Using Machine Learning to Predict Perceptions of a Motorbike Ban in Hanoi

Minh Kieu[1], Alexis Comber[2], Eric Wanjau[3], Kristina Bratkova[3], Hang Nguyen Thi Thuy[4], Thanh Bui Quang[5], Phe Hoang Huu[6] and Nick Malleson*[2,3]

[1]Department of Civil and Environmental Engineering, University of Auckland, New Zealand
[2]School of Geography, University of Leeds, UK
[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

Summary

The dependence on motorbikes has contributed to severe traffic problems in Hanoi, Vietnam. Policymakers have considered a controversial ban on non-electric 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. This paper applies a machine learning algorithm (XGBoost) to a bespoke travel survey to better understand how residents perceive a potential motorbike ban and how their perceptions might change under different policy scenarios. Our results suggest that prior awareness of the ban and shorter distances to public transport both increase peoples' favour.

KEYWORDS: Machine Learning; Transport; Motorbike; Hanoi.

1 Introduction

Hanoi, the capital of Vietnam, is growing rapidly. However, because motorbikes have replaced bicycles as the main means of transport in recent decades (Thuy et al., 2012; Hansen, 2017; Ly et al., 2020; Hansen, 2022), traffic congestion and pollution have also increased dramatically (Pham et al., 2021; Van et al., 2009; Lim, 2018; Ly et al., 2020; Huu et al., 2021). A city that used to be famous for "tranquil" streets full of bicycles is now characterised by "almost constant buzzing and honking" (Hansen, 2022) and was the 2nd most polluted city in South East Asia in 2018 (Huu et al., 2021).

One potential pollution/congestion reduction strategy could be to ban motorbikes from parts of the city. In fact, some cities with similar demographic characteristics have already (or plan to)

*Corresponding author: n.s.malleson@leeds.ac.uk

banned vehicles like motorbikes; including Guangzhou and Shenzen in China (Pucher et al., 2007; Dong and Liu, 2017) and Karachi in Pakistan (Ahmed and Fatmi, 2016). 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, by providing a flexible, independent and, relatively effortless form of mobility, motorbikes have become "absolutely vital" (Hansen, 2022) to most people in Vietnam (Thuy et al., 2012; Hansen, 2016, 2017). Any motorbike-related policy must be designed and implemented carefully to avoid "mobility injustice" (Turner, 2020), where travel and employment opportunities are reduced for some people – as was the case in Yangon, Myanmar, when motorbikes were banned (Gupta, 2019) – or inadvertently increasing the use of cars (Van et al., 2009).

This paper presents the analysis of a large, bespoke household travel survey that was conducted by the authors in Hanoi to better understand citizens' travel behaviour and preferences. In particular, the survey asked people whether they were aware of the Hanoi People's Council Decision No. 5953/QDUBND that outlines a plan to ban motorbikes in specific areas of Hanoi (Huu et al., 2021). The 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). We use a gradient-boosted decision tree model (XGBoost) to both find the characteristics of the respondents that are the most strongly associated with positive or negative opinions of the ban (*variable importance*; not discussed here) and to predict how sentiment might change were (i) people better informed about the ban and (ii) public transport better available. We find that both factors suggest more a positive view on the policy.

2 Methods and Data

The authors conducted a bespoke survey of $N = 26,339$ households in Hanoi. Although the survey is one of the largest of its kind, especially for Hanoi, it still covers only a relatively small part of the wider city, as per Figure 1. 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. The survey collected information about general demographics, travel patterns, behaviours, and especially peoples' perceptions of transportation policies, vehicle ownership and transportation modes.

We use an Extreme Gradient Boosting (XGBoost) decision tree classification model (see Chen and Guestrin, 2016) to make a binary prediction 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 questions in the survey. The main idea behind *boosting* is to build a strong model from an ensemble of weak models (individual decision trees in this case) in series, with each model fitted to the residuals of the previous models.

We train the model on 70% of the data, holding the rest back for testing. Once the model has been trained, we use it to predict whether a specific hypothetical person will favour the ban or not. 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 the Shapley value (Štrumbelj and Kononenko, 2014), using a Monte-Carlo approxi-

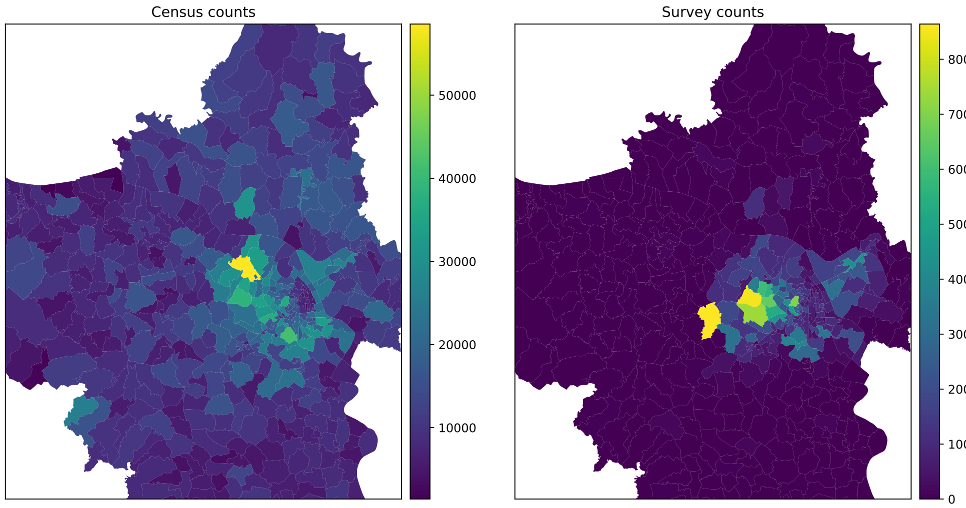


Figure 1: Map of surveyed households, versus the distribution of households in the most recent (2019) Vietnamese census.

mation as described in Štrumbelj and Kononenko (2014) because there are such a large number of variables:

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

where $\hat{f}\left(x_{+j}^m\right)$ 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 $\hat{f}\left(x_{-j}^m\right)$ is also from the same random sample. We calculate the Shapley value for each variable.

3 Results: Exploring the factors that influence opinion about the motorbike ban

We use the classification model to investigate what policymakers might do to make people more or less favourable towards the motorbike ban. We do this by estimating how the model output—the individual probability of acceptance of the motorbike ban—might vary as we adjust the input variables. Because the model is nonparametric, we cannot simply analyse its parameters to understand the impacts of a variable change on its output, so instead we adopt a Monte Carlo simulation approach. We systematically modify the testing dataset to simulate a change in one particular variable, while keeping all other variables consistent, and then use the trained XGBoost model to make predictions. By comparing the model output (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 to the ban.

3.1 Awareness of the motorbike ban

In our variable importance analysis, whether people were already aware of the ban before completing the survey had a strong impact on whether they were opposed or supportive of the ban. We introduce a modifier to the testing data set, M, to change the status of a percentage of the people who were "unaware" of the ban to "aware". For example, a policy scenario with $M = 0.5$ would randomly choose 50% of those who are unaware of the ban and change their awareness to "aware". We then perform a series of experiments, stepping M from 10% to 100%. Note that in addition to those who were "unaware" of the ban, there is a further set of people who "do not care" about the ban; individuals from both of these groups are randomly changed to "aware" according to M.

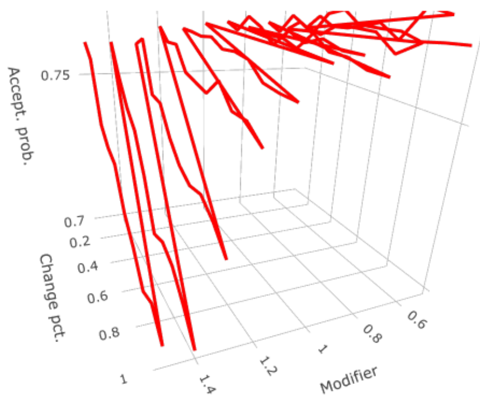
Figure 2 shows the median value of the individual probabilities of acceptance across all respondents. The results suggests that if more people had become aware of the ban before conducting the survey, we would have seen improvements in the probability of acceptance. This points to a potential useful policy recommendation: a government awareness campaign about the ban could greatly help with public acceptance. Whilst this seems obvious at first, we have no way of knowing *how* respondents became aware of the ban. There is no reason to think that they will have become aware of the ban through positive rather than negative messaging, so it is illuminating that they are more favourable and suggests that a *positive* government message may be even more successful than initially conceived.

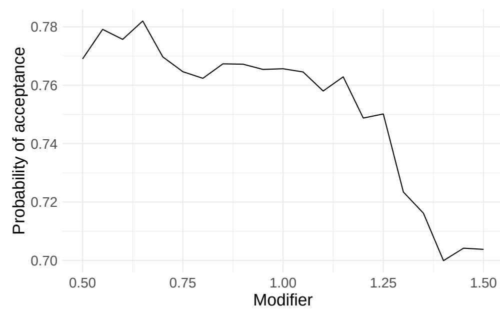
Figure 2: Median probability of acceptance

Figure 3: Scenario 1: How individual probability of acceptance changes as people become more aware of the motorbike ban

3.2 Distance to public transport

Another important variable is *distance to public transport*. In this scenario we repeat the classification prediction process using a modifier, M, that multiplies the current distance for each individual sequentially from $M = 0.5$ to $M = 1.5$. In practice, however, the distance cannot be increased or decreased for everyone, as any new or reduced transit services will only have a direct impact on those who are close to those services. Thus we also introduce a *change percentage* variable, C, to randomly chose an individual with a probability of 10% to 100% and apply M to that individual. Figure 4a shows the changes in the median probability of acceptance to the motorbike ban as M and C vary. To better see the pattern, Figure 4b illsutrates the results where C is fixed at 100%, or every one receives the same change in the distance to their nearest public transport stops.





(b) Without *change percentage* ($C = 100\%$

(a) With *change percentage* (C)

Figure 4: Scenario 2: Distance to public transport

Our results suggest that if public transport is a suitable mobility option, people in Hanoi will be more open to a potential motorbike ban. Figure 4 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 new public transport infrastructure is not created. Policymakers should at least maintain current levels of access to public transport to reduce dependency on motorbikes and other private vehicles.

4 Conclusions

This paper has used the XGBoost decision tree classification model to experiment with peoples' changing attitudes to a potential motorbike ban in the city of Hanoi, Vietnam. Our results suggest that policy makers might be able to increase the popularity of a ban by properly informing the public about it and by decreasing the distance that people have to travel to access public transport.

Acknowledgements

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

References

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

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Dong, X. & Liu, X. (2017). Effect Analysis of banning of motorbikes and limiting of electric scooters. *World Construction*, *6*, https://doi.org/10.18686/wc.v6i2.98.

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

Hanoi People's Council (2017). 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" https://sogtvt.hanoi.gov.vn/documents/1468930/2148465/120171129153810_qdub-5953-2017-01_signed.pdf/2133f507-944f-430e-8fc5-1b6aa801172b.

Hansen, A. (2016). Hanoi on wheels: emerging automobility in the land of the motorbike. *Mobilities*, (pp. 1–18)., https://doi.org/10.1080/17450101.2016.1156425.

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

Hansen, A. (2022). *Driving Doi Moi*, (pp. 486–501). Routledge: London, first edition.

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

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

Ly, B.-T., et al. (2020). Characteristics of roadside volatile organic compounds in an urban area dominated by gasoline vehicles, a case study in Hanoi. *Chemosphere*, *254*, 126749, https://doi.org/10.1016/j.chemosphere.2020.126749 https://www.sciencedirect.com/science/article/pii/S0045653520309425.

Pham, T. X. T., Nguyen, N. T., & Duong, L. B. T. (2021). 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, https://doi.org/10.1007/s12469-020-00256-8 https://doi.org/10.1007/s12469-020-00256-8.

Pucher, J., Peng, Z.-r., Mittal, N., Zhu, Y., & Korattyswaroopam, N. (2007). Urban Transport Trends and Policies in China and India: Impacts of Rapid Economic Growth. *Transport Reviews*, *27*(4), 379–410, https://doi.org/10.1080/01441640601089988 https://doi.org/10.1080/01441640601089988.

Singkham, P. (2016). *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 https://bibalex.org/baifa/en/resources/document/456464.

Thuy, P. C., Kameda, T., Toriba, A., Tang, N., & Hayakawa, K. (2012). 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, https://doi.org/10.1080/10406638.2012.679015.

Tuan, V. A. (2015). 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, https://doi.org/10.11175/easts.11.473.

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

Van, T. H., Schmoecker, J.-D., & Fujii, S. (2009). 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, https://doi.org/10.11175/eastpro.2009.0.335.0.

Štrumbelj, E. & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665, https://doi.org/10.1007/s10115-013-0679-x https://doi.org/10.1007/s10115-013-0679-x.