Enhanced Inventory Demand Forecasting with Machine Learning

Haoyuan Ren

Rensselaer Polytechnic Institute, Troy, America

rhy2756339798@gmail.com

Abstract. Modeling inventory demand is critical for businesses to manage resources and ensure customer satisfaction. Traditional economic models, rooted in utility functions and structural approaches, often face challenges due to stringent assumptions and inability to adapt to real-world data. This research harnesses machine learning, specifically the LightGBM algorithm, to enhance demand prediction. Unlike traditional models tied to Gaussian distribution, LightGBM adapts to actual data distributions, capturing complex, non-linear relationships. The results highlight sales channels and product types as pivotal demand drivers. This study blends traditional econometric techniques with modern machine learning, offering a roadmap for future demand forecasting research.

Keywords: demand estimation, gradient boosting decision trees, demand analytics

1 Introduction

Modeling inventory demand is crucial for businesses. Effective inventory management not only preserves valuable resources but also ensures continuous customer satisfaction. As a result, demand modeling has become a central focus in economic research, aimed primarily at accurately predicting consumer behavior.

Historically, economists have relied on utility functions to model consumer demand. They propose that consumers, behaving rationally, aim to maximize their utility or satisfaction, constrained by their budget. To more accurately reflect reality, numerous studies have modified the utility function to align with established economic principles. Concurrently, another significant strand of research has focused on introducing refined budgetary constraints. Structural models, deeply rooted in economic theory, have gained popularity. They are designed to ensure that the statistical relationships they reveal align well with theoretical expectations.

However, these conventional economic models, particularly the structural ones, encounter inherent challenges. The strict assumptions underpinning many of these models render empirical validation daunting. One significant drawback of traditional economic modeling is its unwavering commitment to the normal distribution assumption, both

X. Ding et al. (eds.), *Proceedings of the 2023 4th International Conference on Big Data and Social Sciences (ICBDSS 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 12,
https://doi.org/10.2991/978-94-6463-276-7_45

for error terms and input features. Often, real-world data deviate from this norm, showing traits like long-tailed distributions and other complexities. These disparities can compromise a model's predictive accuracy. Moreover, the inclination to simplify these models for analytical ease may reduce their effectiveness in real-world forecasting situations. Another major limitation is their inflexibility, often restraining their adaptability to fresh, unseen data.

To address these challenges, the predictive power of machine learning has been enlisted for demand forecasting. Integrating machine learning techniques with conventional economic models can significantly improve demand prediction, surmounting previously daunting hurdles. In our research, we have utilized the LightGBM algorithm, a variant of the Gradient Boosting Decision Trees (GBDT) framework. Rather than being bound to the Gaussian distribution, LightGBM employs feature binning, adapting more closely to the genuine data distribution. This method enables the model to discern intricate, non-linear relationships commonly found in real-world datasets. By fusing the theoretical depth of economic models with the empirical capabilities of machine learning, LightGBM offers a commendable combination of clarity and precision. Our findings indicate that, for this inventory demand dataset, sales channels and product types are the two pivotal factors influencing consumer demand. This study provides insights into how food providers can optimize resource allocation to enhance sales channels and product quality.

Our research culminates in three main contributions. Firstly, we harness Gradient Boosting Decision Trees to enhance predictive precision, bypassing the rigid normal distribution assumptions tied to traditional regression models. Secondly, we explore feature importance, employing cumulative information gain. Our results emphasize the crucial roles diverse channels and product categories play in demand forecasting. Finally, our study bridges machine learning and econometrics. We navigate an innovative path, merging traditional utility-maximization frameworks with state-of-the-art machine learning algorithms, laying down a foundation for future research in this arena.

2 Related Work

Our research is related to two distinct streams of literature: (1) economic models for inventory demand estimation (2) machine learning algorithms for prediction.

2.1 Economical Models for Demand Forecasting

In fundamental terms, the demand for a product tends to decrease as its price increases, and conversely, it increases as the price decreases. This inverse relationship is often depicted by a linear demand function, $Q = a - bP$, where 'Q' stands for the quantity demanded, 'P' denotes the price of the good, and 'a' and 'b' are constants. The concept of demand elasticity plays a pivotal role here, gauging how sensitive the quantity demanded is to price fluctuations or other external variables. This sensitivity is often visualized through the slope of the demand curve.

Within the realm of utility theory, consumer demand is conceptualized via utility functions. These functions posit that consumers, in their rationality, consistently make choices to maximize their utility, which can be equated to their level of satisfaction or happiness, all while adhering to their budgetary constraints.

The overarching theory of utility maximization has, over time, bifurcated into two distinctive research trajectories. One path has been the exploration and design of various utility functions that align with specific economic theories. Notable contributions in this vein include Cobb and Douglas's (1928) utility function, which elucidates the interplay between labor, capital, and product. [1] Subsequent endeavors by Wales and Woodland (1983) [2] shaped the utility in a quadratic mold, while Pollak and Wales (1992) [3] and Du and Kamakura (2008) [4] articulated a CES utility function. On another front, Lee and Pitt (1986) [5] and Millimet and Tchernis (2008) [6] adopted the indirect log utility in a dualistic approach to decipher demand equations.

Concurrently, a different school of thought placed its emphasis on delving deeper into budget constraints. This line of inquiry, bolstered by works such as those by van Soest et al. (1993) [7] and Tamer (2003), sought to unravel the intricate dynamics of financial limitations on demand. [8]

In more recent times, a trend has emerged wherein economists are leaning towards structural models. These models aim to weave economic theory seamlessly into empirical analyses. The end goal is to unearth the foundational structural parameters steering our economic systems. Pioneers in this domain include notable works by Millimet and Tchernis (2008) [6], Allenby et al. (2010) [9], Honka (2014) [10], and Hastings and Shapiro (2018). [11]

2.2 Machine Learning for Forecasting

The recent boom in machine learning technology has ignited a substantial interest in its potential for demand forecasting. Renowned for its versatility, machine learning can detect complex patterns and relationships within extensive historical datasets. Algorithms such as the Support Vector Machine (SVM) [12], Random Forest (RF), and Neural Networks (NN) have established themselves as cornerstones in demand forecasting. Their ability to handle high-dimensional data, identify non-linear data relationships, and swiftly adapt to market changes is unmatched. Recent innovations are now merging traditional economic models with machine learning techniques, boosting both precision and forecasting acumen.

Demand forecasting has traditionally toggled between two primary methodologies: classic economic models and machine learning techniques. Lately, an increasing number of studies have fused machine learning with econometric models. For example, Berry et al. (1993) employed a Logit model to predict product choices. [13] Breiman in his works from 1996 and 2001 introduced the random forest through a simulation experiment. [14] Bajari et al. (2015) [15] leveraged Lasso regression for variable selection and controlled the fixed effect of demand estimation. Adma et al. (2020) formulated a two-stage regression method, incorporating machine learning predictions to refine demand forecasting. [16] Furthermore, Gandhi et al. (2017) developed an objective centered around inequality to rectify prediction bias. [17]

Our research augments existing machine learning studies by integrating them with economic models. By leveraging cutting-edge machine learning algorithms, we aim to sidestep the stringent assumptions embedded in economic theoretical structures concerning demand prediction.

3 Methodology

3.1 Data

We gathered our data from Grupo Bimbo, a food provider in Mexico. The dataset captures the weekly demand and sales of specific food items. It offers a unique ID for each product and client, details of individual sales depots and sales channels, and information on the number of returns. The dataset calculates demand by taking the number of sales and subtracting the number of returns. Consequently, the inventory demand is determined by subtracting return units from sales units.

Table 1. Descriptive statistics

	Sales	Return	Adjust Demand
N	74,180,464	74,180,464	74,180,464
mean	7.31	0.13	7.22
std	21.97	29.32	21.77
min	0.00	0.00	0.00
median	3.00	0.00	3.00
max	7200.00	250000.00	5000.00

As depicted in Table 1, both the average return unit and amount are zero. This suggests that our target demand estimation aligns closely with sales predictions. Consequently, understanding the factors influencing sales becomes crucial. Figure 1 reveals that both sales and returns exhibit a pronounced left-skewed distribution. Such a deviation from the normal distribution assumption could compromise the performance of traditional economic models.

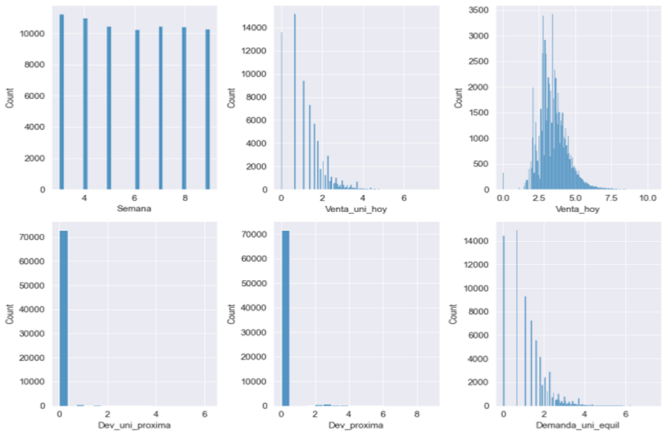

Fig. 1. Data Distribution

3.2 Gradient Boosting Decision Trees (GBDT)

GBDT, or Gradient Boosted Decision Trees [18][19], is a machine learning technique that constructs an ensemble of shallow decision trees sequentially. These trees are then combined to produce a final prediction. Essentially, GBDT comprises multiple "weak learner" trees, trained one after the other. Once a tree is trained, the residuals are determined and subsequently used as the target for the succeeding tree. Each tree in the sequence aims to rectify the errors of its predecessor, employing gradient descent to minimize the target. In the end, the predictions of all the trees are aggregated.

The primary hurdle in inventory demand estimation for economic models lies in the stringent assumption of data features following a normal distribution. In contrast, real-world large-scale datasets often exhibit sparsity. For instance, Table 1 illustrates that the majority of returns are zero. To address this, we employ LightGBM (see algorithm 1), an extension of the GBDT framework. Firstly, continuous feature values are categorized into discrete bins. For example, a feature vector [0,0,0.1,0,0.7,0,0,0.9] can be divided into two bins: 0 to 0.5 (bin 1) and 0.5 to 1 (bin 2). This approach is agnostic and indifferent to specific distribution assumptions.

Next, the model constructs a series of decision trees in sequence to enhance prediction accuracy. In each iteration (individual tree), it identifies the optimal feature split to reduce the cumulative loss across child nodes:

$$l = \frac{1}{N}\left[\sum_{i \in R_{\text{left}}} \left(y_i - \hat{y}_{R_{\text{left}}}\right)^2 + \sum_{i \in R_{\text{right}}} \left(y_i - \hat{y}_{R_{\text{right}}}\right)^2\right] \tag{1}$$

Where $\hat{y}_{R_{\text{left}}}$ and $\hat{y}_{R_{\text{right}}}$ denote the current tree prediction on the data instances of the left branch and right branch, respectively. Notably, LightGBM employs regularization techniques to each tree, such as limiting tree depth, capping the number of leaves, and introducing randomness in feature selection. Successive trees focus on the residual, or prediction error, of the preceding tree. By aggregating predictions from all trees, the combined model significantly diminishes prediction error. While individual trees might offer weak accuracy without overfitting, the boosting process ensures the overall model exhibits robust generalization.

Ultimately, LightGBM leverages the Gradient-based One Side Sampling (GOSS) technique for optimal performance in sparse feature spaces. Specifically, the model keeps the top N_1 samples with the most significant gradients and randomly downsamples N_2 instances from the remaining $N - N_1$ dataset. Given that data instances with larger gradients indicate inadequate training (non-zero values), this strategy emphasizes underrepresented samples, largely overlooking zero instances.

Algorithm 1. LightGBM Implementation of GBDT

Algorithm 1 LightGBM Algorithm

procedure LIGHTGBM(D, T, F)
  Input: Dataset D, Number of trees T, Features F
  Initialize $f_0(x) = \arg\min_{\hat{y}} \sum_{i=1}^{n} l(y_i, \hat{y})$
  Discretize continuous feature values of D into bins to form histograms
  for $t = 1$ **to** T **do**
    Compute the negative gradients $g_i = -\left[\frac{\partial l(y_i, f(x_i))}{\partial f(x_i)}\right]$ for all i
    Build a tree using the pre-computed histograms:
      - Construct histograms using the gradients g_i
      - Determine best splits using histogram information
    Update the model $f_t(x) = f_{t-1}(x) + \alpha_t h_t(x)$ where α_t is the learning rate and $h_t(x)$ is the tree structure.
    Employ Gradient-based One-Side Sampling (GOSS) to:
      - Retain top instances with largest gradients
      - Randomly downsample instances with small gradients
  end for
  Output: Final model $f_T(x)$
end procedure

Our method leverages several advantages to address previously mentioned challenges:

- LightGBM discretizes continuous feature values into discrete bins, forming histograms wherein each bin denotes a range of feature values. During tree growth, LightGBM relies on these bins rather than the actual continuous values. As a result, even long-tailed distributions are quasi-uniformly handled.
- The ensemble models are adept at identifying complex non-linear relationships amongst features.
- LightGBM incorporates various regularization techniques. These include limiting tree depth, shrinkage (scaling down the predictions of each tree), and randomization (sampling features randomly). Such measures counteract overfitting and bolster the model's generalization capabilities with out-of-sample data.

4 Empirical Results

4.1 Benchmarks

To align with models in the fields of economic and machine learning, we employ typical benchmarks to illustrate our model performance. We adopt linear price elasticity demand (PED), regularized price elasticity demand (Lasso and Ridge Regression), Decision Tree and Random Forest as our baselines. These models are widely used and discussed in current research literature.

4.2 Experiment Setup

We randomly divide our dataset into training, validation, testing set by a split ratio of 0.8, 0.1, 0.1. For regularized regression Lasso and Ridge, we set the penalty parameter λ=1. For decision tree, we limit the maximum depth to 5 and maximum number of leaves to 64. For random forest, we keep the parameter same as the decision tree and set number of trees to 100. For LightGBM, we set the learning rate to 0.2, column and row sample fraction to 0.6, regularization parameter λ=100, number of trees to 300.

4.3 Results

We report the results in Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Lower values of MSE and RMSE indicate better model performance.

$$MSE = \frac{1}{n}(y - \hat{y})^2 \tag{2}$$

Table 2. Demand Prediction Error

Metric	MSE	RMSE
Linear PED	348.33	18.66
Lasso PED	348.57	18.67
Ridge PED	348.33	18.66
Decision Tree	298.17	17.27
Random Forest	296.74	17.23
GBDT	**267.18**	**16.35**

From Table 2 we show that tree method is in general better than linear demand function since it captures more complex non-linear relationships. Furthermore, our model (GBDT) outperforms other benchmarks significantly.

4.4 Model Analysis

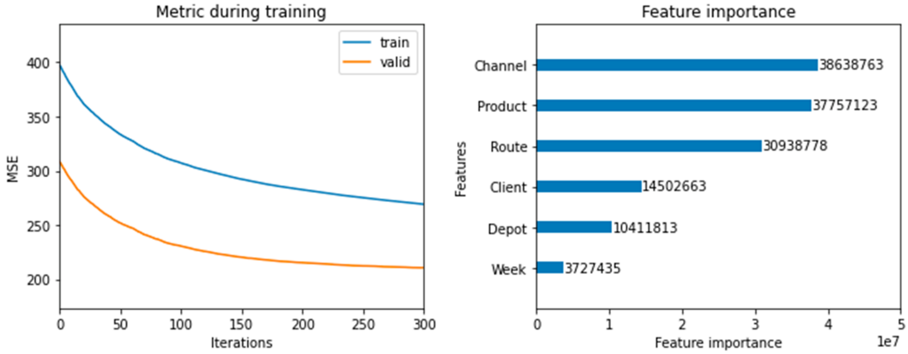

Fig. 2. Demand Prediction Error

Figure 2 elucidates why our model surpasses other baselines. First, the training curve demonstrates convergence on the validation set, achieving a low Mean Squared Error (MSE). Second, we aggregate the reduction in MSE at each feature split across all trees.

A feature's importance is proportional to its contribution to the MSE decrease. The model suggests that the channel and product are the most pivotal factors for demand prediction. Following these are the route, client, and depot associated with the goods.

5 Conclusion

In this paper, we address inventory demand prediction. In response to the challenges of rigid assumptions, oversimplification, and limited generalizability presented by traditional economic demand estimation models, we employ Gradient Boosting Decision Trees. This approach results in more accurate predictions compared to other benchmarks. Additionally, our study bridges the gap between demand estimation literature and contemporary machine learning techniques.

References

1. Cobb, C., & Douglas, P. (1928). A theory of production.
2. Wales, T., & Woodland, A. (1983). Estimation of consumer demand systems with binding non-negativity constraints. Journal of Econometrics, 21(3), 263–285.
3. Pollak, R., & Wales, T. (1992). Demand system specification and estimation. Oxford University Press.
4. Du, R., & Kamakura, W. (2008). Where did all that money go? Understanding how consumers allocate their consumption budget. Journal of Marketing, 72(6), 109–131.
5. Lee, L.F., & Pitt, M. (1986). Microeconometric demand system with binding nonnegativity constraints: the dual approach. Econometrica: Journal of the Econometric Society, 1237–1242.
6. Millimet, D., & Tchernis, R. (2008). Estimating high-dimensional demand systems in the presence of many binding non-negativity constraints. Journal of Econometrics, 147(2), 384–395.
7. Van Soest, A., Kapteyn, A., & Kooreman, P. (1993). Coherency and regularity of demand systems with equality and inequality constraints. Journal of Econometrics, 57(1-3), 161–188.
8. Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. The Review of Economic Studies, 70(1), 147–165.
9. Allenby, G., Garratt, M., & Rossi, P. (2010). A model for trade-up and change in considered brands. Marketing Science, 29(1), 40–56.
10. Honka, E. (2014). Quantifying search and switching costs in the US auto insurance industry. The RAND Journal of Economics, 45(4), 847–884.
11. Hastings, J., & Shapiro, J. (2018). How are SNAP benefits spent? Evidence from a retail panel. American Economic Review, 108(12), 3493–3540.
12. Hearst, M., Dumais, S., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems and their applications, 13(4), 18–28.
13. Berry, S. T., Levinsohn, J. A., & Pakes, A. (1993). Automobile prices in market equilibrium: Part I and II.
14. Breiman, L. (2001). Random forests. Machine learning, 45, 5–32.
15. Patrick Bajari, Denis Nekipelov, Stephen P. Ryan, Miaoyu Yang (2015). Demand estimation with machine learning and model combination.

16. Adam, H., He, P., & Zheng, F. (2020). Machine learning for demand estimation in long tail markets. Columbia Business School Research Paper Forthcoming.
17. Gandhi, A., Lu, Z., & Shi, X. (2023). Estimating demand for differentiated products with zeroes in market share data. Quantitative Economics, 14(2), 381-418.
18. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., & others (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1–4.
19. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.