

FORECASTING DEMAND AT THE OUTLET–PRODUCT LEVEL IN DISTRIBUTION USING MACHINE LEARNING ALGORITHMS: A CONCEPTUAL MODEL

Ne'matov Abdug'ani

*Candidate of Physical and Mathematical Sciences, Docent, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Department of Multimedia Technologies
tuit.9393@gmail.com*

Ismailov Shixnazar Rashid o'g'li

*Docent, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Department of Multimedia Technologies
shixnazar.ismailov@gmail.com*

Ashiraliyev Zokirjon Nosirali o'g'li

*Master's student, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
zokirjonashiraliyev@gmail.com*

ANNOTATSIYA

Tez aylanuvchi iste'mol tovarlari (FMCG) distribyutsiyasida agentlar va van-sellerlarga to'g'ri reja tuzib berishning asosi — har bir savdo nuqtasi va mahsulot uchun kelgusi davr talabini aniq prognoz qilishdir. Uch qisimli turkumning birinchi qismi bo'lgan ushbu maqolada savdo nuqtasi–mahsulot (SKU) darajasida haftalik talabni mashinali o'qitish (ML) algoritmlari yordamida prognozlashning konseptual va uslubiy modeli taklif etiladi. Real distribyutor axborot tizimi (DMS/SFA) ma'lumotlar sxemasiga asoslangan holda xususiyatlar tarkibi, model tanlovi (Random Forest, XGBoost) va baholash metodologiyasi (MAE, RMSE, MAPE, R²) bayon qilinadi. Maqola amaliy joriy etish uchun yaxlit ramka taqdim etadi: empirik tekshirish real ma'lumotlar asosida amalga oshirilishi nazarda tutiladi.

Kalit so'zlar: distribyutsiya, talab prognozlash, mashinali o'qitish, XGBoost, Random Forest, FMCG, SKU, konseptual model.

АННОТАЦИЯ

В дистрибуции товаров повседневного спроса (FMCG) основой составления правильного плана для торговых агентов и ван-селлеров является точный прогноз будущего спроса по каждой торговой точке и товару. В данной статье — первой части цикла из трёх работ — предлагается концептуальная и

методологическая модель прогнозирования недельного спроса на уровне «торговая точка–товар (SKU)» с помощью алгоритмов машинного обучения (ML). На основе схемы данных реальной информационной системы дистрибуции (DMS/SFA) описываются состав признаков, выбор моделей (Random Forest, XGBoost) и методология оценки (MAE, RMSE, MAPE, R^2). Статья представляет целостную методическую основу для практического внедрения; эмпирическая проверка предполагается на реальных данных.

Ключевые слова: дистрибуция, прогнозирование спроса, машинное обучение, XGBoost, Random Forest, FMCG, SKU, концептуальная модель.

ABSTRACT

In the distribution of fast-moving consumer goods (FMCG), the basis for building a correct plan for sales agents and van-sellers is an accurate forecast of future demand for each outlet and product. This article — the first part of a three-part series — proposes a conceptual and methodological model for forecasting weekly demand at the outlet–product (SKU) level using machine learning (ML) algorithms. Based on the data schema of a real distribution information system (DMS/SFA), it describes the feature set, the choice of models (Random Forest, XGBoost), and the evaluation methodology (MAE, RMSE, MAPE, R^2). The article provides an integrated framework for practical implementation; empirical verification is intended to be carried out on real data.

Keywords: distribution, demand forecasting, machine learning, XGBoost, Random Forest, FMCG, SKU, conceptual model.

INTRODUCTION

In a distribution system, a manufacturer’s success in the market depends directly on the activity of sales agents and van-sellers — that is, sales representatives who travel around outlets and sell products. The foundation for planning what they do — “which outlet to visit, how much goods to bring, which product to offer” — is reliable demand forecasting. In traditional practice this forecast relies on simple rules, which

leads to either stockouts or excess inventory. The aim of this article is to develop an integrated, scientifically grounded conceptual model of demand forecasting.

LITERATURE REVIEW

There is extensive research on the use of machine learning in demand forecasting. Carbonneau, Laframboise and Vahidov [3] were among the first to show that machine learning methods (neural networks, support vector machines) outperform traditional approaches such as moving average and linear regression in supply-chain demand forecasting. In food retail, Aichner and Santa [1] likewise demonstrate that machine learning is more accurate than traditional methods because it better captures the effects of seasonality and promotions, and Ceran and colleagues [2] confirm this in practice at the outlet–product level in the FMCG sector. Ali and colleagues [4] specifically studied SKU-level demand forecasting in the presence of promotions and showed that promotional features substantially improve accuracy, while Ma and Fildes [5, 6] used a large number of features and promotional data for retail forecasting. In the well-known M5 competition [7], gradient boosting–type models (including XGBoost) achieved the best results. This article likewise relies mainly on two powerful models — Random Forest [8] and XGBoost [9].

PROBLEM STATEMENT AND DATA SCHEMA

For each (outlet i , product j) pair, the weekly demand $q_{ij}(t)$ is unknown, and the task is to estimate $\hat{q}_{ij}(t+1)$ from the available history. A real DMS/SFA system stores rich data for each customer: location and outlet type, how long it has been a customer, sales days, the assigned agent and deliverer, order history, payment history, order type (bonus/discount), equipment (refrigerator), and returned and exchanged items in orders. This data schema is based on the data structure of SalesDoctor [10], a real DMS/SFA platform used by more than 300 companies in Uzbekistan. These fields are a rich source of features for forecasting.

The data preprocessing stage is important: missing values are imputed, anomalies (incorrectly entered quantities) are detected, and “false zero” sales caused

by stockouts are separated from true zero demand (the censored-demand problem). Bonus or discount orders, as well as returned and exchanged orders, are flagged separately, because they may distort the future demand signal or, conversely, provide additional information.

PROPOSED METHODOLOGY

System architecture

The proposed approach is part of a unified decision-support system in which DMS/SFA data is transformed, through three modules — demand forecasting, visit planning, and product recommendation — into a field representative's daily plan (Figure 1). This article focuses on the first module.

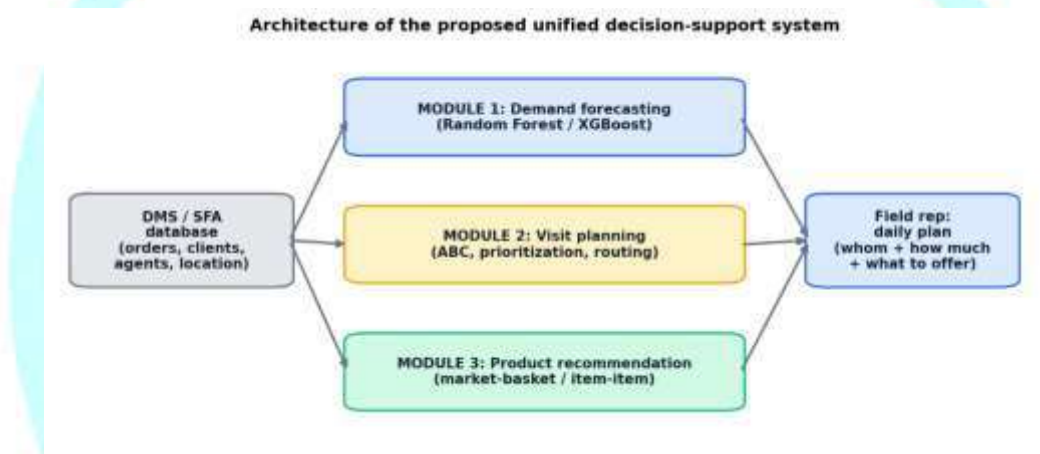


Figure 1. Architecture of the unified decision-support system and the place of the forecasting module.

Data and feature engineering

The forecasting pipeline starts from the data sources and, through feature engineering, forms the model input (Figure 2). Based on the DMS/SFA fields, the features are grouped into five categories: historical (lag, rolling mean, penetration), product, client, operational (promotion, bonus order type), and calendar (season, holiday, payday) attributes (Figure 3).

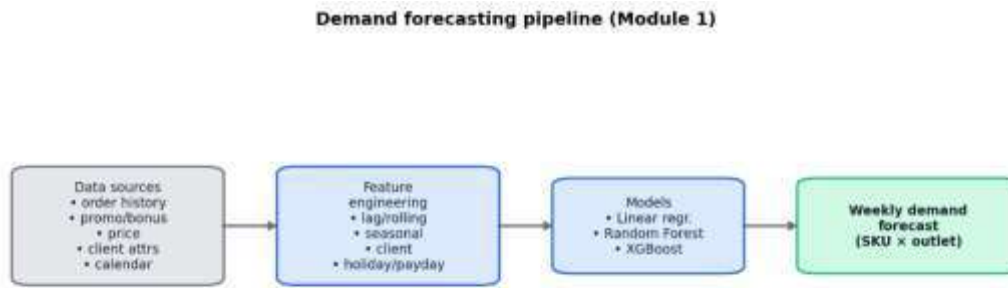


Figure 2. Demand forecasting pipeline: from data to forecast.

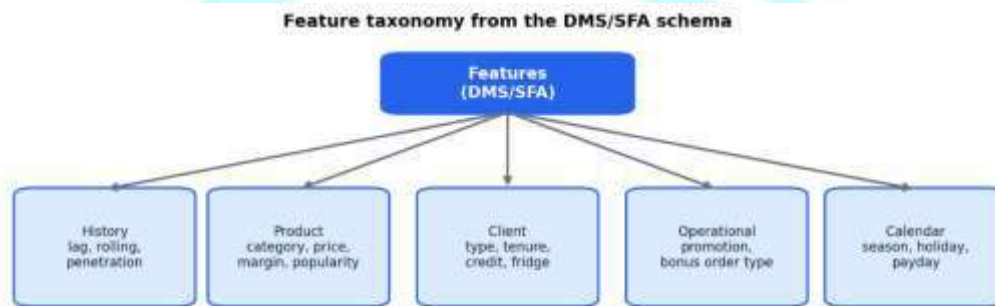


Figure 3. Feature taxonomy derived from the DMS/SFA schema.

Models

As baselines, classical time-series approaches are used (last value, moving average, seasonal naive). As the main models, linear regression, Random Forest [8], and XGBoost [9] are proposed; these tree-based ensemble models effectively learn nonlinear relationships and feature interactions. Validation uses a chronological train/test split (and, for greater reliability, walk-forward validation) so that no future information leaks into the features.

EVALUATION METHODOLOGY

The models are evaluated on a time-based test set using the following metrics (Table 1). These metrics capture both the absolute and the relative accuracy of the forecast as well as its directional bias.

Table 1. Proposed evaluation metrics (definitions)

Metric	Meaning	What it measures
MAE	Mean Absolute Error	Average error magnitude (in units)

RMSE	Root Mean Squared Error	Accuracy sensitive to large errors
MAPE	Mean Absolute Percentage Error	Relative (percentage) error
R ²	Coefficient of determination	Share of variance explained by the model
Bias	Mean directional error	Systematic over-/under-forecasting

Note: validation uses a chronological train/test split so that the forecast horizon always follows the training period, preventing data leakage.

EXPECTED RESULTS AND DISCUSSION

According to the literature, tree-based ML models (Random Forest, XGBoost) are expected to improve forecast accuracy relative to classical methods: the M5 competition demonstrated the superiority of gradient boosting models [7], and Ceran and colleagues confirmed the advantage of ML in the FMCG context [2]. In addition, the calendar (holiday, payday) and operational (bonus order) features drawn from the DMS/SFA system are expected to improve forecast quality. Concrete quantitative results are to be obtained in an empirical stage on real data.

Model interpretability matters in practice: feature importance shows the field manager what drives the forecast, which builds trust in the system. Likewise, estimating forecast uncertainty (interval forecasts) helps manage safety-stock risk, since products with higher uncertainty may require a larger buffer.

CONCLUSION

This article proposed a conceptual and methodological model for demand forecasting at the SKU–outlet level in distribution: the feature set based on the DMS/SFA schema, the choice of models, and the evaluation methodology were presented within an integrated framework. This model serves as a foundation for optimizing visit planning and product recommendation in the subsequent parts of the series, and it is ready for empirical verification on real data.

REFERENCES

1. Aichner, T., & Santa, V. (2023). Demand Forecasting Methods and the Potential of Machine Learning in the FMCG Retail Industry. In: *Serving the Customer*. Springer, Wiesbaden. https://doi.org/10.1007/978-3-658-39072-3_8
2. Ceran, B., Özkan, E., Eskiocak, D.İ., Mert, B., & Yüceoğlu, B. (2024). Machine Learning-Based Demand Forecasting for an FMCG Retailer. *INFUS 2024, LNNS*, vol. 1090. Springer, Cham. https://doi.org/10.1007/978-3-031-67192-0_11
3. Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140–1154.
4. Ali, Ö. G., Sayın, S., Van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10), 12340–12348.
5. Ma, S., & Fildes, R. (2017). A retail store SKU promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, 260(2), 680–692.
6. Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting. *European Journal of Operational Research*, 249(1), 245–257.
7. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364.
8. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
9. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD*, 785–794.
10. SalesDoctor (Sales Doctor) — distribution automation (DMS/SFA) platform, Tashkent. URL: <https://salesdoc.io> (accessed: 2026).