ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 178-184
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Sales Prediction for Fashion Retailer Based on State Space Model

Muhammad Nadzrin Hassan[1], Kek Sie Long[1*]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: slkek@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.022

Article Info

Received: 4 January 2026
Accepted: 20 January 2026
Available online: 9 July 2026

Keywords

State Space Model, Mean Square Error

Abstract

The fashion retail industry has experienced uncertain demand and volatile sales, influenced by new brands, seasonal patterns, and frequent promotional activities. Hence, forecasting demand and sales is a challenging task for which traditional forecasting methods often fail by assuming stable sales patterns over time. To address the dynamic patterns in sales, we propose a state-space model to predict sales of a fashion retailer from 1999 to 2017 in this paper. For this aim, a least-squares optimization problem is introduced, in which the differences between the real data and the model output are minimized. By calculating gradients, the state-space model parameters are estimated, and the model solution is updated at each iteration. The model output shall closely approximate the real data at the end of the computation, yielding a small mean squared error (MSE). The results demonstrate that the proposed model successfully captures the complex dynamics of fashion sales, with predictive accuracy improving dramatically, where the MSE value decreased from an initial value of 0.015775 to a final value of 3.2618×10^{-11}, confirming its ability to generate highly accurate forecasts for the volatile sales in the fashion retail sector. Therefore, the state-space model is an effective tool for sales prediction in fashion retail.

1. Introduction

Accurately forecasting customer demand is challenging in the fast-paced global fashion retail industry. The shifting trends, seasonality, and rapid changes that characterize fashion sales are often overlooked by traditional forecasting techniques like ARIMA, which assume that past patterns remain constant [1]. Fashion retailers face unpredictable, fluctuating demand, especially for seasonal items, leading to costly stockouts or overstock. Traditional time series forecasting methods are popular, but they are not good enough at capturing real-world sales dynamics because they assume static data [2]. A more flexible modelling strategy is required to address this research gap and improve prediction accuracy while reducing industrial financial risk.

To manage complexity, modern sales forecasting requires advanced computational tools. In supply chain demand forecasting, [3] demonstrated that collective machine learning models, such as Voting Regressor, significantly outperform conventional techniques. [4] highlighted the importance of structured adoption curves, using the Bass Model with early sales data to accurately forecast the peak sales period for new items. Additionally, [5] presented the importance of multi-source data by incorporating sentiment analysis from online reviews into a simplified forecasting model for new energy vehicles, achieving high accuracy. It is noticed that the shift from static models to hybrid or adaptive methods that better represent real-world sales dynamics is an ongoing trend in these studies.

A state-space model provides a strong foundation for examining dynamic, time-varying systems. By modelling driver fatigue as a latent state emerging from observable signals, [6] detected it with excellent accuracy using a customized state-space model with Kalman filtering. While [7] examined respiratory system models using the state-space approach to understand parameter sensitivity and stability better. These approaches illustrate the model's main advantages, including managing unobservable variables, controlling temporal dependencies, and producing results that are easy to understand. However, volatility issues and the need for adequate input data are the limitations.

As a more flexible alternative, this paper explores the use of a state-space model to predict sales in fashion retail. The state space model is excellent at replicating dynamic systems and latent trends over time. It offers an effective way to improve prediction accuracy in the volatile fashion industry and to support better inventory and business decisions. Given these reasons, our study proposes an advanced sales forecasting model to improve the accuracy, efficiency, and adaptability of demand forecasting for a fashion retailer. We also analyse the fashion retailer's historical sales data using descriptive statistics. Moreover, we predict historical sales data for a fashion retailer using the state-space model, and assess the accuracy of the sales prediction using mean squared error (MSE).

Since the fashion industry is a fast-paced, trend-driven sector where accurate sales forecasting is essential, the dynamic state-space model would overcome the limitations of static forecasting techniques, making our study significant for fashion sales prediction. Thus, fashion retailers can gain a competitive edge by using the improved forecast to support real-time operational decisions, thereby improving inventory management, reducing waste and boosting profitability.

2. Methodology

Consider a state space model,

$$\dot{x} = Ax + b, \tag{1}$$
$$y = Cx + d, \tag{2}$$

where x is the state variable and y is the model output. While A, b, C and d are the model parameters, with A is the state transition matrix, C is the output coefficient matrix, and b and d are constant vectors. The initial state is given by $x(0) = x_0$. Here, Eq. (1) is the state equation and Eq. (2) is the output equation. The analytical solution for the state equation is given by

$$x(t) = e^{At}x(0) - (I - e^{At})A^{-1}b, \tag{3}$$

and its discrete solution, with the sampling time Δt, is written as

$$x_{k+1} = e^{A\Delta t}x_k - (I - e^{A\Delta t})A^{-1}b, \tag{4}$$
$$y_k = Cx_k + d. \tag{5}$$

2.1 Least Squares Optimization Problem

Denote y_s be the historical sales data for a fashion retailer. We introduce a least squares optimization problem as follows,

$$\text{Minimize} \quad J = \frac{1}{2}(y - y_s)^T(y - y_s), \tag{6}$$

over the state equation (1) and the output equation (2). The aim is to estimate the model parameters A, b, C and d by observing the historical sales data, in turn, to update the model solution $\hat{x}$ so as the model output $\hat{y}$ can approximate the real data y_s as closely as possible.

The gradients with respect to the model parameters A, b, C and d are calculated by

$$\frac{\partial J}{\partial A} = C^T(\hat{y} - y_s)(\hat{x}^T), \tag{7}$$
$$\frac{\partial J}{\partial b} = C^T(\hat{y} - y_s)(1^T), \tag{8}$$
$$\frac{\partial J}{\partial C} = (\hat{y} - y_s)(\hat{x}^T), \tag{9}$$
$$\frac{\partial J}{\partial d} = (\hat{y} - y_s)(1^T). \tag{10}$$

The model parameters are estimated through the following recursion equations,

Enhanced Knowledge in Sciences and Technology Vol. 6 No. 1 (2026) p. 178-184

$$A^{i+1} = A^i - \alpha_1 \left(\frac{\partial J}{\partial A}\right)^i, \tag{11}$$
$$b^{i+1} = b^i - \alpha_2 \left(\frac{\partial J}{\partial b}\right)^i, \tag{12}$$
$$C^{i+1} = C^i - \alpha_3 \left(\frac{\partial J}{\partial C}\right)^i, \tag{13}$$
$$d^{i+1} = d^i - \alpha_4 \left(\frac{\partial J}{\partial d}\right)^i, \tag{14}$$

with α_1, α_2, α_3 and α_4 are the learning rates and i = 0, 1, 2, ... represent the iteration numbers.

In addition, the model solution is updated using the recursion equation below,

$$\hat{x}^{i+1} = x^i - \alpha_5 \left(\frac{\partial J}{\partial x}\right)^i, \tag{15}$$

with the learning rate α_5 and the gradient with respect to the model solution x is given by

$$\frac{\partial J}{\partial x} = C^T(\hat{y} - y_s). \tag{16}$$

Hence, the state estimate, which is represented by the model solution (15), is determined. Referring to (5), the model output estimate is then measured.

It is noticed that the first-order necessary conditions are satisfied when those gradients, (7) to (10) and (16), tend to zero, for which the optimal model parameters and the optimal model solution, including the optimal model output, can be resulted. The convergence will be reached once the stopping rule,

$$||J^{i+1} - J^i|| \quad < \text{tolerance}, \tag{17}$$

is satisfied, where the objective function J from (6) at each iteration i is evaluated. Otherwise, the iteration is repeated.

2.2 Computational Algorithm

From the discussion above, we summarize the calculation procedure as an iterative algorithm as follows.

- Data: Consider the initial values for the model parameters, A^0, b^0, C^0, d^0 and the initial state x^0. Set $i = 0$ to begin the calculation.
- Step 1: Compute the state equation (1) to give the model solution (4), and calculate the output equation (2) to obtain the model output (5).
- Step 2: Evaluate the objective function (6), and the gradients (7), (8), (9) and (10) with respect to the model parameters.
- Step 3: Update the model parameters from (11), (12), (13) and (14).
- Step 4: Evaluate the gradient (16) with respect to the state variable, and update the model solution (15).
- Step 5: Test the convergence. If the stopping rule (17) is satisfied, stop the calculation to give the prediction solution. Otherwise, repeat the iteration, and go to Step 1.

The explanation for the first iteration is given as follows. The first iteration result at i = 0 is obtained by sequentially applying the steps above using initial guesses for A^0, b^0, C^0, d^0 and x^0. First, the state equation $x_{k+1} = e^{A^0\Delta t}x_k^0 - (I - e^{A^0\Delta t})(A^0)^{-1}b^0$ is solved to compute $\hat{x}^0$, and it is followed by the output $\hat{y}^0 = C^0\hat{x}^0 + d^0$. The initial cost $J^{(0)}$ is then evaluated as $\frac{1}{2}(\hat{y}^0 - y_s)^T(\hat{y}^0 - y_s)$. Using $\hat{y}^0$ and $\hat{x}^0$, the gradients $\left(\frac{\partial J}{\partial A}\right)^0$, $\left(\frac{\partial J}{\partial b}\right)^0$, $\left(\frac{\partial J}{\partial C}\right)^0$, and $\left(\frac{\partial J}{\partial d}\right)^0$ are calculated according to the gradient formulas. These gradients are used to update the parameters to A^1, b^1, C^1, d^1 via the updating rules with step sizes α_1 to α_4. Simultaneously, the state gradient $\left(\frac{\partial J}{\partial x}\right)^0 = (C^0)^T(\hat{y}^0 - y_s)$ is computed, and the state estimate is updated to $\hat{x}^1 = x^0 - \alpha_5 \left(\frac{\partial J}{\partial x}\right)^0$. The outcome of the first iteration is therefore the updated parameter set (A^1, b^1, C^1, d^1), the refined state estimate $\hat{x}^1$, and the initial cost value $J^{(0)}$. The algorithm then proceeds to test convergence by comparing $J^{(1)}$ (to be computed in the next iteration) with $J^{(0)}$, and continues iteratively until the stopping criterion is met.

The methodology relies on several critical assumptions to ensure the validity and applicability of the identified model. First, the system is assumed to be linear time-invariant, as expressed by the state-space structure $\dot{x} = Ax + b$ and $y = Cx + d$. This implies that the system dynamics do not change over time and that superposition holds. Second, the discretization process is assumed to be exact and performed using the matrix exponential $e^{A\Delta t}$, which requires that the sampling interval Δt is constant and that the continuous-time dynamics are accurately captured by the zero-order-hold discretization formula provided. Third, regarding noise and disturbances, the method implicitly assumes that the measurement data y_s contains negligible noise or that the noise is unbiased and sufficiently small such that the quadratic objective function J effectively approximates a maximum-likelihood estimator under Gaussian noise. No explicit process noise is considered in the state equation, implying that the model is deterministic and any unmodeled dynamics are absorbed into the parameter estimation error. Finally, the initial state and parameter guesses are assumed to be sufficiently close to the true values to allow gradient-based optimization to converge to the global minimum, as the algorithm does not include explicit mechanisms to avoid local minima.

2.3 Real Data

Fig. 1 shows the historical sales data that have been scaled from zero to one. The raw data, which refers to data directly without any summarization or modification. Raw data itself is the original value collected from its source that has not been processed. This graph aims to maintain complete information transparency where each point represents a specific unit of data. For example, the graph in Figure 1 shows sales results against the year. We can see the entire data distribution including repeated values and anomalies without any hiding that occurs.

Raw data graphs are particularly useful for small to medium-sized data sets where the number of observations is not too large to cause visual congestion. They allow analysts to make direct interpretations based on the reality of the data, avoiding potential bias. However, for large data sets, direct display of raw data can be less effective due to the problem of overlapping points, making it difficult to interpret. In short, raw data graphs serve as the most direct and transparent visual tool for representing raw information. They emphasize the importance of showing every detail of the data, which is especially valuable in the early data exploration stage or when the accuracy of every insight is critical. However, the appropriateness of their use depends greatly on the size and nature of the data and the analysis goals to be achieved.

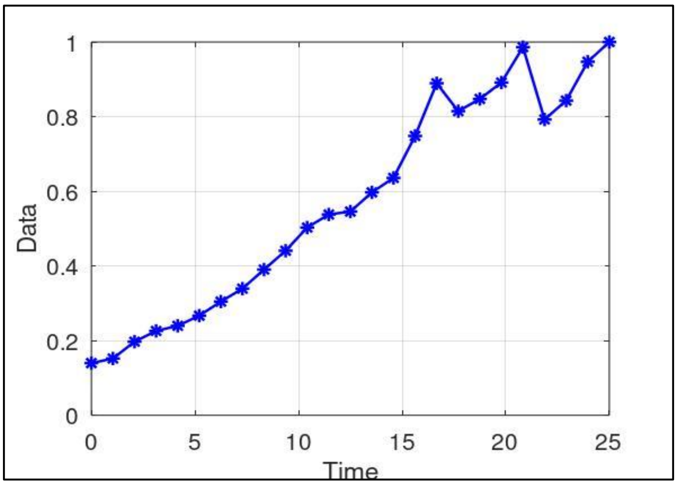

Fig. 1 Real Data

2.4 State Space Model Prediction Curve

Fig. 2 shows the relationship between time and the output of the system, where A = -0.11, b = 0.09, C = 1, d = 0 and x_0 = 0.1 in (1) and (2). At the beginning of time, the output value is low, around 0.1, then increases continuously as time increases. This increase indicates that the system responds positively to the initial input conditions. However, the rate of increase in output does not remain the same. In the early stages, output growth is faster but after some times, the rate of increase becomes more and more flat. This situation shows that the system is approaching a stable value.

From the state space model perspective, the pattern reflects a dynamically stable system. The output does not show any oscillations or overshoots, but moves smoothly towards the maximum value. This indicates that the model used is well controlled and suitable for long-term analysis. Overall, Fig. 2 shows that the system output will continue to increase until it reaches a certain level of stability. In a forecasting context such as sales forecasting, this can be interpreted as increasingly stable sales growth as the market approaches saturation. Therefore, this state space model is suitable for describing and predicting the behaviour of a system that is gradually evolving.

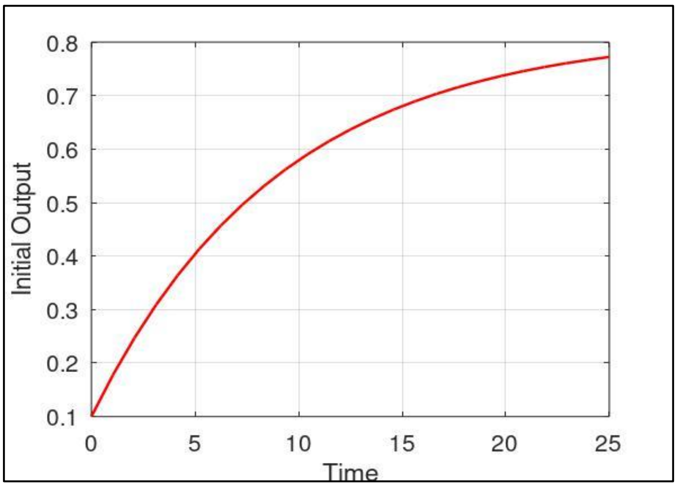

Fig. 2 State Space Model Prediction Curve

3. Results and Discussion

Fig. 3 shows a comparison between the state space model output and actual sales data over the study period. The model performance was evaluated using the mean square error (MSE) measure to see how accurate the estimates produced were. At the initial stage of modelling, the MSE value obtained was 0.015775, indicating that there was still a significant difference between the model output and the actual data. This situation is reasonable because the model was still in the initial adaptation phase and has not yet reached full stability.

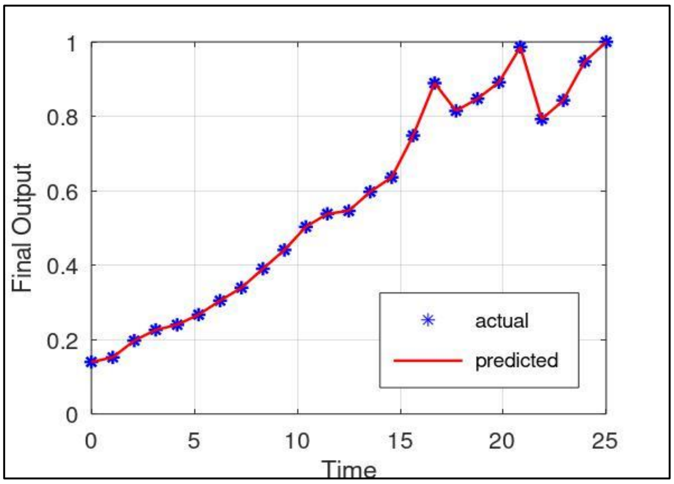

Fig. 3 Results of Sales Prediction

As the modelling process continued, the model accuracy showed a clear improvement. At the end of the process, the MSE value decreased significantly to 3.2618×10^{-11}, which was almost close to zero. This decrease indicated that the model output has almost matched the actual sales data, thus proving the ability of the state space model to capture sales patterns and dynamics well. Table 1 shows the MSE, time elapsed and iteration number when the state space model is used to predict the historical sales data.

Table 1 Simulation Results

Iteration Number	Initial MSE	Final MSE	Elapsed Time (s)
91	0.015775	3.2618×10^{-11}	0.0318651

In terms of computational efficiency, the time elapsed of 0.0318651 seconds was recorded and remained the same throughout the modelling process. In addition, the number of iterations required to reach the final result is 91 iterations, indicating that the model converges quickly without requiring additional iterations. This reflects the efficiency of the algorithm used in the MATLAB environment.

The prediction error's minor variations are probably caused by a number of related factors that are part of the modelling and optimization process. First, fixed learning rates (0.01 for parameters and 0.1 for states) are used in gradient descent optimization. This can cause slight oscillations around the cost function minimum, especially when adaptive tuning mechanisms are not present. Second, residual unmodelled dynamics result from the model's structural simplicity as a first-order linear state space system, which restricts its ability to properly represent nonlinear trends, seasonal components and external influences associated with actual economic data. Third, during times of fluctuating growth rates, data normalisation by the maximum value might amplify modest relative mistakes even while it improves numerical stability. Furthermore, small numerical approximations are introduced by separating the continuous time system using matrix exponentials which build up across time steps. Lastly, the gradient descent may converge to a local minimum as a result of sensitivity to initial parameter values where tiny persistent error fluctuations persist because of the local curve of the cost surface. All these elements work together to produce the residual variability that is shown, which is practically insignificant but represents fundamental trade-offs between convergence behaviour, numerical accuracy and model complexity.

Overall, the combination of the very low final MSE value, the short processing time, and the consistent number of iterations indicate that the developed state space model not only provides accurate estimates, but is also computationally efficient. Therefore, the results in Fig. 3 confirm the suitability of this model for use in the analysis and forecasting of sales data for this study.

Fig. 4 shows the variation of the prediction error measured over the study period. The horizontal axis represents time, while the vertical axis shows the prediction error value. The MSE values shown are on a very small scale, in the range of 10^{-5}, indicating that the difference between the model estimate and the actual data is minimal overall.

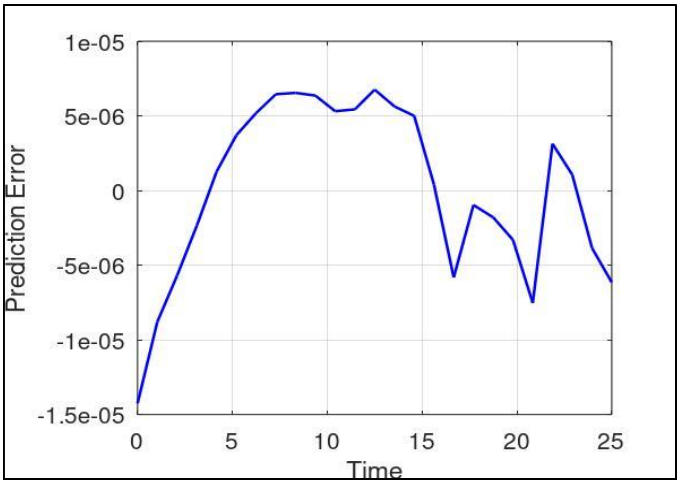

Fig. 4 Prediction Error

In the early stages of the study, it can be observed that the error value is higher and negative, indicating that the model at this stage is still in the process of adapting to the sales data. This situation is likely due to the lack of historical information in the early stages, which causes the state estimation to not be fully stable. However, the error shows a decreasing trend as time increases, indicating an increase in the model's ability to track the actual data.

In the middle period of the time series, the error value is found to be close to zero and remains relatively stable. This situation reflects the good performance of the model, where the estimates produced are almost identical to the observational data. The stability of the error during this period indicates that the state space model has reached a steady-state and is able to model the system dynamics effectively.

However, some spikes in error can be observed towards the end of the study period. This sudden increase in MSE values may be due to sudden changes in sales patterns or external disturbances not included in the model, such as changes in market demand or seasonal factors. However, the magnitude of the error is still within a small range, indicating that the overall performance of the model remains satisfactory.

Overall, the error analysis based on Figure 4 proves that the developed state space model has high prediction accuracy. The low MSE value and the stability of the error over most of the study period confirm the suitability of this model for use in modelling and forecasting future sales data.

4. Conclusion

This study sought to dissect historical sales data with the main goal of grasping sales behaviour. First of all, we did some of the simplest calculations such as the average and the standard deviation of the sales numbers and then; to make it easier to detect clear seasonal patterns along the time, we plotted those numbers in the form of graphs. These seasonal patterns along with the amount of variation were the main ingredients for our state space model that could spot the latent trends and eliminate the noise from the historical sales numbers thus obtaining a neat line that represents the real underlying sales pattern.

To see to what extent the model really functioned, we looked at the model's forecasts against the actual historical sales and calculated the Mean Squared Error which was equal to 3.5032. Such a low number indicates that the model is precise and can be used in sales forecasting, hence, the state space approach is confirmed as a valid tool for the analysis and prediction of sales behaviour in fashion retail. There are some limitations identified during the research. A small set of data will constrain the ability to accurately predict. A limited volume of data increases the risk of overfitting, where the model may capture noise rather than underlying pattern. This method only can be used when there is little noise in the data. Too much noise will distracting the results. The state space model is more suitable for measuring linear data. If the method is used to measure a volatile date, the result will not be so accurate. To improve the research in the future, there are some recommendations we would suggest. Firstly, it is important to use a significantly larger dataset. With a lot of data, it will improve the accuracy of the result and boosting reliability. Secondly, real-world data usually contains noise that will affect the accuracy of prediction. Future study should intentionally test the method with noisy data. This could make the method more accurate and beneficial. Finally, it would be helpful to compare this method with other modern techniques, especially modern machine learning methods. This comparison will show if the new strategy succeeds over the others, performs worse, or is about the same. It could also be helpful in combining the most beneficial aspects of each approach.

Acknowledgement

The authors would also like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Muhammad Nadzrin Hassan, Kek Sie Long;* ***data collection:*** *Muhammad Nadzrin Hassan;* ***analysis and interpretation of results:*** *Muhammad Nadzrin Hassan, Kek Sie Long;* ***draft manuscript preparation:*** *Muhammad Nadzrin Hassan, Kek Sie Long. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Ensafi, Y., Amin, S. H., & Zhang, G. (2022). Time-series forecasting for sustainable fashion management in a fast fashion industry: A systematic literature review. *IEEE Access, 10*, 115401–115415.

[2] Shankar, U., Devanagavi, G. S., Satyanarayana, R. D., & Bharadwaj, S. A. (2022). Sales prediction using machine learning techniques. In *2022 IEEE International Conference on Data Science and Information System (ICDSIS)* (pp. 1–6). IEEE.

[3] Ali, M., Khan, M. M. H., Rana, M. S. M. S., Hossain, S. K. S., & Islam, M. R. (2023). A comparative analysis of machine learning algorithms for demand forecasting in supply chain management. *Annals of Data Science, 10*(5), 1547–1567.

[4] Krishnan, T. V., Feng, S., & Jain, D. C. (2023). Peak sales time prediction in new product sales: Can a product manager rely on it? *Journal of Business Research, 165*, 114054.

[5] Dai, X., Chen, J., & Liu, W. (2022). Sales forecasting based on sentiment analysis and improved support vector machine. *Electronic Commerce Research, 22*(3), 883–901.

[6] Zhang, X., Wang, Y., & Li, J. (2024). State space model detection of driving fatigue considering individual differences and time cumulative effect. *International Journal of Transportation Science and Technology, 15*(2), 123-135.

[7] Saatci, E., Saatci, E., & Akan, A. (2020). Analysis of linear lung models based on state-space models. *Computer Methods and Programs in Biomedicine, 183*, 105094.