positive SHAP values and high (purple) values are associated with negative SHAP values. This means that if the respondent does not use a car ($freq\_car\_X0$ equals 1), then they are less likely to agree with the ban and vice versa.
- The second most important variable is the distance to public transport ($dist\_to\_pub$). As this is a continuous variable, the patterns are not as clear as for $freq\_car\_X0$, but the results appear to suggest that higher distances to public transport are associated with disapproval of the potential motorbike ban. This shows that people in Hanoi are still relying on motorbikes, and that providing better transit accessibility would increase the probability of the motorbike ban to be accepted.

– It is noteworthy to see that the motorbike ownership variable itself does not make it to the top 14 most important variables, perhaps because most households own at least a motorbike. Only the two most important alternative vehicles (*own_car* and *own_bike*) show their importance in individual's perceptions on the motorbike ban. Between them, car ownership has stronger impacts (SHAP value equals 0.361) compared to bike ownership (SHAP value equals 0.19), which shows that car owners have a stronger opinion on the ban compared to bike owners. The small figure at the bottom of Figure 14 provides a more focused look at these two variables. Families who own more cars (higher value of *own_car*) are more likely to support the ban on motorcycles than families who do not own cars. On the other hand, people who owns more bikes (high value of *own_bike*) are either strongly supportive or opposed to the ban. While being an alternative travel mode, motorbikes are often seen as an immediate upgrade from bikes in Vietnam.
– The analysis of future vehicles in Figure 7 shows that cars are becoming one of the most important mode of transport for people in Hanoi. This is confirmed by the fact that *tf_alt_veh_car*, the dummy variable that equals one if a person chooses cars as an alternative vehicle to replace motorcycles, and zero otherwise, is the third most important variable in our prediction model. If a person thinks that cars are their alternative vehicles (high *tf_alt_veh_car*), it is more likely that they would approve the potential ban on motorbikes.
– We can also confirm several earlier findings from the exploratory analysis, such as the importance of the awareness of the ban, by looking at two variables *aware_ban_yes* and *aware_ban_donotcare*. If an individual is aware of the ban (*aware_ban_yes* equals 1), they are more likely to accept the ban and vice versa. If the person does not care about the ban (*aware_ban_donotcare* equals 1), it is likely that they will reject the ban.
– Spatial variables such as home location (*homelon* and *homelat*), as well as the routed distance weighed by mode for the particular trip (*network_dist*) play important roles in individual perceptions of the ban. The SHAP value for *network_dist* shows that further the travel, the more likely a respondent would approve the motorbike ban.
– Socio-demographic attributes such as age, gender, and occupation are not as important as we expected. None of these attributes has made it to the top-14 most important variables. It shows that sociodemographic characteristics do not have significant impacts on how people think about transport policies in Hanoi.

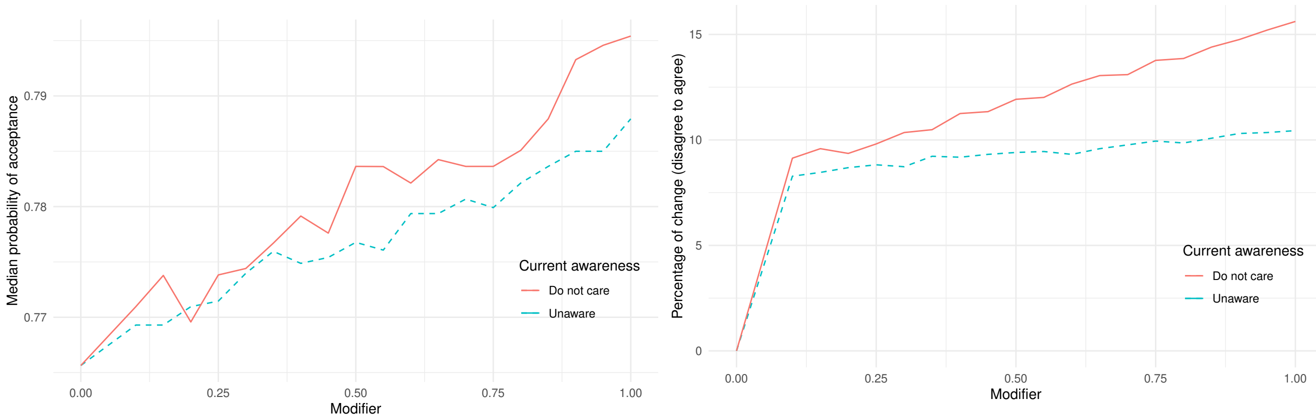
6 Policy scenarios

Having identified which factors are the most strongly associated with positive or negative attitudes towards a ban, this final step in the analysis aims to use the classification model to investigate what policymakers might do to make people more or less favourable towards the motorbike ban. Because the model is nonparametric, we cannot simply analyse its parameters to understand the impacts of a variable change on its output (the individual probability of acceptance of the motorbike ban). We instead adopt here a Monte Carlo simulation approach, to systematically modify the testing dataset (as described in Section 5) to simulate a change in one particular variable, while keeping all other variables as they are. Here we use the same testing dataset for these policy scenarios, to make sure that the model has not seen the data before the scenarios. Finally, the trained XGBoost model is employed to make predictions of individuals' probability of accepting the motorbike ban. By comparing the individual probability of accepting the motorbike ban before and after the change in one variable, we will be able to understand the impact of that change to individual's opinion on the ban. Although almost any variable can be modified, we only focus on the most important variables in Figure 14 as they are the most important to perceptions about motorbike ban.

6.1 Scenario 1: Improving the awareness to the motorbike ban

Both the exploratory data analysis in Section 4 and the classification model in Section 5 have confirmed that awareness of the motorbike ban (before conducting the survey) is one of the important variables to decide whether an individual accepts the ban or not. In this section, we are interested in how the probability of acceptance would change as people become more aware of the motorbike ban. We introduce a Modifier to simulate a change in "aware" of the motorbike ban by 10% to 100% of those who are "unaware" or "do not care" about the ban. Here if Modifier equals 10%, we randomly pick 10% of those who are "unaware" or "do not care" to change to "aware" of the ban. For example, a policy scenario with Modifier equal to 0.5 would randomly pick 50% of those who are "unaware" or "do not care", and change their awareness of the motorbike ban to "aware". The model is then implemented on the new testing dataset with the modified awareness variable to provide the probability of acceptance for the individual respondent. As the classification is probabilistic, the model can return a continuous probability that a respondent accepts the motorbike ban. Figure 15(a) shows the median value of these individual probabilities.

Figure 15(a) shows that if people who are "unaware" and "do not care" change to "aware", we will see significant improvements in the probability of acceptance. If the model returns a higher probability of acceptance than



(a) Median probability of acceptance

(b) Proportion (%) of people who changed from disagreement to agreement with the ban

Fig. 15: Scenario 1: How individual probability of acceptance changes as people become more aware of the motorbike ban

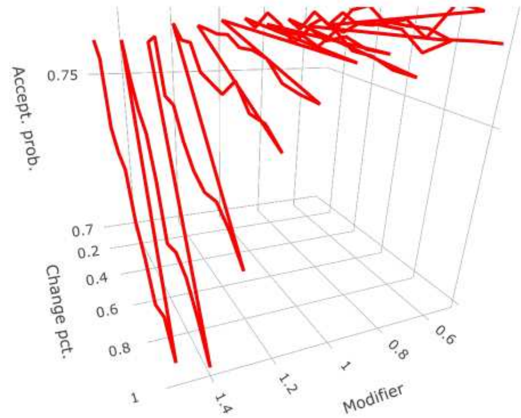
the probability of rejection of the motorbike ban, we predict that the respondent accepts the ban. To this end, Figure 15(b) shows the percentage of people who changed from disagreement to agreement with the ban as they become aware of it. We can see that the first 10% changes in the level of awareness can move around 8% of people who currently disagree with the ban to acceptance.

If we can improve the awareness of people who “do not care” about the ban, the impacts are more significant than simply informing people who are “unaware” about the ban. Making this improvement may potentially increase the median probability of acceptance from around 76% to 80%, and convince up to 15% of people from disagreement to agreement. In addition, our survey does not ask people *how* to become aware of the ban, so some people may have become aware through negative reporting in the media or other sources. If the government was to make people aware through positive messaging, the acceptance probability might be even higher.

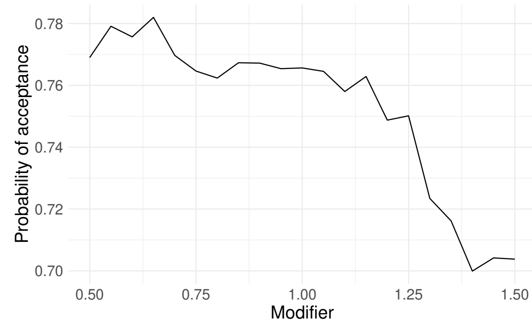
6.2 Scenario 2: Distance to public transport

Distance from public transport is the second most important variable for our proposed XGBoost model, as seen in Figure 14. In this scenario, we simulate a change in the distance to the nearest public transport service for individual respondents in the survey data. Similarly to the previous scenario, we introduce a *Modifier* to change the distance from public transport from 0.5 times to 1.5 times the current distance to public transport for each individual. Thus with the *Modifier*, we simulate the case where the distance to the nearest public transport stop is half to 1.5 times the current value. Similar to the practice, the distance to public transportation generally cannot be increased or decreased for everyone, as any new or reduced transit services will make direct impacts to those who are close to those services. Thus, we also introduce a *Change percentage* variable, to randomly pick an individual in the data set with a probability of 10% to 100% to then apply *Modifier* on its distance from public transport. Figure 16(a) shows the changes in the median probability of acceptance to the motorbike ban, as *Modifier* varies from 0.5 to 1.5, and Change percentage varies from 10% to 100%. To better see the pattern, we also introduce another simplistic scenario in Figure 16(b) where Change percentage is 100%, or every one receives the same change in the distance to their nearest public transport stops.

Generally, we see that if public transport is a suitable mobility option, people in Hanoi will be more open to a potential motorbike ban. Figure 16 shows that, surprisingly, the median probability of acceptance only increases marginally as the distance from public transport is reduced, but would reduce significantly if the distance increases. One possible explanation for this fact is that the nearest transit stop may not be the one the respondent uses, but if the stop is too far from where they live, then transit is no longer a rational option. The *Change percentage* exacerbates this fact, as the acceptance probability reduces significantly when more people are affected by the increase in distance from public transport. As the city expands, urban sprawl may lead to an increase in distance to transit stops if policymakers in Hanoi are not following transit-orientated development, and may then lead to further dependency on private vehicles. Policymakers in Hanoi should at least maintain current access to public transport to reduce dependency on motorbikes and private vehicles in Hanoi.



(a) With Change percentage



(b) Without Change percentage

Fig. 16: Scenario 2: Distance to public transport

6.3 Scenario 3: Individual modal choice

From the previous analysis in Section 4, we can observe that people in Hanoi seem to see cars as the alternative vehicles if motorbikes were banned: people using cars are more likely to accept the ban, people who think cars as the alternative vehicles are also more likely to accept the ban. However, we can all agree that the car is not the answer to motorbike dependency in Hanoi, as the dependency on cars could be worse for traffic congestion and safety issues in the city. In this scenario, we investigate the scenarios in which people in Hanoi use more and less cars, motorbikes, buses, and bikes for their main trip. Figure 17 illustrates the scenarios in which we increase the modal share of a mode from 10% to 100% by randomly changing a respondent who travelled with another mode of transport to a car, motorcycle, bus, or bike.

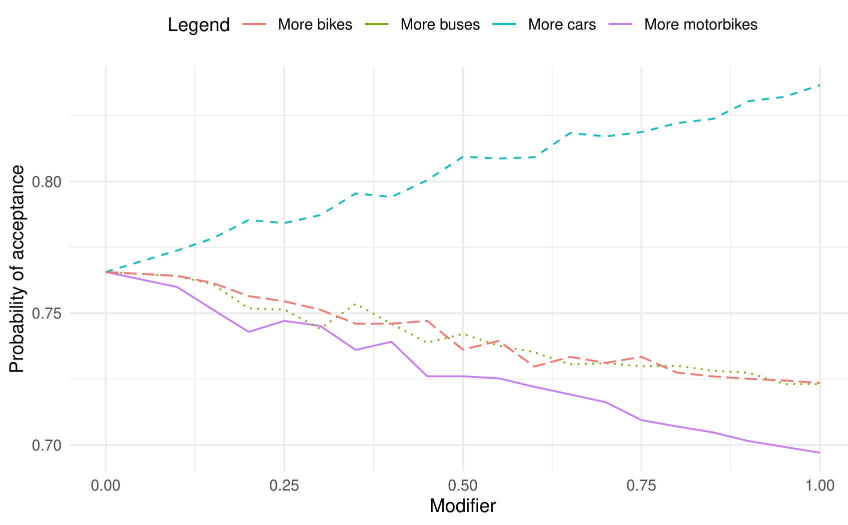


Fig. 17: Scenario 3: Individual modal choice

Figure 17 shows that the probability of acceptance for the ban will be reduced to only 70% if we have more motorbike users. Car users are supportive of the ban, and we can increase the acceptance rate of the ban to nearly 85% if we have more car users on the road. However, as previously discussed, this is not ideal because we may move from motorbike dependency to car dependency, which can lead to more congestion, pollution and crashes.

Surprisingly, if we have more people using bikes and buses, there are less people who are supportive of the ban. This is because motorbikes are considered an immediate upgrade for public transport riders and cyclists in Hanoi. Public consultation is needed to make buses and bikes the alternative vehicles of motorbikes and not cars.

7 Discussion and Conclusions

In recent decades, motorbikes have transformed Hanoi, Vietnam, from a "tranquil" city filled with bicycles to a metropolis that is instead famous for the "constant buzzing and honking" (Hansen 2022) that they create. The rapid increase in motorbike use has been negatively associated with traffic congestion, safety, and pollution, as with many cities in the Global South. At the same time the motorbike has become a vital part of life for most residents which makes reducing motorbike dependency particularly difficult. One potential solution to dependency is to ban motorcycles in some parts of the city. However, with 90% of the traffic in Hanoi on motorbikes, any future ban on them will need the support of the public to be successful. We conducted a travel survey with six main sections, collecting information on general demographics, living conditions, family composition, vehicle ownership, regular journeys, mode choice and especially the perception of a potential motorbike ban in the central business district of Hanoi. We then develop a four-step data-driven modelling framework to answer four research questions:

7.1 Can we predict an individual's perception of a future transport policy?

Our classification model has a high accuracy of nearly 90% on an unseen test dataset.

7.2 What policy implications can we obtain from the data and the modelling?

In the second and third step of our analysis (Section 5) leads to a number of interesting findings and associated policy implications:

- Spatial variables such as home latitude and longitude, and distance to central business district are within the top 25 most important variables for our model. This suggests that the decision of which area to ban motorbikes can be crucial for transport authorities in Hanoi.
- The frequency of usage and ownership of other modes of transport, such as cars and public transport, is essential to decide whether a person agrees with the motorbike ban or not. It shows that while residents are still relying on motorbikes, providing more alternative travel modes may lead to reduction in motorbike usage.
- Sociodemographic attributes are not as important as we expected. Perhaps more data from a wider population is needed in the future development of the survey.

7.3 What factors are most and least important in explaining individuals' perceptions?

The explanatory data analysis – conducted in step 1, prior to modelling (Section 4) – lead to some interesting policy implications directly from the data:

- Motorbikes are the most popular type of vehicle in Hanoi, with 95% of households owning at least one vehicle and some having more than 5. It supports the existing evidence (and the experience of residents) that the city is dependent on motorbikes (Hansen 2016, 2017; Ngoc et al. 2017). However, on top of owning motorbikes, around 60% of the households in the survey also own a car.
- Convenience is the main reason why people in Hanoi chooses private transport modes (e.g motorbikes and cars) as their main mode of travel. This means that efforts to reduce the convenience of these modes through measures such as reduced parking availability or banning motorbikes in certain areas may reduce their attractiveness to road users.
- To promote active transport and encourage people in Hanoi to use more pedal bikes and e-bikes, it is necessary to deal with their main concerns of being 'slow' and 'unsafe'. Bicycle infrastructure such as bike lanes, bike parking facilities and a rental bike system may support these modes and reduce dependency on other modes of transport. Ongoing research by the authors is simulating the implications of changes in transport mode across the network and might be useful here in suggesting the optimal locations of new bike lanes.
- 'Traffic jams' are the main concern of people in Hanoi regarding motorbikes and e-bikes as modes of transport, but surprisingly, traffic jams are not a concern for cars.

– Around a third of motorbike users are planning to buy cars in the future, and one fourth of them plan to buy more motorbikes . While our survey shows that the majority of people has no plan to buy more private vehicles, the road authorities in Hanoi should still plan to encourage people who are currently using private transport to use more public and active transport, rather than buying more private vehicles. However, this requires investment in public transport infrastructure, which is a considerable financial overhead in most lower- and middle-income countries.
– The awareness of the motorbike ban is essential to decide whether an individual would be more likely to accept or reject it. This suggests that public campaigns and consultations are needed for such policies to be widely accepted among people in Hanoi.

Through a 4-step data-driven framework using explanatory data analysis and classification models, we have obtained important policy implications regarding a potential motorbike ban in Hanoi, Vietnam. Future studies may include more policy scenarios to explore what policymakers can do to reduce the motorbike dependency for people in Hanoi.

Acknowledgements

This work has received funding from the British Academy under the Urban Infrastructures of Well-Being programme [grant number UWB190190].

Bibliography

Fareed Ahmed and Zafar Fatmi. 882 impact of three wheeler ban policy on road traffic injuries in karachi, pakistan. *Injury Prevention*, 22(Suppl 2):A314–A315, 2016. ISSN 1353-8047. https://doi.org/10.1136/injuryprev-2016-042156.882. URL https://injuryprevention.bmj.com/content/22/Suppl_2/A314.3.

David Bray and Nicholas Holyoak. Motorcycles in Developing Asian Cities: A Case Study of Hanoi. October 2015.

Leo Breiman. Some properties of splitting criteria. *Machine Learning*, 24(1):41–47, 1996.

Francisco Calvo-Poyo, Adriana Medialdea, and Ramón Ferri-García. Citizens' opinion about investment in public transport projects in cities. *International Journal of Sustainable Transportation*, 14(10):806–818, August 2020. ISSN 1556-8318. https://doi.org/10.1080/15568318.2019.1630529. URL https://doi.org/10.1080/15568318.2019.1630529.

Laure Charleux. Deriving Mobility Archetypes from Household Travel Survey Data. *The Professional Geographer*, 70(2):186–197, April 2018. ISSN 0033-0124. https://doi.org/10.1080/00330124.2017.1338588. URL https://doi.org/10.1080/00330124.2017.1338588.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

Chun-Chen Chou, Kento Yoh, Hiroto Inoi, Tadanori Yamaguchi, and Kenji Doi. Effectiveness evaluation on cross-sector collaborative education programs for traffic safety toward sustainable motorcycle culture in vietnam. *IATSS research*, 2022.

Shreya Das and Debapratim Pandit. Importance of user perception in evaluating level of service for bus transit for a developing country like India: a review. *Transport Reviews*, 33(4):402–420, July 2013. ISSN 0144-1647. https://doi.org/10.1080/01441647.2013.789571. URL https://doi.org/10.1080/01441647.2013.789571. Publisher: Routledge _eprint: https://doi.org/10.1080/01441647.2013.789571.

Xinyue Dong and Xi Liu. Effect Analysis of banning of motorbikes and limiting of electric scooters. *World Construction*, 6, May 2017. https://doi.org/10.18686/wc.v6i2.98.

Elizabeth A Freeman, Gretchen G Moisen, John W Coulston, and Barry T Wilson. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, 46(3):323–339, 2016.

Kriti Gupta. Motorcycles Are Banned In This Myanmar City And Nobody Knows Why!, February 2019. URL https://www.indiatimes.com/news/world/motorcycles-are-banned-in-this-myanmar-city-and-nobody-knows-why-361627.html.

Hanoi People's Council. 5953/qd-ubnd on approving the scheme "strengthening the management of road transport means to reduce traffic congestion and environmental pollution in hanoi city, the period of 2017–2020 vision 2030", 2017. URL https://sogtvt.hanoi.gov.vn/documents/1468930/2148465/120171129153810_qdub-5953-2017-01_signed.pdf/2133f507-944f-430e-8fc5-1b6aa801172b.

Arve Hansen. Hanoi on wheels: emerging automobility in the land of the motorbike. *Mobilities*, pages 1–18, April 2016. https://doi.org/10.1080/17450101.2016.1156425.

Arve Hansen. Transport in transition: Doi moi and the consumption of cars and motorbikes in Hanoi. *Journal of Consumer Culture*, 17(2):378–396, July 2017. ISSN 1469-5405. https://doi.org/10.1177/1469540515602301. URL https://doi.org/10.1177/1469540515602301.

Arve Hansen. *Driving Doi Moi*, pages 486–501. Routledge, London, first edition, 2022. ISBN 978-1-315-76230-2. https://doi.org/10.4324/9781315762302-37.

Prathurng Hongsranagon, Theerachai Khompratya, Surbpong Hongpukdee, Piyalamporn Havanond, and Nathawan Deelertyuenyong. Traffic risk behavior and perceptions of Thai motorcyclists: A case study. *IATSS Research*, 35(1):30–33, July 2011. ISSN 0386-1112. https://doi.org/10.1016/j.iatssr.2011.03.001. URL https://www.sciencedirect.com/science/article/pii/S0386111211000124.

Duc Nguyen Huu, Van Nguyen Ngoc, et al. Analysis study of current transportation status in vietnam's urban traffic and the transition to electric two-wheelers mobility. *Sustainability*, 13(10):5577, 2021.

Tri Basuki Joewono and Hisashi Kubota. User perceptions of private paratransit operation in indonesia. *Journal of Public Transportation*, 10(4):5, 2007.

Hun Kee Kim. Ho chi minh city's urban transport challenges. *Researchers at ISEAS - Yusof Ishak Institute Analyse Current Events*, 65, 2017.

Yuto Kitamura, Makiko Hayashi, and Eriko Yagi. Traffic problems in southeast asia featuring the case of cambodia's traffic accidents involving motorcycles. *IATSS research*, 42(4):163–170, 2018.

Max Kuhn. Futility analysis in the cross-validation of machine learning models. *arXiv preprint arXiv:1405.6974*, 2014.

Pranandang Adi Laksana. A research on the factors that make indonesian government's slow response in regulatory arrangements of motorbike taxi online transportation. Master's thesis, Seoul National University, 2019.

Guanie Lim. Public policy with Vietnamese characteristics: the case of the motorcycle industry. *Journal of Asian Public Policy*, 11(2):226–244, May 2018. ISSN 1751-6234. https://doi.org/10.1080/17516234.2017.1338326. URL https://doi.org/10.1080/17516234.2017.1338326.

Pengfei Lin, Jiancheng Weng, Dimitrios Alivanistos, Siyong Ma, and Baocai Yin. Identifying and Segmenting Commuting Behavior Patterns Based on Smart Card Data and Travel Survey Data. *Sustainability*, 12(12):1–18, 2020. URL https://ideas.repec.org/a/gam/jsusta/v12y2020i12p5010-d373532.html.

Bich-Thuy Ly, Yoshizumi Kajii, Thi-Yen-Lien Nguyen, Koki Shoji, Dieu-Anh Van, Thi-Nhu-Ngoc Do, Trung-Dung Nghiem, and Yosuke Sakamoto. Characteristics of roadside volatile organic compounds in an urban area dominated by gasoline vehicles, a case study in Hanoi. *Chemosphere*, 254:126749, September 2020. ISSN 0045-6535. https://doi.org/10.1016/j.chemosphere.2020.126749. URL https://www.sciencedirect.com/science/article/pii/S0045653520309425.

Harry Manley, Nuttha Paisarnsrisomsuk, Andrew Hill, and Mark S. Horswill. The development and validation of a hazard perception test for Thai drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 71:229–237, May 2020. ISSN 1369-8478. https://doi.org/10.1016/j.trf.2020.04.011. URL https://www.sciencedirect.com/science/article/pii/S1369847820304071.

Christoph Molnar. *Interpretable machine learning*. Lulu. com, 2020.

National Traffic Safety Committee. Implementing road safety strategies and action plans in vietnam, 2021. URL https://eurochamvn.glueup.com/resources/protected/organization/726/event/34373/528d7f8b-2e26-4e5a-a1d6-a4f34656761d.pdf.

A.M. Ngoc, K.V. Hung, and V.A. Tuan. Towards the Development of Quality Standards for Public Transport Service in Developing Countries: Analysis of Public Transport Users' Behavior. *Transportation Research Procedia*, 25: 4560–4579, 2017. ISSN 23521465. https://doi.org/10.1016/j.trpro.2017.05.354.

Ngoc T. Nguyen, Tomio Miwa, and Takayuki Morikawa. Switching to Public Transport Modes for Commuting Trips: Considering Latent Motivations in Ho Chi Minh City. *Asian Transport Studies*, 5(1):117–136, 2018. https://doi.org/10.11175/eastsats.5.117.

Trond Nordfjærn and Torbjørn Rundmo. Perceptions of traffic risk in an industrialised and a developing country. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(1):91–98, January 2009. ISSN 1369-8478. https://doi.org/10.1016/j.trf.2008.08.003. URL https://www.sciencedirect.com/science/article/pii/S1369847808000740.

Thanh Xuan Thi Pham, Nhat Tien Nguyen, and Long Bien Thi Duong. Hierarchy-attribute decision making regarding public buses and private motorbikes: a case study in Ho Chi Minh City, Vietnam. *Public Transport*, 13(1):233–249, March 2021. ISSN 1613-7159. https://doi.org/10.1007/s12469-020-00256-8. URL https://doi.org/10.1007/s12469-020-00256-8.

John Pucher, Zhong-ren Peng, Neha Mittal, Yi Zhu, and Nisha Korattyswaroopam. Urban Transport Trends and Policies in China and India: Impacts of Rapid Economic Growth. *Transport Reviews*, 27(4):379–410, July 2007. ISSN 0144-1647. https://doi.org/10.1080/01441640601089988. URL https://doi.org/10.1080/01441640601089988.

Rusdi Rusli, Oscar Oviedo-Trespalacios, and Suhaila Azura Abd Salam. Risky riding behaviours among motorcyclists in Malaysia: A roadside survey. *Transportation Research Part F: Traffic Psychology and Behaviour*, 74:446–457, October 2020. ISSN 1369-8478. https://doi.org/10.1016/j.trf.2020.08.031. URL https://www.sciencedirect.com/science/article/pii/S1369847820305192.

Washington José dos Santos, Vanessa Maria da Silva Cêlho, Gustavo Barreto Santos, and Albanita Gomes da Costa de Ceballos. Work overload and risk behaviors in motorcyclists. *Revista Brasileira de Enfermagem*, 72: 1479–1484, 2019.

Michael C Shewry and Henry P Wynn. Maximum entropy sampling. *Journal of applied statistics*, 14(2):165–170, 1987.

Phathai Singkham. *Separate lane for motorcycle to reduce severity of road traffic injury among motorcyclist in Thailand*. PhD thesis, KIT - Royal Tropical Institute , VU - Vrije Universiteit Amsterdam, 2016. URL https://bibalex.org/baifa/en/resources/document/456464.

Pham Chau Thuy, Takayuki Kameda, Akira Toriba, Ning Tang, and Kazuichi Hayakawa. Characteristics of Atmospheric Polycyclic Aromatic Hydrocarbons and Nitropolycyclic Aromatic Hydrocarbons in Hanoi-Vietnam, as a Typical Motorbike City. *Polycyclic Aromatic Compounds*, 32(2):296–312, March 2012. ISSN 1040-6638, 1563-5333. https://doi.org/10.1080/10406638.2012.679015.

Long T. Truong, Le-Minh Kieu, and Tuan A. Vu. Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. *Accident Analysis & Prevention*, 94:153–161, September 2016. ISSN 0001-4575. https://doi.org/10.1016/j.aap.2016.05.028. URL `https://www.sciencedirect.com/science/article/pii/S0001457516301944`.

Vu Anh Tuan. Mode Choice Behavior and Modal Shift to Public Transport in Developing Countries - the Case of Hanoi City. *Journal of the Eastern Asia Society for Transportation Studies*, 11:473–487, 2015. https://doi.org/10.11175/easts.11.473.

Sarah Turner. Informal motorbike taxi drivers and mobility injustice on hanoi's streets. negotiating the curve of a new narrative. *Journal of transport geography*, 85:102728, 2020.

Tan Hong Van, Jan-Dirk Schmoecker, and Satoshi Fujii. Upgrading from motorbikes to cars: Simulation of current and future traffic conditions in Ho Chi Minh City. *Proceedings of the Eastern Asia Society for Transportation Studies*, 2009:335–335, 2009. https://doi.org/10.11175/eastpro.2009.0.335.0.

Günter Wallner, Simone Kriglstein, Edward Chung, and Syeed Anta Kashfi. Visualisation of trip chaining behaviour and mode choice using household travel survey data. *Public Transport*, 10(3):427–453, December 2018. ISSN 1613-7159. https://doi.org/10.1007/s12469-018-0183-5. URL `https://doi.org/10.1007/s12469-018-0183-5`.

Xiaoyu Yan and Roy J Crookes. Energy demand and emissions from road transportation vehicles in china. *Progress in Energy and Combustion Science*, 36(6):651–676, 2010.

Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3):647–665, December 2014. ISSN 0219-3116. https://doi.org/10.1007/s10115-013-0679-x. URL `https://doi.org/10.1007/s10115-013-0679-x`.