Exploring the Factors affecting Teleworking Decision Making

Rakibul Hassan*[1], Hamed Naseri[2], Sara Gharavi[3], Jean Dubé[4], Francesco Ciari[5]

[1] PhD candidate, Polytechnique Montréal, Montréal, Canada [ORCiD : 0000-0001-6183-0575]
[2] Postdoctoral fellow, Polytechnique Montréal, Montréal, Canada [ORCiD : 0000-0002-4151-330X]
[3] PhD candidate, Université Laval, Québec City, Canada [ORCiD : 0000-0002-4666-9450]
[4] Professor, Université Laval, Québec City, Canada [ORCiD : 0000-0003-3445-3449]
[5] Associate Professor, Polytechnique Montréal, Montréal, Canada [ORCiD : 0000-0002-6192-7948]

ABSTRACT

Teleworking has become a new standard after the pandemic since its continued higher adoption than the pre-pandemic period. Therefore, it is crucial for both employers and policymakers to investigate the factors that influence workers' preferences for teleworking. While earlier models are largely focused on socio-demographic attributes, the current study extends its scope by adding climate and built environment variables. This study examines the influence of 17 alternative variables on telecommuting propensity precisely in Quebec City, Canada. Since only 7% of the people engage in telecommuting, the dataset is imbalanced, a fact that is counter-balanced by using an under-sampling technique. EXtreme Gradient Boosting (XGB), SHapley Additive exPlanations (SHAP), and Partial Dependency Plots (PDP) have been used for modeling. Results suggest that the commuting distance is the most influential factor in the teleworking decision-making process. Additional variables added from the supplementary dataset, i.e., average daily temperature and walk score, show a moderate correlation.

Keywords: Teleworking, Weather, EXtreme Gradient Boosting, SHapley Additive exPlanations

1. INTRODUCTION

The concept of teleworking was raised in the 1960s by Owen (1962). Since then, it has constantly changed with emerging personal computers, internet access, and finally advanced communication facilities. During the COVID-19 pandemic, the adoption of telework or remote working accelerated due to the need for keeping physical distancing to try to avoid or reduce infection. Wray (2024) observes that the share of Canadians working most of their hours from home increased from 7% in May 2016 to 24% in July 2022 and has declined slightly since then to 21% by July 2023.

Researchers have developed various models to analyze teleworking desirability, the decision-making process, and the factors influencing individuals' teleworking adoption decisions. To this end, various techniques have been applied, such as logistic regression (Pabilonia & Victoria Vernon, 2022), ordinal logistic regression (Stefaniec et al., 2022), random parameter ordered logit model (Rahman Fatmi et al., 2022), latent class analysis (Asmussen, Mondal, & Bhat, 2024). Although those techniques are easy to interpret, their prediction accuracy is generally less than powerful ensemble learning techniques, such as eXtreme Gradient Boosting (XGB) (Naseri, Waygood, et al., 2024b).

However, applying such ensemble methods to analyze teleworking preferences has not received enough attention. Further, due to the scarcity of data, most predictive models generally relied on employee characteristics. These changes are not uniform across populations, as teleworking feasibility depends on factors like profession, income, and household characteristics (Asmussen, Mondal, Batur, et al., 2024). Rationally, an individual's ability to telework is influenced by a broader range of factors, including socio-demographics, weather conditions, mobility options, built environment attributes, and job-specific characteristics. Therefore, incorporating these diverse factors into model development and analysis is critical for a more accurate and holistic understanding of teleworking behavior. Considering these factors, this study aims to

- Develop a model to accurately predict workers' preference toward teleworking
- Create a framework to tackle the issue associated with imbalance datasets
- Identify and analyze the influence of variables on the workers' teleworking probability

2. METHODOLOGY AND DATA

2.1. METHODOLOGY FLOW

The methodology flow consists of steps: (i) Merging various datasets (Weather & Walk score data to Origin destination (OD) data), (ii) Cleaning and filtering dataset, (iii) Tuning hyperparameters (ensemble learning), (iv) Tackling data imbalance (under-sampling), (v) Running XGB & (vi) Perform SHAP and PDP.

2.2. DATA & CLEANING

The study relies on a large-scale origin-destination (OD) survey collected in 2017 that recorded over 222,000 trips from 81,000 residents of Quebec City, Quebec, Canada. The OD survey has been enriched by calculating the distance between home and workplace using Google Maps API, integrating walkability data from the Walk Score dataset (Walkscore, 2021), and incorporating weather conditions from a climate dataset (Government of Canada, 2024).

The combined dataset was thoroughly cleaned and processed to prepare it for model development. During the filtering process, only individuals identified as full-time or part-time workers were retained. Further filtering was applied to include workers aged between 18 and 80. Additionally, trips lacking information about work locations or having unreasonable location data were excluded. For the teleworking variable, four response categories were present. Rows with responses indicating "N/A" or "Refused to answer" were removed, as these did not provide clear information about teleworking decisions. Only rows with definitive responses of "Yes" or "No" for teleworking were retained. Duplicate entries were also removed to ensure data integrity. After filtering and cleaning, the final dataset comprised 389 teleworkers out of 5,563 workers, representing 6.99% of the total dataset.

2.3. MODELING PROCESS AND VARIABLES TESTED

The dependent variable was the individual decision (teleworking or not). Independent variables were selected through a filter-based feature selection approach, eliminating those with Variance Inflation Factors (VIF) exceeding 5 (Akinwande et al., 2015), to address multicollinearity. The final dataset included 17 independent variables, including age, household income, occupation status (i.e., full-time or part-time), driving permit, gender, distance from home to work location, car per adult ratio in the household, average daily temperature, Walk Score of residential location, presence of a disabled person in the household, car-sharing subscription, carpool user (e.g., Uber), monthly transit pass, going out in the day, residential location region, region of occupation location, type of place of occupation.

XGB was used for modeling since it outperforms other ensemble learning techniques when comparing prediction accuracy (Naseri, Aliakbari, et al., 2024). OPTUNA and five-fold cross-validation were used to tune the hyperparameters of XGB. For modeling, 80% of the data was randomly considered training data and 20% was considered testing data. Since the data was imbalanced, the F1-score is considered the metric in OPTUNA (Sun et al., 2022). Moreover, Nearmiss-2 (an under-sampling method) was employed since it avoids information loss of the majority class (Wang et al., 2022). After under-sampling and tuning, XGB was performed and two interpretation techniques were applied to capture the influence of independent variables on teleworking preferences: SHAP and PDP (Naseri, Waygood, et al., 2024a).

3. RESULTS AND DISCUSSION:

3.1. PERFORMANCE OF THE MODEL

The tuned model reached the F1-score of 92.22% for five-fold validation data. Then, the performance of the model on testing data was evaluated, and the testing data accuracy and F1-score were 94.23% and 94.21%, respectively. The confusion matrix shown in Figure 1 demonstrates that the model

		Predicted	
		In-person	Telework
Actual	In-person	78	0
	Telework	9	69

Figure 1: Testing data confusion matrix

accurately predicted all decisions except for nine teleworking cases misclassified as in-person, demonstrating high teleworking prediction accuracy.

3.2. The Relative Influence of Variables on Teleworking Preference

The relative importance of different factors in determining the preference for teleworking was determined based on SHAP values. The distance from home to work location is the most important factor. Other influential factors include age, average daily temperature, and household income, which are influential and thus show that demographic characteristics, environmental conditions, and home environments play a critical role in shaping the preference for teleworking. Besides, other variables such as Walk Score of residential location, car per adult ratio in the household, and type of place of occupation reflect the importance of mobility, accessibility, and workplace characteristics. The less influential factors are the monthly transit pass, gender, and driving permit, while carpooling and car-sharing subscriptions have negligible impacts, suggesting their limited relevance in teleworking decisions.

3.3. Direction influence of variables on the teleworking probability

PDP was applied to demonstrate the non-linear relation of top variables on teleworking probability. The following paragraphs describe how these relationships behave.

Commuting Distance: Commuting distance affects the probability of teleworking the most supporting the hypothesis that longer commuting distances positively correlate with teleworking preference, emphasizing the role of distance in shaping work location decisions.

Age & Gender: Teleworking probability increases with age, reaching its highest point around 40–49 years, before gradually decreasing after the age of 50. This suggests that mid-career professionals may prefer teleworking due to job flexibility and responsibilities, while younger workers may have lower remote work opportunities, and older individuals may prefer in-person work due to the lack of interest and information about new technologies. Results also revealed that women (~49.02%) have a slightly higher chance of teleworking than men (~48.90%), though very marginal.

Average Daily Temperature: Our analysis found that people prefer to telework during cold weather to avoid commuting in poor conditions while mild or warming weather depressed the possibility of teleworking.

Household Income & Car Ownership: Our analysis found that household income and teleworking probability are positively correlated, meaning that individuals with higher household incomes are more likely to telework. Results also reveal that households with fewer cars per person (close to zero) have a higher probability of teleworking (around 56%). However, as the car-per-person ratio increases beyond 0.2, the probability of teleworking declines significantly, reaching its lowest point around 0.6

Walk Score & Employment Status & Disability: Individuals living in highly walkable neighborhoods (walk score >80) are more likely to telework due to better access to local amenities, reducing dependence on commuting. In contrast, those in less walkable areas may need to commute for work and other activities, leading to lower teleworking probability. Our analysis also suggests that part-time workers have a higher likelihood of teleworking (49.5%) than full-time workers (47%). Furthermore, it was also found that the workers living in a household with a disabled person have a significantly higher probability of teleworking (~49%) compared to those without such responsibilities (~44%).

4. LIMITATION AND FUTURE RESEARCH

This study develops a model to predict workers' preferences for teleworking by incorporating environmental factors, walk scores, and socio-demographics in Quebec City, Canada. The results reveal that environmental and walk score variables significantly influence teleworking decisions. However, this study does not include detailed job characteristics due to data limitations. Future research will aim to enhance teleworking participation estimates by incorporating job-specific characteristics.

ACKNOWLEDGEMENTS

This research has been funded by Ministère des Finances du Québec (MFQ).

REFERENCES

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics*, *05*(07), 754–767. https://doi.org/10.4236/ojs.2015.57075

Asmussen, K. E., Mondal, A., Batur, I., Dirks, A., Pendyala, R. M., & Bhat, C. R. (2024). An investigation of individual-level telework arrangements in the COVID-era. *Transportation Research Part A: Policy and Practice*, *179*(November 2022), 103888. https://doi.org/10.1016/j.tra.2023.103888

Asmussen, K. E., Mondal, A., & Bhat, C. R. (2024). The interplay between teleworking choice and commute distance. *Transportation Research Part C: Emerging Technologies*, *165*(June), 104690. https://doi.org/10.1016/j.trc.2024.104690

Government of Canada. (2024). *Historical Climate Data, Evironment and Climate Change Canada, available at: https://climate.weather.gc.ca/index_e.html*.

Naseri, H., Aliakbari, A., Javadian, M. A., Aliakbari, A., & Waygood, E. O. D. (2024). A novel technique for multi-objective sustainable decisions for pavement maintenance and rehabilitation. *Case Studies in Construction Materials*, *20*, e03037. https://doi.org/10.1016/J.CSCM.2024.E03037

Naseri, H., Waygood, E. O. D., Patterson, Z., & Wang, B. (2024a). Which variables influence electric vehicle adoption? *Transportation*, 1–38. https://doi.org/10.1007/s11116-024-10525-1

Naseri, H., Waygood, E. O. D., Patterson, Z., & Wang, B. (2024b). Who Is More Likely to Buy Electric Vehicles? *Transport Policy*, *155*, 15–28. https://doi.org/10.1016/J.TRANPOL.2024.06.013

Owen, W. (1962). Transportation and Technology. *The American Economic Review*, *52*(2), 405–413. http://www.jstor.org/stable/1910908

Pabilonia, S. W., & Victoria Vernon. (2022). Telework, Wages, and Time Use in the United States. *Review of Economics of the Household*, *20*(3), 687–734. https://doi.org/10.1007/s11150-022-09601-1

Rahman Fatmi, M., Mehadil Orvin, M., & Elizabeth Thirkell, C. (2022). The future of telecommuting post COVID-19 pandemic. *Transportation Research Interdisciplinary Perspectives*, *16*(August), 100685. https://doi.org/10.1016/j.trip.2022.100685

Stefaniec, A., Brazil, W., Whitney, W., & Caulfield, B. (2022). Desire to work from home: Results of an Irish study. *Journal of Transport Geography*, *104*(February), 103416. https://doi.org/10.1016/j.jtrangeo.2022.103416

Sun, Y., Dong, Y., D. Waygood, E. O., Naseri, H., Jiang, Y., & Chen, Y. (2022). Machine-Learning Approaches to Identify Travel Modes Using Smartphone-Assisted Survey and Map Application Programming Interface. *Transportation Research Record: Journal of the Transportation Research Board*, 036119812211064. https://doi.org/10.1177/03611981221106483

Walkscore. (2021). *Walk Score [dataset]*.

Wang, M., Song, L., Zhang, Y., Gao, H., Yan, L., & Yu, B. (2022). Malsite-Deep: Prediction of protein malonylation sites through deep learning and multi-information fusion based on NearMiss-2 strategy. *Knowledge-Based Systems*, *240*, 108191. https://doi.org/10.1016/j.knosys.2022.108191

Wray, D. (2024). Telework, time use, and well-being: Evidence from the 2022 Time Use Survey. *Statistic Canada*.