Policy Decision-Making for Population Decline Using AI to Estimate Population Density From Well-Being Indicators

TATSUKI FUKUDA,
Department of Data Science, Faculty of Data Science,
Shimonoseki City University,
2-1-1 Daigakucho, Shimonoseki City, Yamaguchi, 751-8510,
JAPAN

Abstract: - This paper examines whether a model that infers habitable area population density from regional well-being indicators can serve as a guide for policy decision-making to address population decline. The study uses 51 subjective evaluation items from the regional well-being indicators and habitable area population density calculated from e-stat, a Japanese government database. The inference model was created through ensemble learning, generating six weak learners and combining them with a meta-model to form the final model. Using data from Shimonoseki City in Yamaguchi Prefecture, Japan, we varied the value of a single subjective evaluation item to observe changes in the inferred population density. The results showed that subjective evaluations related to public transportation, crime prevention, dining options, and local government initiatives significantly impact habitable area population density. Prioritizing these factors could enhance resident satisfaction and potentially mitigate the issue of population decline.

Key-Words: - Population Decline, Regional well-being indicator, Ensemble learning, Stacking, Neural network, Resident satisfaction.

Received: March 15, 2024. Revised: August 11, 2024. Accepted: September 13, 2024. Published: October 18, 2024.

1 Introduction

Japan is currently experiencing a population decline, coupled with the challenges of an aging society and a declining birthrate[1]. Yamaguchi Prefecture's Shimonoseki City, designated as a core city since October 2005, is no exception. At the time of its designation, core city status required a population of over 300,000, a criterion Shimonoseki met through a merger with surrounding areas. However, in 2000, five years before the merger, the population was approximately 310,000, but by 2005, it had already decreased to about 290,000[2]. This downward trend has continued to the present day[3]. Consequently, promoting settlement among younger generations is an urgent task, and it is equally critical to develop policies for post-child-rearing generations to enhance their satisfaction and prevent population outflow.

Regarding resident satisfaction, the regional well-being indicators[4],[5], which encompass both objective and subjective measures, quantify residents' "livability" and "happiness." A high regional well-being score suggests a place is not only livable but also desirable, serving as a potential means to stem population outflow.

In recent years, advancements in computational power and the development of machine learning have led to numerous innovations across various research fields and everyday life. Specifically, progress in machine learning (ML) has played a crucial role in diverse areas such as disease diagnosis[6], fraud detection[7], text classification[8], and image recognition[9]. Traditional machine learning algorithms primarily focused on improving the performance of individual models. However, ensemble learning, which enhances performance by combining multiple models, has gained significant attention in recent times[10].

The fundamental concept of ensemble learning lies in combining multiple models to offset the shortcomings of each individual model, thereby achieving higher accuracy and stability. This approach allows for performance improvements that single models cannot achieve, making it a viable solution for many real-world problems.

This paper aims to develop a model using ensemble learning to represent the relationship between regional well-being indicators and population dynamics. By doing so, we seek to identify the indicators necessary to maintain population levels and explore strategies to curb population outflow, particularly in rural areas.

The structure of this paper is as follows: Chapter 2 provides a detailed explanation of the fundamental concepts of well-being indicators, neural networks, and ensemble learning. Chapter 3 describes the methodologies employed in this study. Chapter

4 introduces the models used and presents the results obtained. Chapter 5 discusses the findings. Finally, Chapter 6 concludes the paper.

2 Preliminary

This section introduces the preliminary knowledge relevant to this study.

2.1 Regional Well-Being Indicators

The well-being indicators, published by the Digital Agency of the Japanese government, quantify and visualize citizens' "livability" and "happiness"[5]. These indicators are composed of subjective measures based on surveys and objective measures derived from open data. Each indicator is further categorized into three groups of factors: living environment, community relationships, and personal fulfillment. Additionally, the subjective measures include the following four questions to assess happiness and life satisfaction:

- How happy are you currently?
- How happy do you think the people in your neighborhood are overall?
- How satisfied are you with your current living environment?
- Do you feel that both you and the people close to you are generally in a good mood?

These questions are used to calculate both subjective and objective well-being indices.

Regarding the well-being indicators, the age distribution of survey respondents is shown in Figure 1.

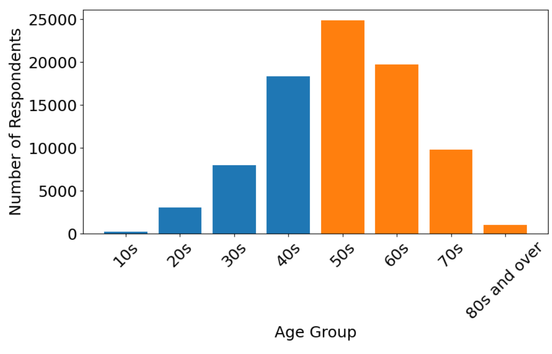


Fig. 1: Number of respondents by age group in the national survey of regional well-being indicators

The data indicates that a relatively high proportion of respondents are from older age groups, with 65.2% being in their 50s or older, typically considered the post-child-rearing generation. In contrast, as of the end of April 2024, the proportion of Shimonoseki City's population aged 50 and above is 56.2%, [11]. However, population projections for Shimonoseki City, as depicted in Figure 2,[12], suggest that the proportion of residents aged 50 and above will gradually increase, reaching 65.3% by 2040. Therefore, current analyses using regional well-being indicators could serve as a useful tool for predicting the well-being of Shimonoseki City's residents in 2040.

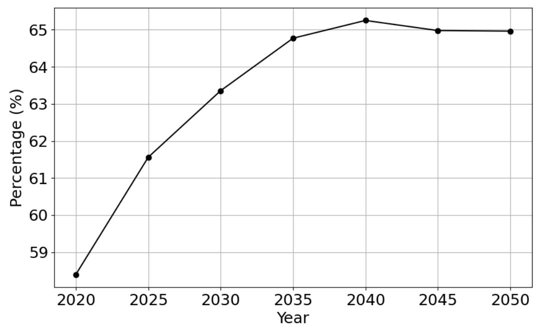


Fig. 2: Projected proportion of the population aged 50 and over in Shimonoseki City (graph created based on reference [6])

2.2 Neural Network

Neural networks are a type of machine learning model designed to mimic the functioning of neurons in the human brain and are widely used in the field of deep learning[13],[14]. A neural network is composed of multiple units called neurons, organized into layers. Each layer processes input data and transmits it to the next layer, enabling the approximation of complex functions and pattern recognition[15].

Neural networks consist of the following three layers[14],[16]:

Input Layer This layer receives the input data.

Hidden Layers These layers process the input data, performing feature extraction and non-linear transformations. Networks with one or more hidden layers are referred to as deep neural networks (DNNs)[17].

Output Layer This layer provides the final prediction or output.

Each neuron receives inputs from the neurons in the preceding layer, applies weights to these inputs, adds a bias, and then transforms the result using an activation function. Mathematically, this can be represented by Equation 1 follows[18]:

$$a_j^{(l)} = f(\Sigma_{i=1}^{n^{(l-1)}} w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}), \qquad (1)$$

where $a_j^{(l)}$ denotes the output of the j-th neuron in the l-th layer, $w_{ij}^{(l)}$ denotes the weight between the i-th neuron in the $(l-1)$-th layer and the j-th neuron in the l-th layer, $b_j^{(l)}$ denotes the bias of the j-th neuron in the l-th layer, f is the activation function, and $n^{(l-1)}$ denotes the number of neurons in the $(l-1)$-th layer.

Activation functions serve to non-linearly transform the output of a neuron, enabling the network to learn complex patterns[19]. Some of the representative activation functions are sigmoid function, ReLU (Rectified Linear Unit) function, and tanh function[20]:

- Sigmoid Function: $\sigma(x) = \frac{1}{1+e^{-x}}$
- ReLU Function: $f(x) = max(0, x)$
- Tanh Function: $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

The training of neural networks is performed to minimize a loss function[14]. The loss function quantifies the error between the predicted values and the actual values. Common examples include Mean Squared Error (MSE) and Cross-Entropy Loss[21]. Minimization of the loss function is typically achieved using Gradient Descent or its variants[22].

The basic equation of gradient descent is as follows:

$$\theta := \theta - \eta \Delta_\theta J(\theta), \tag{2}$$

where θ denotes the parameters of the model, η denotes the learning rate which is a scalar determining the step size, and $\Delta_\theta J(\theta)$ denotes the gradient of the loss function $J(\theta)$ with respect to the parameters θ.

Neural networks are layered learning models, particularly adept at pattern recognition and prediction in complex data. By selecting appropriate structures, activation functions, and learning algorithms, neural networks can achieve high performance across various tasks[23].

2.3 Ensemble Learning

Ensemble learning is a method that combines multiple machine learning models to achieve superior predictive performance compared to individual models. The concept of ensemble learning was proposed in 1979[24]. They introduced a method for partitioning the feature space using multiple component classifiers. Subsequently, in 1990, it was demonstrated that applying an ensemble of artificial neural network classifiers could achieve better predictive performance than a single classifier[25].

Additionally, boosting techniques, which transform weak classifiers into strong ones, have been proposed[26]. These techniques later formed the basis for powerful algorithms such as AdaBoost, Gradient Boosting, and XGBoost.

There are three main techniques in ensemble learning: bagging, boosting, and stacking[27].

Bagging (Bootstrap Aggregating)[28] is a method that trains multiple models using different subsamples of the training data. By averaging or using a majority vote to determine the final prediction, it helps prevent overfitting. When averaging, the method follows Equation 3.

$$\hat{f}(x) = \frac{1}{B}\Sigma_{b=1}^{B}\hat{f}_b(x), \tag{3}$$

where $\hat{f}$ denotes the aggregated prediction, B denotes the number of models, and $\hat{f}_b(x)$ denotes the prediction from the b-th model.

For example, Random Forest is a representative algorithm of bagging. It trains multiple decision trees and combines their predictions through majority voting[29]. Bagging reduces model variance and helps prevent overfitting.

Boosting is a technique that sequentially trains weak learners, improving the model's accuracy by focusing on the samples misclassified by the previous models. The final model is expressed as a weighted average of the weak learners(Equation 4).

$$\hat{f}(x) = \Sigma_{b=1}^{B}\alpha_b\hat{f}_b(x), \tag{4}$$

where α_b denotes the weight of each model.

Representative algorithms include AdaBoost[30], Gradient Boosting[31], and XGBoost[32]. Boosting reduces model bias and achieves higher accuracy. AdaBoost is the first successful boosting algorithm. It sequentially trains simple learners (usually decision stumps) on weighted samples, constructing a strong final classifier.

Gradient Boosting formulates the concept of boosting as an optimization problem. In this method, new learners are added in the direction that minimizes the model's error.

XGBoost further improves on Gradient Boosting. It offers high computational efficiency and predictive performance, winning numerous machine learning competitions.

Stacking uses the predictions of multiple different models as new features to train another meta-model[33]. Thus, the final prediction can be expressed as Equation 5:

$$\hat{f}(x) = g(\hat{f}_1(x), \hat{f}_2(x), \cdots, \hat{f}_B(x)), \tag{5}$$

where g denotes the meta-model and $\hat{f}_i(x)(1 \leq i \leq B)$ denote the predictions from the base models. This method leverages the strengths of each model while compensating for their weaknesses. Logistic regression or linear regression is commonly used as the meta-model.

Each of these methods has its own advantages, and in practical applications, they are often combined to leverage their respective strengths. For example, ensemble learning models using boosting techniques have achieved high predictive accuracy in the prediction of hepatitis C[34]. Additionally, studies on brain tumor detection have demonstrated high accuracy in early detection by combining deep learning with ensemble learning[35].

Thus, ensemble learning improves model generalization and reduces the risk of overfitting by using multiple models. Consequently, when combined with neural networks and deep learning, it provides solutions to more complex problems, making it a powerful tool for maximizing the performance of machine learning models.

3 Methodology

First, we describe the methods used in this study.

3.1 Model Inputs and Outputs

In this study, we aimed to model the relationship between resident satisfaction and population, hypothesizing that subjective evaluations are more directly connected to satisfaction than objective evaluations. Therefore, among the various items used to calculate the well-being indicators, we used subjective evaluations as inputs for the model. Specifically, there are 46 subjective evaluation items used to calculate the well-being indicators, and we also included the 5 items calculated as regional happiness and life satisfaction, resulting in a total of 51 input items for the model. These items are provided as standard scores, but we normalized them using Equation 6 before inputting them into the model.

$$N(x) = \frac{x - x_{min}}{x_{max} - x_{min}}, \qquad (6)$$

where x denotes the original value and $N(x)$ denotes the normalized value of x. x_{min} and x_{max} denote the minimum value and maximum value in the dataset, respectively.

On the other hand, for the values predicted by the model, using the simple population might result in higher absolute numbers for larger areas. Thus, we used habitable area population density. This density refers to the population density of land where people can actually reside. To calculate habitable area population density, we used data from the 2020 survey downloaded from the Government Statistics Portal (e-Stat)[36].

The dataset comprises data at the municipal level across Japan. After excluding municipalities with fewer than 50 responses, the dataset was reduced to 575 entries. Of these, 70% were used for training, while the remaining 30% were reserved for validation.

3.2 Model Training

In this study, we used ensemble learning. Given that the model aimed to predict habitable area population density from 51 input items, we envisioned a relatively simple model structure using neural networks, employing the stacking method as the ensemble learning technique.

Firstly, we prepared L models, NN-1 to NN-L. These models take the normalized subjective evaluation scores (51 items used in the well-being indicators) as input and output the habitable area population density (Figure 3).

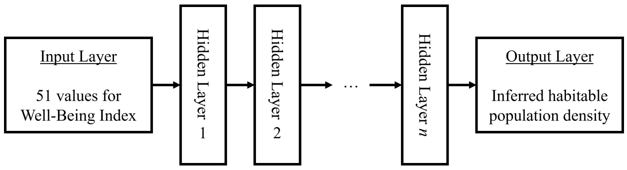


Fig. 3: Model for inferring habitable area population density using well-being indicators as input. This model acts as a weak learner in ensemble learning.

During this process, the training of these six models includes tuning the hyperparameters as well as the model architecture. Once the training of the six models is complete, they are combined in parallel. A meta-model is then used to derive the final predicted values from the predictions made by NN-1 to NN-L (Figure 4).

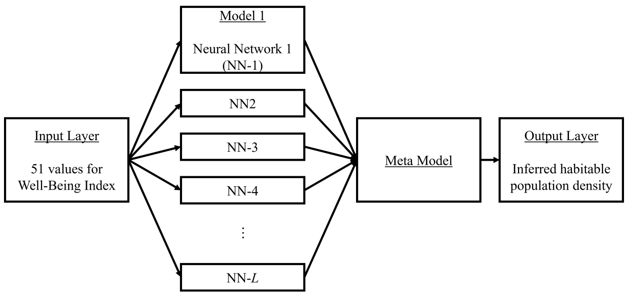


Fig. 4: Model for inferring habitable area population density using well-being indicators as input. The structure combines L weak learners and aggregates them with a meta-model.

3.3 Experimental Objective

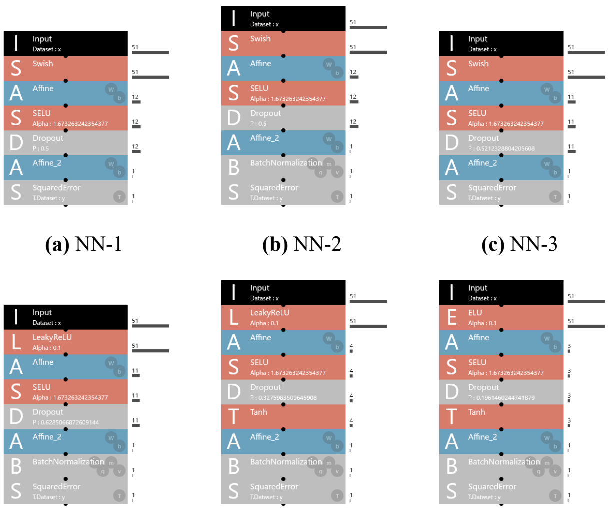
The final model can predict habitable area population density from the normalized values of 51 subjective evaluation items. Therefore, by varying the value of any one of these 51 subjective evaluation items, we investigated which evaluation items significantly influence habitable area population density. For values other than the one being varied, we used the values corresponding to Shimonoseki City, the target area of this study.

4 Results

After reporting on the generated model, we will discuss the findings on how the habitable area population density changes significantly when any single evaluation item is varied.

4.1 Obtained Models

The models obtained from the training process, NN-1 to NN-6, are shown in Figure 5.



(a) NN-1 **(b)** NN-2 **(c)** NN-3

(d) NN-4 **(e)** NN-5 **(f)** NN-6

Fig. 5: The six weak learner models, NN-1 to NN-6, generated through training.

The Mean Absolute Error (MAE) for models NN-1 to NN-6, as well as the final model generated from these models, is presented in Table 1.

Table 1. List of Mean Absolute Errors (MAE) for the generated weak learners and the final inference model aggregated by the meta-model

model	MAE
NN-1	0.0516
NN-2	0.0502
NN-3	0.0516
NN-4	0.0523
NN-5	0.0516
NN-6	0.0496
Meta Model	0.0494

4.2 Changes in Habitable Area Population Density Due to Variation in a Single Item

We investigated the changes in habitable area population density when varying the value of any one of the 51 evaluation items for Shimonoseki City, using the models obtained in section 4.1. Although it is not possible to present all results here, we focus on the most significant findings.

The primary goal of this study is to identify subjective evaluations that significantly influence habitable area population density. Therefore, limiting the discussion to the most impactful results is in line with the study's objectives.

4.2.1 Public Transportation

The most significant change in habitable area population density occurred when the evaluation item related to public transportation was varied. This subjective evaluation is based on a survey question asking, "Can you travel to any place you want at any time using public transportation?"

Figure 6 shows the relationship between the normalized subjective evaluation scores for public transportation and the normalized habitable area population density. The vertical axis represents the normalized subjective evaluation scores for public transportation, while the horizontal axis represents the normalized habitable area population density.

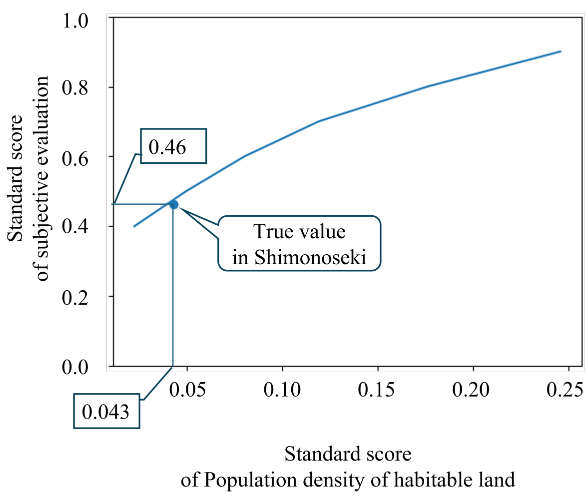


Fig. 6: Inference results of habitable area population density based on the normalized subjective evaluation scores for the question, "Can you travel to any place you want at any time using public transportation?" The horizontal axis represents normalized habitable area population density, and the vertical axis represents normalized subjective evaluation scores.

When the normalized subjective evaluation score for public transportation was adjusted from 0 to 1, the resulting change in habitable area population density was 0.322. This indicates that public transportation accessibility significantly impacts population density in habitable areas, with better transportation options potentially leading to higher population densities.

4.2.2 Crime Prevention and Security

The second most significant impact on habitable area population density was observed when varying the evaluation item related to crime prevention and security. This subjective evaluation is based on a survey question asking, ”Are crime prevention measures (such as police stations, streetlights, security cameras, and neighborhood watch) well established, and is the area safe?”

Figure 7 shows the relationship between the normalized subjective evaluation scores for crime prevention and security and the normalized habitable area population density. The vertical axis represents the normalized subjective evaluation scores for crime prevention and security, while the horizontal axis represents the normalized habitable area population density.

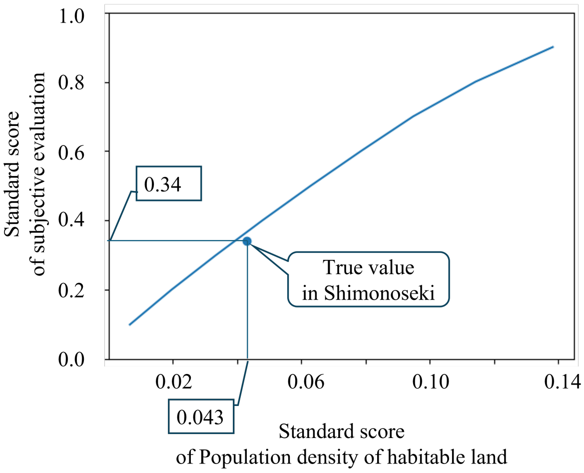


Fig. 7: Inference results of habitable area population density based on the normalized subjective evaluation scores for the question, ”Are crime prevention measures (such as police stations, streetlights, security cameras, and neighborhood watch) well established, and is the area safe?” The horizontal axis represents normalized habitable area population density, and the vertical axis represents normalized subjective evaluation scores.

When the normalized subjective evaluation score for crime prevention and security was adjusted from 0 to 1, the resulting change in habitable area population density was 0.161. This result underscores the importance of safety and crime prevention measures in influencing population density in habitable areas, suggesting that improvements in these areas could enhance population retention.

4.2.3 Availability of Dining Options

The third most significant impact on habitable area population density was observed when varying the evaluation item related to dining options. This subjective evaluation is based on a survey question asking, ”Are there ample places to enjoy dining?”

Figure 8 shows the relationship between the normalized subjective evaluation scores for dining options and the normalized habitable area population density. The vertical axis represents the normalized subjective evaluation scores for dining options, while the horizontal axis represents the normalized habitable area population density.

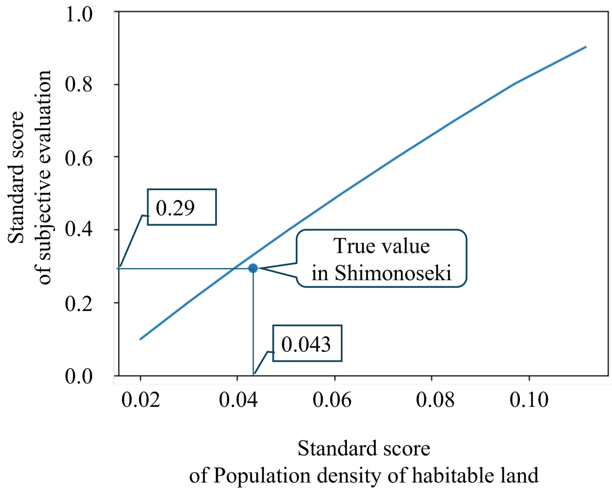


Fig. 8: Inference results of habitable area population density based on the normalized subjective evaluation scores for the question, ”Are there ample places to enjoy dining?” The horizontal axis represents normalized habitable area population density, and the vertical axis represents normalized subjective evaluation scores.

When the normalized subjective evaluation score for dining options was adjusted from 0 to 1, the resulting change in habitable area population density was 0.113. This finding highlights the role of dining options in influencing population density in habitable areas, indicating that a greater variety and availability of dining venues can contribute to higher population densities.

4.2.4 Local Government Initiatives

The impact on habitable area population density was the same when varying the evaluation item related to local government initiatives as when varying the item related to dining options. This subjective evaluation is based on a survey question asking, ”Does the local government take the community’s concerns seriously?”

Figure 9 shows the relationship between the normalized subjective evaluation scores for local government initiatives and the normalized habitable area population density. The vertical axis represents the normalized subjective evaluation scores for local government initiatives, while the horizontal axis represents the normalized habitable area population density.

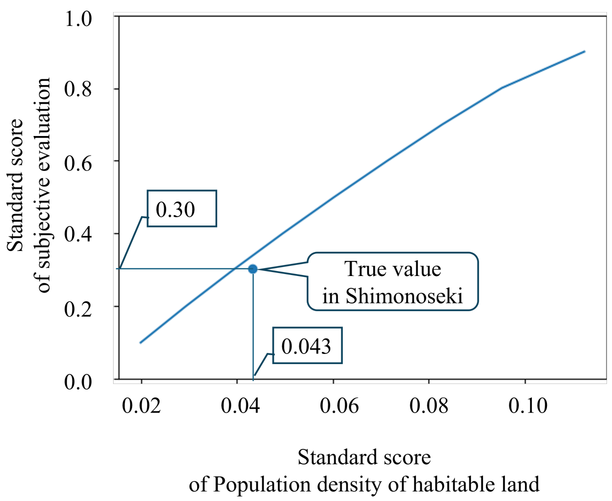


Fig. 9: Inference results of habitable area population density based on the normalized subjective evaluation scores for the question, ”Does the local government take the community's concerns seriously?” The horizontal axis represents normalized habitable area population density, and the vertical axis represents normalized subjective evaluation scores.

When the normalized subjective evaluation score for local government initiatives was adjusted from 0 to 1, the resulting change in habitable area population density was 0.113. This result underscores the importance of effective local government initiatives in influencing population density in habitable areas, suggesting that proactive and community-focused governance can contribute to higher population densities.

5 Analysis

In section 4.2, we identified the factors that significantly influenced habitable area population density: 1) public transportation, 2) crime prevention and security, and 3) dining options. While the results for public transportation and crime prevention are expected, the impact of dining options might be surprising. However, these factors are interconnected. Dining establishments are primarily used at night, often involving alcohol consumption, which increases the need for public transportation. Additionally, employees and part-time workers in dining establishments frequently rely on public transportation. If the last bus departs early, these establishments cannot stay open late. Therefore, it is reasonable to view the development of public transportation and dining establishments as related. Furthermore, to stay open late, it is crucial to maintain security and crime prevention measures. Thus, the top three factors are likely closely related.

In other words, when considering the impact on habitable area population density, improving the convenience of public transportation and enhancing regional crime prevention measures in the medium to long term, creating an environment conducive to going out at night, and promoting the development of the dining industry could potentially help mitigate the decline in habitable area population density.

On the other hand, for residents to evaluate ”local government takes the community's concerns seriously,” ranked fourth, it is essential that they understand the local government's initiatives. This means that the local government must not only implement policies that residents need but also effectively communicate these initiatives to the residents. By doing so, residents can be aware of the local government's efforts and the various supports provided, leading to higher satisfaction.

6 Conclusion

In this paper, we explored potential measures to address the current issue of population decline in Japan. Using machine learning with neural networks, we examined the relationship between 51 subjective evaluation items included in the regional well-being indicators, which reflect resident satisfaction, and habitable area population density. By varying these subjective evaluation items, we assessed how the habitable area population density changes.

The results revealed that the item related to public transportation had the most significant impact on habitable area population density. Additionally, the second most influential item was related to crime prevention and security, while the third was related to dining options. These findings suggest a potential relationship between the availability of public transportation and these factors. Furthermore, the study highlighted the importance of effective communication by local governments. Without proper dissemination of information about implemented measures, resident satisfaction may decline, thereby affecting habitable area population density.

It is important to note that the inherent meaning of the regional well-being indicators suggests that regions with high satisfaction regarding public transportation and crime prevention also exhibit high resident satisfaction. Therefore, caution is necessary in interpreting these results. However, given that the future population of Shimonoseki City, particularly those aged 50 and above, is likely to mirror the demographic of current well-being survey respondents, it is plausible that these factors will remain significant in determining resident satisfaction in the future. Addressing issues related to public transportation and crime prevention from a medium- to long-term perspective may help prevent the outflow of residents in this age group.

This study was based on the regional well-being indicators for Shimonoseki City, indicating population decline measures specific to this area. The methodology used in this study can be easily adapted to other municipalities by changing the base data, thereby contributing to the development of population decline countermeasures for various regions.

Future research should focus on substantiating the identified possibilities with solid evidence to ensure their reliability and applicability.

Acknowledgment:
I would like to thank K. Sugiura, M, Suga, S. Koyanagi, M. Ishimura, and K. Matsuura for useful discussions. I also thank the city of Shimonoseki for providing the seed for this theme.

Declaration of Generative AI and AI-assisted technologies in the writing process
During the preparation of this work the authors used Grammarly for language editing. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] Statistics Bureau, ”Statistics Topics No. 119: Trends in Heisei Era,” Ministry of Internal Affairs and Communications, https://www.stat.go.jp/data/topics/topi1191.html, Apr. 2019(accessed Jun. 2024).

[2] Shimonoseki City Council Secretariat. ”Overview of Shimonoseki City Administration,” https://www.city.shimonoseki.lg.jp/uploaded/attachment/65103.pdf, Feb. 2023(accessed Jun. 2024).

[3] Shimonoseki City. ”(5) Trends in Population and Number of Households (National Census and Estimated Population),” https://www.city.shimonoseki.lg.jp/uploaded/attachment/65103.pdf, Mar. 2024(accessed Jun. 2024).

[4] Smart City Institute Japan. ”Regional Well-Being Indicators,” https://www.sci-japan.or.jp/LWCI/index.html, accessed Dec. 2023.

[5] Digital Agency. ”Regional Well-Being Indicators,” https://well-being.digital.go.jp/, accessed Jun. 2024.

[6] Ahsan, M. M., Luna, S. A., and Siddique, Z., Machine-Learning-Based Disease Diagnosis: A Comprehensive Review, *Healthcare*, Vol.10, No.3, 2022, pp. 541.

[7] Alomari, D. M. and Mirza, S., Machine Learning-Based Detection for Unauthorized Access to IoT Devices, *Journal of Sensor and Actuator Networks*, Vol.12, No.2, 2023, pp. 27.

[8] Damaševičius, R. and Maskeliūnas, R., Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review, *Algorithms*, Vol.16, No.5, 2023, pp. 236.

[9] Krizhevsky, A., Sutskever, I., and Hinton, G. E., ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.

[10] Mienye, I. D. and Sun, Y., A survey of ensemble learning: Concepts, algorithms, applications, and prospects, *IEEE Access*, vol.10, 2023, pp.99129-99149.

[11] Shimonoseki City, (3-1) Population by Age (Total City and District-wise Registered Population), https://www.city.shimonoseki.lg.jp/soshiki/134/1188.html, accessed Jun. 2024.

[12] National Institute of Population and Social Security Research, Japan’s Regional Future Population Projections (2023 Estimates), https://www.ipss.go.jp/pp-shicyoson/j/shicyoson23/t-page.asp, accessed Jun. 2024.

[13] LeCun, Y., Bengio, Y., and Hinton, G., Deep learning, *Nature*, Vol.521, No.7553, 2015, pp. 436-444.

[14] Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, *MIT Press*, 2016.

[15] Schmidhuber, J., Deep learning in neural networks: An overview, *Neural Networks*, Vol.61, 2015, pp. 85-117.

[16] Krizhevsky, A., Sutskever, I., and Hinton, G. E., ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.

[17] Bishop, C. M., Neural Networks for Pattern Recognition, *Oxford University Press*, 1995.

[18] Nair, V. and Hinton, G. E., Rectified linear units improve restricted Boltzmann machines, *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.

[19] Glorot, X. and Bengio, Y., Understanding the difficulty of training deep feedforward neural networks, *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249-256.

[20] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., Learning representations by back-propagating errors, *Nature*, Vol.323, No.6088, 1986, pp. 533-536.

[21] Kingma, D. P. and Ba, J., Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014.

[22] Bottou, L., Large-scale machine learning with stochastic gradient descent, *Proceedings of COMPSTAT'2010*, 2010, pp. 177-186.

[23] Silver, D., et. al., Mastering the game of Go without human knowledge, *Nature*, Vol.550, No.7676, 2017, pp. 354-359.

[24] Dasarathy, B. V. and Sheela, B. V., A Composite Classifier System Design: Concepts and Methodology, *Proceedings of the IEEE*, Vol.67, No.5, 1979, pp. 708-713.

[25] Hansen, L. K. and Salamon, P., Neural network ensembles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, No.10, 1990, pp. 993-1001, doi:10.1109/34.58871.

[26] Schapire, R. E., The strength of weak learnability, *Machine Learning*, Vol.5, 1990, pp. 197-227.

[27] Khan, A. A., Chaudhari, O., and Chandra, R., A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation, *Expert Systems with Applications*, Vol.244, 2024, pp. 122778, doi:10.1016/j.eswa.2023.122778.

[28] Breiman, L., Bagging predictors, *Machine Learning*, Vol.24, 1996, pp. 123-140.

[29] Breiman, L., Random forests, *Machine Learning*, Vol.45, 2001, pp. 5-32.

[30] Freund, Y. and Schapire, R. E., A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences*, Vol.55, No.1, 1997, pp. 119-139, doi:10.1006/jcss.1997.1504.

[31] Friedman, J. H., Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 2001, pp. 1189-1232.

[32] Chen, T., and Guestrin, C., Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[33] Wolpert, D. H., Stacked generalization, *Neural Networks*, Vol.5, No.2, 1992, pp. 241-259.

[34] Edeh, M. O., Dalal, S., Dhaou, I. B., Agubosim, C. C., Umoke, C. C., Richard-Nnabu, N. E., and Dahiya, N., Artificial intelligence-based ensemble learning model for prediction of hepatitis C disease, *Frontiers in Public Health*, Vol.10, 2022, pp. 892371.

[35] Alsubai, S., Khan, H. U., Alqahtani, A., Sha, M., Abbas, S., and Mohammad, U. G., Ensemble deep learning for brain tumor detection, *Frontiers in Computational Neuroscience*, Vol.16, 2022, pp. 1005617.

[36] Statistics Center, Government Statistics Portal (e-Stat), https://www.e-stat.go.jp/, accessed Dec. 2023.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The author solely contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflicts of Interest
The author is affiliated with Shimonoseki City University, a public university corporation located in Shimonoseki City, which is the subject of this research.

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)
This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US