

The Determinants of Hukou Acquisition Intentions among Migrants in Tianjin: A Study Using XGBoost and SHAP

Hongyuan Liu

Tianjin Normal University, Tianjin, China

Email: 1350253500@qq.com

Manuscript received January 6, 2025; revised February 5, 2025; accepted February 11, 2025; published February 21, 2025.

Abstract—In this paper, the author investigates the key factors influencing migrants' intentions to get hukou in Tianjin according to the 2017 China Migrants Dynamic Survey (CMDS), which includes 5,000 questionnaires, the study uses XGBoost to find significant determinants. Uses SHAP (SHapley Additive exPlanations) to interpret the relative importance. The study researches many factors, including characteristics of the migration destination, individual attributes, factors related to the hukou origin city, and stay wish. The machine learning compared to traditional logistic regression methods, XGBoost can capture more complex nonlinear relationships between different periods in one feature, and supply a more precise analysis of how different factors and how the feature change influence Hukou Acquisition Intentions across different levels. Additionally, unlike traditional studies on hukou acquisition intentions, machine learning gives more deeper insights into the different self characteristics of willing to obtain hukou in Tianjin.

Keywords—hukou acquisition, migrants, XGBoost, SHAP, machine learning

I. INTRODUCTION

Hukou system is an internal household registration policy in China. It started in the 1950s. It has shaped mobility between provinces and helped the urban citizen to get urban benefits [1]. The hukou policy was used to control population of immigrants to plan the social economic in China before the open door policy, the system has led to region and urban-country inequalities between residents living in different regions. Most of the time, the rural hukou holders help a lot in urban development by supplying food and labour to near cities, but they face unfair treatment such as less opportunity to public services in health care and education [2]. Because of that makes the gap between rural and urban areas. The changes have been made to loosen hukou rules in a few cities. So this paper will do the study in Tianjin City to find the Hukou Acquisition Intentions in the floating population. We found that in some aspects it is hard to fully eliminate the gap in rural migrants into city life when they work or study in Tianjin.

More and more researches start to focus on why migrants want or do not want to get urban hukou [3]. One would think migrants from rural areas prefer to be quick to get urban hukou because they think the urban hukou has more value in many aspects. But studies show the outcome is the opposite: many migrants do not rush to get hukou when they work or study in a new city, even they can get hukou in a short time without any barriers. This may be because they worry about finance reasons, the high cost of city life especially the price of daily cost and rent cost [4]. The hukou system, most of the time, puts rural migrants at a disadvantage place in many

ways, which affects their choice to stay in cities and their chances to move up in society.

The machine learning helps researchers understand why migrants may or may not want hukou in this dissertation. Other studies about migrants prefer to use methods like logistic regression [5], this way can only find the liner affect and hard to explain the complex reasons. The machine learning models like XGBoost and SHAP (SHapley Additive exPlanations) can give better answers when dealing with the same problem [6]. The XGBoost can help show the local effects of factors and their importance. The studies using XGBoost and SHAP on spatial and socio-economic data show we can use better ways about migration and policy problems [7].

This paper focuses on what factors affect migrants' decisions to get hukou in Tianjin (Fig. 1). It uses XGBoost and SHAP to study data from the 2017 China Migrants Dynamic Survey (CMDS), We use the 5000 questionnaires about the floating people. The study shows what drives hukou decisions and compares its outcome with other studies [8]. It will give suggestions to help hukou reform [9] and improve city policies to absorb the talents.

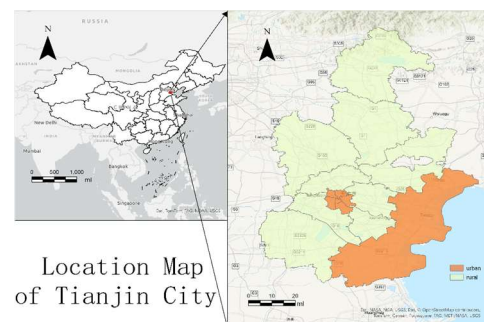


Fig. 1. Location map of Tianjin City.

II. METHODOLOGY

This section explains the method used to study the factors that affect Hukou Acquisition Intentions of migrants in Tianjin. The study uses machine learning to make predictions and SHAP to show how different factors influence the results in Fig. 2. The dataset includes many types of information, such as destination type (urban or non-urban), number of co-residents, gender, age, education level, household registration type (urban or rural), marital status, hukou origin type (agricultural or urban), length of stay in Tianjin, employment status, first-time migration alone, total number of cities migrated to, home ownership, Hukou Acquisition Intentions, and stay willing.

The data was prepared before analysis. Missing information was deleted. Hukou Acquisition Intentions, the

main result to predict, was turned into a binary value where 1 means the person intends to acquire hukou and 0 means they do not.

XGBoost [10] was chosen as the model because it is accurate, handles complex patterns well, and avoids overfitting. The data was split into two parts, 70% for training and 30% for testing. The model was improved by tuning key parameters like learning rate, maximum tree depth, number of trees, and sample ratio. To measure how well the model worked this study used ROC.

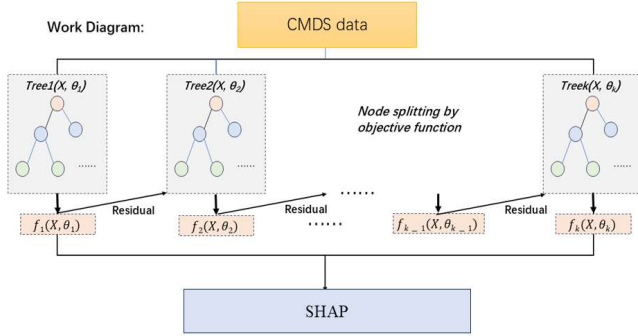


Fig. 2. XGBoost work diagram.

The input data was the feature set from traindata. It was changed into a matrix format by removing the last column, which had the target variable. This made sure XGBoost could use the input features correctly. The target variable traindata target was turned into numbers using `as.numeric`. It was a binary classification problem with labels 0 and 1. This matched the objective function “binary:logistic”. The parameter `max_depth` was set to 3. This limited how deep each tree could grow. It helped control the complexity of the model and reduced the chance of overfitting. The learning rate (`eta`) was set to 0.1. This controlled how big each update step was. A small learning rate made the model more stable but needed more rounds to train. The model used 2 threads (`nthread`) to make the computation faster by running processes at the same time. The total number of rounds (`nrounds`) was set to 100. This meant the model built 100 weak trees. The number of rounds was chosen to balance between underfitting and overfitting based on the other parameters. The objective function “binary” was used for binary classification. It gave a probability between 0 and 1.

To make the results easier to understand, SHAP [11] was used. SHAP helps explain the predictions by showing how much each factor contributed. The SHAP values were calculated for all features, and the features were ranked by their average importance. Several plots were made to show the results clearly. Dependence plots were used to show how specific factors affected Hukou Acquisition Intentions.

III. RESULTS

Fig. 3 shows the importance of different factors in Hukou Acquisition Intentions. Stay wish is the most important factor with a value of 0.421. Hukou status is next with a value of 0.187. Age has an importance of 0.163. Duration of stay follows with a value of 0.148. Own property is also important with a value of 0.132. Distance has a value of 0.123. Urban area is slightly less important with a value of 0.102.

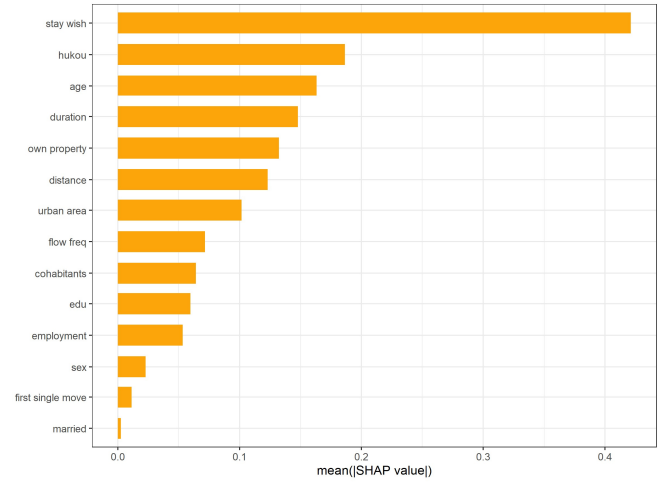


Fig. 3. Importance ranking.

Other factors include flow frequency with 0.072, cohobitants with 0.064, education level with 0.060, and employment status with 0.053. Factors with lower importance are sex with 0.023, first single move with 0.011, and marital status with 0.003. These factors show how each variable affects Hukou Acquisition Intentions differently.

IV. DISCUSSION

In this part, the author focuses on how 12 key features (Figs. 4–6: stay wish, hukou, age, duration, own property, distance, urban area, flow frequency, cohobitant, education, employment, and sex) affect the hukou acquisition intention in migrants.

Among all the features, stay wish had the strongest effect on hukou acquisition intention. This shows that when people plan to live in the city for a long time, they are more likely to want a hukou. Policies that help people feel like they belong and improve life in the city may increase their hukou acquisition intention.

The second important feature is hukou type, which also has a big impact on hukou acquisition intention [12]. Migrants with urban hukou are more likely to get local hukou than those with a rural hukou. This may be because having an urban hukou better to change Hukou to another city when they float. The results show that the population with rural hukou afraid to lose their homestead in hometown so they pretend to give up the chance to get hukou in Tianjin [13].

Age and hukou acquisition intention in an inverted U-shaped trend in Fig. 7. The highest levels are in the 20–24 years and 30–34 years age groups. People in 20–24 most of the time with high education experience and low pressure from family, the 30–34 group more like to support their children to have a better education and environment in a different city. The population is beyond 35 because “Haihe Talent Program” encourages younger people to stay in Tianjin, they have less chance to get hukou. In other age groups, fewer people show interest in getting hukou because they are less likely to move or already have stable lives elsewhere [14].

Stay duration in Tianjin has an important and clear effect on hukou acquisition intention. The longer migrants work or study in Tianjin City, the more chance they have to settle down by getting Tianjin hukou. Over time, they form stronger connections with the city and feel more comfortable

living in Tianjin, that with the time they stay in Tianjin grow will help them decide to increase hukou acquisition intention.

When migrants own property in Tianjin, their average SHAP values stay around 0.2. This shows a positive effect on their hukou acquisition intention [15]. On the other hand, when they do not have property in Tianjin, their SHAP values

stay around -0.1 show a negative effect. The difference means that buying property clearly increases the chance of hukou acquisition intention. Owning property shows that the person plans to stay in the Tianjin city longer. It also lowers concerns about future living situations. So, migrants who buy their own house are more likely to want local hukou.

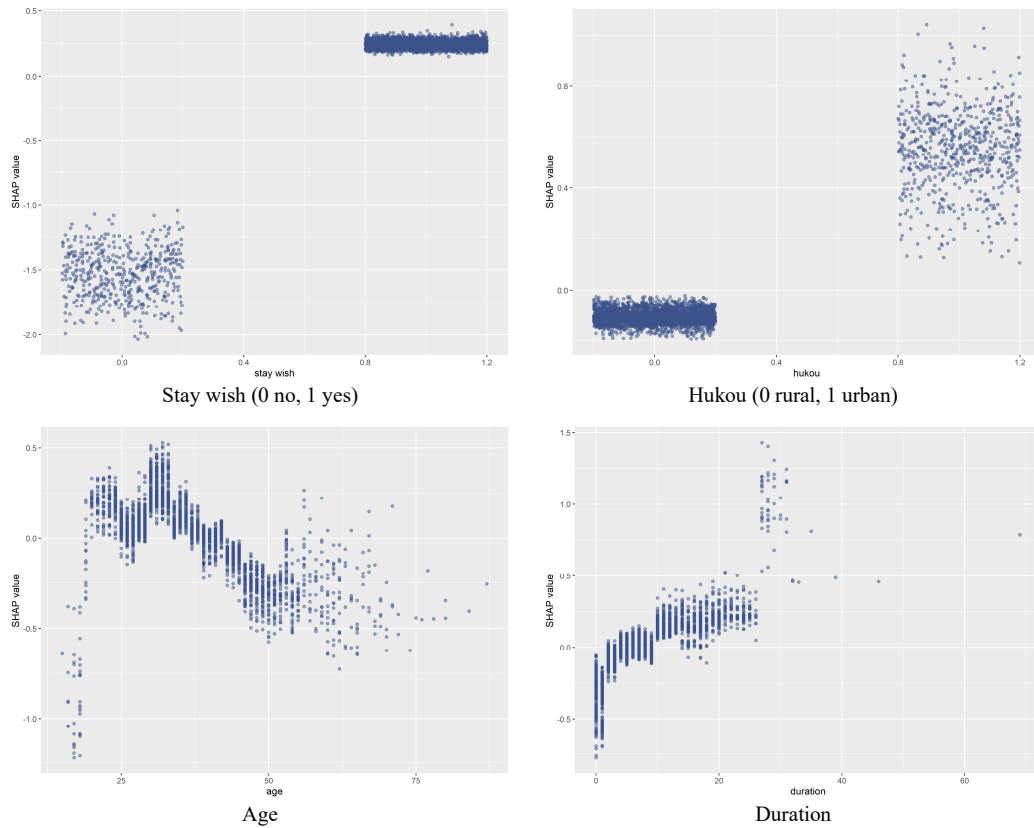


Fig. 4. SHAP value of stay wish, hukou, age, and duration.

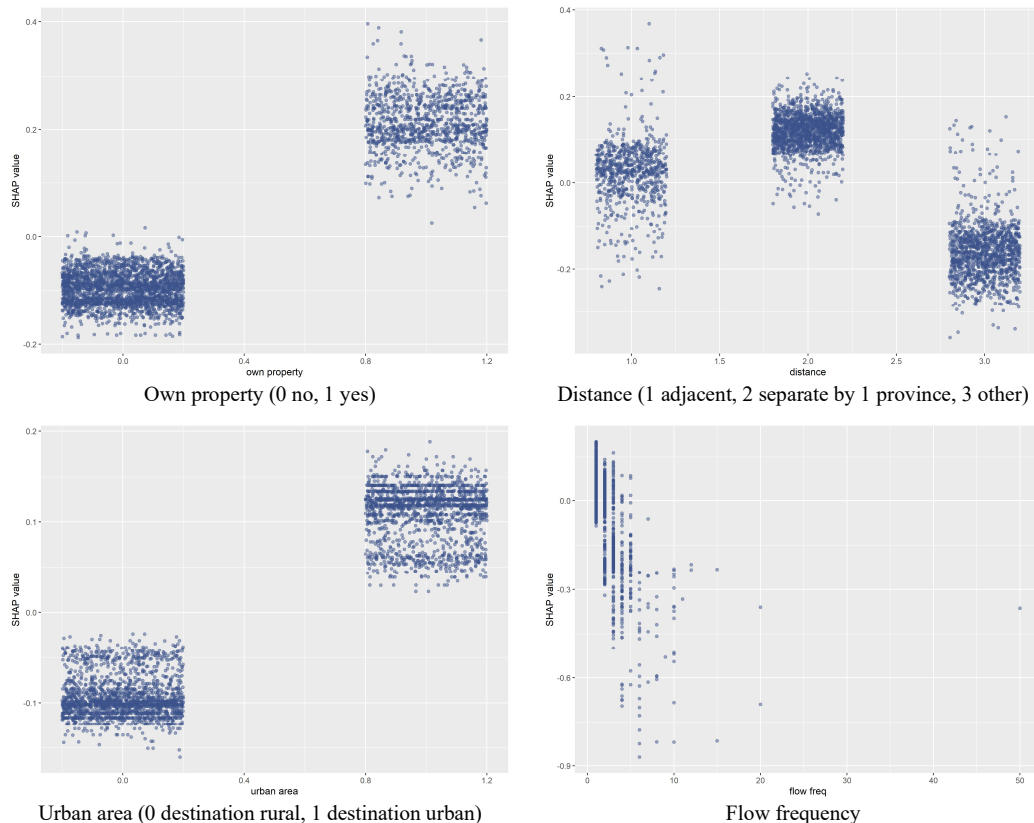


Fig. 5. SHAP value of own property, distance, urban area, and flow frequency.

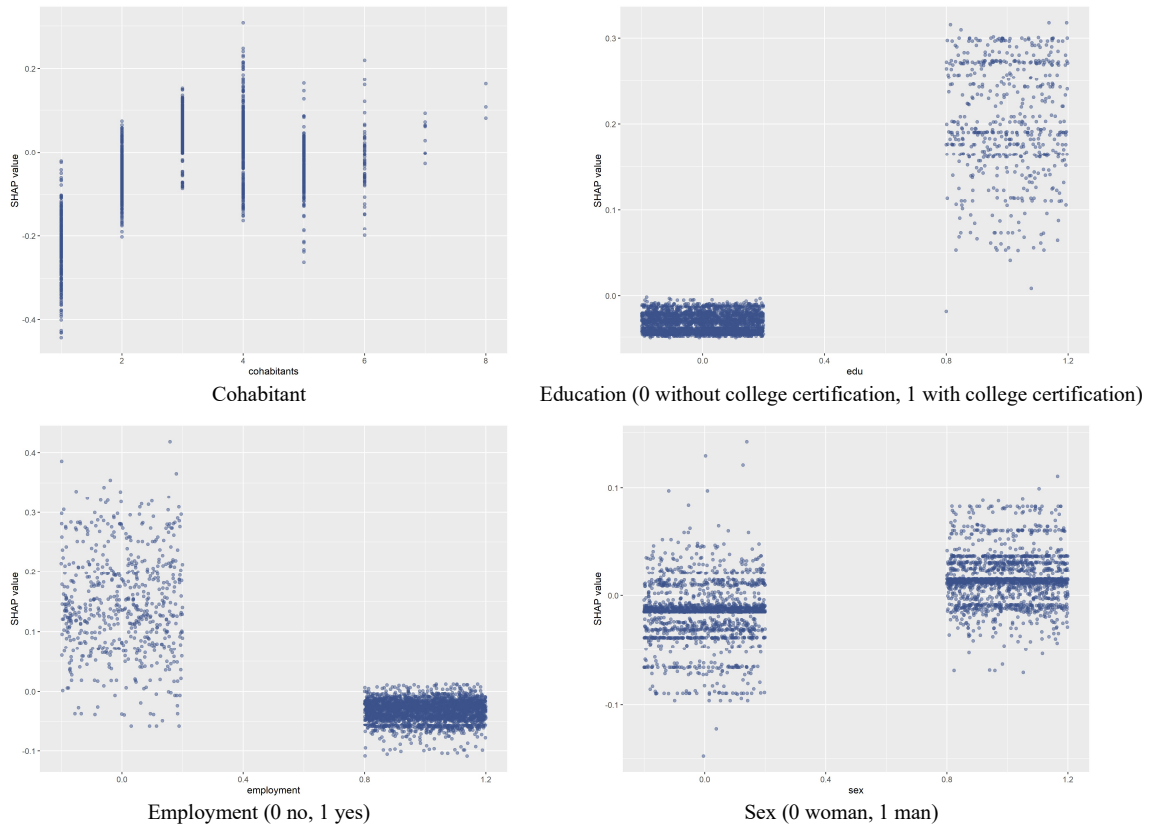


Fig. 6. SHAP value of cohabitant, education, employment, and sex.

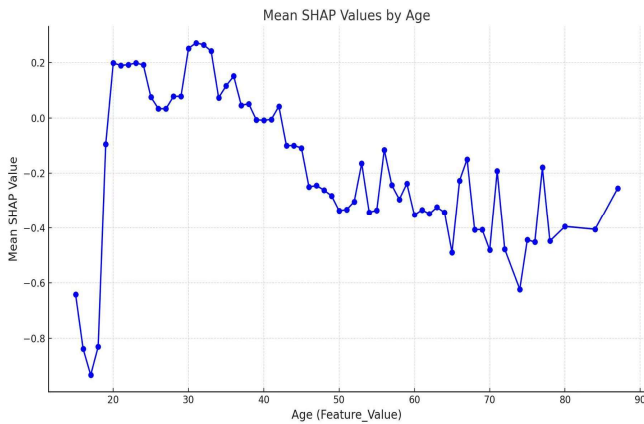


Fig. 7. Mean SHAP values by age.

To Analyze migrants' hukou acquisition intention in Tianjin, we categorized the distance factor into three levels: regions directly bordering Tianjin (like Beijing and Hebei Province), cities separated by one province, and cities farther away. Generally, greater distances between cities often mean more significant differences in their culture and lifestyle. Our analysis shows that migrants from cities one province away have the highest SHAP values, indicating a strong intention to settle in Tianjin and get hukou. In contrast, those from neighboring areas such as Beijing show a lower tendency to settle, possibly because they can commute without moving or their hukou value is better than Tianjin. Migrants from more distant regions have less chance to settle, likely due to both the physical distance and cultural differences.

The different choice of destination affects migrants' Hukou Acquisition Intention. Migrants who move to urban areas like the city center and coastal regions have a stronger intention to get hukou than those whose destination is suburban areas. This may be because center cities offer more

job opportunities, better services, and social networks with friends [16]. Suburban areas may not attract as many migrants due to fewer opportunities and lower-quality public services.

Migration frequency affects Hukou Acquisition Intention. When migrants move more often, their intention to get hukou declines. Migrants who change cities frequently may not want to settle in certain places. They hope to focus more on finding short-term jobs and staying flexible. This shows that people who move frequently care less about long-term stability to get hukou and more about immediate opportunities in Tianjin City.

The number of cohabitants living together is different when they affect migrants' Hukou Acquisition Intention. Migrants who live alone show lower intention to get hukou, to be contrast, those living with three to four people show higher intention. Living with others gives support and helps migrants feel more connected to Tianjin City. This makes them more likely to want hukou and settle in Tianjin. Also, good education raises Hukou Acquisition Intention. Migrants with education have better job chances and more financial stability, so they are more willing to get hukou. Gender does not seem to have a clear effect in this dissertation.

It is shocking that not having a job increases migrants' Hukou Acquisition Intention, which goes against common ideas [17]. Survey results show that migrants without jobs often move for other reasons like education, marriage, or family reasons. These reasons create a stronger need for hukou compared to work-related migration. Migrants who move for school or marriage want long-term stability, so they are more likely to seek hukou. Family relocation also increases their desire to settle by getting hukou. This shows that non-work reasons often lead to a higher need for hukou than job-related reasons.

V. CONCLUSION

This study researches what affects Hukou Acquisition Intentions in Tianjin. It gives ideas to make policies better and helps the government make clear plan to increase the ability to absorb the talents. The study uses XGBoost and SHAP to find these factors.

One idea is to help people stay in the city longer. The study shows that stay willing in Tianjin city is the most important reason people want hukou. The city could give housing help, programs to make people feel welcome, and rewards for staying longer. People with rural hukou often worry about losing their land in their hometown. Tianjin could give them extra help during this change.

Younger people, especially those aged 20–34, are more likely to want hukou. They are a good group to focus on and do future research. The city could make it easier for them to move by giving them more job and study chances. People who own houses in the city are also more likely to want hukou. The city could help migrants buy houses by offering cheaper housing and rent options.

Migrants who move to city centers want hukou more than those in suburban areas because cities have better jobs and services. To fix this, the city could make suburbs better by giving them more jobs and better public services or improving the barrier in the urban area to get hukou. This could help bring more migrants to those areas to shrink the gap between rural and urban areas in Tianjin.

The study author used XGBoost and SHAP to learn Hukou Acquisition Intentions. XGBoost helps to find hidden patterns in the data. SHAP helps explain how important each factor is and how different steps in the same factor affect Hukou Acquisition Intentions. We can figure out how things like age and distance affect migrants' decisions in different levels. These tools give clear ideas to help the city improve its hukou policies.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- [1] R. Cui and J. H. Cohen, "Reform and the hukou system in China," *Migration Letters*, vol. 12, no. 3, pp. 327–335, 2015.
- [2] T. Sicular, X. Yue, B. Gustafsson, and S. Li, "The urban-rural income gap and inequality in China," *Review of Income and Wealth*, vol. 53, no. 1, pp. 93–126, 2010.

- [3] M. Hou, X. Zhou, and R. Jiang, "What influences the family migration decision of China's new generation rural-urban migrants? A multilevel logistic regression analysis," *Journal of Geographical Research*, vol. 5, no. 4, 2022.
- [4] Q. Fu and Q. Ren, "Educational inequality under China's rural-urban divide: The hukou system and return to education," *Environment & Planning*, vol. 42, no. 3, pp. 592–610, 2010.
- [5] C. Chen and C. C. Fan, "China's hukou puzzle: Why don't rural migrants want urban hukou?" *China Review*, vol. 16, no. 3, pp. 9–39, 2016.
- [6] H. Xu, J. Situ, R. Hou, et al., "Predicting body composition in the U.S. population using machine learning models," *Medicine & Science in Sports & Exercise*, vol. 56, no. 10S, pp. 479–479, 2024.
- [7] C. Li and S. Managi, "Impacts of community attachment and community livability on environmental activity according to XGBoost and SHAP," *Cities*, vol. 156, 105559, 2025.
- [8] C. Hu, Y. Zhu, L. Lin, and W. Wang, "Analysis on floating population's hukou transfer intention and its influence factors: Insights from a survey in Fujian Province," *Population and Development*, vol. 17, no. 3, 2011.
- [9] Y. Zhu, "China's floating population and their settlement intention in the cities: Beyond the hukou reform," *Habitat International*, vol. 31, no. 1, 2007.
- [10] B. Pan, "Application of XGBoost algorithm in hourly PM2.5 concentration prediction," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2018, vol. 113, 012127, p. 7.
- [11] W. Laphrom, C. Se, T. Champahom, et al., "XGBoost-SHAP and unobserved heterogeneity modelling of temporal multivehicle truck-involved crash severity patterns," *Civil Engineering Journal*, vol. 10, no. 6, 2024.
- [12] S. Tang, P. Hao, and X. Huang, "Land conversion and urban settlement intentions of the rural population in China: A case study of suburban Nanjing," *Habitat International*, vol. 51, pp. 149–158, 2016.
- [13] Y. Huang, "Urbanization, hukou system and government land ownership: Effects on rural migrant workers and on rural and urban hukou residents," *A Background Paper for the Global Development Outlook*, 2010.
- [14] X. Mu, A. Yeh, X. Zhang, et al., "Moving down the urban hierarchy: Turning point of China's internal migration caused by age structure and hukou system," *Urban Studies*, vol. 59, no. 7, 2021.
- [15] N. E. Coulson and M. Tang, "Institutional and demographic influences on the presence, scale and geographic scope of individual Chinese real estate investment," *Regional Science and Urban Economics*, vol. 43, no. 2, pp. 187–196, 2013.
- [16] D. Xiaom and C. Jianbao, "Empirical analysis on impacts of migration with and without hukou change on China's provincial economy," *China Population Today*, vol. 30, no. 1, pp. 65–65, 2011.
- [17] H. Zhang, "The hukou system's constraints on migrant workers' job mobility in Chinese cities," *China Economic Review*, vol. 21, no. 1, pp. 51–64, 2010.

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](#)).