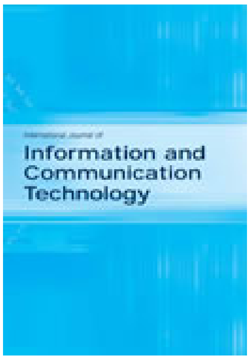


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

An intelligent analysis model for enhancing rural e-commerce sales efficiency in live streaming environments

Rong Zhou

DOI: 10.1504/IJICT.2025.10073714

Article History:

Received:	21 July 2025
Last revised:	01 September 2025
Accepted:	02 September 2025
Published online:	16 October 2025

An intelligent analysis model for enhancing rural e-commerce sales efficiency in live streaming environments

Rong Zhou

Business College,
Zhengzhou Railway Vocational and Technical College,
Zhengzhou, 450000, China
Email: zhourong1361@sina.com

Abstract: Amidst the swift advancement of internet technology, live broadcast e-commerce has emerged as a pivotal driver for rural economic growth. However, the issue of low sales efficiency is becoming increasingly conspicuous, necessitating the development of effective solutions. Against this backdrop, this study investigates strategies to enhance the sales efficiency of rural e-commerce within the live broadcast context, with a focus on devising an intelligent analysis model to address the sales challenges. By precisely analysing the vast amounts of multi-source data generated during live broadcasts, the model employs data mining, machine learning, and deep learning algorithms for in-depth association analysis and feature extraction. Experimental results demonstrate that compared with outstanding deep learning models, the REISE model has reduced errors by approximately 25.60% and 22.31% in MAE metrics, 49.46% and 46.77% in MSE metrics, and 30.12% and 29.27% in RMSE metrics, respectively.

Keywords: rural e-commerce; live broadcast sales; intelligent analysis model; sales efficiency; data mining.

Reference to this paper should be made as follows: Zhou, R. (2025) 'An intelligent analysis model for enhancing rural e-commerce sales efficiency in live streaming environments', *Int. J. Information and Communication Technology*, Vol. 26, No. 37, pp.106–122.

Biographical notes: Rong Zhou received her PhD from Stamford International University in 2024. She is currently an Associate Professor at the Business College of Zhengzhou Railway Vocational and Technical College. Her research interests include e-commerce, online promotion, and channel management.

1 Introduction

In the late 1960s, Thomas, a famous economist, first proposed the concept of e-commerce. At that time, the interpretation of e-commerce was to complete the process of product sales based on advanced information technology, including product design, production and support. With the continuous development of big data, VR, 5G communication and other internet technologies, as well as the popularity of mobile terminals, webcast has risen in the internet field in recent years, showing a trend of breaking through, rapidly penetrating into many fields such as politics, economy, culture, and involving different industries (Wang et al., 2022). Since 2003, the live broadcast of video games has been broadcast on TV, although at that time, the live broadcast was mostly used as a means of real-time traffic communication, and the content and theme of the live broadcast were relatively single. Since then, from 2010 to 2015, live broadcasting has entered the stage of savings and development, and a large number of live broadcasting apps have been launched one after another. Until 2016, the Taobao live studio was born, and the live broadcast function was launched quickly. So far, the first year of live webcast has been ushered in. Hundreds of live webcast products have sprung up, and the number of users watching live webcast has continued to grow (Xin et al., 2023). Since its birth, the live broadcasting industry has been actively seeking diversified development, constantly combining with other content, involving many fields, including games, education and so on, forming the concept of 'live broadcasting+'. In the past two years, it has risen in an all-round way, bursting out huge economic potential. With its brand-new mode, live broadcasting with goods has led to the emergence of a new marketing mode and communication mode. Some scholars believe that this is a kind of choice behaviour, that is, farmers consider whether to adopt e-commerce platform for promotion according to their own and family conditions and combined with agricultural products.

In modern society, the meteoric rise of e-commerce has greatly benefited people by offering convenience and opening up new avenues for increasing income. Regarding the development of e-commerce in rural areas, scholars hold diverse views but generally acknowledge its crucial role in advancing the e-commerce industry. The rise of rural e-commerce sales provides an effective solution to the challenges of long-distance quality inspection and pricing of agricultural products. When consumers need to purchase agricultural products from a long distance, the quality of agricultural products cannot be tested on site, and various market prices cannot be compared. Due to the limited time and resources of buyers, they are unable to visit each farm. Folorunso et al. (2006) highlighted the importance of employing agricultural e-commerce to facilitate suitable business activities between farmers and buyers in the online arena. Moreover, through agricultural e-commerce platforms, consumers are able to gain access to more favourable pricing options. Pramanik et al. (2017) proposed that information and communication technology (ICT) has an important impact on rural development, aiming to study the role and contribution of ICT in promoting rural economic development, social progress and narrowing the digital divide between urban and rural areas in developing countries. Through in-depth analysis of ICT application status and typical cases in rural areas of different developing countries, this study reveals the development opportunities, challenges and constraints of ICT in rural education, health care, agriculture, entrepreneurship and other fields, such as inadequate infrastructure, lack of digital skills and imperfect policy support. Rural e-commerce sales pay a low cost and realise the connection between a large number of production enterprises and a large number of consumer groups. It not only enables agricultural products to enter the city, but also enables industrial products to flow into the countryside, which not only enables farmers to earn money, but also allows farmers to save money. Taking China as the research object, Kshetri (2018) pointed out that the governments of less developed countries must pay attention to the development of rural e-commerce sales, because rural e-commerce sales can bring employment opportunities and optimise rural people's livelihood. Ruiz-Garcia et al. (2010) analysed the real-time logistics problems and route planning of online trading products in rural areas, and proposed using information network system to improve the efficiency of rural e-commerce. Ruch and Sackmann (2012) found that most of the decline in orders for online platforms is due to the risk of deception brought to people during payment. Sellers of online platforms should pay attention to this problem. Canavari et al. (2010) mentioned that consumers' perceived risks of products can cause complications in agricultural product e-commerce development. To ensure smooth online sales of agricultural products, it is essential to address this issue. Sun and Zhang (2006) proposed that there are moderating factors in the process of user's technology acceptance, aiming to study how these moderating factors affect user's technology acceptance behaviour. Through empirical research, this paper reveals the impact mechanism of moderating factors on user's technology acceptance in different situations, such as the impact of individual characteristics, task characteristics and environmental characteristics on technology acceptance. Davis (1989) argued that perceived usefulness and ease of use are crucial in determining users' acceptance of information technology. He aimed to examine how these elements connect to users' acceptance behaviours. Zhou (2022) analysed the innovative approach to poverty alleviation through agricultural e-commerce live streaming. He looked at its benefits, obstacles, and practical implementation, aiming to discover how e-commerce live streaming can aid the economic growth of impoverished rural regions. Luo et al. (2021) examined the effect of persuasive language style on the sale of products in live-streaming within social e-commerce. Through empirical research, they revealed how different language styles affect consumers' purchase behaviour, providing a theoretical basis for improving the effect of live broadcast delivery. Wang et al. (2023) analysed the influencing factors of sustainable supply chain development for agricultural and food products in cross-border e-commerce live-streaming. They concentrated on the cross-border e-commerce live-streaming environment and studied the key influencing factors of each segment of the supply chain. The goal was to establish an efficient and stable cross-border e-commerce supply chain for agricultural goods. Using Wenchuan County as a case study, Liang and Cheng (2024) explored how e-commerce fosters agricultural innovation and entrepreneurship in a platform economy. They demonstrated that e-commerce can invigorate agricultural innovation, advance entrepreneurial efforts, and stimulate rural economic growth and farmers' income through diverse approaches. Li and Gan (2025) explored China's rural e-commerce sustainability from multiple dimensions. They looked at how economic, social, and environmental factors affect rural e-commerce growth. The goal was to create a scientific evaluation system and improvement path for rural e-commerce sustainability.

Live e-commerce, with its instant interaction and live presentation, has become a new channel for agricultural product sales. However, accurately grasping consumer demand and optimising product presentation and sales strategies remain pressing issues. Thus, designing an intelligent analysis model is crucial for enhancing the sales efficiency of rural e-commerce in live broadcasting. The model will comprehensively consider the multi-dimensional data in the live broadcast process, such as audience interaction, product display and anchor sales scripts, and use advanced data analysis technology to provide real-time decision support for rural e-commerce live broadcast. Specifically, by analysing the real-time feedback of the audience, the model accurately grasped the needs of consumers and provided data support for the optimisation of product display and sales strategy. At the same time, the model will also analyse the sales script of the anchor to help the anchor improve sales skills and enhance the interaction effect with the audience. The model also aims to boost agricultural product sales and support the sustainable growth of rural e-commerce in the live-streaming era. It achieves this by forecasting sales trends to inform product selection, improving audience conversion and average spending. Additionally, the model optimises inventory to cut overstock costs and provides momentum for rural e-commerce development. Although specific quantitative metrics are not provided in this study, the REISE model contributes to inventory cost reduction through accurate sales trend forecasting and optimised inventory planning, effectively addressing overstock issues.

The main innovations and contributions of this work include:

1. This paper creatively constructs a multi-dimensional data analysis model, which comprehensively considers multi-source data such as audience interaction, product display and anchor sales scripts, and deeply mines consumer demand and behaviour characteristics through machine learning algorithm. Compared with the traditional single dimension data analysis, this model can more accurately describe consumer behaviour patterns, provide data support for optimising product display strategies and adjusting sales scripts, and effectively improve the live broadcast sales efficiency and conversion rate.

2. Aiming at the problem that the live broadcast sales strategy is difficult to adjust in real-time, this paper proposes a dynamic optimisation mechanism based on real-time feedback. By monitoring the interaction behaviour and emotional changes of the audience, the mechanism can quickly capture the changes of consumer preferences, and adjust the product recommendation and promotion strategies in real-time. For example, when the audience shows high attention to the specific selling points of a certain agricultural product, it can immediately strengthen the publicity and launch relevant discounts to stimulate the desire to buy and reduce the loss of sales opportunities.

3. This study introduces a deep learning based predictive approach to enhance agricultural product stock control. It leverages live-streaming sales data and market patterns to forecast agricultural product sales. This helps rural e-commerce businesses plan their stock better and cut overstock costs. Also, when combined with logistics data, this approach can optimise supply chain management, boost the circulation efficiency of agricultural products, guarantee their timely supply, meet consumer needs, and strengthen the competitiveness of rural e-commerce in the live-streaming setting.

2 Relevant technologies

2.1 GBDT gradient lifting tree

Ensemble learning refers to the combination and integration of several base learners to achieve better prediction results than a single learner (Zhang and Jung, 2020). The idea of boosting algorithm of ensemble learning is to promote the base learning machine with poor prediction effect to a strong learning machine step by step through a certain learning mechanism, which is a serial learning method. The boosting algorithm first trains the dataset to get a weak learner, and calculates the training error at the same time. In the next round of learning, it increases the weight of the sample with large training error, so that the learner pays more attention to the learning of this part of the sample. It iterates continuously in the case of changing the sample weight in each round, and the error becomes smaller and smaller. Finally, it fuses all the learners to get more accurate prediction results. The representative algorithm of boosting is GBDT algorithm and various improved algorithms based on GBDT algorithm. GBDT is an ensemble algorithm based on boosting principles. It typically employs classification and regression tree decision trees as base classifiers, training them through iterative refinement of residuals (Du et al., 2024). Ultimately, these classifiers are combined, with each assigned a weight, to create a robust and powerful classifier. The t^{th} base learner in GBDT targets the residual from the $t - 1$st learner, creating interdependence among trees. GBDT's objective function is its loss function, with optimisation aiming to minimise this loss. Gradient descent shows the loss decreases fastest when moving against its gradient. Unlike AdaBoost, which adjusts sample weights based on classification accuracy, GBDT maintains uniform sample weights. For each GBDT tree, fitting the negative gradient achieves the steepest loss function decline (Zhang et al., 2019).

Assume that the training set sample is $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, where t, l, and $f(x)$ denote the model's maximum iterations, loss function, and final output strong learner, respectively. The GBDT algorithm steps are as follows:

Initially, the weak classifier is setup, with c denoting the classifier's output value. In the context of mean square loss, c corresponds to the sample dataset's average value:

$$f_0(x) = \arg\min_c \sum_{i=1}^{m} L(y, c) \tag{1}$$

where $L(y, f(x))$ signifies the GBDT algorithm's loss function. For each iteration $t = 1$ to T:

Calculate the negative gradient for each sample $i = 1$ to M:

$$r_{ti} = -\left[\frac{\partial L\left(y_i, f\left(x_i\right)\right)}{\partial f\left(x_i\right)}\right]_{f(x)=f_{t-1}(x)} \tag{2}$$

Fit the dataset $\{(x_i, r_{ti})\}$ $(i = 1, 2, \ldots, m)$ using a decision tree to obtain the t^{th} tree. Assume this tree has j leaf nodes, with the j^{th} leaf node set denoted as R_{tj} $(j = 1, 2, \ldots, J)$.

Determine the optimal fitting values for all leaf nodes j:

$$c_{tj} = \arg\min_c \sum_{x_i \in R_{tj}} L\left(y_i, f_{t-1}\left(x_i\right) + c\right) \tag{3}$$

The optimal fitting value of the j^{th} leaf node is c_{tj}. If the loss function is quadratic loss, then c_{tj} is the average of all samples at the j^{th} leaf node.

Get c_{tj} update strong learner:

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^{J} c_{tj} I\left(x \in R_{tj}\right) \tag{4}$$

The function $I(x \in R_{tj})$ equals 1 if the sample is in set R_{tj}; otherwise, it equals 0. This helps form the final strong learner model.

$$f(x) = f_0(x) + \sum_{t=1}^{T} \sum_{j=1}^{J} c_{tj} I\left(x \in R_{tj}\right) \tag{5}$$

2.2 SHAP feature selection

Feature selection aims to identify features closely tied to prediction outcomes, isolating key influencing factors. It removes less impactful, redundant features to achieve dimensionality reduction. Effective feature selection enhances model speed, simplifies complexity, and boosts generalisation and prediction performance. SHAP, a game theory-based method, explains any model's output by assessing feature marginal effects (Wang et al., 2024). Using Shapley values, it clarifies individual sample contributions to predictions, differing from overall dataset importance analyses.

SHAP assesses feature importance by calculating each feature's Shapley value for predictions. The Shapley value of a feature is the difference between the original set's contribution and the contribution after adding new features (Le et al., 2022). To calculate the Shapley value of feature D, given an existing feature set s and its contribution $V(s)$, the Shapley value of D is determined as follows:

$$\phi_D = v(S \cup \{D\}) - v(S) \tag{6}$$

The simplest model explanation is the model itself. For instance, in a linear model, the coefficient of each feature serves as its weight, making it easy to comprehend. However, complex models like ensemble models and neural networks require simpler interpretive models for approximation. SHAP offers a model agnostic interpretive approach using additive feature attribution methods, defined as a linear function of binary variables.

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \tag{7}$$

in this context, g represents the interpretation model. M is the collection of features. ϕ_0 stands for the average predicted value based on the input data. ϕ_i is the Shapley value for the i^{th} feature, which is binary (0 or 1) to indicate the feature's presence or absence. The model's core goal is to compute these Shapley values. Per equation (6), this involves the feature set and all combinations of other features. The Shapley value is derived from the weighted sum of contributions across these combinations, reflecting each feature's prediction impact.

$$\phi_j = \sum_{S \subset \{x_1, \ldots, x_p\} / \{x_j\}} \frac{|S|!\left(p - |S| - 1\right)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S)\right) \tag{8}$$

in this equation, S denotes a subset of features utilised in the model. x represents the vector of eigenvalues for the sample being interpreted. p indicates the total number of features. $val(S)$ is the model's output value given the feature combination S. $(p)!$ signifies the number of combinations possible with p features. When a feature j is fixed, the remaining combinations are represented by $(p - 1)!/(|S|)!$ species. SHAP feature selection method has a solid theoretical basis and solves the problem of feature multi-collinearity. When calculating the Shapley value, it not only considers the influence of a single variable, but also considers the possible synergy between variables (Gramegna and Giudici, 2022).

2.3 Graph neural network

Graph neural network (GNN) aims to capture the complex association and structural characteristics between nodes in graph structure data (Scarselli et al., 2008). GNN typically employs a directed graph $G = (V, E)$ for data representation. In this structure, V symbolises the collection of nodes, while E denotes the set of edges. Each node $v_i \in V$ encapsulates node attributes within its eigenvector. After multi-layer representation learning, GNN can capture the neighbourhood information of different ranges, and apply the node representation to specific graph tasks. GNN learns the node representation through the message passing mechanism between nodes (Wu et al., 2020). This process can be formulated as follows:

$$a_v^{(l)} = Aggregate\left(\left\{h_u^{(l-1)} : u \in N(v)\right\}\right) \tag{9}$$

where $a_v^{(l)}$ is the aggregate neighbour information of the target node v at the l^{th} layer; $h_u^{(l-1)}$ represents the feature representation of neighbour node u at layer $l - 1$; $N(v)$ is the neighbour node set of target node v; the $Aggregate(.)$ function is an operation to aggregate the information of neighbour nodes. Common methods include summation or average. Through message passing, the node feature representations are updated:

$$h_v^{(l)} = Update\left(h_v^{(l-1)}, a_v^{(l)}\right) \tag{10}$$

Graph convolution network (GCN) learns the relationship between nodes through multi-level convolution operation, transmits neighbour node information, and updates its own node representation accordingly (Bhatti et al., 2023). The node representation of layer l is defined as follows:

$$H^{(l)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l-1)}W^{(l)}\right) \tag{11}$$

where $\hat{A} = A + I$ denotes the adjacency matrix with self-loops added; $\hat{D}$ represents the degree matrix associated with $\hat{A}$; $W^{(l)}$ is the weight matrix for layer l; and $\sigma(.)$ stands for the activation function, such as ReLU or Sigmoid. Through this formula, each layer realises the normalisation and linear transformation of neighbour information, and gradually learns to a higher level of node representation.

Graph attention network (GAT) introduces attention mechanism on the basis of GNN, so that different nodes can pay different attention to their neighbours. By dynamically calculating the weight between the target node and its neighbour nodes, the weighted neighbour information is integrated into the target node representation to achieve more accurate information transmission (Vrahatis et al., 2024). The calculation formula represented by GAT node is as follows:

$$h_v^{(l)} = \sigma\left(\sum_{u \in N(v)} \alpha_{vu} W^{(l)} h_u^{(l-1)}\right) \tag{12}$$

where α_{vu} is the weight between node v and its neighbour node u calculated through the attention mechanism, and the mathematical expression is as follows:

$$\alpha_{vu} = \frac{\exp\left(LeakyReLU\left(V^T\left[W^{(l)}h_v \parallel W^{(l)}h_u\right]\right)\right)}{\sum_{k\in N(v)} \exp\left(LeakyReLU\left(V^T\left[W^{(l)}h_v \parallel W^{(l)}h_k\right]\right)\right)} \tag{13}$$

where V is the learnable weight vector and '||' represents the connection operation of the matrix. Through this mechanism, gat can dynamically adjust the attention weight according to the relationship between nodes, so as to more effectively combine the neighbour information in the node representation.

Seasonal-trend decomposition using loess (STL) algorithm is a technology for time series decomposition, which decomposes the original time series data Y_t into three main components: periodic component S_t, trend component T_t and residual component Rt. therefore, the original data of each time point t can be expressed as:

$$Y_t = S_t + T_t + R_t \tag{14}$$

This decomposition method helps to reveal the intrinsic dynamics of time series data, including periodic changes, long-term trends, and irregular fluctuations. The uniqueness of the STL algorithm lies in its dual layered loop structure, consisting of an outer loop and an inner loop. The outer loop is mainly responsible for enhancing the robustness of the algorithm, reducing the impact of residuals on the decomposition results, and ensuring the stability of the decomposition. The inner loop is responsible for executing specific decomposition steps. The STL algorithm decomposes original time series data Y into periodic component S, trend component T and residual component R after several rounds of inner loop iteration. The residual component is calculated by $R = Y - S - T$. This decomposition method reveals the basic characteristics of time series, including periodic patterns, long-term trend changes, and random fluctuations, which are of great significance for time series analysis and prediction.

3 Model design and implementation

3.1 Definition of network traffic issues

The successful live streaming of rural e-commerce relies on precise guidance and efficient conversion of network traffic. In depth analysis of the transmission rules and node characteristics of network traffic is the key to optimising the live streaming effect (Nam et al., 2021). The spatial topology of network traffic nodes is defined as an undirected graph $G = (V, E, A)$. The spatial topology of each time node in the spatiotemporal structure of network traffic data is defined as graph G, where V represents the set of all network nodes in the topology, $V = \{v_1, v_2, \ldots, v_N\}$, Among them, $|V| = N$ represents that the number of nodes in the topology graph is N, E represents the set of edges connected by nodes in graph G, in the field of network flow, it represents the connectivity between network nodes, and $A \in R^{N\times N}$ is the matrix of connected nodes in network topology graph G. Each node in topology G will receive N time series data at a fixed time, forming a feature vector of length N. The traffic values between all nodes obtained at this time form an $N - N$ time dimension traffic matrix (Chen et al., 2020). This graph-structured representation of network traffic is directly integrated into the REISE model's GNN module, where it enables the capture of complex spatiotemporal dependencies and connectivity patterns essential for analysing live-streaming traffic dynamics.

Assuming there are N router nodes in the network, the $N \times N$ traffic matrix is represented as TM. The actual measured traffic matrix sequence is $TM_t(t \in [1, T])$, where T is the total measurement time period. Using historical data (TM_{t-k}, TM_{t-k+1}, …, TM_{t-1}) ($k \in [1, t]$), the predicted TM_t is obtained:

$$TM_t = \begin{bmatrix} T_{1,1}^t & T_{1,2}^t & \cdots & T_{1,n}^t \\ T_{2,1}^t & T_{2,2}^t & \cdots & T_{2,n}^t \\ \cdots & \cdots & \cdots & \cdots \\ T_{n,1}^t & T_{n,2}^t & \cdots & T_{n,n}^t \end{bmatrix} \tag{15}$$

among them, TM_t is a two-dimensional array representing the traffic matrix TM of the t^{th} time interval. From the above equation, it can be seen that each term of the TM_t matrix is defined as T_{ij}^t, T_{ij}^t representing the traffic from the source node i to the destination node j at time t.

The topology information of the network can be represented by graph G, where A represents the adjacency matrix of the network routing nodes at a certain moment, A includes all node information and link information of the network routing, and the degree matrix of the topology nodes is D, which has spatial characteristics. The adjacency matrix contains node information of the network topology and information about the edges between nodes. The traffic of the network over time t is represented as $TM(t)$. The prediction problem of network traffic can be expressed as:

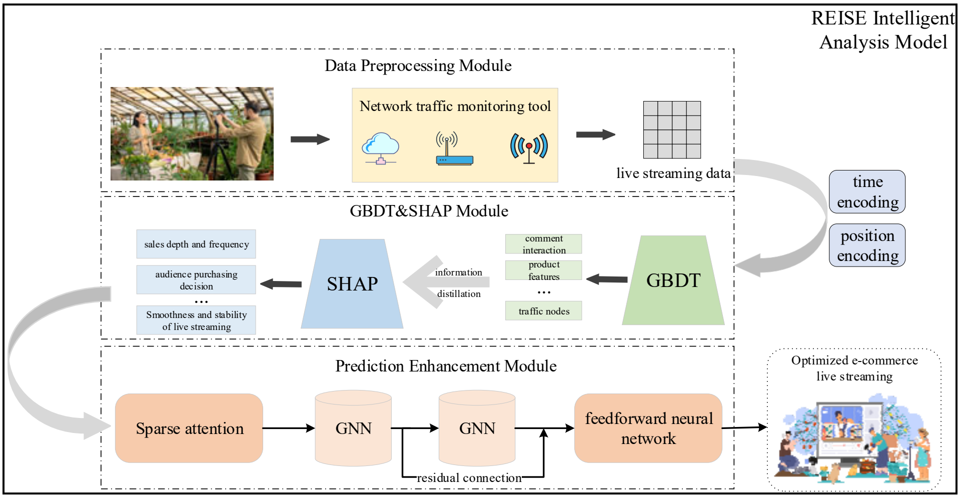
$$TM(t+f) = P\big(A, TM(t-1), TM(t-2), \ldots, TM(t-h)\big) \tag{16}$$

among them, P represents the traffic prediction model, h represents the historical flow interval time that needs to be analysed to predict future flows, and f refers to the interval time that needs to be predicted for future flows. Moreover, the traffic of a certain node at a certain time will have a certain impact on the traffic values of nodes that have a correlation with this node in the future. Therefore, network traffic has spatiotemporal correlation, and prediction models need to be able to fully and comprehensively obtain the spatiotemporal characteristics of network traffic. Since TM is a two-dimensional matrix, it can be treated as a greyscale image. Thus, the TM prediction issue can be turned into a time-series prediction problem for greyscale image sequences, enabling direct prediction of the overall flow matrix.

3.2 Intelligent analysis model design

The intelligent analysis model REISE designed in this study aims to improve the sales efficiency of rural e-commerce in live streaming environments. It integrates advanced technologies such as gradient boosting tree (GBDT), SHAP feature selection, and GNN, and deeply analyses the transmission rules and node characteristics of network traffic. This model starts from multidimensional data of rural e-commerce live streaming, including audience interaction behaviour, product display effects, anchor sales language, network traffic node topology structure, etc. Through a hierarchical and modular architecture, it achieves accurate identification, quantitative analysis, and real-time optimisation of sales efficiency influencing factors and sales strategies. The specific model structure is shown in Figure 1.

Figure 1 REISE model framework diagram (see online version for colours)



The data collection module of the model extensively collects various types of data during rural e-commerce live streaming, and the network traffic monitoring tool obtains real-time spatial topology data of network traffic nodes, laying a data foundation for subsequent analysis. The data pre-processing stage performs cleaning, normalisation, feature engineering, and other operations on the collected multi-source heterogeneous data, removes noisy data and outliers, unifies data formats and dimensions, extracts preliminary feature vectors to meet the requirements of model input. The pre-processed data is first input into the GBDT module. GBDT, as a powerful GBDT algorithm, can effectively handle nonlinear data relationships and capture complex patterns in data through iterative construction and weighted combination of decision trees. In this model, GBDT is responsible for preliminary learning and analysis of multidimensional data from rural e-commerce live streaming, and identifying key features and factors that affect sales efficiency. In addition, GBDT can analyse the preliminary characteristics of network traffic nodes and screen out node features that have a significant impact on traffic transmission efficiency and stability. By optimising the loss function through gradient descent algorithm, the GBDT model continuously adjusts the parameters and weights of the decision tree to improve the accuracy and generalisation ability of prediction, providing high-quality feature representation for subsequent feature selection and deep learning modules.

The feature data processed by the GBDT module is further input into the SHAP feature selection module. The SHAP method adopts a game-theoretic approach to quantify the marginal contribution of each feature to model predictions by computing its Shapley value. This mechanism not only identifies the most predictive features but also minimises the influence of redundant or irrelevant variables, thereby reducing noise and enhancing model interpretability. By emphasising features that significantly impact sales efficiency – such as audience engagement levels, product exposure duration, and anchor speech patterns – the SHAP module ensures that only the most relevant and influential attributes are retained for subsequent processing. In the model, the SHAP module interprets and refines the importance of the features output by GBDT, accurately identifying the subset of features that have the most impact on improving sales efficiency. At the same time, the SHAP module can also explain which node connection relationships and traffic transmission indicators play a key role in the smoothness and stability of the live streaming process in the characteristics of network traffic nodes. By filtering and assigning weights to these key features, the SHAP feature selection module provides more focused and efficient feature inputs for subsequent GNN modules, improving the overall performance and interpretability of the model. After completing feature selection, the selected feature data is fed into the GNN module. The GNN module further enhances the model's capability to model nonlinear and spatially-dependent structures by propagating information between connected nodes in the graph. Through message passing and feature aggregation within the network topology, GNN captures high-order dependencies and nonlinear influence patterns among network traffic nodes, which are critical for understanding viewer access patterns, interaction hotspots, and real-time traffic dynamics in live e-commerce scenarios. The GNN module fully considers the spatial topology structure of network traffic nodes and the complex relationships between nodes in rural e-commerce live streaming. Define network traffic nodes as nodes in the graph, and represent the connection relationships between nodes as edges to construct graph structured data that reflects the transmission patterns and node characteristics of network traffic. GNN can capture high-order neighbourhood information and global topology characteristics between nodes through message passing and node state updates in the graph. In the model, the GNN module combines the graph structure data of network traffic nodes with the business characteristics of rural e-commerce live streaming, deeply analyses the propagation path and aggregation mode of network traffic in the network, and how these traffic characteristics affect the viewing experience, interactive effect, and final sales efficiency of live streaming. At the same time, it can also discover optimised transmission paths of network traffic between different nodes, providing guidance for improving the smoothness and stability of live streaming. In addition, the GNN module can also model the social network relationships in rural e-commerce live streaming, analyse the interactive communication patterns between viewers and the influence of opinion leaders, and further explore potential sales opportunities and marketing strategies. The real-time feedback and strategy optimisation module of the model is the core link to improve the efficiency of live streaming sales. This module continuously processes live multi-source data streams – including audience interactions (e.g., comments, clicks, and dwell time), real-time sales conversions, and network traffic metrics – to detect shifts in consumer interest and engagement. Based on these inputs, the system dynamically refines sales tactics, prioritises high-response products, and personalises recommendation sequences. Simultaneously, it optimises video stream parameters and content delivery paths according to network node topology and traffic conditions, ensuring smooth playback and enhanced viewing experiences. Furthermore, inventory and promotion strategies are adjusted in sync with sales velocity and stock levels, enabling agile and efficient operations. Through close collaboration with previous modules, it achieves dynamic adjustment and optimisation of sales strategies. At the same time, based on the transmission rules and node characteristics of network traffic, optimise the encoding parameters and transmission paths of live streaming videos to ensure that viewers can watch live streams smoothly, enhance their viewing experience and purchase intention. In addition, the real-time feedback module can adjust promotional strategies and product recommendation sequences in real-time based on sales conversion and inventory levels, achieving precise marketing and efficient inventory turnover. Through real-time monitoring and strategy optimisation of the live streaming process, this module can maximise the sales efficiency and economic benefits of rural e-commerce live streaming.

4 Experimental results and analyses

To assess REISE's effectiveness in enhancing rural e-commerce sales efficiency in live-streaming settings, we compared it with various classic time series prediction models. By comparing the performance of REISE with these models in predicting agricultural product sales trends, optimising inventory management, and adjusting sales strategies in real-time, the aim is to verify the accuracy and robustness of REISE in handling complex dynamic data in live streaming e-commerce.

1 Recurrent neural network (RNN): It is a traditional neural network model commonly used for processing sequential data, but it faces the challenge of capturing long-term dependencies when dealing with long sequences.

2 Gated reacurrent unit (GRU): A variant of RNN that improves the way information is transmitted by updating and resetting gates, effectively solving the gradient problem of RNN when processing long sequences.

3 LSTM and ConvLSTM: Widely used in the field of time series prediction, LSTM focuses on capturing long-term dependencies, while ConvLSTM incorporates convolutional structures and is suitable for processing data with spatial structures.

4 PredRNN: Designed ST-LSTM spatiotemporal units that can simultaneously handle spatial and temporal changes in a unified storage space and transfer states between different layers.

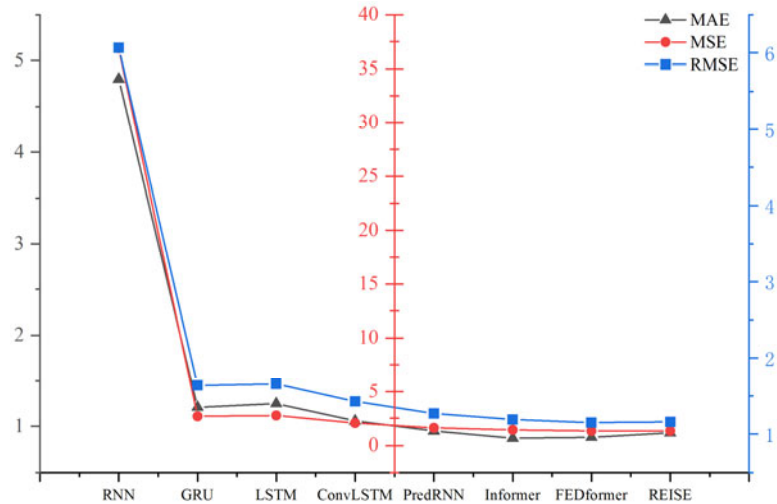
5 Informer and FEDformer: Both are attention based models and are state-of-the-art models in the field of time series prediction in recent years.

Table 1 Comparison results of performance evaluation between different baseline models and REISE model

Models	*MAE*	*MSE*	*RMSE*	*Fitting state*
RNN	4.80	36.89	6.07	Under-fitting
GRU	1.21	2.69	1.64	Fitting
LSTM	1.25	2.77	1.66	Fitting
ConvLSTM	1.06	2.06	1.43	Fitting
PredRNN	0.95	1.61	1.27	Fitting
Informer	0.87	1.42	1.19	Fitting
FEDformer	0.88	1.33	1.15	Fitting
REISE	0.93	1.34	1.16	Fitting

The selection of these comparative models covers different categories from traditional statistical methods to modern deep learning techniques, providing a comprehensive and representative benchmark for the performance evaluation of REISE. The specific experimental results are shown in Table 1.

Figure 2 Comparison of model performance experimental results (see online version for colours)



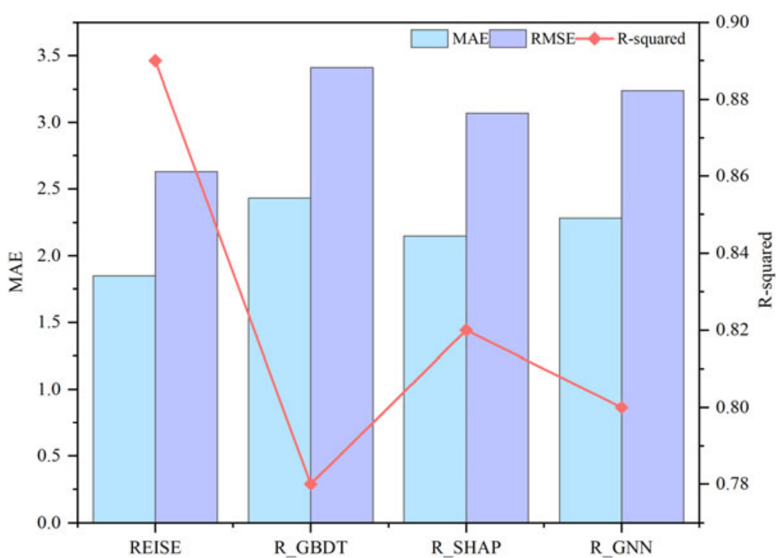
From the experimental results, it can be seen that the REISE model performs well in terms of prediction accuracy, with MAE, MSE, and RMSE reaching 0.93, 1.34, and 1.16, respectively. Comparative analysis demonstrates REISE's superior predictive performance with significantly reduced error rates across all metrics when benchmarked against conventional sequential models including RNN, GRU, and LSTM. Compared with traditional RNN models, REISE reduces errors by nearly 76.46% in MAE indicators, about 63.86% in MSE indicators, and about 67.48% in RMSE indicators. This indicates that the REISE model can more accurately predict sales trends when processing complex live streaming e-commerce data, providing a more reliable basis for inventory management and sales strategy optimisation. Compared with outstanding deep learning models such as LSTM and GRU in recent years, the REISE model has reduced errors by approximately 25.60% and 22.31% in MAE metrics, 49.46% and 46.77% in MSE metrics, and 30.12% and 29.27% in RMSE metrics, respectively. This further demonstrates the advantages of the REISE model in capturing complex patterns and nonlinear relationships in live streaming e-commerce data. In addition, the REISE model has shown strong competitiveness compared to the attention based Informer and FEDformer models. Although slightly higher than the Informer model in MAE indicators, it is comparable to the Informer model in MSE indicators and only 0.03 higher than the Informer model in RMSE indicators. In terms of fitting state, the REISE model performs more stably without overfitting or underfitting. This indicates that the REISE model has better generalisation ability and stability while maintaining high prediction accuracy. The selection of these comparative models covers different categories from traditional statistical methods to modern deep learning techniques, providing a comprehensive and representative benchmark for the performance evaluation of REISE. In summary, the REISE model outperforms other comparative models in terms of prediction accuracy, generalisation ability, and stability, verifying its effectiveness and superiority in improving rural e-commerce sales efficiency in live streaming environments.

Meanwhile, in order to verify the contribution of each key component in the REISE model to overall performance, this study designed ablation experiments. The ablation experiment evaluates the impact of each module on model performance by gradually removing key modules from the model to obtain R_GBDT, R_SHAP, and R_GNN. The experimental dataset selected live sales data from a rural e-commerce platform in the past year, including multi-dimensional information such as audience interaction, product display, anchor language, and network traffic. Divide the dataset into training and testing sets in a 7:3 ratio to ensure consistency in data distribution during the training and testing process of the model. The outcomes of the ablation experiment are presented in Table 2.

Table 2 Comparison of ablation experiment results

Models	*MAE*	*RMSE*	*R-squared*
REISE	1.85	2.63	0.89
R_GBDT	2.43	3.41	0.78
R_SHAP	2.15	3.07	0.82
R_GNN	2.28	3.24	0.80

Figure 3 Comparison chart of ablation experiment results (see online version for colours)



From Table 2, it can be seen that the complete REISE model performs the best in prediction accuracy, with the lowest MAE and RMSE, and the highest R^2. After removing the GBDT module, the model's MAE and RMSE rose considerably. This shows GBDT is key for capturing nonlinear relationships and enhancing the model's predictive power. Similarly, after removing the SHAP feature selection module, the model performance also decreased, indicating that SHAP feature selection can effectively filter out the key features that have the greatest impact on sales efficiency, improving the model's focus and efficiency. After removing the GNN module, the model cannot fully utilise the topology information of network traffic nodes, resulting in a decrease in prediction accuracy. This validates the key value of the GNN module in processing network traffic data and capturing complex relationships between nodes. Through ablation experiments, we clearly see that each key component in the REISE model makes an indispensable contribution to overall performance, further demonstrating the rationality and effectiveness of the model design.

In summary, the experimental evaluation demonstrates the clear advantages of the REISE model in enhancing sales efficiency prediction within rural e-commerce live-streaming contexts. By integrating multi-algorithm capabilities – including gradient boosting, interpretable feature selection, and graph-based learning – the model achieves a significant reduction in prediction errors across key metrics such as MAE, MSE, and RMSE. These improvements confirm its ability to handle complex, nonlinear relationships and dynamic data patterns inherent in live-streaming environments. Furthermore, the model exhibits strong generalisation and stability, avoiding common pitfalls such as overfitting or underfitting. Compared to both classical and state-of-the-art baseline models, REISE not only delivers higher accuracy but also provides a more robust and interpretable framework for real-time decision support, ultimately contributing to more effective sales strategy optimisation and inventory management in rural e-commerce.

5 Conclusions

This paper focuses on the core issue of how to improve the sales efficiency of rural e-commerce in the live broadcast environment, and proposes an innovative intelligent analysis model. The model integrates multi-source data such as audience interaction, product display effects and anchor sales scripts, and uses advanced machine learning algorithms for in-depth analysis, so as to accurately grasp the needs and behaviour characteristics of consumers and provide a solid basis for optimising sales strategies. The innovation of this paper is to propose a multi-dimensional data analysis model, which can comprehensively and accurately describe the potential needs and behavioural preferences of consumers. This model not only considers the real-time interaction trajectory of the audience, but also pays attention to the multi-dimensional effect of product display and the personalised sales script of the anchor. Through in-depth learning technology, the model can analyse and predict the live broadcast sales data and market trends, and provide forward-looking sales forecasts for rural e-commerce. This forecasting ability can help enterprises reasonably plan inventory, optimise supply chain management, reduce inventory costs, and ensure timely supply of agricultural products to meet the needs of consumers. This paper also presents a real-time feedback-based dynamic optimisation mechanism. It adjusts sales strategies in line with audience emotion and preference shifts, improving sales efficiency and conversion rates. Overall, this study offers a novel approach to enhancing rural e-commerce sales in live-streaming settings. The intelligent analysis model helps rural e-commerce better understand market trends, optimise sales tactics, and boost operational efficiency. The study is theoretically significant and provides practical guidance for rural e-commerce, promoting its sustainable live-streaming development.

Declarations

All authors declare that they have no conflicts of interest.

References

Bhatti, U.A., Tang, H., Wu, G., Marjan, S. and Hussain, A. (2023) 'Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence', *International Journal of Intelligent Systems*, Vol. 2023, No. 1, p.8342104.

Canavari, M., Fritz, M., Hofstede, G.J., Matopoulos, A. and Vlachopoulou, M. (2010) 'The role of trust in the transition from traditional to electronic B2B relationships in agri-food chains', *Computers and Electronics in Agriculture*, Vol. 70, No. 2, pp.321–327.

Chen, Liu, B., Wan, S., Qiao, P. and Pei, Q. (2020) 'An edge traffic flow detection scheme based on deep learning in an intelligent transportation system', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22, No. 3, pp.1840–1852.

Davis, F.D. (1989) 'Perceived usefulness, perceived ease of use, and user acceptance of information technology', *MIS Quarterly*, pp.319–340.

Du, L., Song, H., Xu, Y. and Dai, S. (2024) 'An architecture as an alternative to gradient boosted decision trees for multiple machine learning tasks', *Electronics*, Vol. 13, No. 12, p.2291.

Folorunso, O., Sharma, S.K., Longe, H. and Lasaki, K. (2006) 'An agent-based model for agriculture e-commerce system', *Information Technology Journal*, Vol. 5, No. 2, pp.230–234.

Gramegna, A. and Giudici, P. (2022) 'Shapley feature selection', *FinTech*, Vol. 1, No. 1, pp.72–80.

Kshetri, N. (2018) 'Rural e-commerce in developing countries', *IT Professional*, Vol. 20, No. 2, pp.91–95.

Le, N.Q.K., Ho, Q-T., Nguyen, V-N. and Chang, J-S. (2022) 'BERT-Promoter: an improved sequence-based predictor of DNA promoter using BERT pre-trained model and SHAP feature selection', *Computational Biology and Chemistry*, Vol. 99, p.107732.

Li, F. and Gan, Y. (2025) 'Research on the sustainable development capability of Chinese rural e-commerce based on multidimensional perspective', *Scientific Reports*, Vol. 15, No. 1, p.11547.

Liang, S. and Cheng, H. (2024) 'Research on the mode of e-commerce assisting agricultural innovation and entrepreneurship under the background of platform economy – a case study of e-commerce assisting farmers in Wenchuan County', *Economics & Management Information*, pp.1–10.

Luo, H., Cheng, S., Zhou, W., Yu, S. and Lin, X. (2021) 'A study on the impact of linguistic persuasive styles on the sales volume of live streaming products in social e-commerce environment', *Mathematics*, Vol. 9, No. 13, p.1576.

Nam, B.K-H., Jiho, C. and Hongsuk, Y. (2021) 'Spatial-temporal graph neural network for traffic forecasting: an overview and open research issues', *Applied Intelligence*, Vol. 52, No. 3, pp.2763–2774.

Pramanik, J., Sarkar, B. and Kandar, S. (2017) 'Impact of ICT in rural development: perspective of developing countries', *American Journal of Rural Development*, Vol. 5, No. 4, pp.117–120.

Ruch, M. and Sackmann, S. (2012) 'Integrating management of customer value and risk in e-commerce', *Information Systems and e-Business Management*, Vol. 10, pp.101–116.

Ruiz-Garcia, L., Steinberger, G. and Rothmund, M. (2010) 'A model and prototype implementation for tracking and tracing agricultural batch products along the food chain', *Food Control*, Vol. 21, No. 2, pp.112–121.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G. (2008) 'The graph neural network model', *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, pp.61–80.

Sun, H. and Zhang, P. (2006) 'The role of moderating factors in user technology acceptance', *International Journal of Human-Computer Studies*, Vol. 64, No. 2, pp.53–78.

Vrahatis, A.G., Lazaros, K. and Kotsiantis, S. (2024) 'Graph attention networks: a comprehensive review of methods and applications', *Future Internet*, Vol. 16, No. 9, p.318.

Wang, G., Zhang, Z., Li, S. and Shin, C. (2023) 'Research on the influencing factors of sustainable supply chain development of agri-food products based on cross-border live-streaming e-commerce in China', *Foods*, Vol. 12, No. 17, p.3323.

Wang, H., Liang, Q., Hancock, J.T. and Khoshgoftaar, T.M. (2024) 'Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods', *Journal of Big Data*, Vol. 11, No. 1, p.44.

Wang, Y., Lu, Z., Cao, P., Chu, J., Wang, H. and Wattenhofer, R. (2022) 'How live streaming changes shopping decisions in e-commerce: a study of live streaming commerce', *Computer Supported Cooperative Work (CSCW)*, Vol. 31, No. 4, pp.701–729.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S. (2020) 'A comprehensive survey on graph neural networks', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 1, pp.4–24.

Xin, B., Hao, Y. and Xie, L. (2023) 'Strategic product showcasing mode of e-commerce live streaming', *Journal of Retailing and Consumer Services*, Vol. 73, p.103360.

Zhang, B., Ren, J., Cheng, Y., Wang, B. and Wei, Z. (2019) 'Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm', *IEEE Access*, Vol. 7, pp.32423–32433.

Zhang, Z. and Jung, C. (2020) 'GBDT-MO: gradient-boosted decision trees for multiple outputs', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 7, pp.3156–3167.

Zhou, G-J. (2022) 'Analysis and research on the new model of e-commerce poverty alleviation: e-commerce live broadcast of agricultural products', *Journal of Accounting, Finance & Management Strategy*, Vol. 17, No. 2, pp.33–59.