PeerJ Computer Science

Design of a consumer behavior prediction model integrating reinforcement learning and time series analysis in online e-commerce reviews

Zongping Lin[1], Yingyi Huang[2], Jing Yang[3], Chunhu Cui[4], Yabin Lian[5], Honglei Zhang[6] and Fadi Al-Turjman[7]

[1] School of Economics and Management, Quanzhou University of Information Engineering, Quanzhou, Fujian, China
[2] School of Business, Ningbo Tech University, Ningbo, Zhejiang, China
[3] School of Economics and Business, Xiamen City University, Xiamen, Fujian, China
[4] School of Economics and Management, Tsinghua University, Beijing, China
[5] Academy of Art & Design, Minnan Science and Technology College, Quanzhou, Fujian, China
[6] Beijing Meitu Home Technology Co., Ltd, Beijing, China
[7] Artificial Intelligence, Software, and Information Systems Engineering Departments, Research Center for AI and IoT, AI and Robotics Institute, Near East University, Nicosia, Cyprus

ABSTRACT

This study focuses on the design and optimization of consumer behavior prediction models in online e-commerce reviews. To address the issues of slow convergence and insufficient robustness in the traditional Q-learning reinforcement learning algorithm, this article introduces a probabilistic action-selection algorithm. This algorithm employs a multi-step, iterative mechanism that uses instantaneous differencing to increase the likelihood of selecting high-Q actions during model iteration, thereby accelerating the solution process and ensuring robust network optimization. Given the nonlinear and high-noise characteristics of consumer behavior time-series data in e-commerce reviews, we propose a hybrid intelligent prediction model, Q-learning-Artificial Neural Network-Hidden Markov Model (QL-ANN-HMM), that effectively reduces the impact of systematic random errors and significantly improves prediction accuracy. Experimental results demonstrate that the improved Q-learning algorithm achieves 2.71% and 5.96% improvements in mean absolute percentage error (MAPE) and normalized mean squared error (NMSE), respectively, compared to the traditional Q-learning algorithm on the Amazon Reviews 2023 and Flipkart Reviews datasets. Additionally, the QL-ANN-HMM model achieves lower mean absolute error (MAE), MAPE, and NMSE values on both datasets, recorded at 0.0195, 0.019, and 0.0189, respectively. This research not only provides novel theoretical support and technical methods for predicting consumer behavior in online e-commerce reviews but also enables e-commerce platforms to more accurately track market dynamics, optimize resource allocation, and achieve sustainable development by comprehensively analyzing consumer behavioral data.

Subjects Algorithms and Analysis of Algorithms, Artificial Intelligence, Data Mining and Machine Learning, Software Engineering, Neural Networks

Keywords Reinforcement learning algorithms, Hidden Markov, Time series analysis, Artificial neural networks, Consumer behavior prediction

Submitted 9 September 2025
Accepted 13 November 2025
Published 12 January 2026

Corresponding author
Yingyi Huang,
huangyingyi198205@163.com

Academic editor
Osama Sohaib

Additional Information and Declarations can be found on page 17

DOI **10.7717/peerj-cs.3451**

How to cite this article Lin Z, Huang Y, Yang J, Cui C, Lian Y, Zhang H, Al-Turjman F. 2026. Design of a consumer behavior prediction model integrating reinforcement learning and time series analysis in online e-commerce reviews. **PeerJ Comput. Sci. 12:e3451 DOI 10.7717/peerj-cs.3451**

INTRODUCTION

With the rapid advancement of Internet technology, e-commerce has become an integral part of everyday life. The growing trend of consumers purchasing goods and services through online platforms not only offers significant convenience for users but also presents unprecedented business opportunities for e-commerce enterprises (*Pan, Liu & Pan, 2022*). However, as the number of e-commerce platforms rises and market competition intensifies, accurately predicting consumer behavior to optimize inventory management, improve marketing strategies, and enhance user experience has become a critical challenge for these businesses. In this context, developing an efficient consumer behavior prediction model is essential.

The surge in e-commerce activity has generated vast amounts of user behavioral data, including online reviews, browsing records, and purchase histories. These data offer a valuable resource for e-commerce enterprises, enabling them to gain deeper insights into consumer needs and behavioral patterns. Online reviews, as a key form of consumer feedback, contain crucial information about product quality, price, and user experience, while also reflecting consumers' emotional tendencies and decision-making processes (*Chen et al., 2022*). Therefore, conducting a comprehensive analysis of online reviews to uncover hidden consumer behavior patterns is highly significant for enhancing e-commerce enterprises' predictive capabilities and market competitiveness.

Time series analysis (*Dai, Tong & Jia, 2024*), a critical statistical method, has been extensively applied in e-commerce demand forecasting. By analyzing and forecasting data collected continuously over time, time series analysis effectively captures trends, seasonality, and other patterns, while also integrating external factors (*e.g.*, holidays, promotions) to enhance forecasting accuracy. However, traditional time series analysis methods face limitations when dealing with the complexity and dynamic nature of the e-commerce environment (*Costa & Rodrigues, 2024*). To further improve prediction accuracy, this article introduces reinforcement learning techniques. Reinforcement learning (*Ernst et al., 2024*) continuously optimizes an intelligent agent's behavior by simulating its learning and decision-making processes within a given environment, enabling it to achieve specific goals. In the context of e-commerce, reinforcement learning can simulate consumers' purchasing decision-making processes, enabling the prediction model to learn and optimize, thus providing more accurate predictions of future consumer behavior. Combining reinforcement learning with time series analysis leverages the strengths of both methods, resulting in a more robust and accurate prediction model.

Despite the advancements in e-commerce consumer behavior prediction and the numerous machine learning and data mining methods proposed by scholars, several challenges remain in practical applications. For instance, e-commerce review data often exhibits nonlinearity and high levels of noise. This article seeks to address these challenges by designing a consumer behavior prediction model that integrates reinforcement learning and time series analysis. The proposed model aims to enhance prediction accuracy and improve the market competitiveness of e-commerce enterprises.

The specific contributions of this article are as follows:

1. The traditional reinforcement learning Q-learning algorithm is improved, and an algorithm for selecting actions based on probability distributions is designed by combining multi-step iterations with instantaneous differencing, which ensures the probability of selecting higher Q-value actions during the iteration process of the model, improves the convergence speed when solving the model, and guarantees the robustness of the network optimization search.
2. A time series forward multi-step hybrid intelligent prediction model Q-learning-Artificial Neural Network-Hidden Markov Model (QL-ANN-HMM) is proposed, which combines the improved reinforcement learning with an artificial neural network and a Hidden Markov Model, to reduce the systematic stochastic error.
3. The experimental results on Amazon Reviews 2023 and Flipkart Reviews datasets show that the improved Q-learning algorithm improves the MAPE metric by 2.71% and the NMSE metric by 5.96% compared to the traditional Q-learning algorithm, which significantly improves the prediction accuracy and effectiveness.

In this article, we introduce the current research status of reinforcement learning and time series analysis algorithms, and analyze their limitations for consumer behavior prediction in e-commerce reviews in 'Related Works'. In 'Model Design', the improved reinforcement learning algorithm and the time-series consumer behavior intelligent prediction algorithm, the QL-ANN-HMM model, are introduced. 'Experiments and Analysis' describes the experimental results and discusses the performance of the improved reinforcement learning algorithm and the QL-ANN-HMM model on the Amazon Reviews 2023 and Flipkart Reviews datasets. 'Conclusion' provides a summary of the impact of the improved reinforcement learning algorithm and the QL-ANN-HMM model developed in this article on the prediction of consumer behavior in online e-commerce reviews.

RELATED WORKS

Reinforcement learning

Reinforcement learning is a critical class of control and decision-making methodologies, grounded in the principle that intelligent agents learn to make optimal decisions in a given environment through interaction. It operates on the concept of trial-and-error learning, where agents optimize their behavior by performing actions, observing the outcomes, and aiming to maximize long-term rewards. In recent years, reinforcement learning has achieved remarkable breakthroughs in solving complex decision-making problems and has seen significant success in domains such as robotics, autonomous driving, recommender systems, and gaming.

For instance, *Li et al. (2024)* models the output feedback synchronization problem *via* robust output regulation and reinforcement learning, depicting the interactions between agents through a zero-sum game framework. It proposes an output-feedback learning algorithm that uses input-output system data to achieve distributed, robust, optimal

synchronization for heterogeneous multi-agent systems. In the context of discrete-time multi-agent systems subject to external disturbances, *Ito & Fujimoto (2022)* examines the impact of perturbations from both local neighbors and the agents themselves. It transforms the optimization problem into a zero-sum game with control and disturbance strategies. It introduces a data-driven iterative algorithm based on strategy gradients to solve the Hamilton-Jacobi-Isaacs (HJI) equations. *Liu et al. (2023)* presents a zero-sum game Q-learning algorithm for both state feedback and output feedback in discrete fractional-order multi-agent systems.

Furthermore, the *Ren, Jiang & Ma (2022)* addresses actuator faults in second-order multi-agent systems. It proposes a fault-tolerant control strategy based on a zero-sum differential game to ensure system stability and performance optimization. To address server denial-of-service attacks in multi-agent systems, the *Jin et al. (2025)* introduces a neural network-based reinforcement learning scheme grounded in a multi-player hybrid zero-sum game strategy.

However, the majority of existing reinforcement learning research focuses on synchronization, optimization, and game-theoretic problems within multi-agent systems. These scenarios are not fully applicable to predicting consumer behavior on e-commerce platforms, which primarily focus on identifying patterns in individual user activity. Furthermore, the studies above do not adequately examine the role of time-series analysis in this context. Time series analysis is vital for capturing temporal trends in user behavior and constitutes a key component of consumer behavior prediction models based on e-commerce reviews.

Therefore, our research aims to demonstrate that integrating Q-learning with artificial neural networks and the Hidden Markov Model (HMM) can more effectively address the challenges of predicting consumer behavior in online e-commerce reviews. This integration leverages Q-learning's strengths in reinforcement learning by approximating the Q-value function using neural networks, overcoming the difficulties traditional Q-learning faces in high-dimensional state and action spaces. Simultaneously, the HMM captures the temporal features of user behavior, further enhancing prediction accuracy.

Time series analysis

In recent years, with the growing prominence of artificial intelligence and increasing computational power, more and more research on time series forecasting has been published. *Jin & Xu (2024)* proposed an autoregressive model to forecast oil and gas market pricing. The introduction of the autoregressive (AR) model also means the birth of traditional time series analysis methods. Since then, scholars in various countries have proposed a series of extensions and improvements based on AR. The most widely used model is AutoRegressive integrated moving average (ARIMA) (*Aminullah, 2024*), which combines the autoregressive model and moving average model (MA) with the difference operation (I), and determines the time series data by observing the time series data, and then determines the time series data of AR, I, and MA, and then determines the time series data of AR, I, and MA. Order of the three processes of the data model AR, I, and MA. Then, model training and parameter estimation are performed using historical

observations based on the determined orders. Finally, forecast evaluation is carried out. The rapid development of ARIMA methods also marks a significant leap in time-series analysis for forecasting. To date, many scholars still use ARIMA models for time series forecasting. *Dong et al. (2024)* used the ARIMA model to forecast influenza in Chongqing to provide a reference for influenza disease prevention and control. *Liu et al. (2024)* used an ARIMA model to predict the exhaust temperature of a turboshaft engine, and predicted the remaining life of the engine by monitoring the change in engine exhaust temperature. *Harrou et al. (2024)* combines a convolutional neural network (CNN) with an ARIMA model to predict short-cycle traffic flow at traffic intersections.

Exponential smoothing is also widely used in time series analysis. This method belongs to the moving average family, mainly using weighted averages to capture the trend and pattern of data changes. It is suitable for data with small fluctuations and obvious periodicity and trend. *Ahmed & Kumar (2023)* utilized exponential smoothing to forecast the average nodal tariffs collected in Boston, New England. *Lyu et al. (2025)* used the Holt-Winters exponential smoothing model to predict the demand for Achillea antibiotics during the COVID-19 pandemic. The seasonal decomposition method is generally not used as a separate model for forecasting in today's time-series forecasting problems. Its main principle is to split the time series data. For example, the SARIMA (*Cheng et al., 2024*) model, which combines it with the ARIMA model, is widely used for time series forecasting. *Okorie, Afuecheta & Nadarajah (2023)* used the SARIMA model to analyze and forecast the market price of red beans in Canada. In addition, seasonal decomposition methods can be combined with various models; for example, *Wang et al. (2024)* combines seasonal decomposition methods with graph convolutional neural networks to forecast traffic flow.

Traditional time series analysis methods can perform well on relatively smooth and straightforward linear time series data. However, methods such as ARIMA are primarily designed for univariate time series. They cannot be directly applied to multivariate time series, which often exhibit complex nonlinearities and hidden dependencies among variables.

MODEL DESIGN

In this article, we employ time-series analysis techniques to capture temporal dependencies and trends in consumer behavior data. By forecasting values such as historical sales figures, number of views, and number of comments, the prediction results serve as inputs to the reinforcement learning module. This integrated approach allows us to conduct a more comprehensive and accurate predictive analysis of consumer behavior. Ultimately, the combined model leverages both time-series forecasting and reinforcement learning to provide deeper insights into consumer behavior and trends.

Improved reinforcement learning algorithm

Q-learning is a classic reinforcement learning algorithm. Let S represent the set of states, A is the set of actions, R is the reward function, and P is the transition probability between states. The value function, denoted as Q, is used to estimate the expected future rewards for

each state-action pair. The goal of Q-learning is to learn an optimal policy that maximizes the expected cumulative reward by iteratively updating the Q-value, $Q(s, a)$, according to the following equation:

$$Q(s,a) = R(s,s',a) + \gamma \sum\nolimits_{s' \in S} P(s'|s,a)\left(\max_{a \in A} Q(s',a)\right) \tag{1}$$

where $Q(s, a)$ denotes the value of Q after the execution of action a in state s, s′ represents the transfer process of s, γ is the discount factor of the transfer process, which signals the influence of R.

Introducing this equation into the qualification trajectory, the iterative update equation for $Q(\lambda)$ can be obtained as Eq. (2).

$$Q_{k+1}(s,a) = Q_k(s,a) + \alpha \delta_k e_k(s,a) \tag{2}$$

$$\delta_k = R(s_k, s_{k+1}, a_k) + \gamma\left(\max_{a' \in A} Q(s',a) - Q(s_k,a_k)\right) \tag{3}$$

where k is the number of iterations, α is the learning factor for reinforcement learning, $e_k(s, a)$ is the qualification trajectory, and R(−) is the reward function at the k-th iteration. During multiple rounds of $Q(\lambda)$ iterations, the action with the highest value is selected based on the greedy principle, denoted as:

$$a_g = \arg\max_{a \in A} Q_k(s,a). \tag{4}$$

The strategy in making the action selection is shown in Eqs. (5) to (7):

$$P_s^{k+1}(a_g) = P_s^{k+1}(a_g) + \beta\left(1 - P_s^{k+1}(a_g)\right) \tag{5}$$

$$P_s^{k+1}(a) = P_s^{k+1}(a)(1-\beta), \forall a \in A, a \neq a_g \tag{6}$$

$$P_{s'}^{k+1}(a) = P_{s'}^{k}(a), \forall a \in A, \forall s' \in S, s \neq s' \tag{7}$$

where $P_s^k(a)$ is the probability of a state s being selected at k iterations, and β is the update factor of the probability. When the optimal function converges, the control policy of the learning system will be obtained. In this study, the probabilistic action-selection mechanism follows an ε-greedy strategy with softmax-based probability updates. Specifically, the probability of selecting an action is determined by:

$$\mathrm{P(a|s)} = \frac{e^{\mathrm{Q(s,a)}/\tau}}{\sum_{\mathrm{a'}} e^{\mathrm{Q(s,a')}/\tau}} \tag{8}$$

where τ denotes the temperature parameter controlling exploration. The eligibility trace is updated as

$$e_{t+1}(s,a) = \gamma\lambda e_t(s,a) + 1. \tag{9}$$

which accumulates the recent state–action pairs to accelerate convergence. This probabilistic structure ensures smoother exploration compared to the deterministic greedy selection.

Figure 1 illustrates the algorithmic flow of optimal carbon flow prediction within the framework of trend computing, utilizing reinforcement learning. In each iteration, the

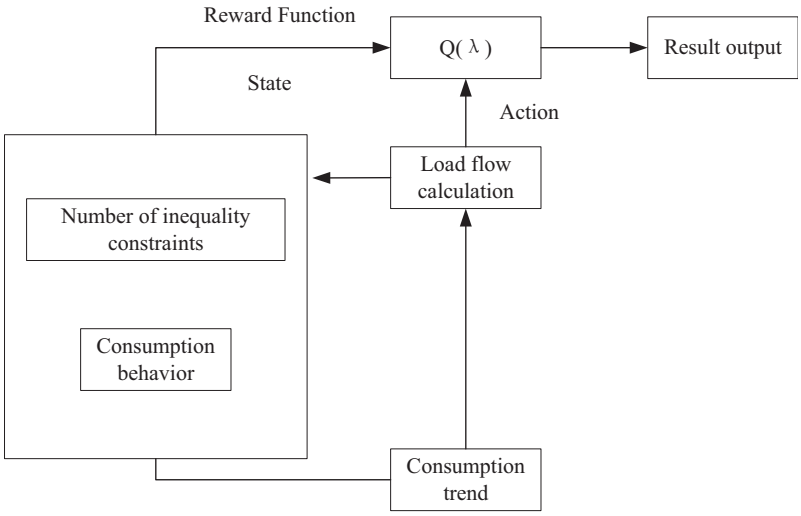


Figure 1 **Improved Q-learning algorithm.** Full-size DOI: 10.7717/peerj-cs.3451/fig-1

learner obtains the reward function value through a multi-step iteration process. This value is then combined with the learner's current state to continuously update the reward function R. As iterations proceed, the model refines its understanding of the environment, progressively improving decision-making. Ultimately, the learner identifies the optimal action a, which corresponds to the most efficient carbon flow based on the learned behavior and observed trends.

Construction of QL-ANN-HMM

The time series of online e-commerce reviews is a sequence of random variables arranged in time order $x(t), t = 1, 2, \ldots$ t. Generally, samples of observations are used to represent the observed values of a time series of random variables. The primary objective of time series forecasting is to develop a statistical model that best fits the characteristics of the observed data. This process can be described as follows: for $k > 0$, the goal is to use the sequence data at time $t + k$ to recursively predict the sequence data at time $t + k + 1$. This prediction is made by calculating the posterior probability, which can be expressed as Eq. (10):

$$p(x_{t+k+1}|e) = \sum_{x_{t+k}} p(x_{t+k+1}|x_{t+k})p(x_{t+k}|e). \tag{10}$$

The sequence data and evidence denote key aspects of the analysis. It has been observed that time-series predictions of consumer behavior in online e-commerce reviews exhibit nonlinearity and high noise levels. To address this, we combine the improved Q-learning algorithm with the HMM. The improved Q-learning algorithm leverages historical observation data as rewards, emphasizing the varying influence of both recent and distant historical data. By iterating through this process, the role of historical observations in the model is enhanced. This enhanced historical data is then integrated into a neural network and HMM, thus benefiting from the neural network's strong data-fitting capabilities and the HMM's ability to reduce random system errors.

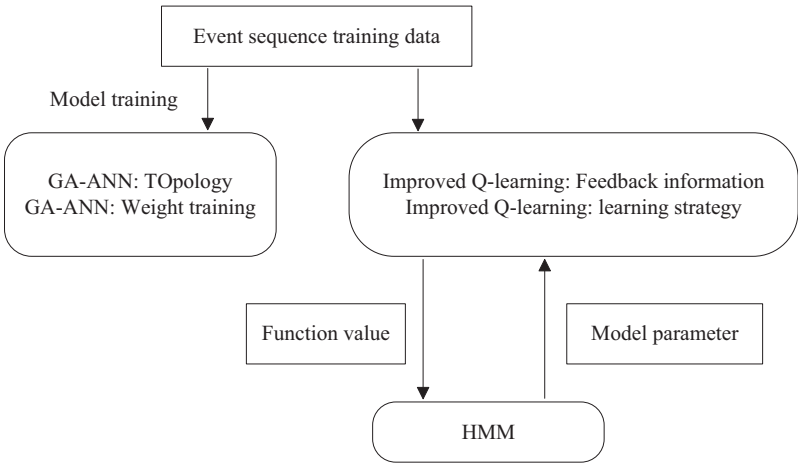


Figure 2 **Running process of training model.** Full-size DOI: 10.7717/peerj-cs.3451/fig-2

Among the models used, the artificial neural network (ANN) is particularly effective at managing variable correlations. However, in time series forecasting problems, determining these correlations can be challenging. The HMM, on the other hand, is used in pattern recognition tasks, where it learns from an observation sequence to estimate model parameters and then decodes to uncover the hidden state patterns corresponding to that sequence. This approach allows the HMM to predict nonlinear, multidimensional time series effectively and reduces errors caused by system randomness. To address the high noise, non-smoothness, and nonlinearity in e-commerce review time series, this article proposes a hybrid intelligent prediction model that integrates an ANN and an HMM. The output of the ANN is used as the observation probability input to the HMM emission layer. This approach effectively transforms the ANN's nonlinear feature outputs into the HMM's emission distribution and improves the accuracy of forward multi-step time series predictions.

The hybrid intelligent prediction model comprises two stages: a training process and a prediction process. During training, historical data is used to train both the ANN and HMM models independently. In the prediction process, the model combines historical data with predictions generated by the trained ANN, integrating them through an improved reinforcement learning mechanism to produce final results. The overall process of training the intelligent prediction model is illustrated in Fig. 2, while the consumer behavior prediction model is depicted in Fig. 3.

In this article, the ANN model employs a three-layer feed-forward network, with the Genetic Algorithm (GA) utilized to determine and optimize the network topology. To address the common limitations of the backpropagation algorithm, such as its tendency to fall into local optima and its slow convergence, the GA is also used as a training algorithm for adjusting the neuron connection weights. The input layer takes multiple historical time series for comparison, with two configurations in this study: one with eight nodes and another with 60 nodes. The neuron activation function is the sigmoid function, and the

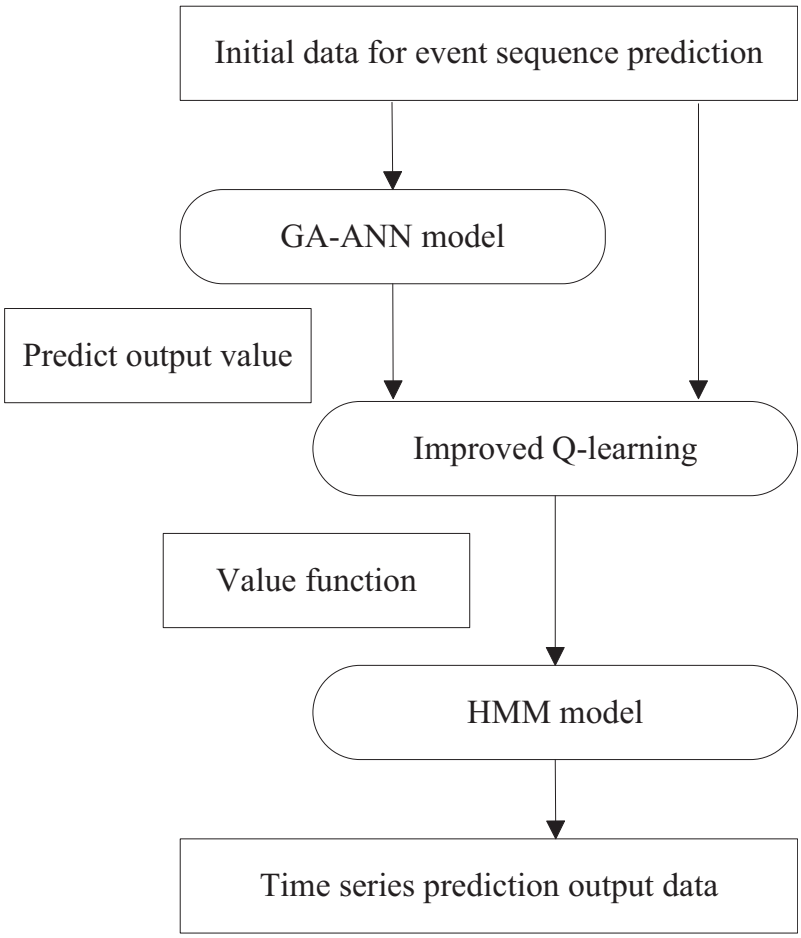


Figure 3 **Running process of prediction model.** Full-size DOI: 10.7717/peerj-cs.3451/fig-3

output layer consists of a single neuron that generates the GA-ANN model's prediction. The output layer provides predictions for the next day's data at each iteration. During the training of the hybrid intelligent prediction model, the GA-ANN submodel learns to determine the network topology and weights, achieving a more accurate fit to the historical time-series data and better reflecting the volatility of the actual data series.

During the prediction process, the trained GA-ANN sub-model is used alongside the initial prediction data as inputs to the HMM model. The improved reinforcement learning algorithm acts as a bridge, combining the GA-ANN sub-model with the HMM model through its greedy strategy. This approach ultimately yields a more accurate multi-step forward-time-series prediction, as depicted in Fig. 4. The parameter estimation process within the model remains unchanged, with the hidden state corresponding to the probability distribution of the consumption strategy and its performance within the e-commerce scenario. Let c denote the number of Gaussian components, w the mixture weight, $b_{ji}(w_i(t))$ a normal distribution with a mean vector u_{ji} and a covariance matrix Σ_{ji}.

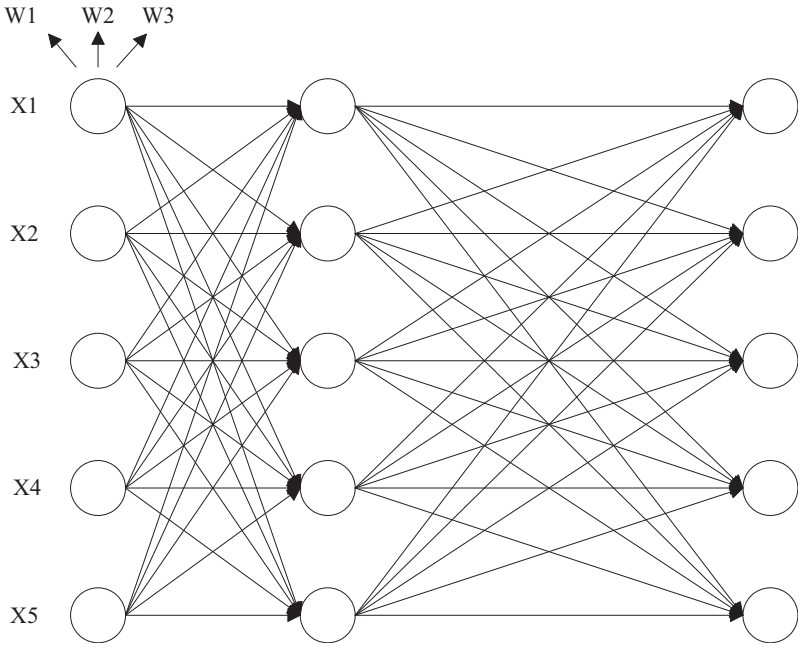


Figure 4 **HMM model structure.** Full-size DOI: 10.7717/peerj-cs.3451/fig-4

The emission function of the hidden state is modeled using a Gaussian mixture distribution, with its probability density function given by Eqs. (11) to (12):

$$b_{ij} = \sum_{i=1}^{c} w_{ji} b_{ji}(w_i(t)), j = 1, 2, \ldots, n \tag{11}$$

$$\sum_{i=1}^{c} w_{ji} = 1, w_{ji} \geq 0, j = 1, 2, \ldots, n \tag{12}$$

$$b_{ji}(w_i(t)) = N\left(w_i(t), u_{ji}, \Sigma_{ji}\right). \tag{13}$$

After determining the parameters of the HMM model by performing the learning task, the probability of the observation sequence that is most likely to reach the hidden state $\delta_j(t)$ and the hidden state corresponding to the most likely observation sequence $\psi_j(t)$ are selected, and then, based on the initial state, the sequence of hidden states and the predicted observation sequence are iteratively calculated as shown in Eqs. (12) to (13):

$$\delta_j(t) = \max_{1 \leq i \leq n} \delta_i(t-1)\alpha_{ji} \cdot \sum_{i=1}^{m} w_{ji}\left(w(t), u_{ji,\Sigma_{ji}}\right) \tag{14}$$

$$\psi_j(t) = \arg \max_{1 \leq i \leq n} \delta_i(t-1)\alpha_{ji}. \tag{15}$$

Finally, we integrate the improved Q-learning algorithm into the HMM submodel, yielding the QL-ANN-HMM model. In this model, the payoff returns of the improved Q-learning algorithm are derived from historical observation data, and the action that maximizes the payoff is executed through a greedy learning strategy. This approach

leverages the ANN's ability to fit output data to the volatility in the prediction results, while simultaneously mitigating the random systematic errors inherent in the HMM.

Data preprocessing

Before model training, both the Amazon Reviews 2023 and Flipkart Reviews datasets underwent systematic preprocessing to address issues of noise, imbalance, and sparsity:

Text cleaning—All review texts were tokenized, converted to lowercase, and stripped of stop words, punctuation, and special symbols. Emojis and non-ASCII characters were removed to maintain encoding consistency.

Normalization—Numerical variables (*e.g.*, review length, rating scores, helpfulness counts) were normalized to the [0, 1] range using Min-Max scaling to reduce bias in the Q-learning reward function.

Time series structuring—Review entries were ordered chronologically. A sliding-window segmentation strategy (window size = 30 days, stride = 7 days) was applied to transform review streams into multivariate time series suitable for ANN-HMM modeling.

Noise filtering—Extremely short reviews (<5 words) and duplicated entries were discarded to reduce random noise. In addition, z-score filtering ($|z| > 3$) was applied to numerical attributes to mitigate outliers.

Train–test split—For Flipkart, the dataset was split into 30,000 for training and 10,000 for testing; for Amazon Reviews, random stratified sampling was used to construct a comparable 80/20 train–test split.

Feature encoding—Sentiment polarity and semantic embeddings were extracted using a pre-trained BERT encoder, providing input features for the ANN. Simultaneously, categorical variables were one-hot encoded to be compatible with the HMM emission probabilities.

Computing infrastructure

All experiments were conducted on a workstation running Ubuntu 22.04 LTS (64-bit) as the primary operating system. The hardware environment consisted of an Intel(R) Core (TM) i9-13900K CPU @ 3.00 GHz, 128 GB DDR5 RAM, and an NVIDIA RTX 4090 GPU with 24 GB VRAM to accelerate neural network training. The experimental framework was implemented in Python 3.10 using open-source libraries, including PyTorch 2.0 for model development, NumPy 1.25 and Pandas 2.1 for numerical computation and data management, and scikit-learn 1.3 for preprocessing and evaluation metrics. The HMM modules were developed using the hmmlearn package, while evolutionary optimization procedures (Genetic Algorithm) were implemented using the deap library. Visualization of results was carried out using Matplotlib 3.7 and Seaborn 0.12.

EXPERIMENTS AND ANALYSIS

In this section, we evaluate the performance of the proposed QL-ANN-HMM model by comparing it with other models, focusing on the convergence of the improved QL algorithm and additional performance metrics.

Experimental data

The Amazon Reviews 2023 and Flipkart Reviews datasets were analyzed experimentally. The Amazon Reviews 2023 dataset, collected by McAuley Labs, is an updated version of the original Amazon Reviews dataset. It contains over 570 million reviews and 48 million items spanning 33 different categories. Flipkart, a prominent e-commerce platform in India, is the source of the Flipkart Reviews dataset, which contains user reviews from the platform. This dataset is divided into a training set of 30,000 samples and a test set of 10,000 samples. The total dataset size is 8,012,850 bytes, with a download size of 1,355,637 bytes.

Evaluation criteria

To evaluate the performance of the improved Q-learning algorithm and the QL-ANN-HMM model, we utilize three key evaluation metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Normalized Mean Squared Error (NMSE). These metrics are used to optimize prediction accuracy and are defined as follows.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i'| \tag{16}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i'|}{y_i} \tag{17}$$

$$NMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(y_i - y_i')^2} \tag{18}$$

where y_i is the actual optimal consumer behavior and y_i' is consumer behavior solved by reinforcement learning.

Comparison of improved Q-learning performance

In this article, $Q(\lambda)$ is used for consumer behavior prediction optimization. To assess the iteration efficiency of the algorithms, we compare the traditional Q-Learning algorithm (*Ernst et al., 2024*) with the improved Q-learning algorithm presented in this article. Figure 5 illustrates the optimization of the reward function for both algorithms over iteration steps. From Fig. 5, it is evident that the improved Q-learning algorithm shows a steep decline in the iteration curve at the initial stages. After approximately 50 iterations, the R-value stabilizes, and the final convergence of the reward function is significantly better than that of the traditional Q-Learning algorithm.

Additionally, this article compares the performance of four algorithms in predicting e-commerce consumer behavior: the Genetic Algorithm (GA) (*Alhijawi & Awajan, 2024*), Quantum-Inspired Genetic Algorithm (QGA) (*Sadeghi Hesar & Houshmand, 2024*), Q-learning (*Ernst et al., 2024*), and the Improved Q-learning (IQ-learning) proposed in this study. To verify the effectiveness of these algorithms, we conducted experiments on two datasets: Amazon Reviews 2023 and Flipkart Reviews. We statistically analyzed the performance metrics of each model on these datasets, and the results are presented

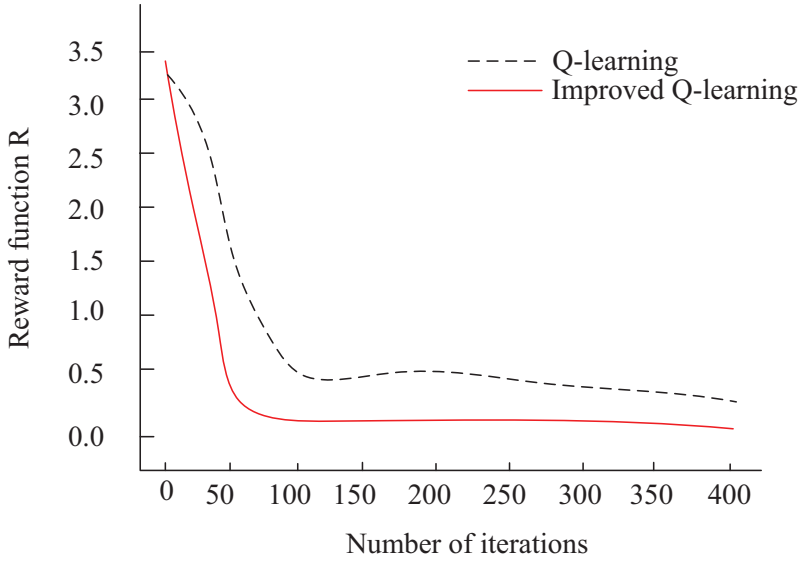
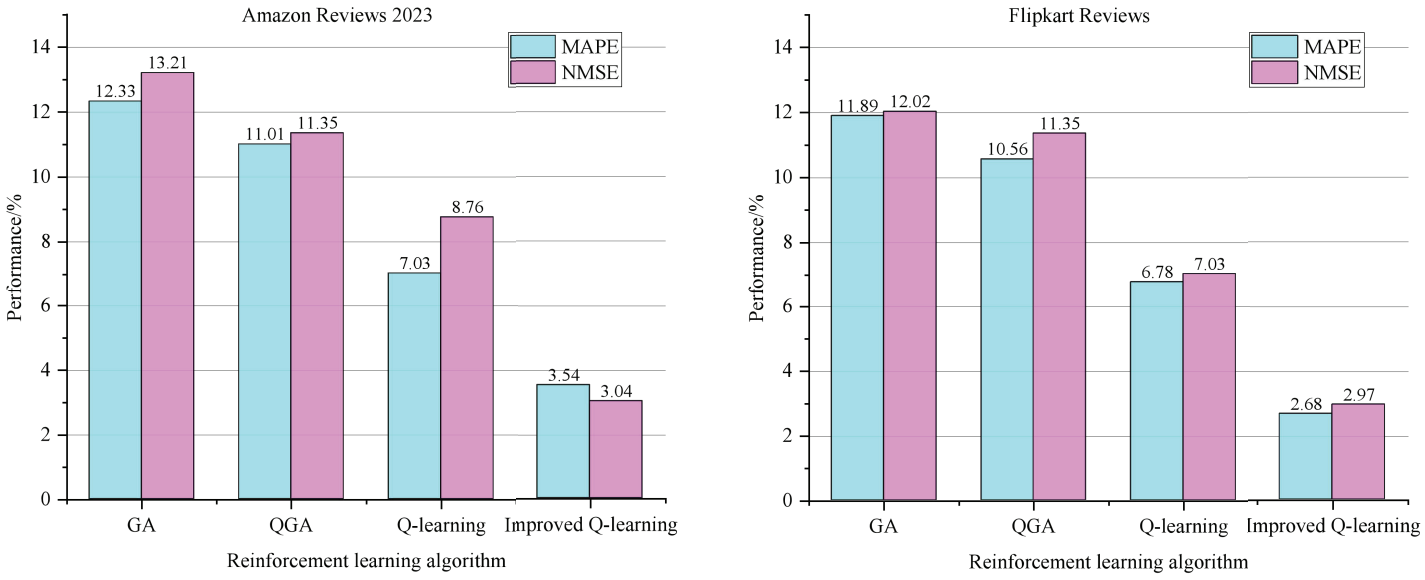


Figure 5 **Iteration *vs.* Reward (R-value).** Full-size DOI: 10.7717/peerj-cs.3451/fig-5



（a）Amazon Reviews 2023　　（b）Flipkart Reviews

Figure 6 **The performance index of different algorithms.** (A) Amazon Reviews 2023. (B) Flipkart Reviews.
Full-size DOI: 10.7717/peerj-cs.3451/fig-6

in Fig. 6. Performance indices (MAE, MAPE, NMSE, %) of different algorithms. Specifically, for the Amazon Reviews 2023 dataset, the Reinforcement Learning Algorithm (and its improved version) proposed in this article achieved the lowest values for both MAPE and NMSE, at 3.54% and 3.04%, respectively, demonstrating a significant advantage over the other algorithms. On the Flipkart Reviews dataset, although QGA

performed similarly on some metrics, the reinforcement learning algorithm maintained its leading position, with MAPE and NMSE at 2.68% and 2.97%, respectively.

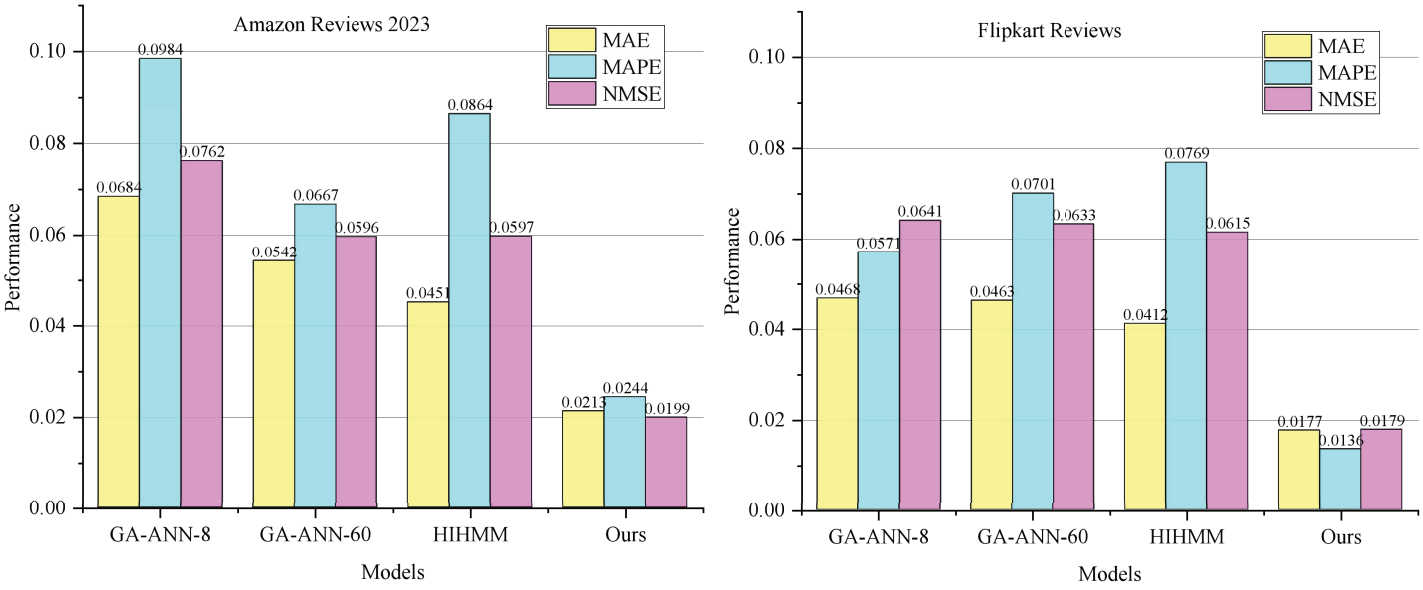
In contrast, the GA algorithm performs the poorest on both datasets. This is primarily due to the limitations of traditional optimization algorithms, such as genetic algorithms, which are constrained by factors such as population evolution strategies and population size selection. These parameters often require significant engineering expertise from algorithm designers, making it challenging to achieve optimal configurations in practical applications. As a result, such algorithms face challenges when applied to complex scenarios, such as predicting consumer behavior in e-commerce. In comparison, reinforcement learning algorithms demonstrate greater adaptability and flexibility. In practice, it is only necessary to standardize the mapping relationships between variables and algorithm parameters within a trend framework, enabling independent calculation of consumer behavior. This adaptability makes reinforcement learning algorithms particularly well-suited for complex scenarios, such as predicting consumer behavior in e-commerce.

Moreover, when comparing the traditional Q-learning algorithm with the improved Q-learning algorithm proposed in this article, we observe a significant improvement in prediction accuracy after introducing multi-step iteration. Specifically, on the Amazon Reviews 2023 dataset, the improved Q-learning algorithm enhances the MAPE metric by 2.71% and the NMSE metric by 5.96% compared to the traditional Q-learning algorithm. These results suggest that the performance of reinforcement learning algorithms for predicting consumer behavior in e-commerce can be further improved by refining the algorithms and optimizing the iteration strategy.

Performance analysis of QL-ANN-HMM

Figure 7 shows the prediction results for MAE, MAPE, and NMSE of the four models—GA-ANN-8 (*Dhawale, Kamboj & Anand, 2023*), GA-ANN-60 (*Słowik & Cpałka, 2021*), HiHMM (*Sohn et al., 2015*; *Wu et al., 2025*), and QL-ANN-HMM—on the two datasets.

Figure 7 demonstrates that the QL-ANN-HMM model proposed in this article exhibits clear advantages across all three evaluation metrics on both datasets. For the same prediction sequence, varying the training length has relatively little impact on the QL-ANN-HMM model's prediction outcomes. Preliminary analysis suggests that this is primarily because the datasets provide sufficient training data to fit the GA-ANN model accurately. Specifically, the prediction results from the GA-ANN-60 model, with 60 input layer nodes, significantly outperform those from the GA-ANN-8 model, which has only 8 input layer nodes. The choice of 8 and 60 input nodes was determined empirically through cross-validation on the training dataset. A grid search over {8, 16, 32, 60, 100} input nodes showed that 60 provided the best trade-off between overfitting risk and prediction accuracy. This indicates that increasing the length of the historical time series used in the ANN can yield more valuable information. Consequently, in the QL-ANN-HMM model using GA-ANN-60 prediction data, even with different data sequence characteristics, the improved Q-learning algorithm's fitting effect enhances performance. This hybrid intelligent prediction model outperforms both the GA-ANN and HMM models



（a）Amazon Reviews 2023　　（b）Flipkart Reviews

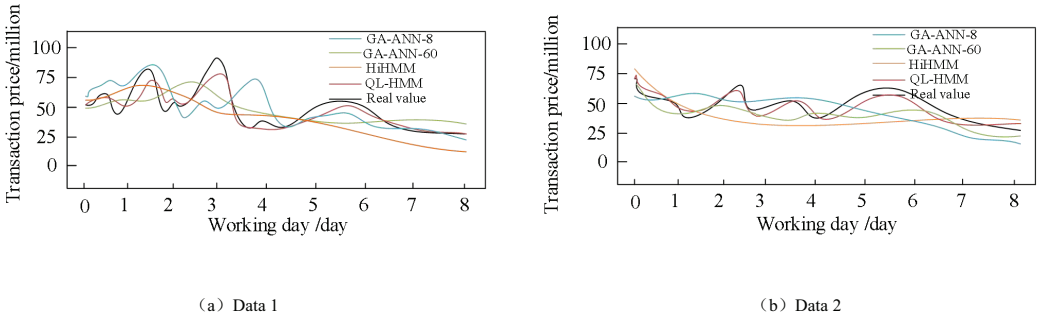
Figure 7 Comparative MAE, MAPE, NMSE results. (A) Amazon Reviews 2023. (B) Flipkart Reviews.
Full-size DOI: 10.7717/peerj-cs.3451/fig-7

individually. The test results from both datasets confirm that the QL-ANN-HMM model, which integrates GA-ANN and HMM using Q-learning, achieves superior error performance when employing GA-ANN-60.

To quantify the reduction of stochastic errors, we compared the residual distributions before and after applying the HMM. The residual variance decreased from 0.027 to 0.011 on the Amazon dataset and from 0.031 to 0.012 on the Flipkart dataset, confirming that the HMM effectively smooths random noise and enhances model stability.

Figure 8 visualizes the fit between the actual e-commerce platform business data and the predicted values from multiple time-series models for eight working days backward. From the figure, it is clear that the QL-ANN-HMM model in this article shows significant advantages in predicting e-commerce platform business. The strong alignment between its predicted curves and the actual values indicates that QL-ANN-HMM possesses an excellent ability to capture the business trends of e-commerce platforms. In contrast, although the traditional HiHMM demonstrates some predictive capability, its results tend to be relatively smooth and lack sufficient volatility, which somewhat limits its adaptability to complex market fluctuations.

It is worth noting that both GA-ANN-60 and QL-ANN-HMM demonstrate similar strengths in predicting trend changes. Both models are capable of capturing the fluctuating trends in e-commerce platform operations with greater accuracy, resulting in superior predictive performance. This further underscores the significant potential of intelligent algorithms for time series prediction. Overall, the QL-ANN-HMM model outperforms other models in terms of both prediction accuracy and stability, providing more reliable



（a）Data 1　　（b）Data 2

Figure 8 Comparison between predicted and actual transaction prices. (A) Data 1. (B) Data 2.
Full-size DOI: 10.7717/peerj-cs.3451/fig-8

and precise predictive support for e-commerce platforms. These findings not only validate the effectiveness and reliability of the ARTHMS model but also provide robust data to support the development of future operational strategies for e-commerce platforms.

To accelerate convergence and reduce estimation variance, this work introduces a multi-step instantaneous difference iteration method on top of standard Q-learning. The multi-step return integrates reward information over several future steps in the target computation, thereby reducing bias caused by single-step bootstrap errors. Meanwhile, because the discount factor attenuates long-term returns, the variance of the multi-step target is confined within a finite range, achieving a balance between bias and variance. Experimental results show that this method significantly improves learning efficiency while maintaining accuracy.

CONCLUSION

This study proposed an improved reinforcement learning algorithm and a hybrid QL-ANN-HMM model for predicting consumer behavior from online e-commerce reviews. By incorporating a probabilistic action-selection mechanism into Q-learning, the model's convergence and stability were enhanced. Furthermore, integrating ANN and HMM with reinforcement learning improves prediction stability. Experimental results on the Amazon Reviews 2023 and Flipkart Reviews datasets confirmed that the improved Q-learning algorithm achieved a 2.71% reduction in MAPE and 5.96% reduction in NMSE, while the hybrid QL-ANN-HMM model consistently outperformed baseline methods across MAE, MAPE, and NMSE. These findings demonstrate the potential of combining reinforcement learning with time-series modeling to enhance e-commerce consumer behavior prediction.

LIMITATIONS

Despite these contributions, certain limitations should be acknowledged. The model was validated on only two datasets, which may limit its generalizability. The performance gains, although consistent, were relatively modest, and the hybrid architecture increased computational costs without a systematic analysis of efficiency–accuracy trade-offs. Moreover, while the results showed close alignment with actual consumer behavior trends,

the model's robustness to sudden anomalies or irregular behavioral patterns remains insufficiently examined.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers whose comments and suggestions helped to improve the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work is funded by "2025 Fujian Province Social Science Strength Construction of Hundred Experts Survey Research Project Stage Research Results", the project number is FJSK25DY028. This work is also funded by Interim Research Findings of the 2025 Fujian Provincial Department of Science and Technology Natural Science Foundation Program, the project number is 2025J08355. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
2025 Fujian Province Social Science Strength Construction of Hundred Experts Survey Research Project Stage Research Results: FJSK25DY028.
Interim Research Findings of the 2025 Fujian Provincial Department of Science and Technology Natural Science Foundation Program: 2025J08355.

Competing Interests

The authors declare that there are no financial or personal relationships that could be viewed as potential competing interests. Honglei Zhang is employed by Beijing Meitu Home Technology Co., Ltd. This affiliation is disclosed in accordance with the journal's transparency policy, and it does not give rise to any conflicts of interest with the research presented in this manuscript.

Author Contributions

- Zongping Lin conceived and designed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Yingyi Huang conceived and designed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Jing Yang performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Chunhu Cui analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Yabin Lian performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Honglei Zhang analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

- Fadi Al-Turjman conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The Amazon Reviews are available at Zenodo: Kashnitsky, Y. (2022). Amazon product reviews (mock dataset) (1.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6657410.

The Flipkart Reviews are available at Zenodo: None. (2025). Flipkart Review Dataset [Data set]. Zenodo. https://doi.org/10.5281/zenodo.17612521.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3451#supplemental-information.

REFERENCES

Ahmed MI, Kumar R. 2023. Nodal electricity price forecasting using exponential smoothing and Holt's exponential smoothing. *Distributed Generation & Alternative Energy Journal* **38**:1505–1530 DOI 10.13052/dgaej2156-3306.3857.

Alhijawi B, Awajan A. 2024. Genetic algorithms: theory, genetic operators, solutions, and applications. *Evolutionary Intelligence* **17(3)**:1245–1256 DOI 10.1007/s12065-023-00822-6.

Aminullah AA. 2024. *Penerapan metode seasonal autoregressive intergrated moving average pada peramalan penjualan barang toko bahan bangunan sinar pagi*. Indonesia: UPN Veteran Jawa Timur.

Chen T, Samaranayake P, Cen XY, Qi M, Lan YC. 2022. The impact of online reviews on consumers' purchasing decisions: evidence from an eye-tracking study. *Frontiers in Psychology* **13**:865702 DOI 10.3389/fpsyg.2022.865702.

Cheng J, Tiwari S, Khaled D, Mahendru M, Shahzad U. 2024. Forecasting Bitcoin prices using artificial intelligence: combination of ML, SARIMA, and Facebook Prophet models. *Technological Forecasting and Social Change* **198(44)**:122938 DOI 10.1016/j.techfore.2023.122938.

Costa P, Rodrigues H. 2024. The ever-changing business of e-commerce-net benefits while designing a new platform for small companies. *Review of Managerial Science* **18(9)**:2507–2545 DOI 10.1007/s11846-023-00681-6.

Dai Y, Tong X, Jia X. 2024. Executives' legal expertise and corporate innovation. *Corporate Governance: An International Review* **32(6)**:954–983 DOI 10.1111/corg.12578.

Dhawale D, Kamboj VK, Anand P. 2023. An optimal solution to unit commitment problem of realistic integrated power system involving wind and electric vehicles using chaotic slime mould optimizer. *Journal of Electrical Systems and Information Technology* **10(1)**:4 DOI 10.1186/s43067-023-00069-2.

Dong X, Dang B, Zang H, Li S, Ma D. 2024. The prediction trend of enterprise financial risk based on machine learning arima model. *Journal of Theory and Practice of Engineering Science* **4(1)**:65–71.

Ernst D, Louette A, Feuerriegel S, Hartmann J, Janiesch C, Zschech P. 2024. Introduction to reinforcement learning. 111–126. *Available at https://damien-ernst.be/wp-content/uploads/2025/04/introduction_to_reinforcement_learning_v250407.pdf*.

Harrou F, Zeroual A, Kadri F, Sun Y. 2024. Enhancing road traffic flow prediction with improved deep learning using wavelet transforms. *Results in Engineering* **23**:102342 DOI 10.1016/j.rineng.2024.102342.

Ito Y, Fujimoto K. 2022. Kernel-based hamilton-jacobi equations for data-driven optimal control: the general case. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **105(1)**:1–10 DOI 10.1587/transfun.2021eai0002.

Jin J, Liu Q, Yang Y, Ma B, Pan Y. 2025. How social crowding impacts mobile shopping: a perspective from information processing. *Information & Management* **62(7)**:104197 DOI 10.1016/j.im.2025.104197.

Jin B, Xu X. 2024. Price forecasting through neural networks for crude oil, heating oil, and natural gas. *Measurement: Energy* **1(1)**:100001 DOI 10.1016/j.meaene.2024.100001.

Li H, Xia C, Wang T, Wang Z, Cui P, Li X. 2024. GRASS: learning spatial-temporal properties from chainlike cascade data for microscopic diffusion prediction. *IEEE Transactions on Neural Networks and Learning Systems* **35(11)**:16313–16327 DOI 10.1109/tnnls.2023.3293689.

Liu M, Cai Q, Li D, Meng W, Fu M. 2023. Output feedback Q-learning for discrete-time finite-horizon zero-sum games with application to the H∞ control. *Neurocomputing* **529(7587)**:48–55 DOI 10.1016/j.neucom.2023.01.050.

Liu S, Zhou N, Song C, Chen G, Wu Y. 2024. Exhaust gas temperature prediction of aero-engine via enhanced scale-aware efficient transformer. *Aerospace* **11(2)**:138 DOI 10.3390/aerospace11020138.

Lyu D, Wang Z, Kumar A, Jin J. 2025. Morality is for social being: the role of morality in social-adjustive functional attitudes toward counterfeit luxury consumption. *Journal of Business Ethics* **44(5)**:687 DOI 10.1007/s10551-025-06027-4.

Okorie IE, Afuecheta E, Nadarajah S. 2023. Time series and power law analysis of crop yield in some east African countries. *PLOS ONE* **18(6)**:e0287011 DOI 10.1371/journal.pone.0287011.

Pan CL, Liu Y, Pan YC. 2022. Research on the status of e-commerce development based on big data and Internet technology. *International Journal of Electronic Commerce Studies* **13(2)**:27–48 DOI 10.7903/ijecs.1977.

Ren H, Jiang B, Ma Y. 2022. Zero-sum differential game-based fault-tolerant control for a class of affine nonlinear systems. *IEEE Transactions on Cybernetics* **54(2)**:1272–1282 DOI 10.1109/tcyb.2022.3215716.

Sadeghi Hesar A, Houshmand M. 2024. A memetic quantum-inspired genetic algorithm based on tabu search. *Evolutionary Intelligence* **17(3)**:1837–1853 DOI 10.1007/s12065-023-00866-8.

Sohn K-A, Ho JWK, Djordjevic D, Jeong H-H, Park PJ, Kim JH. 2015. hiHMM: Bayesian non-parametric joint inference of chromatin state maps. *Bioinformatics* **31(13)**:2066–2074 DOI 10.1093/bioinformatics/btv117.

Słowik A, Cpałka K. 2021. Hybrid approaches to nature-inspired population-based intelligent optimization for industrial applications. *IEEE Transactions on Industrial Informatics* **18(1)**:546–558 DOI 10.1109/tii.2021.3067719.

Wang T, Ngoduy D, Li Y, Lyu H, Zou G, Dantsuji T. 2024. Koopman theory meets graph convolutional network: learning the complex dynamics of non-stationary highway traffic flow for spatiotemporal prediction. *Chaos, Solitons & Fractals* **187(9)**:115437 DOI 10.1109/itsc58415.2024.10919501.

Wu X, Li L, Tao X, Yuan J, Xie H. 2025. Towards the explanation consistency of citizen groups in happiness prediction via factor decorrelation. *IEEE Transactions on Emerging Topics in Computational Intelligence* **9(2)**:1392–1405 DOI 10.1109/TETCI.2025.3537918.