

Research Letter

Prediction of Prefecture-Level Subjective Well-Being in Japan by Using Google Trends and Socioeconomic Data: Machine Learning Model Development and Validation Study

Kenichi Kishi, MOT; Hisashi Hayashi, PhD; Shigeomi Koshimizu, PhD

Advanced Institute of Industrial Technology, Tokyo, Japan

Corresponding Author:

Kenichi Kishi, MOT

Advanced Institute of Industrial Technology

1-10-40 Higashi-ooi

Tokyo, 140-0011

Japan

Phone: 81 3 3472 7831

Email: b2410kk@aiit.ac.jp

Abstract

Incorporating prespecified Google Trends indicators into leakage-controlled stacked-ensemble models improved a 2025 holdout prediction of subjective well-being by using 2022-2025 data from Japan's 47 prefectures, reducing the mean squared error from 0.0050 to 0.0045.

(*JMIR Form Res* 2026;10:e88696) doi: [10.2196/88696](https://doi.org/10.2196/88696)

KEYWORDS

infodemiology; infoveillance; subjective well-being; Google Trends; Japan; machine learning; stacked ensemble; forecasting; socioeconomic factors; digital trace

Introduction

Subjective well-being (SWB) increasingly complements traditional economic indicators in evidence-based policymaking. Although Japan's Digital Agency has published annual prefecture-level SWB aggregates (0-10 scale) since 2022 [1], the low reporting frequency limits real-time monitoring and timely policy intervention. Infodemiology involves the study of online information patterns, while infoveillance applies these patterns for health and social monitoring [2]. Notably, the Google Trends search activity can be leveraged as a demand-side infoveillance signal for evaluating prefecture-level SWB.

Google Trends signals have been linked to well-being and SWB nowcasting [3,4], but these data streams can exhibit high volatility at finer geographic scales [5]. Digital-trace SWB studies have also leveraged social media streams (eg, Twitter) [6]. In Japan, prefecture-level keyword queries are often limited by inadequate information [7]; we hypothesized that prespecified indicators using standardized categories and topic IDs can be implemented to improve reproducibility.

Herein, we extend prior digital-trace SWB nowcasting work by focusing on subnational (prefecture-level) prediction in Japan by using prespecified Google Trends category/topic identifiers

to improve reproducibility and by evaluating performance under a strict temporal holdout year (2025). This study aims to evaluate whether prespecified Google Trends indicators provide incremental predictive value for prefecture-level evaluative SWB in Japan beyond socioeconomic and temporal predictors by using leakage-controlled stacked-ensemble modeling and strict 2025 holdout validation.

Methods

This study uses publicly available, aggregated, nonidentifiable prefecture-level indicators; therefore, ethical review was not required.

We analyzed a prefecture-year panel (47 prefectures, 2022-2025) of published prefecture-year mean evaluative SWB (0-10) from the Digital Agency's opt-in online survey [1] with leakage-controlled walk-forward stacked-ensemble modeling and a strict 2025 holdout (training 2022-2024, n=141 prefecture-years; holdout 2025, n=47).

Predictors were evaluated in three nested stages: stage 1 included socioeconomic indicators [8], stage 2 added temporal controls, and stage 3 incorporated Google Trends features. All variables are listed in Table S1 of [Multimedia Appendix 1](#).

Model specification and validation reporting items aligned with TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis with Artificial Intelligence) are summarized in Table S2 of [Multimedia Appendix 1](#).

We prespecified 25 Google Trends category/topic series by mapping OECD (Organization for Economic Co-operation and Development) well-being domains [9] to Google Trends categories and topic IDs and retrieving indices via PyTrends (Table S3 in [Multimedia Appendix 1](#)).

Trends were summarized using principal component analysis (>90% variance), and the outcomes and descriptive comparability summaries are presented in Tables S4 and S5 of [Multimedia Appendix 1](#).

For interpretability, we ranked raw series by association with the first Trends component (principal component 1) and refit

ElasticNet and extreme gradient boosting (XGBoost) by using the top 8 series.

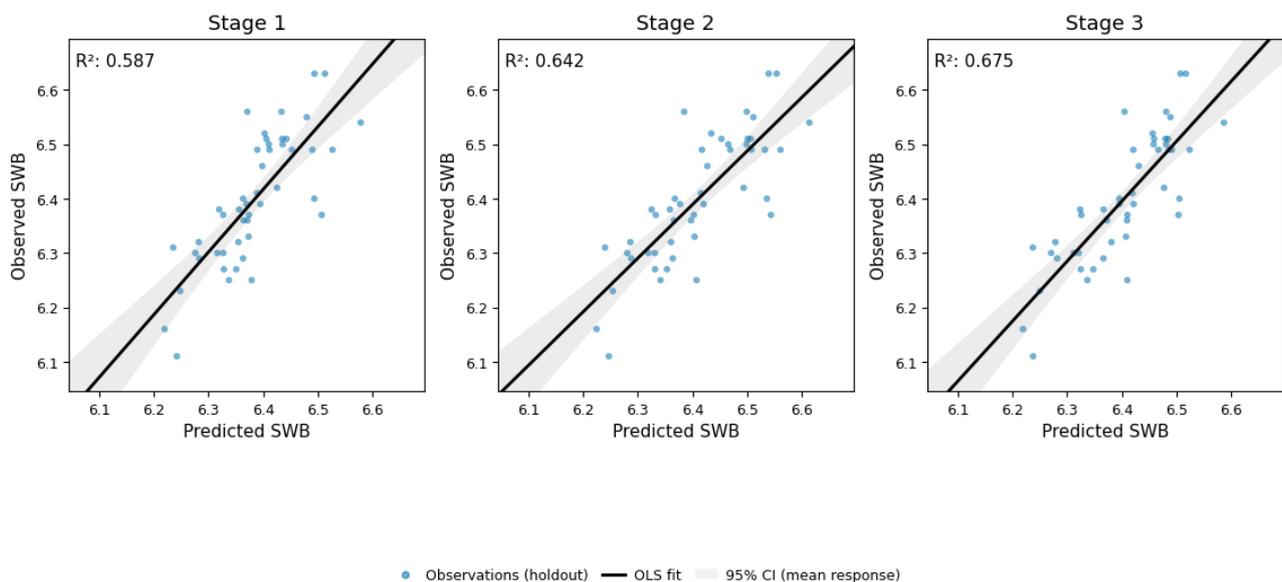
For validation, walk-forward stacking (2022→2023, 2022-2023→2024) was employed with within-window preprocessing and tuning to avoid leakage; we report adjusted R^2 and mean squared error (MSE) with 95% bootstrap CIs for 2025 (B=4000), following TRIPOD+AI guidance [10].

Algorithm S1 ([Multimedia Appendix 2](#)) and the analysis repository ([Multimedia Appendix 3](#)) support the replication.

Results

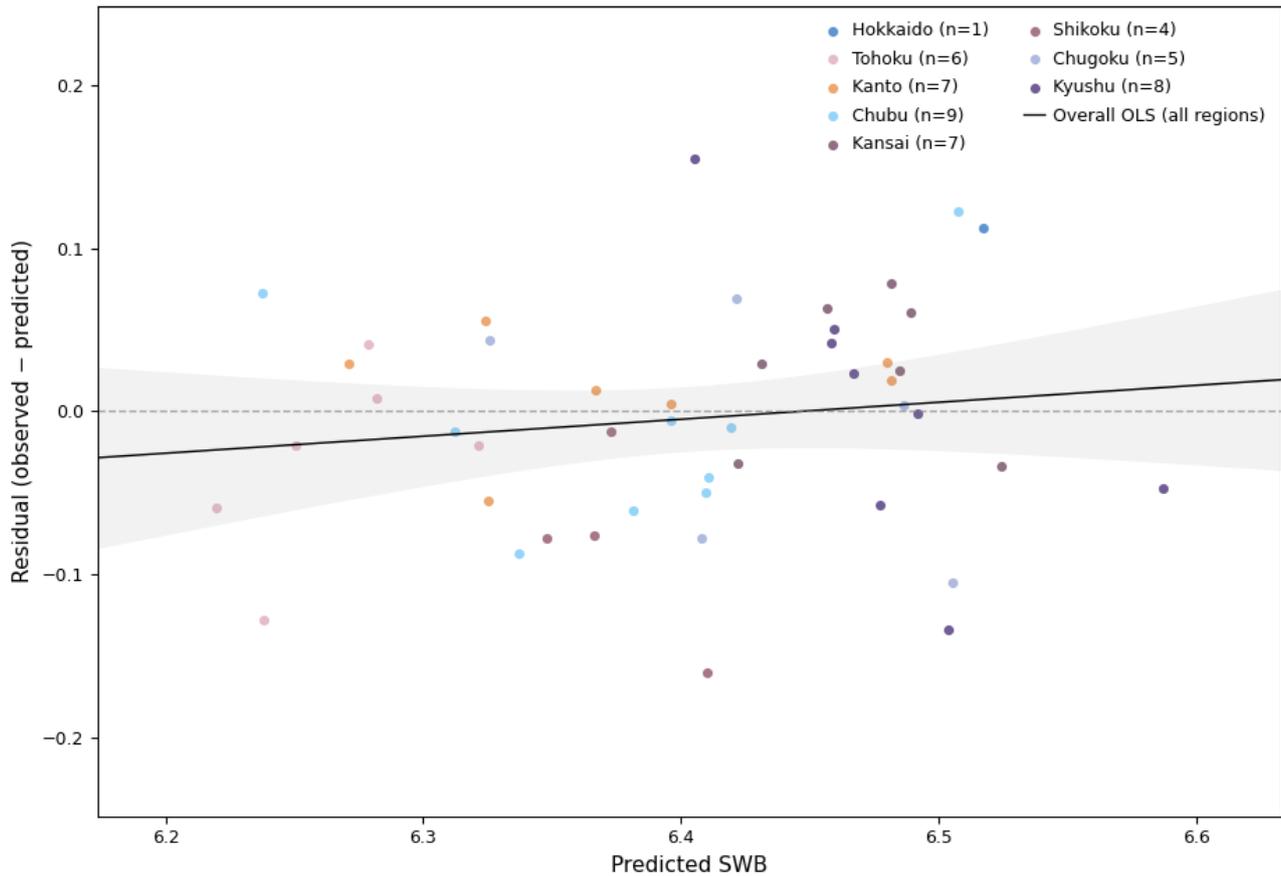
On the 2025 holdout, the adjusted R^2 increased from 0.587 (stage 1) to 0.642 (stage 2) and 0.675 (stage 3), and the MSE decreased from 0.0050 to 0.0045 ([Figure 1](#)). The bootstrap 95% CIs for the adjusted R^2 overlapped.

Figure 1. Observed vs predicted 2025 prefectural subjective well-being (SWB) at stages 1-3. OLS: ordinary least squares.



Sensitivity refits using the 8 principal component 1-aligned raw series did not improve performance compared with the principal component analysis-based Trends specification (Table S6 in [Multimedia Appendix 1](#)).

Calibration yielded a slope of 1.10 (95% CI 0.88-1.33) and an intercept of -0.67 (95% CI -2.12 to 0.78) ([Figure 2](#)). Furthermore, the prefecture-level MSE ranged from 0.000003 to 0.0257 (median 0.0022, $n=47$), with no strong regional bias ([Figure S1](#) in [Multimedia Appendix 4](#)).

Figure 2. Residuals vs predicted 2025 subjective well-being (SWB) for the final model, by region. OLS: ordinary least squares.

Discussion

Adding prespecified Google Trends components to socioeconomic and temporal predictors modestly improved the strict 2025 holdout prediction (adjusted $R^2=0.642-0.675$, $MSE=0.0050-0.0045$). This result demonstrates the incremental value of search-derived indicators as an exploratory signal for prefecture-level SWB inveillance [2].

Compared with national SWB nowcasting using Google Trends [4], subnational prediction with annual outcomes may be limited by lower signal-to-noise ratios. Standardized categories and topic IDs were exploited to improve reproducibility and reduce sparsity relative to keyword queries [7]. The dominant component reflected a broad lifestyle/mobility/consumption factor, and the ranked categories are provided in Table S6 of [Multimedia Appendix 1](#) for hypothesis-driven follow-up. Note

that these outputs should be treated as a cautious predeployment screening signal rather than a ranking or policy decision tool.

Limitations include the short panel (4 years), the ecological outcome, and reliance on a nonprobability opt-in survey with limited publicly disclosed quality-control details and uncertain year-to-year comparability (Tables S4 and S5 in [Multimedia Appendix 1](#)). Because only aggregated prefecture-year means are publicly available, we could not independently assess instrument reliability or test measurement invariance across years. Google Trends indices are normalized and may be sampled, and unmeasured time-varying factors may confound associations. Accordingly, findings should not be interpreted as causal.

Under leakage-controlled temporal validation, Google Trends signals added modest predictive value for prefecture-level evaluative SWB in Japan. Future work should test higher-frequency outcomes, lagged designs, and external validation to clarify when such signals are the most informative.

Acknowledgments

We used generative artificial intelligence tools under full human supervision for code optimization.

Data Availability

All data are publicly available (Digital Agency [1], e-Stat [8], Google Trends); the code and materials are described in [Multimedia Appendix 4](#).

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Authors' Contributions

Conceptualization, data curation, formal analysis: KK, SK

Visualization, writing – original draft: KK

Methodology, software, validation: KK, HH

Writing – review and editing: HH, SK

Supervision: SK

All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed variable definitions, TRIPOD-AI reporting checklist, Google Trends identifiers, subjective well-being outcome transparency documentation, year-to-year comparability summaries, and sensitivity analyses.

[[PDF File \(Adobe PDF File\), 536 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Algorithm S1 (walk-forward stacking pseudocode): step-by-step pseudocode for the leakage-controlled walk-forward training and evaluation pipeline, including preprocessing, feature engineering, and stacked-ensemble integration for predicting prefectural subjective well-being.

[[PDF File \(Adobe PDF File\), 135 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Full analysis repository (code and project root): this archive contains the full project root, including source code, configuration files, and the analysis pipeline used to reproduce the walk-forward stacked-ensemble models for prefectural subjective well-being prediction.

[[ZIP File \(Zip Archive\), 484 KB-Multimedia Appendix 3](#)]

Multimedia Appendix 4

Prefecture-level mean squared error for 2025 holdout.

[[PNG File , 57 KB-Multimedia Appendix 4](#)]

Multimedia Appendix 5

TRIPOD-AI-oriented reporting checklist. This checklist maps key aspects of the study (data sources, predictors, outcomes, model evaluation, robustness, fairness, reproducibility, and intended use) to TRIPOD-AI items to support transparent reporting of the prediction model.

[[PDF File \(Adobe PDF File\), 100 KB-Multimedia Appendix 5](#)]

References

1. Utilizing the liveable well-being city indicator: towards the realization of the digital garden city-state vision. Digital Agency, Government of Japan. URL: <https://well-being.digital.go.jp/?locale=en> [accessed 2025-07-09]
2. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res*. Mar 27, 2009;11(1):e11. [[FREE Full text](#)] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
3. Algan Y, Murtin F, Beasley E, Higa K, Senik C. Well-being through the lens of the internet. *PLoS One*. 2019;14(1):e0209562. [[FREE Full text](#)] [doi: [10.1371/journal.pone.0209562](https://doi.org/10.1371/journal.pone.0209562)] [Medline: [30633759](https://pubmed.ncbi.nlm.nih.gov/30633759/)]
4. Murtin F, Salomon-Ermel M. Nowcasting subjective well-being with Google Trends: a meta-learning approach. *OECD Papers on Well-being and Inequalities*. Jul 28, 2024;27. [[FREE Full text](#)] [doi: [10.1787/4ca48f7c-en](https://doi.org/10.1787/4ca48f7c-en)]
5. Rovetta A. Google trends in infodemiology: methodological steps to avoid irreproducible results and invalid conclusions. *Int J Med Inform*. Oct 2024;190:105563. [doi: [10.1016/j.ijmedinf.2024.105563](https://doi.org/10.1016/j.ijmedinf.2024.105563)] [Medline: [39043059](https://pubmed.ncbi.nlm.nih.gov/39043059/)]

6. Carpi T, Hino A, Iacus SM, Porro G. The impact of COVID-19 on subjective well-being: evidence from Twitter data. *J Data Sci.* 2023;21(4):761-780. [doi: [10.6339/22-jds1066](https://doi.org/10.6339/22-jds1066)]
7. Yang MS, Taira K. Predicting prefecture-level well-being indicators in Japan using search volumes in internet search engines: infodemiology study. *J Med Internet Res.* Nov 11, 2024;26:e64555. [FREE Full text] [doi: [10.2196/64555](https://doi.org/10.2196/64555)] [Medline: [39527805](https://pubmed.ncbi.nlm.nih.gov/39527805/)]
8. Statistics Bureau of Japan. e-Stat: Portal Site of Official Statistics of Japan. URL: <https://www.e-stat.go.jp/en> [accessed 2025-08-19]
9. OECD guidelines on measuring subjective well-being. Organisation for Economic Co-operation and Development. Mar 20, 2013. URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2013/03/oecd-guidelines-on-measuring-subjective-well-being_g1g28641/9789264191655-en.pdf [accessed 2025-08-19]
10. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI Group. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* Apr 18, 2024;385:q902. [FREE Full text] [doi: [10.1136/bmj.q902](https://doi.org/10.1136/bmj.q902)] [Medline: [38636956](https://pubmed.ncbi.nlm.nih.gov/38636956/)]

Abbreviations

MSE: mean squared error

OECD: Organization for Economic Co-operation and Development

SWB: subjective well-being

TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis with Artificial Intelligence

XGBoost: extreme gradient boosting

Edited by A Mavragani, I Steenstra; submitted 30.Nov.2025; peer-reviewed by SM Iacus, M Iqhrammullah; comments to author 19.Feb.2026; accepted 06.Mar.2026; published 20.Mar.2026

Please cite as:

Kishi K, Hayashi H, Koshimizu S

Prediction of Prefecture-Level Subjective Well-Being in Japan by Using Google Trends and Socioeconomic Data: Machine Learning Model Development and Validation Study

JMIR Form Res 2026;10:e88696

URL: <https://formative.jmir.org/2026/1/e88696>

doi: [10.2196/88696](https://doi.org/10.2196/88696)

PMID:

©Kenichi Kishi, Hisashi Hayashi, Shigeomi Koshimizu. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.