Machine Learning-Based Forecasting of U.S. State Unemployment Rates for 2026

Jiahao Ma[1], Tingyu Xie[2,*]

{2023202653@ruc.edu.cn[1], xietingyu1@mails.gdut.edu.cn[2,*]}

Renmin University of China, Sino-French Institute, Suzhou, 215123, China[1], School of International Education, Guangdong University of Technology, Guangzhou, 510006, China[2]

Abstract. This paper forecasts U.S. state unemployment rates for 2026 using quarterly data from 2020–2025 and a macroeconomic predictor, real GDP growth by state. Descriptive statistics document the 2020 pandemic shock, the recovery through 2022, and a mild uptick during 2023–2025, alongside large cross-state differences in peak unemployment. To quantify the output–labor link, a fixed-effects regression for Michigan and Hawaii indicates a countercyclical relationship between GDP growth and unemployment ($\beta \approx -0.48$; $R^2 \approx 0.20$). For prediction, linear regression is compared with random forest and gradient boosting. In cross-validation and a 2025 hold-out test, the linear model achieves the lowest mean squared error, while the ensemble models show signs of overfitting. Using a ±0.25 percentage-point rule relative to end-2025, most 2026 forecasts are classified as stable.

Keywords: Unemployment; State-level forecasting; Machine learning; Panel data; GDP growth

1 Introduction

The COVID-19 pandemic generated the sharpest U.S. labor-market shock in decades. Unemployment rose abruptly in 2020 and then declined as the economy reopened, but the magnitude and pace of recovery differed across states. Tourism- and service-intensive states such as Hawaii experienced especially large spikes, while industrial states such as Michigan also saw pronounced volatility. Understanding these state-level dynamics matters for policy design and for anticipating near-term labor-market risks.

Okun's Law provides a convenient starting point for thinking about unemployment: when output expands, unemployment typically moves in the opposite direction [1]. Yet the GDP–unemployment connection is rarely perfectly linear, and its strength can vary across places and periods. This motivates forecasting approaches that can flexibly accommodate nonlinear responses and complex interactions [2,3]. A growing empirical literature finds that machine-learning models often deliver better predictive accuracy than simple econometric benchmarks, especially when relationships are unstable or heterogeneous [4,5,6]. Other work further shows that combining time-series information with modern learning methods can improve short- and

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365158

medium-horizon forecasts [7,8,9].

This paper applies that perspective to state unemployment forecasting. Using quarterly data covering 2020–2025, it generates predictions for state unemployment rates in 2026 and incorporates state real GDP growth as a macroeconomic signal. Michigan and Hawaii are highlighted to make the results concrete, but model performance is evaluated against the full panel of states to provide a systematic benchmark.

The empirical design proceeds in three steps. First, the data are reorganized into a consistent quarterly panel by mapping monthly unemployment measures into the quarterly frequency used for GDP. Second, a simple fixed-effects regression is estimated to summarize how unemployment co-moves with GDP growth across states over time. Third, the forecasting exercise compares linear regression with two nonlinear alternatives—random forests and gradient boosting—under an out of sample evaluation scheme, and then uses the preferred specifications to produce 2026 forecasts together with a straightforward rule that classifies the predicted trend.

2 Methodology

2.1 Data acquisition and preprocessing

Unemployment rates are taken from the U.S. Bureau of Labor Statistics Local Area Unemployment Statistics (LAUS), and quarterly real GDP by state is taken from the U.S. Bureau of Economic Analysis [10,11].

Because unemployment is reported monthly while GDP is quarterly, monthly unemployment rates are averaged to quarterly values to match the GDP frequency. A panel dataset is constructed with quarterly unemployment as the target variable, state fixed effects (state dummy variables) to capture time-invariant characteristics, and quarter-on-quarter real GDP growth as the macro predictor. A time-ordered split is used for evaluation, with 2020–2024 as the training period and 2025 held out for testing before producing 2026 forecasts.

2.2 Models

Three regression models are estimated to forecast unemployment. Their roles in the comparison are summarized below.

The paper first estimates a linear regression as a transparent baseline for the relationship between GDP growth and unemployment. The paper then applies a random forest regressor to capture nonlinear interactions and heterogeneous state patterns using an ensemble of decision trees. Finally, the paper use a gradient boosting regressor, which sequentially fits trees to reduce residual error and can flexibly model nonlinearities.

2.3 Evaluation

Model accuracy is assessed using mean squared error (MSE) on the 2025 hold-out test set. MSE measures the average squared difference between predicted and observed unemployment rates, so lower values indicate better predictive accuracy while penalizing larger errors more heavily. Within the 2020–2024 training window, time-series cross-validation is also used to tune hyperparameters and to diagnose overfitting (large gaps between training error and validation error).

3 Results

3.1 Descriptive analysis (2020-2025)

National pattern (average across U.S. states). As shown in Figure 1, from 2020 to 2025, the average state unemployment rate displays a clear shock–recovery–stabilization pattern. In early 2020, unemployment is low (around 3.7–4%), followed by a sudden surge in mid-2020 to a peak close to 14%, consistent with a large, short-lived macroeconomic shock. After the spike, unemployment declines rapidly through late 2020 and continues trending downward into 2022, reaching a low range of roughly 3.3–3.6%. Beginning in 2023, the series shows a mild upward drift, rising gradually to about 4.0–4.1% by the end of 2025. Overall, the post-2022 period remains relatively stable compared with the extreme volatility observed in 2020.

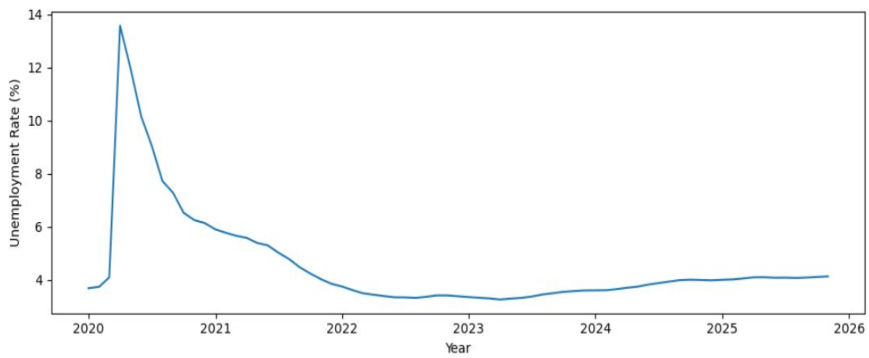

Fig. 1. Average U.S. state unemployment rate (quarterly average), 2020–2025. (Picture credit: Original)

Cross-state heterogeneity (peak unemployment rates). Figure 2 and Table 1 summarize peak unemployment rates across states over 2020–2025 and make the cross-state variation immediately clear. Extreme spikes are concentrated in only a handful of states rather than being broadly shared. Michigan (MI) and Hawaii (HI) record the highest peaks—about 22.7% and 22.5%, respectively—well above the national peak. A second tier of states (including IL, MA, IN, KY, CA, MS, FL, and AZ) reaches peak unemployment mostly in the mid- to high-teens. The wide spread in these peaks implies that the initial shock and its labor-market consequences differed markedly across states, plausibly due to differences in sectoral composition and the degree of exposure to industries most affected by disruptions. For this reason, MI and HI are used as focal examples in the forecasting exercise: both experienced large shock magnitudes and unusually high volatility.

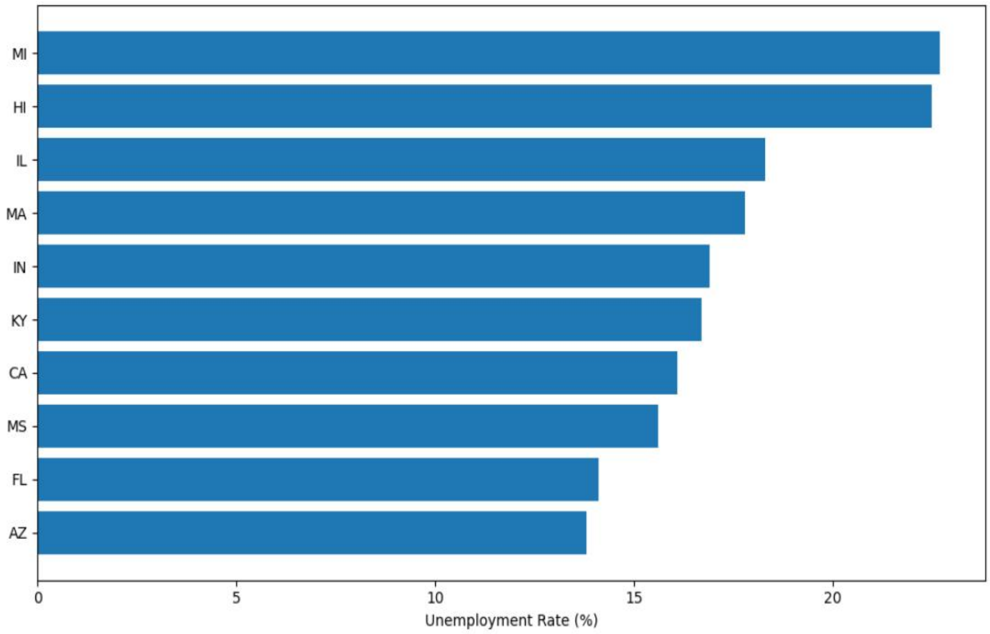

Fig. 2. Top 10 states by peak unemployment rate, 2020–2025. (Picture credit: Original)

Table 1. Summary statistics for the top 10 states by peak unemployment rate, 2020–2025.

State	Peak unemployment (%)	Average unemployment (%)	Volatility (std. dev.)	Unemployment end-2025 (%)
MI	22.7	5.647	3.262	5.0
HI	22.5	4.994	4.396	2.2
IL	18.3	5.720	2.497	4.4
MA	17.8	5.074	2.853	4.7
IN	16.9	4.319	2.213	3.7
KY	16.7	4.883	1.795	4.7
CA	16.1	6.234	2.672	5.5
MS	15.6	4.560	2.138	3.8
FL	14.1	4.324	2.404	4.2
AZ	13.8	4.673	1.924	4.3

3.2 GDP–Unemployment relationship (correlation and fixed-effects regression)

To characterize how labor-market conditions move with output, the analysis relates quarterly-average unemployment (urateq) to the level of real GDP (realgdp) and quarter-on-quarter GDP growth (gdpqoqpct). at the state level. For both Michigan and Hawaii, the correlations are negative, consistent with unemployment rising in weaker economic conditions. In Hawaii, the correlation between unemployment and the level of real GDP is −0.523. The inverse association is even stronger when GDP growth is used: the unemployment–growth correlation equals −0.883 for Hawaii and −0.813 for Michigan, indicating substantial countercyclical co-movement over the sample period. Correlation statistics for MI and HI are reported in Table 2, while Table 3 presents the fixed-effects regression estimates.

Building on these descriptive patterns, the paper then fits a parsimonious fixed-effects model to link GDP growth to unemployment, while absorbing time-invariant differences across states.

A simple fixed-effects specification is estimated.

$$urate_{q,s}=\alpha+\beta * gdp_{qoqpct,s}+\gamma * I_{s=MI}+\varepsilon_{q,s} \tag{1}$$

Equation (1) represents the state fixed-effects regression specification.

Table 2. Correlation between quarterly unemployment and GDP indicators (Michigan and Hawaii).

State	Corr($urate_q$, gdp_{qoqpct})	Corr($urate_q$, $real_{gdp}$)
HI	-0.883	-0.523
MI	-0.813	-0.384

Table 3. Fixed-effects regression results for Michigan and Hawaii.

Term	Estimate
Intercept	5.495
GDP growth (gdp_{qoqpct})	-0.483
State fixed effect (MI)	0.513
R^2	0.204

where the state fixed effect captures time-invariant differences between MI and HI. The estimated coefficient on GDP growth is (β = -0.483), implying that a 1 percentage-point increase in quarterly GDP growth is associated with approximately a 0.48 percentage-point decrease in the quarterly unemployment rate, holding constant state fixed effects.The estimated Michigan fixed effect is positive (around +0.513). Interpreted directly, this means that holding GDP growth constant, Michigan tends to exhibit an unemployment rate that is about 0.51 percentage points higher than Hawaii's on average. Model fit is modest: $R^2 \approx 0.204$, so GDP growth together with state fixed effects accounts for roughly one-fifth of the quarterly variation in unemployment. The remaining dispersion is likely driven by omitted influences such as differences in industrial mix, shifts in labor-force participation, policy interventions, and short-run disturbances or measurement noise. Taken together, the estimates align with the usual countercyclical pattern and justify treating GDP growth as a useful input for the forecasting exercise.

3.3 Forecasting 2026 and model comparison

Predictions for 2026 differ noticeably across specifications, largely because the models impose different functional forms and balance bias against variance in different ways. Table 4 summarizes performance using mean squared error, and Table 5 lists the implied 2026 forecasts for Michigan and Hawaii.

A practical limitation of the linear regression baseline is that it effectively projects a single, stable linear relationship from the historical sample forward—an approach that can perform poorly if the underlying relationship shifts or if the period contains structural breaks. Random forest and gradient boosting allow for nonlinearities and interactions, which can improve fit but may also overfit in short samples if the signal-to-noise ratio is low.

Table 4. Mean squared error (MSE) comparison for Michigan and Hawaii.

Model	MSE (cross-validation)	MSE (training)	MSE (test, 2025)
Linear regression	0.017	0.006	0.010
Random forest	0.469	0.023	0.051
Gradient boosting	0.431	0.000	0.047

Table 5. Forecasted 2026 unemployment rates (%) for Michigan and Hawaii by model.

State	Gradient boosting	Linear regression	Random forest
Hawaii	2.232	2.291	2.519
Michigan	5.043	4.604	5.220

Although ensemble methods are often reported to perform well in the literature [3,4], in this setting linear regression achieves the lowest validation and test MSE, while the ensemble models show weaker generalization (very low training error but higher validation and test error).

Using a ±0.25 pp threshold relative to the 2025 year-end value, Hawaii shows an 'Up' signal only under Random Forest, while Michigan shows a 'Down' signal under Linear Regression; the other model–state combinations are classified as 'Stable'.

4 Conclusion

This study develops a forecasting exercise for U.S. state unemployment rates in 2026 using a data-driven framework built on quarterly state panels. Rather than relying purely on mechanical trend extrapolation, the approach introduces state GDP growth as an economic signal and evaluates whether modern prediction methods add value in this setting. The results point to a broad post–COVID recovery at the national level, but they also underscore that labor-market conditions remain uneven across regions.

Model comparisons in the 2025 out-of-sample test indicate that the linear regression baseline is the most stable performer. By contrast, the ensemble methods fit the 2020–2025 training window extremely closely—reflected in very low in-sample error—yet deliver weaker test performance, suggesting overfitting and limited generalization in a short and unusual sample period.

Two constraints are central. First, the available window (2020–2025) is brief and dominated by an extraordinary shock, which restricts what can be learned about "normal" dynamics. Second, the predictor set is narrow. Extensions could add further macroeconomic and labor-market covariates—such as inflation, interest rates, vacancy measures, and sectoral employment composition—and compare performance against established time-series benchmarks (e.g., ARIMA, mixed-frequency specifications). With a longer historical span, it would also be feasible to test more flexible deep-learning models for longer-horizon forecasting.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Okun, A. M.: Potential GNP: Its Measurement and Significance. Proceedings of the Business and Economic Statistics Section, American Statistical Association. (1962)

[2] Kim, K.: Unemployment Dynamics Forecasting with Machine-Learning Regression Models. arXiv:2505.01933. (2025)

[3] Abd El-Aal, M. F.: Determinants of Egypt's Unemployment Rate with Machine Learning Algorithms. Data Science in Finance and Economics 4(3), 333–349. (2024)

[4] Katris, C.: Prediction of Unemployment Rates with Time Series and Machine Learning Techniques. Computational Economics 55(2), 673–706. (2020)

[5] Grybauskas, A., Pilinkienė, V., Lukauskas, M., Stundžienė, A., Bruneckienė, J.: Nowcasting Unemployment Using Neural Networks and Multi-Dimensional Google Trends Data. Economies 11(5), 130. (2023)

[6] Costa, E. A., Silva, M. E., Galvão, A. B.: Real-time Nowcasting the Monthly Unemployment Rates with Daily Google Trends Data. Socio-Economic Planning Sciences, 101963. (2024)

[7] Magazzino, C., Mele, M., Mutascu, M.: An Artificial Neural Network Experiment on the Prediction of the Unemployment Rate. Journal of Policy Modeling. (2024)

[8] Güler, M., Kabakçı, A., Koç, Ö., Eraslan, E., Derin, K. H., Ünlü, R., Türkan, Y. S., Namlı, E.: Forecasting of the Unemployment Rate in Turkey: Comparison of the Machine Learning Models. Sustainability 16(15), 6509. (2024)

[9] Sun, W., Wang, Y., Zhang, L., Chen, X. H., Hoang, Y. H.: Enhancing Economic Cycle Forecasting Based on Interpretable Machine Learning and News Narrative Sentiment. Technological Forecasting and Social Change 215, 124094. (2025)

[10] U.S. Bureau of Labor Statistics: Local Area Unemployment Statistics (LAUS). Retrieved from: https://www.bls.gov/lau/ (2026)

[11] U.S. Bureau of Economic Analysis: Real GDP by State. Retrieved from: https://www.bea.gov/data/gdp/gdp-state (2026)