Universe Scientific Publishing

Predicting Demand in Food Delivery: A Hybrid Model for Promotional and Post-Promotional Periods

Xu LIU

Beijing Landscape Architecture School, Beijing 102445, CHINA

Abstract: Accurate forecasting of daily order volumes is critical for businesses facing rapid fluctuations in demand, particularly during promotional events such as the Subsidy War. This paper proposes a hybrid forecasting model combining log-normal distribution, ARIMA, and LSTM to predict order volumes during and after the promotional period for a milk tea shop. For the promotional period, we model the demand using a log-normal distribution to capture the skewed and heavy-tailed nature of the order volume. Post-promotion, we use an ARIMA model to account for trend in the order data, augmented with external temperature data to handle residual errors via an LSTM neural network. A weighted sum of the ARIMA forecast and a decaying trend function is used to generate the final predictions, ensuring stationarity and reducing overfitting. The proposed hybrid approach provides a robust solution for managing demand fluctuations, offering valuable insights into both promotional and non-promotional periods.

Keywords: Demand Forecasting; Time-Series Analysis; Hybrid Model; Machine Learning; Forecasting Accuracy.

1. Introduction

In mid-2025, a dramatic "Food Delivery Subsidy War" erupted across major platforms in China, with Meituan, JD.com, and Taobao Flash Sale unleashing aggressive coupon campaigns, some even offering "zero-yuan" milk tea promotions, to capture rapid-growth market share in the burgeoning instant-retail economy. On flagship promotional days, platforms processed well over 100 million food delivery orders in a single day, causing app outages and overwhelming participating merchants. They also created unpredictable demand swings, making next-day volume forecasts essential to avoid shortages, waste, and service disruptions.

This study highlights the importance of time-series based demand forecasting amid global competition and economic uncertainty. Gartner reports that demand driven decision making can cut inventory costs by 15~30% and increase product availability by the same margin. Applying the Pareto 80/20 rule to focus on the top five products, we propose a two-phase hybrid framework: a log-normal distribution to model demand during the subsidy war, and an ensemble of LSTM and ARIMA models augmented with temperature data to capture post-war non-linear effects. This approach aims to deliver robust, high-accuracy forecasts that empower more effective sales planning and cost control.

2. Literature Review

In today's competitive environment, especially in food delivery, where external factors such as promotions cause rapid demand swings. Accurate forecasting is vital to optimize inventory, control costs, and satisfy customers.

LSTM use gated memory cells to learn long-term dependencies and non-linear patterns in sequential data (e.g., incorporating multivariate inputs like temperature or promotions). They handle complex, non-stationary dynamics effectively but demand large datasets, substantial computation, and suffer from limited interpretability. ARIMA models decompose a series into autoregressive, differencing, and moving-average components, transparently capturing trend. However, ARIMA models rely on linear relationships and require stationarity after differencing, and they may yield suboptimal forecasts when confronted with sudden demand fluctuations. Hybrid LSTM-ARIMA approaches combine ARIMA's structured forecasts with LSTM's flexibility, delivering more robust predictions across both regular cycles and extreme events.

The log-normal distribution models positive demand whose logarithm is normally distributed, making it ideal for capturing event-driven surges and the ensuing decay with asymmetric, percentile based forecasts. It provides a clear probabilistic trend for post-event demand but does not account for temporal dependence, so its static parameter estimates must be carefully managed to avoid bias from outliers or limited data.

We employ a two-phase forecasting approach: during the subsidy war, a log-normal distribution models volatile, heavy-tailed demand spikes; afterward, an LSTM-ARIMA hybrid captures stable patterns, with LSTM learning non-linear dependencies and ARIMA modeling trend. To improve performance, temperature is included as an exogenous variable, reflecting the impact of increasingly frequent extreme weather on beverage demand and enhancing post-war forecast accuracy.

3. Scenario Analysis and Algorithm Architecture

This chapter describes how to adapt the predictive models for application during and after subsidy wars by researching the underlying drivers of sales trends while accounting the impacts of extreme weather events. In Section 3.1, the chart of the whole algorithm architecture will show a complete workflow for the study. The requirements and pain points of processing data will be shown in Section 3.2. Section 3.3 will first employ the log-normal distribution to simulate the war order volumes, then superimpose a trend-compliant LSTM+ARIMA model for post-war periods, with additional extreme temperature effects.

3.1. Algorithm Architecture

The following chat in Figure 1 shows the flow of order volume prediction in this study. First, collect the required information and filter the invalid data and sort the data by time and shop, and then divide them into 2 parts by the period. Second, use the data to train the algorithm models and choose a suitable method to fuse their prediction results. Third, evaluate the model performance. Change the parameters or retrain the models if the evaluation metrics don't meet the expectations. If the evaluation results meet expectations, deploy this algorithmic framework.

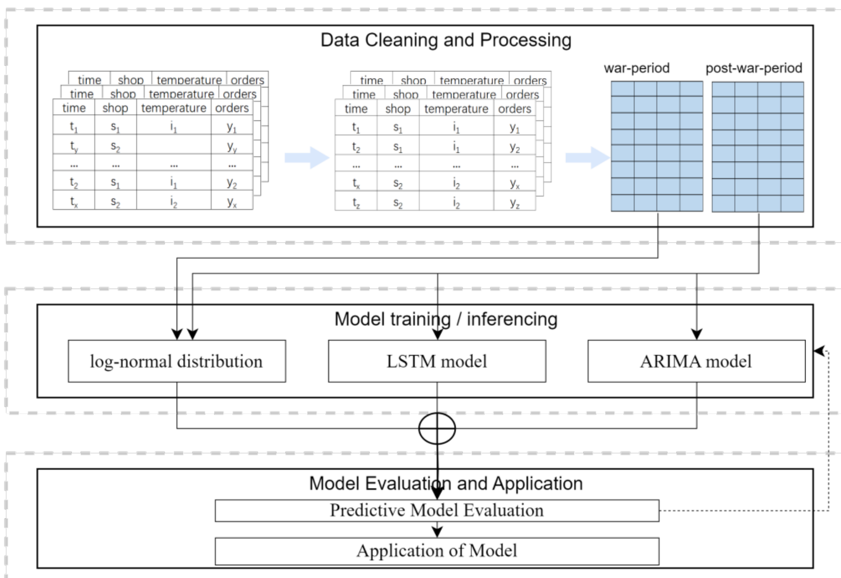

Figure. 1

3.2. Data Collection and Processing

The training data should be drawn from shops with similar baseline order volumes and comparable characteristics - for example, all located in shopping districts favored by young people. The time window should cover the same season over the past three years, focusing on promotional campaigns of similar scale, duration, and intensity. For each campaign, extract data from its first day through two weeks after it ends. According to the above criteria, we collected daily statistics from 50 shops covering Anniversary Celebration and Summer Promotion campaigns in 2023/2024/2025 (total of 6300 raw data of order volume).

1. Timeline: In order to meet the market demand and enhance competitiveness through rapid response, the data must be capable of supporting time granularity at least at the daily level or finer. First, split the six promotions into two tables: one for the promotion period and one for the post-promotion period. Next, convert each date into relative day labels (Event Day 1, Day 2, … and Post-Event Day 1, Day 2, …).

Finally, use store as the primary key and promotion as the secondary key, and sort all records in chronological order.

2. Missing value: Given that we have 50 comparable sales records for each campaign, we can still effectively depict the sales trend even with a small proportion of missing values. And this is the most efficient solution to prepare the data.

3. Temperatures: Temperature is the most influential environmental factor affecting the sales. A new column with temperatures retrieved by date have to be added to the existing date-based tables to adjust the algorithms.

The figure 2 below is the curve obtained by averaging the data. For campaigns shorter than seven days, we pad the missing days up to seven using the median value; for campaigns longer than seven days, we truncate to the first seven days before appending the 2-weeks period following the campaign.

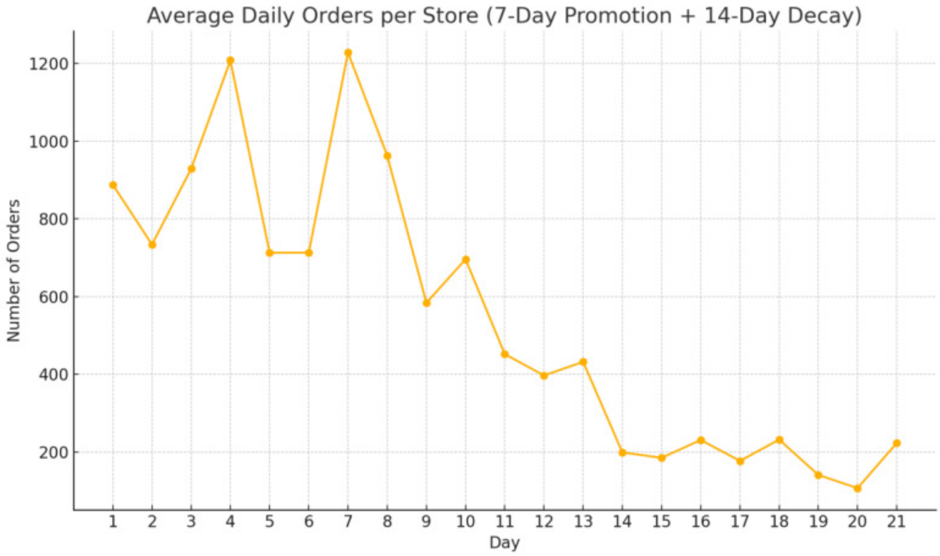

Figure. 2

3.3. Model Creation, Adjustment and Fusion

3.3.1. Log-normal Distribution Model

Temperature is not included in this model, as it has little to no impact on the results. And we have:

$x_{i,j,k} > 0$ (i=1,…,50 shops, j=1,…,21 days, k=1,…,6 events)

for a total of N = 50×21×6 = 6300 observations, defined as

x_l (l=1,2,…,N).

a) Log Transformation

To harness the tractability of the normal distribution, we first apply a logarithmic transformation to the strictly positive daily order volumes x_l. Specifically, for each day i in the promotion period, define

$y_l = \ln(x_l)$ (l=1,…,N).

This transformation normalizes the right-skewed distribution of x_l, since a log-normal random variable X satisfies $\ln X \sim N(\mu,\sigma^2)$. Working in the log-domain also stabilizes variance and makes additive-model assumptions more appropriate.

b) Parameter Estimation via Maximum Likelihood

Once we have y_l, the goal is to estimate the normal parameters μ and σ^2.Under the log-normal assumption, the log-likelihood of the sample is

$$L(\mu,\sigma) = \sum_{i=1}^{N}\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(y_i-\mu)^2}{2\sigma^2}\right].$$

Maximizing this with respect to μ and σ^2 yields the well-known closed-form maximum-likelihood estimators:

$$\hat{\mu} = \frac{1}{N}\sum_{l=1}^{N} y_l, \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{l=1}^{N}(y_l - \hat{\mu})^2 .$$

c) Predictive Distribution

With $(\hat{\mu}, \hat{\sigma}^2)$ in hand,the forecast for the next day's order volume follows the fitted log-normal law.Its probability density function is given by

$$f_X(x) = \frac{1}{x\hat{\sigma}\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \hat{\mu})^2}{2\hat{\sigma}^2}\right).$$

The corresponding cumulative distribution function (CDF) is

$$F_X(x) = \Phi\left(\frac{\ln x - \hat{\mu}}{\hat{\sigma}}\right).$$

where Φ denotes the standard normal CDF. This fully characterizes our uncertainty about the next day's volume, capturing both central tendency and tail risk from extreme promotional orders.

d) Volume Forecast and Prediction Intervals

From the fitted log-normal law, several forecast statistics can be extracted.

Median forecast (50% quantile):

$$\tilde{x} = \exp(\mu).$$

Mean forecast:

$$\mathbb{E}[X_{\text{next}}] = \exp\left(\hat{\mu} + \frac{1}{2}\hat{\sigma}^2\right).$$

(1−α)×100% prediction interval: using the normal quantile $z_{\alpha/2}$, the lower and upper bounds on X_{next} are

$$\left[\exp\left(\hat{\mu} - z_{\alpha/2}\hat{\sigma}\right), \exp\left(\hat{\mu} + z_{\alpha/2}\hat{\sigma}\right)\right].$$

These outputs provide a full probabilistic forecast for the next day's order volume under the log-normal model, enabling data-driven inventory and staffing decisions during the promotion period.

3.3.2. LSTM+ARIMA Model

a) ARIMA trend modeling

We first aggregate by taking the cross-shop average for each promotion event:

$$\bar{x}_t = \frac{1}{50*6} \sum_{i=1,j=1}^{i=6,j=50} x_{t,i,j}.$$

And then all subsequent modeling steps are applied by $\bar{x}_t$. To stabilize the mean of the series and promote stationarity, apply a decaying trend adjustment, denoted by dt, to approximate the natural falloff in customer demand following a subsidy event.

$$d_t = \alpha e^{-\beta t}, \quad \alpha, \beta > 0,$$

where t represents the number of days since the end of the subsidy period. The trend-adjusted series is:

$$X_t' = \frac{\bar{x}_t}{d_t}.$$

And then we can use Augmented Dickey-Fuller test to check the stationarity of the trend adjusted series X_t'.

An ARIMA(p,d,q) model is then fitted to the stationary series X_t'. Define the d-th differenced series:

$$Y_t = (1 - B)^d X_t',$$

where B is the back-shift operator ($BX'_t = X'_{t-1}$). The ARIMA model then takes the form

$$\phi(B)Y_t = c + \theta(B)\epsilon_t,$$

with

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \quad \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q,$$

where ϵ_t is white noise and c is a constant.

In order to select the parameter (p,d,q), it's imperative to examine the autocorrelation and partial autocorrelation functions to identify plausible values for the autoregressive order p and moving-average order q, and then select the combination that minimizes the Akaike and Bayesian information criteria. After estimating the model, it is necessary to confirm no systematic patterns remain, which means to verify its adequacy by applying the Ljung-Box test to the residuals and by visually inspecting the residual series to ensure they behave like white noise.

Once fitted, the one-step ARIMA forecast on the adjusted scale is

$$\hat{X}'_{t,\text{ARIMA}} = c + \sum_{i=1}^{p} \phi_i Y_{t+1-i} + \sum_{j=1}^{q} \theta_j \hat{\epsilon}_{t+1-j},$$

and is then re-scaled by the decay trend to obtain

$$\hat{X}_{t,\text{ARIMA}} = d_t \hat{X}'_{t,\text{ARIMA}} \cdot$$

While ARIMA effectively models linear and seasonal components, it cannot fully capture complex non-linear relationships, particularly those driven by external variables such as temperature. To address this, compute the residuals between the observed data and the ARIMA predictions:

$$r_t = x_t - \hat{x}_{t,ARIMA} \cdot$$

An LSTM model is then trained to predict these residuals using sequences of prior residuals and temperature as an exogenous input. Each training sample at day t is a tuple ([rt−L,…,rt−1],[Tt−L,…,Tt−1]), where Tt is the recorded daily average temperature and L is the look-back window.

The LSTM model outputs a predicted residual $\hat{r}_t$, and the final forecast for day t is:

$$\hat{x}_t^{\text{final}} = w \cdot \hat{x}_t + (1-w) \cdot d_t, \quad 0 \le w \le 1 \cdot$$

To further stabilize predictions and avoid over-reliance on a potentially overfitted ARIMA component, the final result must be retrieved from a weighted sum of the ARIMA prediction and the original decaying trend model:

$$\hat{x}_t^{\text{final}} = w \cdot \hat{x}_t + (1-w) \cdot d_t, \quad 0 \le w \le 1 \ .$$

The weight w can be optimized via cross-validation on a hold-out set, using metrics such as MAE or RMSE.

4. Result Analysis

The ARIMA component ensures structure and interpretability, while the LSTM, augmented with temperature, allows the model to flexibly capture external, non-linear drivers of demand. The use of decaying trend adjustment and blending further promotes stability and robustness in post-promotion forecasting.

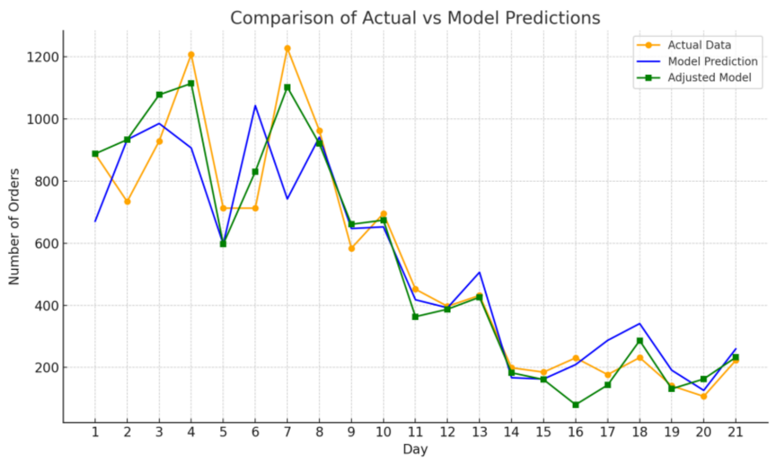

Figure. 3

Figure 3 shows the comparison between model prediction and the data from test data, besides it contains another line named 'adjusted model'. The adjusted model is updated after using the first-day data average of the test data. And then we can see the accuracy rate is higher than the original one. For the original and adjusted models, the RMSE values were 166.53 and 85.71, respectively. Meanwhile, the MAPE values were 21.29% and 16.26%, respectively. They indicate that the prediction model can provide a more accurate next-day volume forecasts compared to the company's empirical law.

5. Discussion and Conclusions

The proposed two-phase framework, which uses a log-normal model for the promotion period and a trend-adjusted ARIMA+LSTM ensemble (with temperature inputs) for post-promotion period. It demonstrates flexibility in capturing both extreme surges and regular demand patterns.

However, several limitations remain:

- Distributional fit: Heavy promotional shocks may produce multi-modal volumes that violate the log-normal assumption, biasing forecasts.
- Aggregation loss: Averaging across shops masks local demand heterogeneity and shop-specific effects.
- Static blending: A fixed weight between ARIMA/LSTM and the decay trend may not adapt quickly to sudden market shifts.
- Limited exogenous inputs: Relying solely on temperature ignores other factors (holidays, competitor actions) that influence residual demand.

Addressing these issues, through mixture distributions, shop-level models, adaptive ensemble weights, and richer external variables offers promising avenues for enhancing forecast accuracy and operational usability.

References

[1]Chien-Chih Wang; Chun-Hua Chien; Amy J.C.. Trappey On the Application of ARIMA and LSTM to Predict Order Demand Based on Short Lead Time and On-Time Delivery Requirements. Processes 2021, 9(7), 1157. [MDPI]

[2]Liyun Su; Lang Xiong; Jialing Yang. Multi-Attn BLS: Multi-head attention mechanism with broad learning system for chaotic time series prediction. Applied Soft Computing, Volume 132, January 2023, 109831. [ScienceDirect]

[3]Fotios Petropoulos; Daniele Apiletti; Vassilios Assimakopoulos. Forecasting: theory and practice. arXiv:2012.03854v4 [stat.AP] 5 Jan 2022. [arXiv]

[4]Cristina Carvallo. Improve Demand Planning With Consumption Data. Gartner December 17, 2024. [Gartner]

[5]Philipp K. Janert. Data Analysis with Open Source Tools. O’Reilly (2010). [O’Reilly]