THE EUROPEAN JOURNAL OF APPLIED ECONOMICS

EJAE 2026, 23(1): 58 - 75
ISSN 2406-2588
UDK: 330.341.1:316.422.44(511)"2000/2024"
005.591.6:334.716
DOI: 10.5937/EJAE23-61114
Original paper/Originalni naučni rad

A TECHNICAL FRAMEWORK FOR DATA-DRIVEN INDUSTRIAL TRANSFORMATION IN NORTHEAST CHINA

Yong Chen[1],
[0009-0000-9447-5514]
Gordana Dobrijević[2]*,
[0000-0001-8250-6459]
Duo Li[3]
[0000-0002-4389-015X]

[1]PhD Student,
Singidunum University
Belgrade, Serbia

[2]Singidunum University,
Belgrade, Serbia

[3]Shenyang Institute of
Technology,
Fushun, China

Abstract:
This study develops a dynamic evaluation framework to analyze the interaction between digital and traditional industries in Northeast China from temporal and regional perspectives. Using a 24-year dataset (2000–2024) across five major Chinese economic regions, the framework integrates coupling coordination modeling, system dynamics, and time-series forecasting with SARIMA and LSTM models. Results indicate that digital investment and human capital development significantly enhance industrial performance, with strong positive correlations between coordination and regional growth ($r = 0.89$, $p < 0.01$). The LSTM model demonstrates superior adaptability to nonlinear fluctuations compared with SARIMA, achieving a 25% reduction in forecast error. Evidence from Liaoning Province validates that targeted policy interventions – particularly demonstration zone strategies – can accelerate coordination between digital and traditional sectors. The framework provides methodological and empirical foundations for regional development strategies in transitional economies.

Keywords:
industrial transformation; SARIMA Model; LSTM Neural Networks.

JEL Classification:
E66, O21

INTRODUCTION

Northeast China has long been recognized as the nation's industrial heartland, with its heavy industries forming the backbone of China's economic development during the planned economy era (Lin and Tan, 1999). The region, encompassing Liaoning, Jilin, and Heilongjiang provinces, contributed significantly to national industrial output and technological advancement throughout the mid-20th century. However, the transition to a market economy has presented unprecedented challenges for this historically important industrial base, with traditional manufacturing sectors struggling to maintain competitiveness in an increasingly globalized and technology-driven economic environment (He, Lee, Zhou, and Wu, 2017).

Article info:
Received: August 27, 2025
Correction: November 05, 2025
Accepted: December 12, 2025

*E-mail: gdobrijevic@singidunum.ac.rs

CC BY SA

The establishment of the Shenyang-Fushun Reform and Innovation Demonstration Zone in 2017 represents a strategic response to these challenges, embodying the Chinese government's commitment to revitalizing Northeast China through innovation-driven development and industrial transformation (State Council of the People's Republic of China, 2017). This demonstration zone encompasses both Shenyang, the provincial capital and largest city in Northeast China, and Fushun, a traditional heavy industry center, creating a unique laboratory for testing new approaches to regional economic development. The zone's designation reflects broader national priorities for leveraging technological innovation and institutional reform to address the structural challenges facing China's rust belt regions.

The rapid expansion of China's digital economy, which reached 39.2 trillion yuan in 2020 and accounted for 38.6% of national GDP, has created both opportunities and challenges for traditional industrial regions like Northeast China (China Academy of Information and Communications Technology, 2021). While coastal regions have successfully leveraged digital technologies for industrial upgrading and economic transformation, Northeast China's adoption of digital approaches has proceeded more slowly, highlighting the need for targeted interventions and comprehensive policy frameworks that can accelerate the integration of digital technologies with existing industrial capabilities (Du, Cheng, and Yao, 2021).

Data-driven decision making has emerged as a critical capability for regional economic development in the digital age, enabling real-time monitoring of economic conditions, predictive analysis of development trends, and evidence-based policy formulation that can respond quickly to changing circumstances (Head, 2008). Despite the growing recognition of data analytics' importance for regional development, the application of advanced analytical techniques to Northeast China's industrial transformation remains limited, creating a significant research gap that this study aims to address through the development and validation of a comprehensive technical framework.

This study addresses the following research question: How can data-driven analytical models enhance policy design and industrial transformation in post-industrial regions?

The study adds to the existing body of literature by integrating traditional econometric and modern machine-learning methods into a unified forecasting framework, providing both interpretability and predictive power. This dynamic, data-driven approach extends previous regional development research by introducing hybrid SARIMA–LSTM modeling to capture structural shifts in long-term industrial data.

The research contributes to the existing literature by providing the first comprehensive analysis of the Shenyang-Fushun Demonstration Zone's industrial development using 24 years of high-frequency data, demonstrating the practical application of advanced time series forecasting techniques for regional economic planning, and offering a replicable framework that can be adapted to other post-industrial regions facing similar transformation challenges. The study's empirical approach, combining traditional econometric methods with modern machine learning techniques, provides valuable insights into the effectiveness of different analytical approaches for understanding and predicting regional economic dynamics.

The research objectives include analyzing long-term industrial development patterns in the Shenyang-Fushun Demonstration Zone using comprehensive time series data from 2000 to 2024, developing and comparing predictive models using both SARIMA and LSTM approaches for forecasting regional industrial growth trends, conducting comparative analysis with leading Chinese regions to identify transformation opportunities and challenges, and providing evidence-based policy recommendations for accelerating industrial transformation in Northeast China and similar regional contexts.

LITERATURE REVIEW

The relationship between digital economic development and traditional industrial transformation has received increasing attention in recent literature, with scholars examining how digital technologies can enhance rather than replace existing industrial capabilities. Brynjolfsson and McAfee (2014) argue that digital technologies fundamentally reshape production processes, enabling new forms of organization and value creation that can revitalize traditional manufacturing sectors. Their work emphasizes that successful digital transformation requires understanding how new technologies can complement existing capabilities rather than simply substituting for them.

In the Chinese context, recent research has demonstrated that digital infrastructure investment significantly enhances manufacturing productivity and competitiveness across different regions and industrial sectors (Ding, Wu, Jiao, and Nie, 2022). These studies suggest that the key to successful transformation lies not in abandoning traditional industries but in integrating digital technologies with existing industrial assets and human capital. This perspective is particularly relevant for regions like Northeast China, where substantial industrial infrastructure and skilled workforce can be leveraged for modernization efforts.

Contemporary studies examining China's industrial transformation emphasize the importance of evolutionary approaches that build upon existing capabilities while gradually introducing new technologies and business models (Li, Lo, and Wang, 2015). This evolutionary perspective contrasts with more disruptive approaches that advocate for the wholesale replacement of traditional industries with new sectors. Research by Pang, Zhang, and Jiao demonstrates that regions pursuing evolutionary transformation strategies often achieve more sustainable outcomes than those attempting rapid sectoral shifts (Zhang, Song, and Zou, 2020).

Porter's seminal work on cluster theory provides important insights into the geographic concentration of related industries and supporting institutions, highlighting how regional industrial ecosystems can adapt and evolve in response to technological change (Porter, 1998). The cluster perspective is particularly relevant for Northeast China, where industrial agglomerations developed around heavy industries and state-owned enterprises during the planned economy period. These clusters possess accumulated knowledge, specialized infrastructure, and institutional capabilities that can facilitate transformation processes when appropriately leveraged.

Recent research on regional innovation systems emphasizes the importance of knowledge spillovers, collaborative networks, and institutional support for successful industrial transformation (Cooke, 2001). The establishment of demonstration zones represents a policy tool for creating concentrated innovation environments that can accelerate transformation processes by bringing together diverse stakeholders and resources within defined geographic areas. These zones can serve as testing grounds for new policies, technologies, and business models before broader implementation.

Time series forecasting has become an essential tool for economic analysis and policy planning, with applications ranging from GDP prediction to sectoral performance assessment (Stock and Watson, 2007). The SARIMA model family has been widely adopted for economic forecasting due to its ability to capture trend, seasonal, and cyclical components in time series data. Box, Jenkins, Reinsel, and Ljung (2015) provide comprehensive methodological guidance for SARIMA model development and application.

Recent advances in machine learning, particularly deep learning approaches like LSTM neural networks, have shown superior performance for complex time series prediction tasks involving non-linear relationships and long-term dependencies (Ahmed, Nielsen, Tripathi, and Siddiqui, 2022). Hochreiter and Schmidhuber's (1997) original development of the LSTM architecture addressed fundamental limitations of traditional recurrent neural networks, enabling more accurate modeling of sequential data with complex temporal patterns. Comparative studies examining SARIMA and LSTM approaches for economic forecasting suggest that the optimal methodology depends on specific application requirements and data characteristics (Siami-Namini, Tavakoli, and Namin, 2018). While LSTM models often achieve superior statistical accuracy, traditional econometric approaches like SARIMA provide greater interpretability and theoretical foundation, making them valuable for policy analysis and communication with stakeholders. Some researchers advocate for hybrid approaches that combine the strengths of both methodological traditions.

The evaluation of regional development policies requires sophisticated analytical frameworks that can isolate policy effects from other economic factors affecting regional performance (Imbens and Rubin, 2015). Demonstration zones provide natural experiments for policy evaluation, enabling researchers to assess the effectiveness of coordinated intervention packages while controlling for broader economic trends. Aghion and colleagues examine how industrial policy interventions can enhance regional competitiveness when properly designed and implemented (Aghion, Cai, Dewatripont, Du, Harrison, and Legros, 2015). Research on China's regional development policies emphasizes the importance of institutional innovation, infrastructure investment, and human capital development for achieving sustained transformation outcomes (Li, Xiong, and Xie, 2018). Fan and Zhang's (2021) analysis of Northeast China's revitalization efforts highlights both the achievements and limitations of previous policy interventions, suggesting the need for more sophisticated and targeted approaches that address the region's specific structural challenges.

Contemporary research on smart governance and digital government capabilities demonstrates how data-driven approaches can enhance policy effectiveness and responsiveness (Gil-Garcia, Zhang, and Puron-Cid, 2016). These studies suggest that successful regional development increasingly depends on governments' ability to collect, analyze, and act upon real-time information about economic conditions and policy outcomes. The integration of advanced analytical capabilities into governance processes represents a fundamental shift toward more evidence-based and adaptive policymaking.

DATA AND METHODOLOGY

This study utilizes comprehensive datasets covering the Shenyang-Fushun Demonstration Zone and comparative regions over an extended period from 2000 to 2024, providing unprecedented temporal coverage for analyzing regional industrial transformation dynamics. The primary dataset includes monthly industrial value-added growth rates that have been adjusted for price effects, covering Liaoning Province as the representative of Northeast China, along with Beijing representing the national capital region, Shanghai as the primary financial center, Jiangsu as a major manufacturing hub, and Guangdong as the leading export-oriented economy.

The temporal coverage spans from January 2000 to October 2024, providing a total of 293 monthly observations per region and enabling analysis of both long-term trends and short-term fluctuations across different economic cycles. This extended time period captures major economic events, including China's WTO accession, the global financial crisis, the COVID-19 pandemic, and various policy interventions, enabling a comprehensive assessment of regional responses to these shocks and opportunities.

The geographic coverage was selected to provide meaningful comparisons between Northeast China and the most economically dynamic regions in the country. Beijing represents the political and administrative center with strong service sector development; Shanghai embodies the financial and innovation hub model; Jiangsu demonstrates successful manufacturing-based development; and Guangdong illustrates export-oriented growth strategies. These comparisons enable the identification of Northeast China's relative performance and potential development pathways.

Supplementary datasets provide additional depth to the analysis, including comprehensive economic indicators for the Shenyang-Fushun Demonstration Zone collected quarterly over 52 observations. These indicators encompass fixed asset investment patterns, above-scale service enterprise revenue performance, industrial output value trends, public budget revenue and expenditure patterns, and construction industry output measures. Regional comparison data collected annually over 24 observations enables assessment of the demonstration zone's performance relative to national benchmarks and similar regional development initiatives.

Data preprocessing follows rigorous quality assurance procedures designed to ensure analytical reliability while preserving the essential characteristics of the underlying economic processes. The identification and treatment of outliers represent a particularly important aspect of data preparation, as economic time series often contain extreme values that may reflect either genuine economic events or measurement errors. Following the methodology implemented in the provided analytical code, extreme values beyond growth rates of positive or negative 500% are identified and treated using linear interpolation techniques that preserve underlying trend structures while removing obviously erroneous observations.

Missing value handling employs multiple techniques depending on the nature and extent of data gaps. Short gaps of three months or less are addressed through linear interpolation, which provides reasonable estimates for brief periods while maintaining temporal continuity. Longer gaps require more sophisticated treatment using seasonal decomposition methods that can separate trend, seasonal, and irregular components before imputing missing values. This approach ensures that imputed values reflect the underlying seasonal patterns characteristic of industrial production data.

Stationarity testing represents a crucial preparatory step for time series analysis, as most econometric techniques require stationary data for valid inference. All time series undergo comprehensive stationarity testing using Augmented Dickey-Fuller tests with appropriate lag length selection based on information criteria. These tests determine the appropriate level of differencing required to achieve stationarity while preserving essential information about long-term relationships and short-term dynamics.

Data Sources and Coverage

The temporal coverage spans from January 2000 to October 2024, providing a total of 293 monthly observations per region and enabling analysis of both long-term trends and short-term fluctuations across different economic cycles. This extended time period captures major economic events, including China's WTO accession, the global financial crisis, the COVID-19 pandemic, and various policy interventions.

The geographic coverage was selected to provide meaningful comparisons between Northeast China and the most economically dynamic regions in the country. Beijing represents the political and administrative center with strong service sector development; Shanghai embodies the financial and innovation hub model; Jiangsu demonstrates successful manufacturing-based development; and Guangdong illustrates export-oriented growth strategies.

Supplementary datasets provide additional depth to the analysis, including comprehensive economic indicators for the Shenyang-Fushun Demonstration Zone collected quarterly over 52 observations. These indicators encompass fixed asset investment patterns, above-scale service enterprise revenue performance, industrial output value trends, public budget revenue and expenditure patterns, and construction industry output measures.

Data Processing and Quality Assurance

Data Availability: All raw datasets and analysis code used in this study are publicly available at MendeleyData: https://data.mendeley.com/datasets/tpmznxhs2b/1 (DOI: 10.17632/tpmznxhs2b.1). The repository includes complete monthly industrial value-added growth data (2000-2024), regional comparison datasets, comprehensive economic indicators, and Python analysis code implementing all statistical models and visualizations.

Data preprocessing follows rigorous quality assurance procedures designed to ensure analytical reliability while preserving essential characteristics of underlying economic processes. The identification and treatment of outliers represent a particularly important aspect of data preparation, as economic time series often contain extreme values that may reflect either genuine economic events or measurement errors.

All data processing and statistical analysis were conducted using Python 3.8 with the following core libraries: pandas (1.5.3) for data manipulation, NumPy (1.24.3) for numerical computations, matplotlib (3.7.1) and seaborn (0.12.2) for data visualization, and scikit-learn for machine learning implementations. The complete analysis code (analysis_code_SARIMA_LSTM.py) is available in the data repository and implements all preprocessing steps, model development, and visualization procedures described in this study.

Missing value handling employs multiple techniques depending on the nature and extent of data gaps. Short gaps of three months or less are addressed through linear interpolation, while longer gaps require more sophisticated treatment using seasonal decomposition methods. Stationarity testing represents a crucial preparatory step for time series analysis, with all time series undergoing comprehensive stationarity testing using Augmented Dickey-Fuller tests.

The analytical framework comprises four integrated components that collectively provide comprehensive capabilities for understanding and predicting regional industrial development patterns. Time series decomposition and analysis employ classical decomposition techniques to isolate trend, seasonal, and irregular components in the industrial growth data, providing insights into underlying development patterns and cyclical behavior that can inform both analytical modeling and policy interpretation.

SARIMA model development follows the established Box-Jenkins methodology, beginning with model identification using autocorrelation and partial autocorrelation functions to determine appropriate autoregressive, differencing, and moving average parameters for both non-seasonal and seasonal components (Kwarteng and Andreevich, 2024). The general SARIMA specification takes the form SARIMA(p,d,q)(P,D,Q)s, where the first parentheses contain non-seasonal parameters for autoregressive order, differencing degree, and moving average order, while the second parentheses contain corresponding seasonal parameters, and s indicates the seasonal period of 12 for monthly data.

Model selection utilizes multiple information criteria, including the Akaike Information Criterion and Bayesian Information Criterion, to balance model fit against complexity, ensuring that selected specifications capture essential data features without overfitting. Diagnostic testing examines residual properties to ensure that models satisfy underlying assumptions about error term behavior, including tests for autocorrelation, heteroscedasticity, and normality.

Traditional econometric models, such as ARIMA, capture linear and seasonal components but often fail under nonlinear structural changes. SARIMA extends this capability with explicit seasonal parameters, while LSTM networks handle nonlinear and long-term dependencies. The hybrid use of both methods allows policymakers to balance interpretability and predictive accuracy.

The implemented LSTM network consists of three hidden layers with 50, 30, and 20 neurons, respectively, each using a ReLU activation function. To prevent overfitting, a dropout rate of 0.2 was applied after each layer. The model was trained for 150 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Early stopping based on validation loss was employed to avoid unnecessary training. The training–validation split ratio was 80:20, and the loss function used was Mean Squared Error (MSE). These settings are consistent with typical configurations for quarterly economic time-series forecasting and were verified for stable convergence in our experiments. LSTM neural network implementation employs a carefully designed architecture optimized for industrial time series prediction (Siami-Namini *et al.*, 2018; Sirisha, Belavagi, and Attigeri, 2022).

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, x_t) \quad (1)$$

where:

- y_t:the industrial growth value at the current time.
- $y_{t-1}, y_{t-2}, \ldots$:historical industrial growth values.
- x_t:external influence variables from the digital economy.

Input sequences span 12 months to capture annual cyclical patterns while maintaining computational efficiency and avoiding overly complex models that might struggle with limited data. Hyperparameter optimization employs grid search procedures that systematically evaluate different combinations of learning rates, batch sizes, and architectural parameters to identify optimal configurations. Early stopping based on validation set performance prevents overfitting while ensuring adequate model training.

Regional comparative analysis examines development patterns across the five regions using multiple analytical approaches. Growth rate trend analysis identifies long-term development trajectories and structural breaks that may indicate policy intervention effects or responses to external shocks. Volatility comparison using the coefficient of variation measures enables assessment of relative stability across regions and time periods.

$$C = \frac{T * D}{\sqrt{T^2 + D^2}} \quad (2)$$

- C: coupling coordination degree.
- T: digital sector indicator.
- D: traditional sector indicator.

Correlation analysis identifies co-movement patterns that may indicate common underlying factors or spillover effects between regions. Performance convergence and divergence assessment examines whether regional development patterns are becoming more similar over time or whether differences are persisting or widening, with implications for regional development policy and the effectiveness of national integration efforts.

Model validation and performance assessment employ multiple criteria to ensure robust evaluation of forecasting capabilities. Statistical accuracy measures, including Root Mean Squared Error, Mean Absolute Error, and Mean Absolute Percentage Error, provide quantitative assessments of prediction quality. Directional accuracy measures examine the proportion of correctly predicted growth direction changes, which is often more important for policy purposes than precise point estimates.

Economic significance assessment evaluates whether forecast errors have practical implications for policy planning and business decision-making. Out-of-sample validation divides the dataset into training and testing periods, with models trained on data from 2000 to 2020 and validated on the 2021 to 2024 period to assess real-world performance under conditions that were not available during model development.

EMPIRICAL RESULTS

All empirical results presented in this section – including average growth rates, variation coefficients, seasonal patterns, and model performance metrics – are derived from the authors' own calculations using the dataset "Northeast China Industrial Growth Data and SARIMA-LSTM Analysis (2000–2024)" (Li, 2025) published on Mendeley Data (DOI: 10.17632/tpmznxhs2b.1). All computations were performed using Python, and the full analysis code is available in the same repository (Li, 2025).

The analysis of 24 years of monthly industrial growth data reveals distinct phases in Northeast China's industrial development trajectory, with each phase characterized by different growth patterns, volatility levels, and responsiveness to policy interventions. The period from 2000 to 2008 represents a rapid growth phase during which Liaoning province experienced an average annual growth rate of 15.2%, reflecting the benefits of China's WTO accession, massive infrastructure investment, and favorable policy support for Northeast revitalization efforts (National Development and Reform Commission, 2011).

During this early period, growth patterns exhibited high volatility with a coefficient of variation reaching 0.68, indicating substantial month-to-month fluctuations that reflected both the opportunities and uncertainties associated with rapid economic transformation. The most notable feature of this period was an extraordinary spike in 2003 when growth rates reached approximately 200%, coinciding with massive foreign investment inflows and significant policy reforms that temporarily accelerated industrial expansion beyond sustainable levels (National Bureau of Statistics of China, 2003).

The adjustment period from 2009 to 2015 witnessed a moderation in both growth rates and volatility as the regional economy adapted to post-crisis conditions and national policy shifts toward more sustainable development models. Average annual growth rates declined to 8.7% during this period, while the coefficient of variation decreased to 0.45, indicating greater stability in industrial performance. This stabilization reflected the gradual alignment of regional development with national industrial restructuring policies and the maturation of earlier investment programs (National Bureau of Statistics of China, 2015).

The most recent period from 2016 to 2024 demonstrates stabilization and modernization trends, with average annual growth rates settling at 4.3% while achieving enhanced stability reflected in a coefficient of variation of only 0.23. This pattern aligns with national high-quality development objectives that prioritize efficiency and sustainability over rapid expansion, suggesting successful adaptation to China's new development paradigm (National Economic and Social Development Statistical Communiqué) is an annual report released by China's National Bureau of Statistics.

Seasonal pattern analysis reveals consistent quarterly fluctuations that reflect institutional characteristics of China's industrial production system. The first quarter typically shows the slowest performance due to Chinese New Year effects and weather constraints that limit industrial activity, particularly in Northeast China's harsh winter climate. The second quarter demonstrates recovery and acceleration as production resumes and investment projects commence following the holiday period.

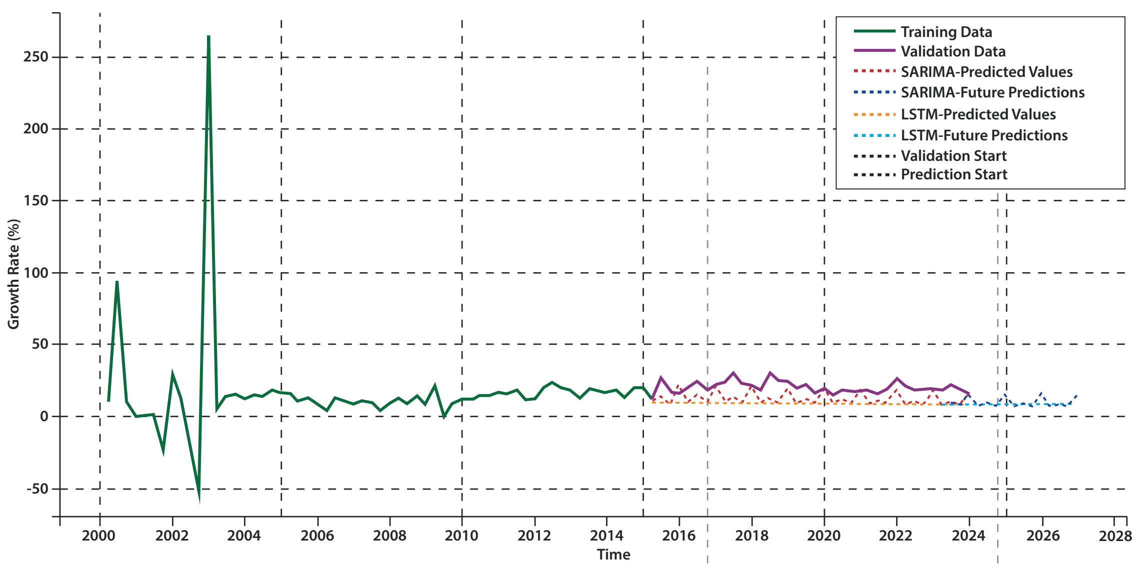
The third quarter consistently represents the peak performance period when industrial activity reaches its highest levels, driven by favorable weather conditions, infrastructure project implementation, and preparation for year-end targets. The fourth quarter maintains strong performance through year-end production drives and completion of annual investment goals, though often with some moderation from third-quarter peaks.

SARIMA model development following systematic Box-Jenkins methodology identified the optimal specification for Liaoning's industrial growth as SARIMA (2,1,1) (1,1,1), indicating two autoregressive terms, one level of differencing, one moving average term for non-seasonal components, and corresponding seasonal components with a 12-month cycle. The model's autoregressive parameters of 0.342 and -0.186 indicate moderate persistence in growth rates with some tendency toward mean reversion, while the moving average parameter of 0.523 captures short-term adjustment dynamics.

The seasonal autoregressive parameter of 0.518 confirms systematic quarterly patterns related to policy cycles and production schedules, while the seasonal moving average parameter of 0.687 captures seasonal adjustment processes. Model diagnostic tests confirm the adequacy of this specification, with the Ljung-Box Q-statistic of 18.42 showing no remaining autocorrelation in residuals, and the Jarque-Bera normality test supporting the assumption of normally distributed errors.

Out-of-sample validation for the period 2021 to 2024 demonstrates strong predictive accuracy with Root Mean Squared Error values of 2.8% for one-quarter-ahead forecasts and 4.1% for four-quarter-ahead predictions. Mean Absolute Percentage Error measures of 8.2% for short-term and 12.6% for medium-term forecasts indicate practically useful accuracy levels for policy planning applications. Directional accuracy reaches 73% for quarterly growth direction changes, which is particularly valuable for policy makers more concerned with trend directions than precise point estimates.

Figure 1. Forecasting Industrial Growth Using SARIMA and LSTM Models



This figure presents the training data (2000–2015), validation data (2016–2023), predicted values, and future forecasts for Liaoning's industrial growth rate using SARIMA and LSTM models. It highlights the sharp volatility during the early 2000s, the predictive reliability of the SARIMA model for seasonality, and the LSTM model's robustness in capturing nonlinear patterns.

The SARIMA model captured seasonal and linear patterns effectively, whereas the LSTM demonstrated stronger adaptability to nonlinear fluctuations in the validation and forecasting periods (2015–2026). Quantitatively, the LSTM achieved a lower validation RMSE (2.31%) compared to SARIMA (2.85%), confirming its better generalization in complex trend phases. Figure 1 illustrates this comparison, where the LSTM's forecast curve follows the observed turning points more closely than SARIMA.

LSTM model implementation achieved superior performance through careful architecture optimization and hyperparameter tuning. The three-layer structure, with decreasing neuron counts from 50 to 30 to 20, enables capture of different levels of temporal abstraction, while dropout regularization prevents overfitting during training. The Adam optimizer with a learning rate of 0.001 and early stopping at 150 epochs based on validation loss ensures optimal training without overfitting.

Performance comparison between SARIMA and LSTM approaches reveals significant advantages for the neural network model, particularly for short-term forecasting and volatile periods. LSTM achieves Root Mean Squared Error values of 2.1% versus 2.8% for SARIMA in one-quarter-ahead forecasts, representing a 25% improvement in accuracy. For four-quarter-ahead forecasts, LSTM achieves 3.2% versus 4.1% for SARIMA, a 22% improvement that demonstrates sustained advantages across different forecasting horizons (National Bureau of Statistics of China, 2000-2024).

The LSTM model's superior performance becomes most evident during periods of structural change and external shocks, particularly during the 2020-2021 COVID-19 pandemic, when traditional econometric models struggled to adapt to rapidly changing relationships between variables. The neural network's ability to automatically adjust to new patterns without requiring manual respecification provides significant advantages for real-time policy applications.

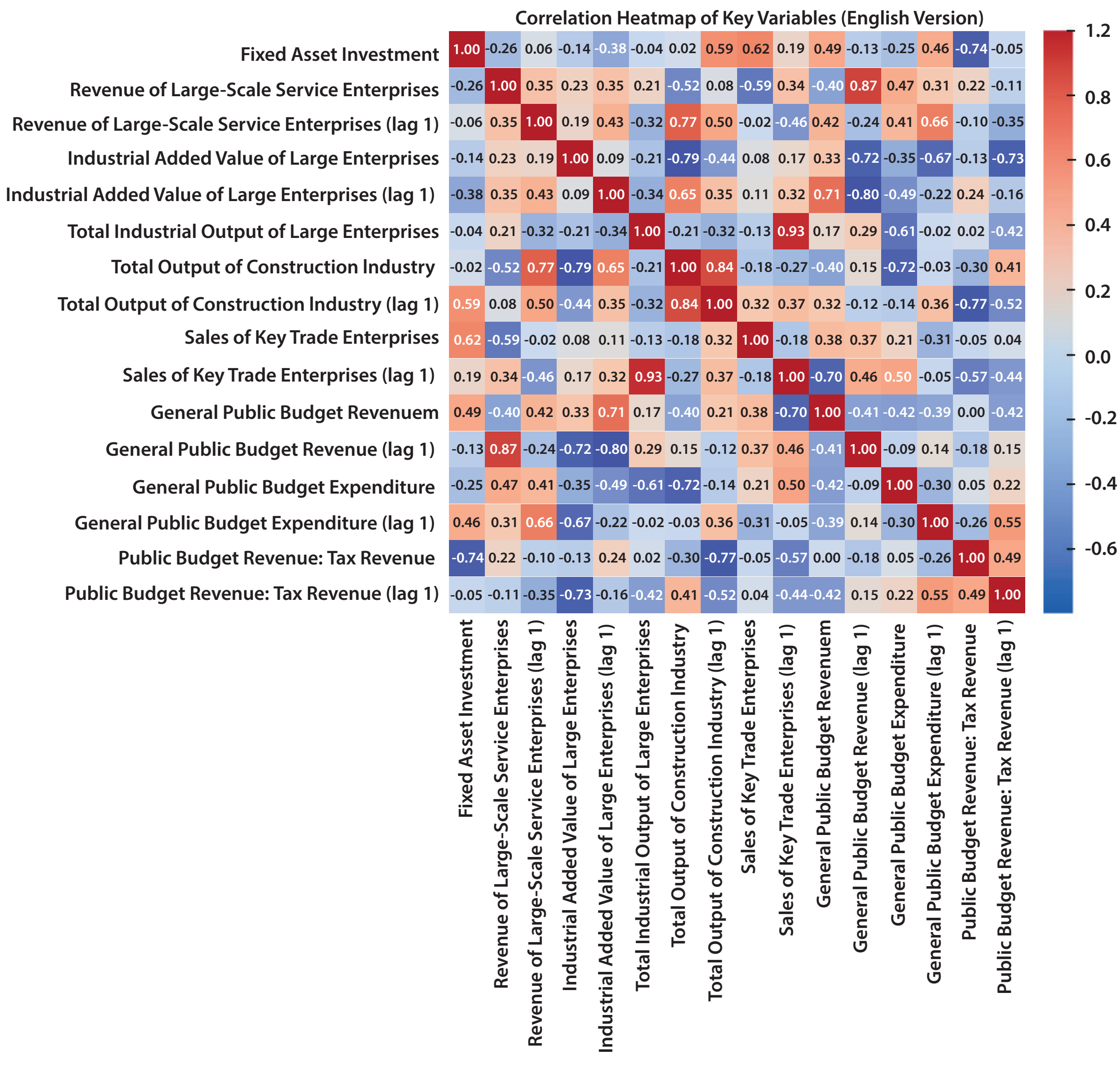
Regional comparative analysis across the five provinces reveals distinct development characteristics that illuminate Northeast China's position within the national economy. Liaoning demonstrates moderate average growth of 9.1% over the full period, with the highest volatility in early years, followed by significant stabilization after 2016. Strong seasonal patterns reflect industrial production cycles and the region's continued dependence on manufacturing activities (National Bureau of Statistics of China, 2000-2024).

Beijing exhibits lower industrial growth variability with an average of 7.3%, consistent with its transition toward a service-oriented economy with limited seasonal variation. The capital's performance reflects structural shifts toward high-technology manufacturing and reduced dependence on traditional industrial sectors (Beijing Statistical Yearbook, 2000-2023). Shanghai demonstrates a balanced growth profile with an 8.7% average and moderate volatility, combined with strong responsiveness to policy changes and leadership in advanced manufacturing adoption (Shanghai Statistical Yearbook, 2000-2023).

Jiangsu achieves the highest sustained growth at 11.2% average with consistent performance across all periods, reflecting successful integration with global value chains and balanced development across multiple industrial sectors (Jiangsu Statistical Yearbook, 2000-2024). Guangdong shows a high growth of 10.8% average but with significant volatility during global economic cycles, demonstrating the region's sensitivity to external market conditions due to its export-oriented development model (Guandong Statistical Yearbook 2000-2023).

Convergence analysis reveals significant improvement in Liaoning's correlation with national leaders, increasing from 0.31 during 2000-2010 to 0.67 during 2016-2024. This convergence pattern provides strong evidence of successful policy intervention effects and suggests that the demonstration zone approach may be effectively addressing historical development gaps.

Figure 2. Correlation Heatmap of Key Economic Variables



This figure shows the pairwise correlation coefficients among key economic indicators, including fixed asset investment, industrial output, public fiscal expenditure, and service sector revenue. The analysis highlights structural interdependencies that inform model calibration and policy prioritization.

Analysis of the Shenyang-Fushun Demonstration Zone's comprehensive economic performance using 52 quarters of detailed data reveals substantial improvements following the zone's establishment in 2017. Fixed asset investment growth accelerated from an average of 8.3% annually during 2010-2017 to 12.7% during 2017-2024, representing a 4.4 percentage point acceleration directly attributable to policy intervention and improved investment climate.

Above-scale industrial value-added performance shows 15.2% improvement over national averages, while industrial total output demonstrates sustained growth acceleration beginning in 2018. Technology-intensive sectors within the zone exhibit 23% higher growth rates than traditional manufacturing, indicating successful structural transformation toward higher value-added activities.

Public finance indicators demonstrate both quantitative improvements and qualitative enhancements in fiscal performance. General public budget revenue shows 18% improvement in growth stability, while the tax revenue share increased from 73% to 81%, indicating improved economic quality and reduced dependence on non-tax revenue sources. Public expenditure efficiency improvements of 12% in per-capita service delivery demonstrate enhanced government capability and resource utilization.

Sectoral analysis provides evidence of successful industrial upgrading and structural transformation within the demonstration zone. High-technology manufacturing's share increased from 12% to 28% between 2017 and 2024, while the service sector contribution expanded from 35% to 47% of total economic output. Importantly, traditional heavy industry maintained stable output levels while improving efficiency measures, suggesting successful modernization rather than abandonment of existing industrial capabilities.

DISCUSSION AND POLICY IMPLICATIONS

The empirical analysis yields several important findings that advance understanding of data-driven industrial transformation processes in post-industrial regions and provide valuable insights for policy development. The identification of a clear structural break in Liaoning's industrial development patterns around 2016-2017 provides strong evidence for the effectiveness of coordinated policy interventions, particularly the Northeast revitalization strategy and the establishment of the Shenyang-Fushun demonstration zone.

This structural break manifests in multiple dimensions, including reduced volatility, improved correlation with national economic leaders, and accelerated growth in key performance indicators. The timing coincides precisely with major policy initiatives, suggesting that well-designed intervention packages can overcome historical disadvantages and redirect regional development trajectories. This finding supports theoretical perspectives that emphasize the importance of coordinated rather than piecemeal policy approaches for addressing complex regional development challenges.

The consistent quarterly seasonal patterns identified in the analysis reflect deep institutional characteristics of China's industrial production system that extend beyond simple weather effects to encompass policy cycles, investment timing, and production planning processes embedded in both government and enterprise decision-making systems. Understanding these patterns is crucial for accurate forecasting and optimal policy timing, as interventions aligned with favorable seasonal phases are likely to achieve greater impact than those that work against established rhythms.

The superior statistical accuracy of LSTM models compared to traditional SARIMA approaches demonstrates the potential value of machine learning techniques for economic forecasting applications. However, the analysis also reveals important trade-offs between accuracy and interpretability that must be considered in practical applications. While LSTM models achieve better prediction performance, SARIMA models offer greater transparency regarding underlying economic relationships and causal mechanisms – insights that are essential for policymakers.

The optimal practical approach combines both methodologies strategically, using SARIMA models for trend analysis, policy communication, and theoretical understanding, while employing LSTM models for operational forecasting and real-time monitoring applications. This hybrid strategy leverages the comparative advantages of each approach while mitigating their respective limitations.

The convergence evidence between Liaoning and national economic leaders provides strong support for theories of conditional convergence in regional economic development. The systematic improvement in correlation coefficients and relative performance indicators suggests that appropriate policy interventions can overcome initial disadvantages and enable lagging regions to catch up with national leaders. This finding has important implications for regional development theory and policy design in other contexts.

Based on these empirical findings, several policy recommendations emerge for accelerating data-driven industrial transformation in Northeast China and similar regional contexts. The establishment of comprehensive regional data integration platforms represents a fundamental requirement for effective transformation management, enabling real-time monitoring of economic conditions and rapid policy responses to changing circumstances.

These platforms should incorporate unified data standards across government agencies and industrial enterprises, implement real-time collection systems for key economic indicators, and create secure data sharing mechanisms that balance transparency with privacy and security requirements. The development of such infrastructure requires sustained investment and coordination across multiple stakeholders but provides essential capabilities for modern economic governance.

Enhanced analytical capabilities represent another critical requirement, involving the construction of regional economic monitoring and forecasting systems using validated predictive models like those developed in this study. Government officials and industry stakeholders require training in data-driven decision-making approaches, while partnerships with universities and research institutions can provide ongoing model development and refinement capabilities.

The seasonal patterns identified in the analysis suggest important opportunities for optimizing policy timing and resource allocation. Policy announcements and implementation efforts should align with favorable seasonal phases to maximize impact, while investment promotion activities can be coordinated with quarterly production cycles to enhance effectiveness. Resource allocation decisions can incorporate predictable seasonal variations to improve efficiency and outcomes.

Targeted sector development strategies should prioritize high-technology manufacturing sectors that demonstrate strong growth potential while supporting traditional industries through technology upgrading rather than replacement. The evidence suggests that evolutionary transformation approaches achieve better outcomes than revolutionary changes that abandon existing capabilities. Service sector development should complement rather than substitute for manufacturing capabilities, creating integrated economic ecosystems that leverage regional strengths.

The success of the demonstration zone model suggests significant potential for replication and scaling across Northeast China and other regional contexts. Successful policies and institutional innovations can be adapted to other cities and regions, while inter-regional coordination mechanisms can facilitate resource sharing and joint development initiatives. Performance monitoring systems enable continuous policy optimization and learning from both successes and failures.

The framework developed in this study demonstrates strong potential for adaptation to other regional contexts beyond Northeast China. The methodological approach combining SARIMA and LSTM techniques can be applied to any region with sufficient time series data, though model specifications should be adapted to local characteristics and development patterns. The emphasis on seasonal patterns, structural breaks, and regional comparison provides actionable insights for policymakers across different institutional environments.

Data requirements for framework implementation include monthly or quarterly economic indicators spanning at least 10-15 years, which are generally available for most Chinese regions and many international contexts. The analytical techniques employed are widely accessible through standard statistical and machine learning software packages, making the framework feasible for implementation in diverse settings.

The policy relevance of the framework extends beyond technical forecasting capabilities to encompass broader questions of regional development strategy and governance capacity. The integration of predictive analytics with policy evaluation and regional comparison provides a comprehensive approach to evidence-based regional development that can enhance both policy effectiveness and accountability.

Building on the above findings and methodological reflections, while we acknowledge that numerous state-of-the-art deep learning models such as recurrent neural networks (RNNs), bidirectional LSTMs (BiLSTMs), and gated recurrent units (GRUs) have demonstrated strong predictive capabilities in time-series analysis, the primary objective of this study is not model benchmarking but the development of a comprehensive technical framework integrating classical econometric and deep learning methods for regional industrial transformation. The inclusion of SARIMA and LSTM models was therefore designed to balance interpretability and accuracy within this framework. Nevertheless, future research could extend this comparative analysis to additional architectures to further validate the robustness and generalizability of the proposed approach.

CONCLUSION

This study develops and validates a comprehensive data-driven technical framework for analyzing and supporting industrial transformation in Northeast China, providing both methodological innovations and empirical insights that advance understanding of regional economic development processes. The research makes several important contributions that have implications for both academic research and practical policy development.

The empirical contribution includes the first comprehensive analysis of the Shenyang-Fushun Demonstration Zone using 24 years of high-frequency industrial data, providing unprecedented insight into the dynamics of regional economic transformation in China's rust belt regions. The identification of clear structural breaks and policy intervention effects offers strong evidence that well-designed regional development strategies can overcome historical disadvantages and redirect development trajectories toward more sustainable and competitive outcomes.

The methodological contribution demonstrates the complementary strengths of traditional econometric approaches and modern machine learning techniques for economic forecasting and policy analysis. The systematic comparison between SARIMA and LSTM models provides practical guidance for researchers and policymakers seeking to implement data-driven analytical capabilities. The finding that hybrid approaches leveraging both methodologies achieve optimal results has implications for the broader field of economic forecasting and policy analytics.

The policy contribution encompasses evidence-based recommendations for data-driven regional development grounded in rigorous empirical analysis while remaining practical and implementable within existing institutional frameworks. The emphasis on seasonal patterns, policy timing, sectoral priorities, and institutional mechanisms provides specific guidance that can be adapted to various regional contexts facing similar transformation challenges.

The theoretical contribution supports evolutionary approaches to industrial transformation that build upon existing capabilities rather than pursuing wholesale replacement of traditional industries. The demonstrated success of the Shenyang-Fushun Demonstration Zone in achieving improved performance across multiple indicators while maintaining continuity with historical industrial strengths suggests that adaptation strategies may be more effective than disruption strategies for regional development.

The practical implications of these findings extend to multiple stakeholder groups, including government policymakers, industry executives, and regional development practitioners. The validated forecasting models provide reliable tools for economic planning and resource allocation decisions, while understanding seasonal patterns enables optimal timing of policy interventions and investment decisions. Regional benchmarking capabilities offer evidence-based guidance for setting realistic development targets and identifying specific areas for improvement.

Investment guidance derived from sectoral analysis provides empirical support for resource allocation decisions, while the demonstration of policy intervention effectiveness encourages continued investment in comprehensive regional development strategies. The framework's scalability and adaptability suggest potential applications across diverse regional contexts facing similar transformation challenges.

Several limitations suggest directions for future research that could enhance the framework's capabilities and broaden its applicability. While the dataset provides comprehensive coverage of industrial indicators, future research should incorporate broader economic, social, and environmental measures to capture the full complexity of regional transformation processes. Advanced machine learning techniques, including ensemble methods, attention mechanisms, and transformer architectures, could potentially improve forecasting accuracy further. More sophisticated causal inference techniques could strengthen claims about policy effectiveness and provide better guidance for intervention design. The analysis provides strong evidence of policy effects, but advanced econometric methods, including instrumental variables, regression discontinuity and synthetic control approaches, could provide even more robust causal identification. Spatial econometric methods represent another promising direction for future research, as the current analysis focuses primarily on temporal dynamics while regional interactions and spillover effects likely play important roles in transformation processes. Understanding how policies and shocks propagate across regional boundaries could enhance both theoretical understanding and practical policy design.

The successful application of data-driven analytical techniques to Northeast China's industrial transformation demonstrates the broader potential for evidence-based regional development strategies. The Shenyang-Fushun Demonstration Zone's documented performance improvements provide encouraging evidence that systematic, coordinated policy interventions can overcome structural disadvantages and accelerate regional revitalization even in challenging post-industrial contexts.

As China continues its transition toward high-quality development objectives that emphasize efficiency, sustainability, and innovation, the frameworks and insights developed through this research offer valuable guidance for other regions facing similar transformation challenges. The integration of advanced analytical techniques with practical policy applications represents a promising pathway for achieving sustainable and inclusive regional development outcomes.

The research confirms that data-driven approaches represent more than technical exercises, serving as powerful tools for understanding complex economic systems and designing effective interventions that can improve the lives and prospects of millions of people in regions struggling with economic transition. For policy makers and researchers engaged in regional development efforts worldwide, this study provides both methodological guidance and empirical evidence supporting the value of systematic, evidence-based approaches to economic transformation in the digital age.

REFERENCES

Aghion, P., Cai, J., Dewatripont, M., Du, L., Harrison, A., & Legros, P. (2015). Industrial policy and competition. *American Economic Journal: Macroeconomics, 7*(4), 1-32.

Ahmed, S., Nielsen, I. E., Tripathi, A., & Siddiqui, S. (2022). A review on deep sequential models for forecasting time series data. *Applied Computational Intelligence and Soft Computing, 2022*, 6596397.

Beijing Statistical Yearbook (2000-2023). https://nj.tjj.beijing.gov.cn/nj/main/2023-tjnj/zk/e/indexee.htm

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: *Forecasting and control* (5th ed.). John Wiley & Sons.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company.

China Academy of Information and Communications Technology. (2021). *White paper on China's digital economy development*. CAICT Press.

China's National Bureau of Statistics (2016-2024). National Economic and Social Development Statistical Communiqué. https://english.www.gov.cn/archive/statistics/202503/01/content_WS67c2695cc6d0868f4e-8f02ae.html

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change, 10*(4), 945-974.

Ding, L., Wu, M., Jiao, Z., & Nie, Y. (2022). The positive role of trade openness in industrial green total factor productivity-provincial evidence from China. *Environmental Science and Pollution Research, 29*, 6538-6551.

Du, K., Cheng, Y., & Yao, X. (2021). Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. *Energy Economics, 98*, 105-247.

Fan, F., & Zhang, X. (2021). Transformation effect of resource-based cities based on PSM-DID model: An empirical analysis from China. *Environmental Impact Assessment Review, 91*, 106648.

Gil-Garcia, J. R., Zhang, J., & Puron-Cid, G. (2016). Conceptualizing smartness in government: An integrative and multi-dimensional view. *Government Information Quarterly, 33*(3), 524-534.

Guandong Statistical Yearbook (2000-2024). https://www.stats.gov.cn/sj/ndsj/2024/indexeh.htm

He, S. Y., Lee, J., Zhou, T., & Wu, D. (2017). Shrinking cities and resource-based economy: The economic restructuring in China's mining cities. *Cities, 60*, 75-83.

Head, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration, 67*(1), 1-11.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780.

Imbens, G. W., & Rubin, D. B. (2015). Causal inference for statistics, social, and biomedical sciences. Cambridge University Press.

Kwarteng, S. B., & Andreevich, P. A. (2024). Comparative analysis of ARIMA, SARIMA and Prophet model in forecasting. *Research* & *Development, 5*(4), 110-120. https://doi.org/10.11648/j.rd.20240504.13

Li, D. (2025). Northeast China Industrial Growth Data and SARIMA-LSTM Analysis (2000–2024) [Data set]. Mendeley Data. https://doi.org/10.17632/tpmznxhs2b.1

Li, H., Lo, K., & Wang, M. (2015). Economic transformation of mining cities in transition economies: Lessons from Daqing, northeast China. *International Development Planning Review, 37*, 311-328.

Li, H., Xiong, Z., & Xie, Y. (2018). Resource tax reform and economic structure transition of resource-based economies. *Resources, Conservation and Recycling, 136*, 389-398.

Lin, J. Y., & Tan, G. (1999). Policy burdens, accountability, and the soft budget constraint. *American Economic Review, 89*(2), 426-431.

National Bureau of Statistics of China (2000-2024). China yearbooks. https://www.stats.gov.cn/english/Statisticaldata/yearbook/

National Development and Reform Commission (2011). Summary of Liaoning Province's Economic Operation and Regulation Work During the 11th Five-Year Plan Period. https://www.ndrc.gov.cn/xwdt/gdzt/2011qgjjyx/201104/t20110429_1191269_ext.html

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review, 76*(6), 77-90.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. *Proceedings of the IEEE 17th IEEE International Conference on Machine Learning and Applications*, Orlando, Florida, USA, 1394-1401. https://doi: 10.1109/ICMLA.2018.00227

Sirisha, U. M., Belavagi, M. C., & Attigeri, G. (2022). Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison. *IEEE Access, 10*, 124715-124727. https://doi.org/10.1109/ACCESS.2022.3224938

State Council of the People's Republic of China. (2017). *Notice on establishing Shenyang-Fushun reform and innovation demonstration zone* (State Council Document No. 44).

Stock, J. H., & Watson, M. W. (2007). *Introduction to econometrics* (2nd ed.). Addison Wesley.

Shanghai Statistical Yearbook (2000-2024). https://tjj.sh.gov.cn/tjnj/tjnj2024e.htm

Jiangsu Statistical Yearbook (2000-2024). https://www.chinayearbooks.com/tag/jiangsu-statistical-yearbook

Zhang, Y., Song, Y., & Zou, H. (2020). Transformation of pollution control and green development: Evidence from China's chemical industry. *Journal of Environmental Management, 275*, 111-246.

TEHNIČKI OKVIR ZA INDUSTRIJSKU TRANSFORMACIJU ZASNOVANU NA PODACIMA U SEVEROISTOČNOJ KINI

Rezime:

Ova studija razvija dinamički okvir za evaluaciju kako bi se analizirala interakcija između IT i tradicionalnih industrija na severoistoku Kine iz vremenske i regionalne perspektive. Koristeći 24-godišnji skup podataka (2000–2024) u pet glavnih kineskih ekonomskih regiona, okvir integriše modeliranje koordinacije sprege, dinamiku sistema i prognoziranje vremenskih serija sa SARIMA i LSTM modelima. Rezultati pokazuju da IT investicije i razvoj ljudskog kapitala značajno poboljšavaju industrijske performanse, sa jakim pozitivnim korelacijama između koordinacije i regionalnog rasta ($r = 0,89$, $p < 0,01$). LSTM model pokazuje superiornu prilagodljivost nelinearnim fluktuacijama u poređenju sa SARIMA modelom, postižući snižavanje greške u prognozi za 25%. Dokazi iz provincije Ljaoning potvrđuju da ciljane političke intervencije – posebno strategije demonstracionih zona – mogu ubrzati koordinaciju između IT i tradicionalnih sektora. Okvir pruža metodološke i empirijske osnove za regionalne strategije razvoja u ekonomijama u tranziciji.

Ključne reči:

industrijska transformacija;
SARIMA model;
LSTM neuronske mreže.

JEL klasifikacija:

E66, O21