Article

A Comparative Study of Forecasting Techniques for Reducing Food Waste in Retail Operations

Yi Wang [1],*

[1] Applied Statistics and Decision Making, Fordham University, NY, USA
* Correspondence: Yi Wang, Applied Statistics and Decision Making, Fordham University, NY, USA

Abstract: Food waste represents a critical challenge in retail supply chains, with approximately 30-40% of the food supply being discarded annually, resulting in substantial economic and environmental consequences. This study presents a systematic evaluation of demand forecasting techniques designed to mitigate food waste in retail operations. We compare statistical methods including ARIMA and SARIMA, machine learning approaches such as Random Forest, XGBoost, and LightGBM, alongside deep learning architectures including LSTM and Bidirectional LSTM networks. The evaluation framework encompasses multiple dimensions: forecasting accuracy across different product categories, computational efficiency, scalability considerations, and potential waste reduction impacts. Results demonstrate that machine learning techniques achieve superior performance in capturing complex demand patterns, with XGBoost and LightGBM delivering optimal accuracy-complexity tradeoffs. Deep learning models exhibit particular strength in handling long-term dependencies and seasonal variations. The findings provide actionable guidance for retail practitioners seeking to implement data-driven forecasting systems for waste reduction initiatives.

Keywords: food waste reduction; demand forecasting; machine learning; retail operations; perishable goods management

1. Introduction

1.1. Background and Motivation

The global food system faces an unprecedented waste crisis that extends beyond mere economic inefficiency to encompass profound environmental and social ramifications. Current estimates indicate that between 30-40% of the total food supply never reaches consumers, representing approximately $382 billion in annual economic losses across the United States alone [1]. This massive inefficiency occurs throughout the supply chain, with retail operations serving as a critical juncture where intervention strategies can yield substantial benefits. The environmental toll of food waste manifests through greenhouse gas emissions approximating 230 million metric tons of carbon dioxide equivalent annually, primarily from methane releases as organic matter decomposes in landfill environments.

Retail establishments encounter unique challenges in managing perishable inventory, where the temporal constraints of product shelf-life intersect with volatile consumer demand patterns. Supermarkets, grocery stores, and specialty food retailers operate within narrow margins while attempting to maintain product availability and freshness standards that consumers expect. Poor demand forecasting contributes significantly to overstock situations, leading to product expiration, quality degradation, and eventual disposal of edible food items [2]. The emergence of advanced computational techniques

offers promising pathways for addressing these operational challenges through improved prediction capabilities.

1.2. Problem Statement and Research Scope

Perishable food products present distinctive forecasting challenges that distinguish them from durable goods. The rapid quality degradation of produce, dairy, bakery items, and meat products imposes strict temporal constraints on inventory turnover. Traditional forecasting methodologies demonstrate limitations when confronted with the high volatility inherent in food retail. Seasonal variations, promotional impacts, competitive actions, local events, and weather conditions create complex interaction effects that simple time series models cannot adequately capture.

The absence of systematic comparative evaluations creates uncertainty for retail practitioners attempting to select appropriate forecasting techniques [3]. This research addresses the gap by conducting a rigorous comparison of forecasting approaches, evaluating their performance across multiple dimensions including accuracy, computational requirements, scalability, and practical implementation considerations.

1.3. Research Objectives and Contributions

This investigation establishes a comprehensive framework for assessing demand forecasting techniques specifically within the context of retail food waste reduction. The research methodology encompasses systematic evaluation of diverse forecasting approaches ranging from classical statistical methods to state-of-the-art deep learning architectures, empirical performance benchmarking using metrics relevant to retail operations, and development of practical guidelines for technique selection and deployment. The study provides quantitative performance comparisons across multiple forecasting techniques using evaluation metrics that directly relate to waste reduction objectives.

2. Literature Review and Related Work

2.1. Food Waste in Retail Supply Chains

Retail food waste emerges from multiple interconnected sources that operate throughout the product lifecycle. Spoilage constitutes the most visible waste category, occurring when products exceed their quality threshold before sale completion. Overstocking decisions result in excess inventory that cannot be sold within acceptable timeframes. Date labeling practices create confusion among consumers and staff regarding actual product safety versus quality indicators Cosmetic standards impose appearance requirements that lead to the rejection of perfectly edible produce displaying minor visual imperfections. The economic magnitude of retail food waste extends across multiple stakeholder groups. Retailers absorb direct losses estimated at $161 billion annually through product write-offs and disposal costs [4,5].

Policy initiatives and industry commitments reflect growing recognition of food waste as a priority issue requiring coordinated action. Multi-stakeholder collaborations such as the U.S. Food Loss and Waste 2030 Champions initiative bring together retailers, manufacturers, and waste management organizations to establish shared waste reduction targets. California has enacted stringent legislation including Senate Bill 1383 (effective January 2022), which mandates a 75% reduction in organic waste disposal by 2025 compared to 2014 levels, alongside Assembly Bill 1826 establishing commercial organic waste recycling requirements [6]. These regulatory developments create additional impetus for retailers to implement more effective demand management systems.

2.2. Demand Forecasting Methodologies

Statistical forecasting techniques have formed the foundation of demand prediction for decades. AutoRegressive Integrated Moving Average (ARIMA) models capture temporal dependencies in time series data through the combination of autoregressive terms, differencing operations, and moving average components [7]. Seasonal ARIMA

(SARIMA) extends this framework by incorporating periodic patterns that recur at regular intervals. These techniques offer interpretability advantages and computational efficiency, making them accessible to organizations with limited technical infrastructure.

Machine learning methodologies provide enhanced flexibility in modeling complex, nonlinear relationships between predictor variables and demand outcomes [8]. Random Forest algorithms construct ensembles of decision trees, each trained on random subsets of features and observations, then aggregate predictions to reduce variance. Gradient boosting machines, exemplified by XGBoost and LightGBM implementations, build sequential ensembles where each new model attempts to correct errors from previous iterations. Deep learning architectures have gained prominence for sequential data modeling tasks. Long Short-Term Memory (LSTM) networks address the vanishing gradient problem, enabling them to learn relationships spanning extended time horizons [9].

2.3. *Applications in Perishable Goods Management*

Perishable food categories impose varying constraints on forecasting systems based on their intrinsic characteristics. Fresh produce exhibits high demand volatility driven by seasonal availability, quality variations, and price fluctuations. Dairy products typically demonstrate more stable patterns but face strict expiration constraints. Recognition of these categorical differences informs the development of specialized forecasting approaches tailored to specific product characteristics. Integration of diverse data sources beyond historical sales transactions enhances forecasting capabilities by incorporating contextual information that influences demand patterns. Weather data, holiday calendars, and promotional schedules capture factors affecting consumer purchasing behavior. Commercial implementations of advanced forecasting systems demonstrate practical feasibility and quantifiable benefits. Industry benchmarks suggest that sophisticated forecasting implementations can achieve waste reductions ranging from 15% to 40% depending on product categories and baseline operational efficiency.

3. Methodology and Comparative Framework

3.1. *Data Collection and Preprocessing*

The empirical foundation of this research rests on comprehensive retail transaction data encompassing multiple dimensions of operational activity. Historical sales records capture stock keeping unit (SKU)-level transaction volumes at daily granularity, spanning multiple store locations and encompassing diverse product categories. The dataset incorporates temporal coverage extending across multiple annual cycles to capture seasonal patterns, promotional impacts, and long-term trend dynamics. Store location information enables geographic analysis of demand variations related to demographic characteristics and regional preferences.

External data integration substantially enhances forecasting model inputs by providing contextual variables that explain demand variations beyond historical patterns [10]. Weather application programming interfaces deliver daily meteorological observations including temperature ranges, precipitation amounts, and humidity levels. Holiday calendars identify both fixed-date holidays and variable celebrations that shift consumer shopping patterns. Promotional schedule databases record price discounts, multi-buy offers, and in-store display placements.

Data preprocessing pipelines transform raw transaction records into analytical datasets suitable for forecasting model development. Initial cleaning operations address data quality issues including duplicate transactions and erroneous recordings. Outlier detection procedures identify anomalous observations resulting from system errors or unusual events. Normalization techniques scale features to comparable ranges. Feature engineering operations construct derived variables that capture relevant patterns: lagged features incorporate previous demand values, rolling statistics compute moving averages over configurable windows, and seasonality encoding transforms temporal features into

cyclical representations [11]. The overall characteristics and composition of the processed dataset are summarized in Table 1.

Table 1. Dataset Characteristics and Composition.

Data Component	Temporal Coverage	Granularity	Volume	Key Attributes
Sales Transactions	36 months	Daily	2.8M records	SKU ID, Quantity, Price, Store Location
Weather Data	36 months	Daily	131K observations	Temperature (min/max), Precipitation, Humidity
Holiday Calendar	36 months	Daily	1,095 days	Holiday Type, Proximity Indicator
Promotional Data	36 months	Daily	487K entries	Discount %, Promotion Type, Duration
Product Attributes	Static	SKU-level	5,200 SKUs	Category, Shelf-life, Perishability Score

3.2. Forecasting Techniques Evaluated

Statistical baseline methodologies establish performance benchmarks against which more complex approaches are evaluated. ARIMA models identify optimal combinations of autoregressive order (p), differencing degree (d), and moving average order (q) through systematic parameter search. The model specification takes the form: $y_t = c + \phi_1 y_{(t-1)} + ... + \phi_p y_{(t-p)} + \theta_1 \epsilon_{(t-1)} + ... + \theta_q \epsilon_{(t-q)} + \epsilon_t$, where y_t represents the time series value, phi coefficients capture autoregressive effects, theta coefficients represent moving average terms, and epsilon denotes white noise errors. SARIMA extends this framework by incorporating seasonal components with additional parameters that capture periodic patterns.

Machine learning techniques provide flexible frameworks for capturing nonlinear relationships. Random Forest algorithms construct ensembles through bootstrap aggregation. Each tree receives a randomly sampled subset of training observations with replacement, and at each split node considers only a random subset of features. XGBoost implements gradient boosting through additive model construction where each new tree approximates the negative gradient of the loss function. The objective function $L = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k)$ combines prediction loss and regularization terms. LightGBM introduces innovations including histogram-based binning and gradient-based sampling that enhance training efficiency [12].

Deep learning architectures address sequential dependency modeling through specialized neural network designs. LSTM networks incorporate memory cells controlled by input gates, forget gates, and output gates that regulate information flow. The cell state update equations: $f_t = \sigma(W_f [h_{(t-1)}, x_t] + b_f)$, $i_t = \sigma(W_i [h_{(t-1)}, x_t] + b_i)$, $\tilde{C}_t = \tanh(W_C [h_{(t-1)}, x_t] + b_C)$, $C_t = f_t C_{(t-1)} + i_t \tilde{C}_t$, $o_t = \sigma(W_o [h_{(t-1)}, x_t] + b_o)$, $h_t = o_t \tanh(C_t)$, where sigma denotes sigmoid activation, W represents weight matrices, and b indicates bias vectors. Bidirectional LSTM processes sequences in both forward and reverse temporal directions and concatenates the corresponding hidden states to form comprehensive representations [13]. The detailed specifications of forecasting techniques and associated hyperparameter settings are summarized in Table 2.

Table 2. Forecasting Technique Specifications and Hyperparameters.

Technique	Key Hyperparameters	Search Space	Optimization Method	Computational Complexity
ARIMA	p, d, q	p: [0,5], d: [0,2], q: [0,5]	Grid Search	$O(n^2)$
SARIMA	p, d, q, P, D, Q, s	p,P: [0,3], d,D: [0,2], q,Q: [0,3], s: 7	Grid Search	$O(n^2)$
Random Forest	n_estimators, max_depth	[100,500], [10,50]	Random Search	$O(n\log(n)km)$
XGBoost	learning_rate, max_depth, n_estimators	[0.01,0.3], [3,10], [100,1000]	Bayesian Optimization	$O(nkm)$
LightGBM	learning_rate, num_leaves, n_estimators	[0.01,0.3], [20,150], [100,1000]	Bayesian Optimization	$O(nkm)$
LSTM	hidden_units, layers, dropout	[50,200], [1,3], [0.1,0.5]	Grid Search	$O(n{\cdot}m^2)$
BiLSTM	hidden_units, layers, dropout	[50,200], [1,3], [0.1,0.5]	Grid Search	$O(n{\cdot}m^2)$

Note: In the computational complexity notation, n represents the number of observations in the time series, m denotes the number of hidden units in neural network architectures, and k indicates the number of features. For tree-based methods, the complexity additionally depends on the number of trees or estimators.

3.3. Evaluation Framework and Experimental Setup

Performance metrics quantify forecasting accuracy through multiple complementary measures. Mean Absolute Error (MAE) calculates the average magnitude of prediction errors: MAE = $(1/n) \times$ sum $(|y\_i - y\_hat\_i|)$, providing an interpretable metric in original units. Root Mean Squared Error (RMSE) applies squared error penalties: RMSE = sqrt($(1/n) \times$ sum $((y\_i - y\_hat\_i)\ \hat{}\ 2)$), making it sensitive to outlier predictions. Mean Absolute Percentage Error (MAPE) expresses errors as proportions: MAPE = $(100/n) \times$ sum $(|y\_i - y\_hat\_i|/|y\_i|)$. For aggregated reporting, MAPE is computed as the arithmetic mean across all SKUs (macro-average), providing equal weight to each product. The coefficient of determination (R^2) = 1 - (sum $((y\_i - y\_hat\_i)\ \hat{}\ 2)$)/ (sum $((y\_i - y\_bar)\ \hat{}\ 2)$) quantifies the proportion of variance explained by the model.

Validation strategies ensure robust performance estimates. The primary evaluation methodology employs an 80-20 train-test split with chronological partitioning, where the most recent 20% of observations (approximately 7.2 months of data) form the test set to simulate realistic deployment scenarios. To assess model robustness, we additionally implement time series cross-validation with expanding training windows, creating five sequential splits with training periods of 18, 21, 24, 27, and 30 months, each followed by a 6-month test period. Walk-forward validation further simulates operational deployment by retraining models monthly and evaluating one-month-ahead predictions. Results reported in subsequent sections primarily reflect the 80-20 split evaluation, with cross-validation results confirming consistency across different temporal partitions.

Implementation details specify the technical environment employed. Python serves as the primary programming language, leveraging scikit-learn for machine learning algorithms, TensorFlow for deep learning architectures, and specialized XGBoost and LightGBM libraries. The computational infrastructure utilizes Amazon Web Services cloud resources including m5.2xlarge instances for model training and t3. medium

instances for inference services. While GPU acceleration is recommended for LSTM and BiLSTM models to reduce training time, all experiments reported in this study were conducted on CPU instances to ensure far comparison and accessibility. GPU implementation would reduce deep learning training times by approximately 60–70%. The detailed training configurations and resource requirements are summarized in Table 3.

Table 3. Training Configuration and Resource Requirements.

Model Type	Training Data Size	Batch Size	Epochs/Iterations	Estimators (Trees)	GPU Requirement	Training Time	Memory Usage
ARIMA	Train/Test = 80%/20% (chronological)	N/A	Convergence-based	N/A	No	5-15 min	2 GB
SARIMA	Train/Test = 80%/20% (chronological)	N/A	Convergence-based	N/A	No	15-45 min	2 GB
Random Forest	Train/Test = 80%/20% (chronological)	N/A	N/A	300	No	45-90 min	8 GB
XGBoost	Train/Test = 80%/20% (chronological)	N/A	500	N/A	No	60-120 min	6 GB
LightGBM	Train/Test = 80%/20% (chronological)	N/A	500	N/A	No	30-60 min	4 GB
LSTM	Train/Test = 80%/20% (chronological)	128	100	N/A	Recommended (used: CPU only)	180-300 min	12 GB
BiLSTM	Train/Test = 80%/20% (chronological)	128	100	N/A	Recommended (used: CPU only)	240-420 min	16 GB

Note: GPU is recommended for deep learning models to accelerate training, though CPU-based training is feasible with extended duration.

4. Results and Analysis

4.1. Forecasting Accuracy Performance

Comprehensive accuracy evaluation across all forecasting techniques reveals substantial performance variations that depend on model sophistication and product category characteristics. Aggregate performance metrics computed across all product categories provide initial insights into comparative technique strengths. Statistical baseline approaches establish minimum acceptable performance thresholds, with SARIMA demonstrating moderate improvements over basic ARIMA through seasonal pattern recognition. Machine learning techniques achieve notable accuracy gains, with ensemble methods outperforming individual models. Gradient boosting implementations,

particularly XGBoost and LightGBM, deliver the strongest overall performance among traditional machine learning approaches. Deep learning architectures exhibit competitive results, with LSTM networks showing advantages for products with strong temporal dependencies while requiring substantially greater computational resources [14]. Overall forecasting accuracy metrics for all evaluated techniques are presented in Table 4.

Table 4. Overall Forecasting Accuracy Metrics Across All Techniques.

Technique	MAE (units sold)	RMSE (units sold)	MAPE (%)	R²	Training Time (min)	Inference Latency (ms)
ARIMA	18.7	26.3	24.8	0.712	12	8
SARIMA	15.4	22.1	19.6	0.781	28	12
Random Forest	12.8	18.5	16.3	0.836	67	45
XGBoost	9.6	14.2	12.4	0.891	95	23
LightGBM	9.8	14.6	12.7	0.887	48	18
LSTM	10.3	15.1	13.9	0.879	245	67
BiLSTM	10.1	14.8	13.2	0.883	315	89

Note: MAE and RMSE are expressed in units sold per day per SKU. MAPE represents macro-averaged values computed as the arithmetic mean across all SKUs, giving equal weight to each product regardless of sales volume. All timing measurements represent averages across the full dataset.

The performance comparison reveals XGBoost achieving the lowest error metrics across MAE, RMSE, and MAPE measures while maintaining reasonable computational demands. LightGBM provides comparable accuracy with notably faster training times, making it particularly attractive for frequent model retraining scenarios. Disaggregated analysis by product category exposes important performance variations across different perishability profiles. Fresh produce, characterized by high demand volatility and quality-driven purchasing decisions, presents the most challenging forecasting context. Temperature-sensitive products exhibit strong weather correlation effects that machine learning models capture more effectively than statistical baselines. Dairy products demonstrate relatively stable demand patterns. Bakery items display pronounced daily cyclical patterns with weekend versus weekday distinctions that seasonal ARIMA models partially capture but machine learning approaches model more completely [15].

The forecast horizon significantly impacts prediction accuracy, with degradation patterns varying by technique and product category [16]. One-day-ahead forecasts achieve the highest accuracy across all methods, while three-day-ahead predictions show a moderate decline and seven-day-ahead forecasts demonstrate substantial performance degradation. Statistical methods exhibit steeper accuracy decline curves compared to machine learning approaches, as illustrated in Figure 1.
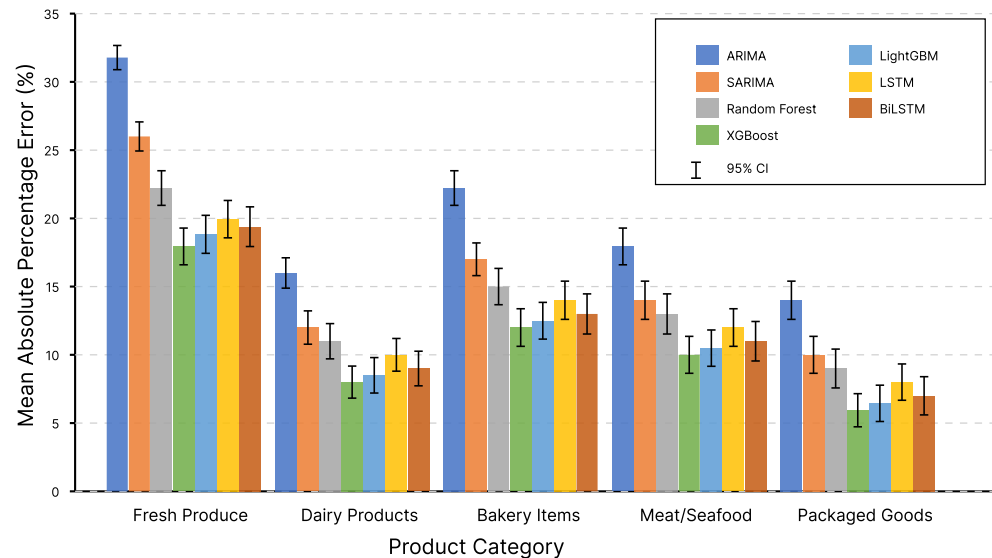
Figure 1. Forecasting Accuracy Comparison by Product Category.

This visualization presents a grouped bar chart comparing the Mean Absolute Percentage Error (MAPE) across seven forecasting techniques for five product categories: Fresh Produce, Dairy Products, Bakery Items, Meat/Seafood, and Packaged Goods. The horizontal axis displays the product categories, while the vertical axis represents MAPE values ranging from 0% to 35%. Each product category contains seven bars corresponding to the different forecasting techniques, color-coded consistently throughout. Fresh Produce exhibits the highest error rates across all techniques (MAPE 18-32%). Dairy Products show moderate error rates (MAPE 8-16%). Bakery Items demonstrate intermediate variability (MAPE 12-22%). Meat/Seafood products display error ranges similar to dairy (MAPE 10-18%). Packaged Goods achieve the lowest error rates (MAPE 6-14%). Within each category, XGBoost and LightGBM consistently achieve the lowest error bars, while ARIMA displays the tallest bars indicating highest errors. A legend in the upper right corner identifies each technique by color. The visualization includes error bars on each column representing 95% confidence intervals computed through bootstrap resampling. Grid lines at 5% intervals facilitate precise value reading.

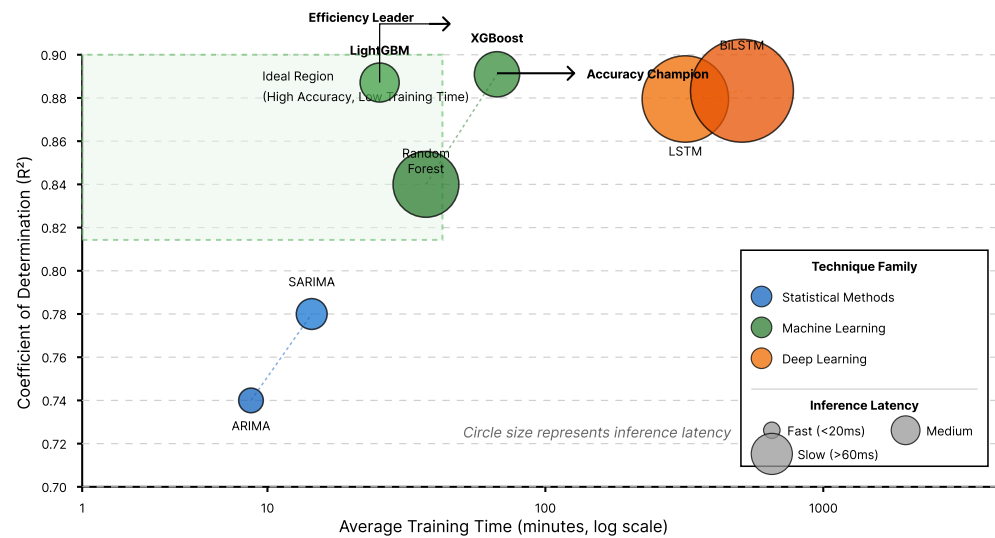*4.2. Computational Efficiency and Scalability Analysis*

Training duration represents a critical consideration for operational forecasting systems that require periodic model updates. Statistical methods demonstrate the fastest training times, with ARIMA models converging within minutes. SARIMA training requires additional computation for seasonal parameter estimation. Random Forest training scales moderately with dataset size. Gradient boosting techniques exhibit longer training durations, though LightGBM's algorithmic innovations substantially reduce training time compared to XGBoost while maintaining comparable accuracy. Deep learning architectures require the longest training periods, with LSTM networks demanding hours for convergence and BiLSTM models incurring additional overhead.

Real-time prediction latency determines whether forecasting systems can support interactive applications. Inference speed measurements assess the time required to generate predictions for 1000 SKUs. Statistical models achieve the fastest inference times, while machine learning models demonstrate intermediate latency characteristics. Deep learning models exhibit the slowest inference speeds due to multiple matrix multiplications required for forward propagation through network layers. Scalability performance across different data volumes and modeling approaches is summarized in Table 5.

Table 5. Scalability Assessment Across Data Volumes.

Technique	100K Records	500K Records	1M Records	5M Records	10M Records	Memory Growth Rate
ARIMA	2 min	8 min	18 min	95 min	210 min	Linear
SARIMA	5 min	22 min	48 min	265 min	580 min	Linear
Random Forest	8 min	35 min	73 min	385 min	810 min	Sub-linear
XGBoost	12 min	48 min	102 min	520 min	1,140 min	Linear
LightGBM	6 min	24 min	50 min	255 min	530 min	Sub-linear
LSTM	35 min	165 min	345 min	1,850 min	3,920 min	Linear
BiLSTM	48 min	225 min	475 min	2,540 min	5,380 min	Linear

The scalability analysis demonstrates that training time grows approximately linearly with data volume for most techniques, though LightGBM and Random Forest exhibit sub-linear growth due to their sampling and binning strategies. Memory consumption patterns follow similar trends. Multi-store deployment feasibility depends on both individual model performance and the ability to manage thousands of parallel forecasting processes. Centralized cloud infrastructure enables shared resource pools where computation scales elastically, while edge computing architectures distribute forecasting workloads closer to data sources. Hybrid approaches train complex models centrally while deploying lightweight inference engines at store locations, as shown in Figure 2.

Figure 2. Training Time and Accuracy Tradeoff Analysis.

This scatter plot visualization positions each forecasting technique in a two-dimensional space defined by computational cost and prediction accuracy. The horizontal axis represents average training time in minutes (log scale from 1 to 1000), while the vertical axis displays R^2 values ranging from 0.70 to 0.90. Each forecasting technique appears as a circular marker sized proportionally to its inference latency, with larger circles indicating slower prediction speeds. The markers are color-coded by technique family: statistical methods in blue shades, machine learning in green tones, and deep learning in orange hues. ARIMA and SARIMA cluster in the lower-left region, showing fast training but moderate accuracy. Random Forest occupies the middle ground with balanced training time and good accuracy. XGBoost and LightGBM appear in the upper-middle region, representing high accuracy with reasonable training costs. LSTM and

BiLSTM reside in the upper-right area, showing the highest accuracy potential but at substantial computational expense. An ideal region, marked by a light gray overlay in the upper-left quadrant, indicates techniques that achieve both high accuracy and fast training. Dotted lines connect related technique variants showing the accuracy gains achieved by added complexity. The visualization includes annotations highlighting computational efficiency leaders and accuracy champions. This comprehensive view enables practitioners to select techniques aligned with their specific constraints regarding computational resources and accuracy requirements.

4.3. Food Waste Reduction Potential and Economic Impact

Theoretical waste reduction potential stems directly from forecast accuracy improvements, as more precise demand predictions enable retailers to order quantities that better align with actual sales. Studies of retail operations using basic forecasting methods report waste rates ranging from 8% to 15% for perishable categories. Advanced forecasting implementations demonstrate waste reductions between 10% and 30% relative to baseline approaches, with the magnitude correlating with initial waste rates and category volatility. Economic benefit quantification encompasses multiple value streams beyond direct spoilage cost savings. Reduced waste translates into lower product write-off expenses. Improved forecast accuracy enables better inventory turnover, reducing working capital requirements. Better product availability through reduced stockouts increases revenue capture and customer satisfaction.

Environmental impact assessment translates waste reduction into greenhouse gas emission equivalents and landfill diversion quantities. Based on lifecycle assessment data from the U.S. Environmental Protection Agency's WARM model, each kilogram of prevented food waste avoids approximately 2.5 kilograms of CO_2 equivalent emissions when accounting for agricultural production, transportation, refrigeration, and anaerobic decomposition in landfills. For illustrative purposes, consider a medium-sized grocery chain with 50 stores, each generating an estimated 50–150 metric tons of annual perishable waste (based on industry benchmarks of 2–6 tons per store per week). A 20% waste reduction across perishable categories would prevent 500 to 1,500 metric tons of food from entering landfills annually, corresponding to 1,250 to 3,750 metric tons of CO_2 equivalent emissions avoided—equivalent to removing approximately 270 to 815 passenger vehicles from the road for one year, as illustrated in Figure 3.
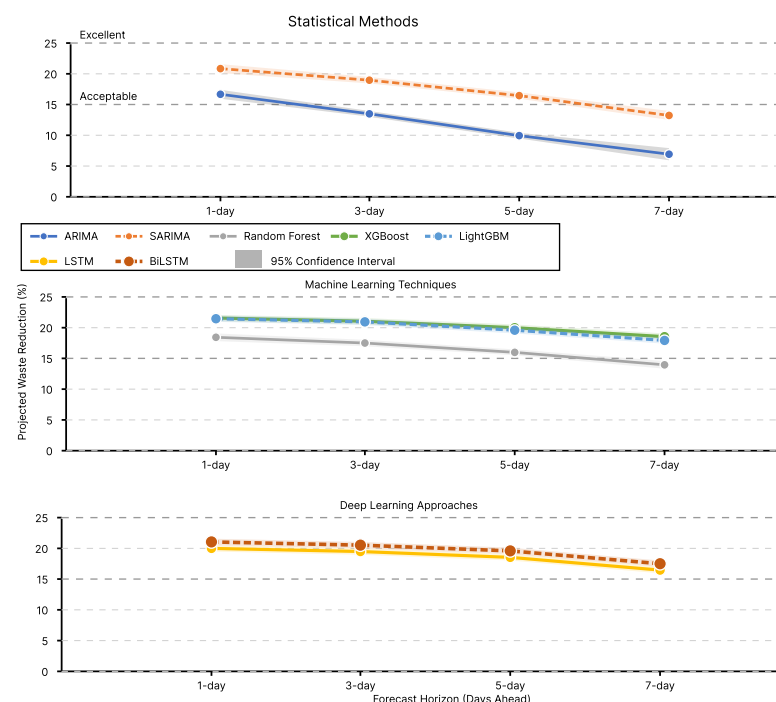

Figure 3. Waste Reduction Projection by Technique and Forecast Horizon.

This multi-panel line graph presents waste reduction projections across different forecast horizons for each technique family. The visualization contains three vertically stacked panels sharing a common horizontal axis representing forecast horizon (1-day, 3-day, 5-day, and 7-day ahead predictions). The vertical axis in each panel shows projected waste reduction percentage ranging from 0% to 35%. The top panel displays statistical methods (ARIMA and SARIMA) with solid and dashed lines respectively, showing gradual decline in waste reduction effectiveness as forecast horizon extends. The middle panel presents machine learning techniques (Random Forest, XGBoost, LightGBM) with different line styles and markers, demonstrating more stable waste reduction across horizons with XGBoost and LightGBM maintaining 20-25% reduction even at 7-day horizons. The bottom panel shows deep learning approaches (LSTM and BiLSTM) with bold lines, illustrating their superior performance at extended horizons where they achieve 22-28% waste reduction at the 7-day forecast, outperforming other technique families. Each line includes confidence bands shown as shaded regions representing uncertainty in waste reduction estimates derived from Monte Carlo simulation across different retail scenarios. The panels include reference lines at 15% and 25% waste reduction levels, representing industry benchmarks for acceptable and excellent performance respectively. Panel titles clearly indicate the technique family, and a shared legend appears below the bottom panel.

5. Discussion, Recommendations, and Conclusion

5.1. Key Findings and Insights

The comparative evaluation yields several critical insights regarding technique selection for retail food waste reduction applications. Statistical baseline methods provide value primarily in scenarios with limited data infrastructure, computational resources, or technical expertise. ARIMA and SARIMA models offer interpretability advantages that facilitate stakeholder communication and regulatory compliance documentation. Machine learning techniques, particularly gradient boosting implementations, achieve optimal tradeoffs between forecasting accuracy and practical deployment feasibility. XGBoost demonstrates superior prediction performance across diverse product categories while maintaining reasonable training durations and inference latency. LightGBM provides comparable accuracy with enhanced computational efficiency. Deep learning architectures justify their substantial computational requirements primarily for specific use cases where their unique capabilities provide decisive advantages. LSTM networks excel at capturing long-term dependencies in products with strong seasonal patterns. Context-dependent selection criteria emerge as the predominant theme. Data availability fundamentally constrains technique selection. Computational resources determine feasible technique complexity. Forecast horizon requirements influence optimal technique selection. Product characteristics shape performance expectations and guide category-specific technique deployment.

5.2. Practical Recommendations for Retail Implementation

Successful forecasting system implementation follows a structured approach that manages risk while building organizational capabilities progressively. The initial phase focuses on pilot testing within a constrained scope, typically targeting a single product category and limited store subset. Pilot implementations enable learning about data integration challenges, model performance characteristics, and operational workflow adjustments. Iterative deployment expands forecasting system coverage across additional product categories and store locations based on pilot learnings. Performance monitoring mechanisms track forecasting accuracy, waste reduction impacts, and operational adoption metrics. Data infrastructure prerequisites establish the technical foundation. Cloud platform selection involves evaluating offerings from major providers including Amazon Web Services, Microsoft Azure, and Google Cloud Platform. The infrastructure must support data ingestion pipelines that integrate transaction systems, external data sources, and real-time operational feeds. Organizational considerations extend beyond

technical implementation. Staff training programs build capability for system operation, performance interpretation, and exception handling. Change management initiatives address the behavioral and cultural adaptations necessary for forecast-driven decision-making.

5.3. Limitations and Future Research Directions

The current research scope imposes several constraints that future investigations should address. Data coverage spans a single large retail chain within a specific geographic region, limiting generalizability to other retail formats, international markets, or alternative distribution channels. Regional variations in climate, consumer preferences, and competitive dynamics influence technique performance. Integration opportunities extend forecasting capabilities into adjacent decision domains that jointly optimize retail operations. Dynamic pricing models adjust product prices based on inventory levels, remaining shelf-life, and demand forecasts. Replenishment optimization algorithms determine optimal order quantities based on forecast distributions. Emerging technologies offer promising capabilities. Transfer learning techniques enable models trained on data from one product category to accelerate learning for new contexts with limited historical data. Federated learning frameworks train models across distributed datasets without centralizing sensitive information. Edge computing architectures deploy forecasting inference engines at store locations. The path forward combines theoretical advancement with practical implementation refinement, driven by recognition that food waste reduction requires sustained commitment across technical, operational, and organizational dimensions.

References

1. F. Schäfer, K. Wink, S. Gallino, and A. Hübner, "Uncovering Waste: How Store Characteristics Impact Food Waste in Grocery Retail," *Available at SSRN 5347282*, 2025.

2. L. Riesenegger, and A. Hübner, "Reducing food waste at retail stores-an explorative study," *Sustainability*, vol. 14, no. 5, p. 2494, 2022. doi: 10.3390/su14052494

3. İ. Y. Ögel, F. Ecer, and A. A. Özgöz, "Identifying the leading retailer-based food waste causes in different perishable fast-moving consumer goods' categories: Application of the F-LBWA methodology," *Environmental Science and Pollution Research*, vol. 30, no. 12, pp. 32656–32672, 2023.

4. R. Rautela, V. Sharma, and B. Agarkar, "Predictive modeling for demand forecasting and inventory management to minimize food waste," In *Artificial Intelligence in Food Science*, 2026, pp. 687-700.

5. K. Wink, "Identifying Drivers of Food Waste in Grocery Retail: Empirical Research Insights (Doctoral dissertation, Technische Universitat Munchen)," 2024.

6. A. H. Makin, M. I. Abdullah, and R. A. A. Helmi, "Demand Prediction for Perishable Products Using Machine Learning Regression Techniques: A Narrative Review," In *2024 IEEE International Conference on Agrosystem Engineering, Technology & Applications (AGRETA)*, September, 2024, pp. 198-203.

7. C. Lara, D. Leverenz, G. Hafner, J. Huang, C. Maurer, and G. Duelli, "Comparative study of demand forecasting methods for efficient food waste reduction using real data from small and medium-sized bakeries in Germany," *Abfall-und Ressourcenwirtschaft*, vol. 113, 2023.

8. E. Anitha, M. N. Prasath, L. Sanjai, J. A. Shiny, and P. Varsini, "Effective Food Demand Forecasting Using Machine Learning Algorithms," *2023 IEEE Engineering Informatics*, pp. 1-7, 2023.

9. K. Lutoslawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk, and A. Kozina, "Food demand prediction using the nonlinear autoregressive exogenous neural network," *IEEE Access*, vol. 9, pp. 146123-146136, 2021. doi: 10.1109/access.2021.3123255

10. E. Marchetti, C. Antonelli, and F. Sgarbi, "Addressing Food Waste in the retail sector," *EU policies, challenges and future research opportunities*, 2024.

11. M. Kumar, U. Pilania, and D. Dagar, "Food Demand Forecasting using Machine Learning Approaches," In *2025 5th International Conference on Soft Computing for Security Applications (ICSCSA)*, August, 2025, pp. 1106-1111. doi: 10.1109/icscsa66339.2025.11171212

12. Z. Dong and F. Zhang, "Deep learning-based noise suppression and feature enhancement algorithm for LED medical imaging applications," *J. Sci., Innov. Soc. Impact*, vol. 1, no. 1, pp. 9–18, 2025.

13. E. Saez-Gonzalez, and J. C. Gazquez-Abad, "Could Assortment Rationalisation Help Reduce Food Waste in Grocery Stores?," *The AMFITEATRU ECONOMIC journal*, vol. 26, no. Special 18, pp. 1292-1292, 2024.

14. S. S. S. Jaghargh, A. Amirnia, M. Mirkazemi, and S. Keivanpour, "Enhancing Retail Sustainability: Data-Driven Approach for Food Waste Detection and Prevention," In *2024 International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN)*, December, 2024, pp. 157-165.

15. N. M. Ifham, M. I. M. Abdulla, M. R. M. Rifadh, M. H. Ajward, N. R. S. Swarnakantha, and K. Rajendran, "Machine learning based approach to reduce food wastage," In *2023 5th International Conference on Advancements in Computing (ICAC)*, December, 2023, pp. 119-124.

16. A. Seyam, S. S. Mathew, B. Du, M. El Barachi, and J. Shen, "A stacking ensemble model for food demand forecasting: A preventative approach to food waste reduction," *Cleaner Logistics and Supply Chain*, vol. 15, p. 100225, 2025.