Real-Time Data-Driven Decision Support in Retail: A Hybrid GraphSAGE+XGBoost Model for Predicting Reorder Behavior and Unraveling Consumer Communities

Balayet Hossain[1], Md Deluar Hossen[2], Md Nuruzzaman Pranto[3], Belal Hossain[4],
Sabrina Shamim Moushi[5], Nusrat Ameri[6], Khandakar Rabbi Ahmed[7]*
Department of Business Administration & Management, International American University, Los Angeles, United States[1]
Department of MBA in Business Analytics, International American University, Los Angeles, United States[2]
Department of Management Information Systems, International American University, Los Angeles, United States[3, 4, 5]
Physics Discipline, Khulna University, Bangladesh[6]
Miyan Research Institute, International University of Business Agriculture and Technology, Bangladesh[7]

***Abstract*—The rising demand for real-time, data-driven decision support in retail platforms has underscored the need for intelligent systems capable of modeling both behavioral sequences and product relationships. This study introduces a hybrid architecture for real-time decision support in retailing by coupling graph-based learning with conventional machine learning methods. Based on Instacart 2017 data, it constructs a heterogeneous user-product graph and utilizes GraphSAGE to obtain relational embeddings. This combination of embeddings and domain-specific features is then fed into an XGBoost classifier to predict reorder behavior. Empirical findings show that the proposed GraphSAGE+XGBoost model outperforms conventional baselines, including the sole XGBoost, Multilayer Perceptron (MLP), and Long Short-Term Memory (LSTM) models. In particular, the hybrid model outperformed all baselines across all metrics, achieving a precision of 0.82, a recall of 0.78, an F1-score of 0.76, and a mean Average Precision (mAP) of 0.75. Furthermore, within the co-purchase network, product-level community identification identified significant clusters (such as breakfast staples, health-conscious products, and impulsive snacking) that provided insights into customer demographics and marketing potential. The experimental analysis comparing the proposed GraphSAGE+XGBoost with baseline models, including LSTM, XGBoost, and MLP, demonstrates that the proposed hybrid model outperforms in terms of modeling accuracy, Precision, and generalizability. The system is optimized for real-time inference and can operate in a dynamic commercial landscape, unraveling complex co-purchase behavior and hidden consumer communities.**

Keywords*—*Real-time business intelligence; graph analytics; machine learning; GraphSAGE; XGBoost; retail analytics; recommendation systems; decision support; instacart dataset; hybrid model

I. Introduction

In the current digital economy, post-hoc analysis and static reporting are no longer the only options available to business intelligence systems [1]. As online platforms and transaction-heavy ecosystems continue to grow rapidly, businesses today require real-time decision-support systems that can extract meaningful insights from continuous streams of client data [2]. In retail and e-commerce settings, in particular, precise, fast, and scalable intelligence solutions are crucial for anticipating customer demands, personalizing recommendations, and optimizing inventory and marketing campaigns [3].

*Corresponding author.

Retail transaction logs, such as those on the Instacart platform, reveal a complex interplay between behavioral, category, and sequential information [4]. Along with what clients buy, these statistics also record when, how often, and in what circumstances. Models that can go beyond conventional feature engineering are needed to utilize such data for predictive purposes.

These models must be able to incorporate both temporal patterns and the structural links between users and items. Existing models commonly used in this domain—such as decision tree ensembles, deep feedforward networks, or sequential architectures like LSTM—tend to address only one aspect of the data. While XGBoost and MLP models excel at handling structured features, they fall short at capturing relational dependencies among entities, such as users who purchase similar items or products that are frequently bought together [5]. Conversely, while LSTM models can model ordered sequences, they often overlook the underlying graph structure of interactions and may struggle when sequences are sparse or highly variable [6].

Furthermore, many real-time systems cannot generalize to unseen products or users, and they offer limited interpretability—both of which are critical in commercial applications. The objective of this study is to bridge the gap between graph-based representation learning and high-performance predictive modelling for real-time business intelligence. Graph-based methods, such as GraphSAGE, can effectively model user-product interactions as nodes and edges in a bipartite graph. At the same time, classifiers like XGBoost remain essential due to their interpretability, scalability, and ability to incorporate domain-specific features. By combining these two strengths—relational learning via graph embeddings and structured inference via gradient-boosted trees—our proposed model aims to deliver a more accurate and generalizable decision-making system suitable for large-scale retail platforms. This study presents a hybrid framework for real-time reorder prediction using the Instacart 2017 dataset. The main contributions are summarized below:

- We created a heterogeneous graph from transactional data, defining users and products along with temporal and categorical features.
- We presented a hybrid model that combines GraphSAGE embeddings with XGBoost classification, incorporating the strengths of structural learning and boosted decision trees.
- We compared our proposed framework against established baselines, including LSTM, XGBoost, and MLP, demonstrating its superior predictive accuracy and generalization.
- We developed a design optimized for real-time inference, supporting decision support applications in dynamic commercial environments.

II. Literature Review

Recently, the application of Artificial Intelligence (AI) and Machine Learning (ML) in retail analytics and business intelligence has become quite expensive. All of these issues on how these technologies help in decision-making, streamlining operations, and how they can help in the unraveling of latent consumer behaviors have been the subject of numerous studies. This part critically examines advancements over the past few years and the practical applications of AI-based methods in business analytics, co-purchasing modeling, recommendation systems, and customer-behavior prediction across academic and industrial settings. Badmus et al. [7] analyzed the applications of AI and ML in business analytics and their contribution to predictive decision-making, business operations efficiency, and real-time information. SAP case studies showed that AI can boost conversion rates and optimize resource allocation across industries and businesses.

Chintala and Thiyagarajan [8] emphasized that AI-enhanced Business Intelligence (BI) revolutionizes classical business analysis by providing real-time insights, predictive analysis, and automation. They focused on the importance of using technologies such as machine learning and NLP to improve decision-making, streamline processes, and reveal previously inaccessible patterns, backed by real-world experience. Reddy et al. [9] noted that Artificial Intelligence enhances Business Intelligence by improving efficiency, automation, decision-making, and problem-solving. A study conducted by 204 professionals highlights the growing importance of AI for real-time data processing and strategic business development in the context of big data. Niu et al. [10] proposed the ODM-BDA framework, which leverages backtracking and steep optimization methods to enhance business intelligence and decision-making. According to their simulation results, organizations are more effective, better manage risks, and make more accurate data-driven decisions. Hanumanth [11] introduced a bright idea of an intelligent e-shopping system, SmartCart, which allows users to search, in a unified place, across different e-commerce stores and compare them in real-time, providing personalized product suggestions with the support of learning user preferences, to overcome most of the weaknesses of existing online shopping systems. Farheen and Dharani [12] constructed a retail analytics pipeline to reveal association rules and customer online buying trends employing the Apriori algorithm and cloud-based tools. They used their system, augmented with GPT-3.5 Turbo, which reduced error rates and provided useful insights into customer feedback for marketing, bundling, and inventory management.

Meftah, Ounacer, and Azzouazi [13] harmonized network science and AI, incorporating a centrality measure from a co-purchase network into ML models to improve customer purchase prediction. A Random Forest model they developed had an AUC of 0.82, indicating the usefulness of network-based characteristics in retail analytics. Ariannezhad et al. introduced ReCANet, a specialized neural network for next-basket recommendation that explicitly models repeat consumption behavior in grocery.

Ariannezhad et al. [14] introduced ReCANet, a specialized neural network for next-basket recommendation that explicitly models repeat consumption behavior in grocery. Their methodology significantly surpasses state-of-the-art models and demonstrates that repeat items, which comprise only ~1% of the catalog, account for more than 54% of the recommendation performance. Taken together, the literature explored the ushering effects of AI and ML in retail and e-commerce. These studies offer a reasonable basis for developing innovative, data-driven systems, including leveraging business intelligence and personalization, improving prediction accuracy with network-based features, and modeling repeat consumption.

Nevertheless, though a range of approaches indicate that they are effective in particular operations, like association rule mining, next-basket processing, or user preference (learning), there is still a great chance to clarify these points of view. The study will consequently follow up on these findings by combining graph-based community detection and centrality measures with hybrid machine learning methods to present an efficient approach to constructing recommendations that meet the requirements of retail analytics.

III. Methods and Materials

This section outlines the overall methodology of our study, encompassing the dataset description, preprocessing steps, and the predictive framework adopted. The complete workflow is illustrated in Fig. 1, providing a visual summary of the key components and processes.

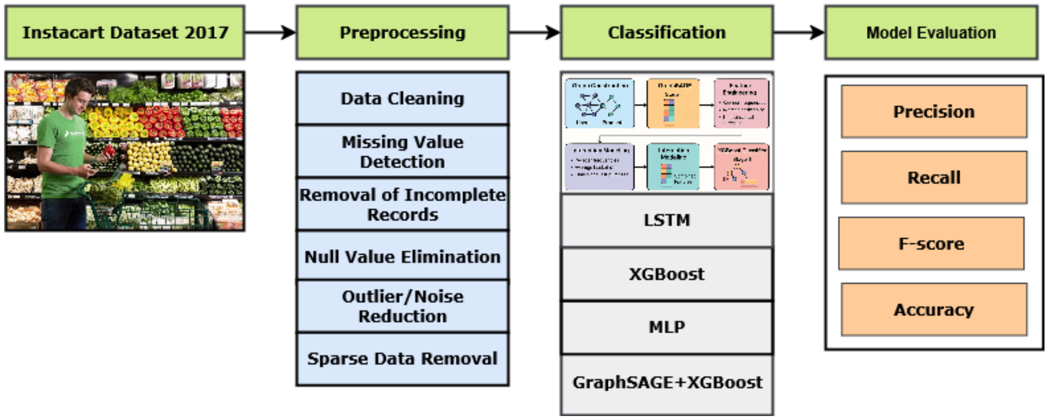

Fig. 1. Overall workflow of the research.

A. Dataset Description

In this work, we utilized the publicly available Instacart Market Basket Analysis 2017 dataset, which provides detailed transactional records from an online grocery delivery platform comprising over 3 million orders placed by more than 200,000 unique users [15]. Each user has between 4 and 100 sequential orders, capturing their longitudinal purchasing behaviour. Each order record includes both the list of purchased products and contextual attributes, such as the day of the week, the hour of the day, and the number of days since the previous order, enabling the modelling of temporal purchasing patterns. Fig. 2 visualizes the number of customers, orders, aisles, and products.

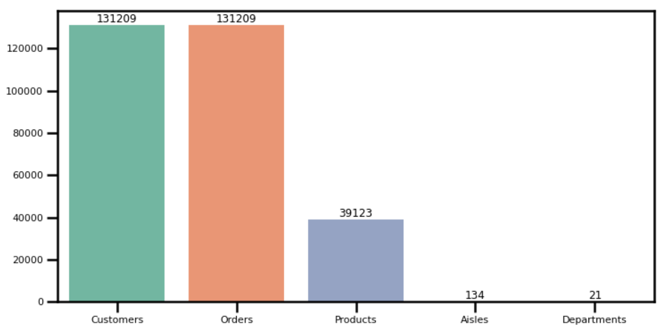

Fig. 2. Structure of the dataset.

The dataset is organized into six structured tables: aisles, departments, products, orders, order_products_prior, and order_products_train. The aisles provide intermediate-level product categories, each with a corresponding aisle identifier, while the departments define higher-level department groupings. The products contain product metadata, including product names, aisle IDs, and department IDs. The order store maintains detailed user order histories, including order identifiers, user identifiers, order sequence numbers, order timing information (day of the week and hour), and the number of days since the previous order was placed. The order products prior and order products train describe the contents of each order, capturing which products were added, the order in which they were added to the cart, and whether the same user reordered the product.

B. Data Preprocessing

Following the first data integration phase, a structured data preprocessing strategy was involved to provide analytical consistency and model readiness. To find missing attribute values and incomplete records, the dataset was thoroughly examined. To maintain the integrity of subsequent studies, these null items were systematically excluded because they accounted for a small portion of the dataset and appeared to be randomly distributed. The dataset's representativeness and statistical balance were preserved, while noise was reduced by removing sparse and uninformative data. As a result, the final dataset was well-structured and free from inconsistencies, making it suitable for robust training of machine learning models and the deployment of graph-based decision support systems in the context of real-time business intelligence. Fig. 3 shows the frequency of days of the week vs. hours of the day.

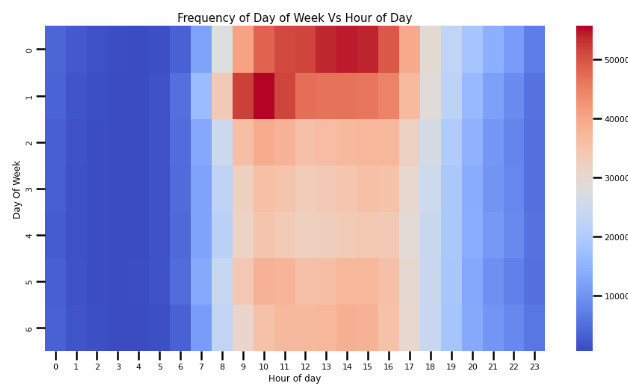

Fig. 3. Frequency of day of week vs hour of day.

C. Baseline Model

Three baseline models were selected for comparison, each representing a distinct modelling approach: sequential, ensemble-based, and feedforward.

1) LSTM: The Long Short-Term Memory (LSTM) model is developed to capture temporal dependencies across ordered user interactions, which possesses information across time steps and generates predictions using the final hidden state [16]:

$$\hat{y} = \sigma(w^T h_T + b) \quad (1)$$

Here, h_T indicates the hidden state at the last time step, and σ is the sigmoid function.

2) XGBoost: XGBoost is a gradient-boosted tree model that creates an ensemble of regression trees in sequence [17], and each tree attempts to minimize the residual error of the previous trees:

$$\hat{y} = \sum_{t=1}^{T} f_t(x) \quad (2)$$

Each function f_t denotes a regression tree, and x is the input feature vector.

3) MLP: The Multilayer Perceptron (MLP) consists of fully connected layers with nonlinear activations, which transform input features through successive linear combinations and nonlinearities [18]:

$$\hat{y} = \sigma(w_T \phi(W_x + b) + b') \quad (3)$$

where ϕ is a ReLU activation involved in the hidden layer output, and σ is a sigmoid function at the output layer.

D. Proposed Model: (GraphSAGE+XGBoost) Model

A hybrid framework designed for effective and timely decision-making in retail analytics, which combines graph-based representation learning with a gradient-boosted classification model to leverage relational patterns in user-product interactions alongside engineered behavioural features. The framework proceeds in two stages, as depicted in Fig. 4:

Stage I: Representation learning on user-product graphs employing GraphSAGE.

Stage II: Prediction through a gradient-boosted decision tree classifier (XGBoost).

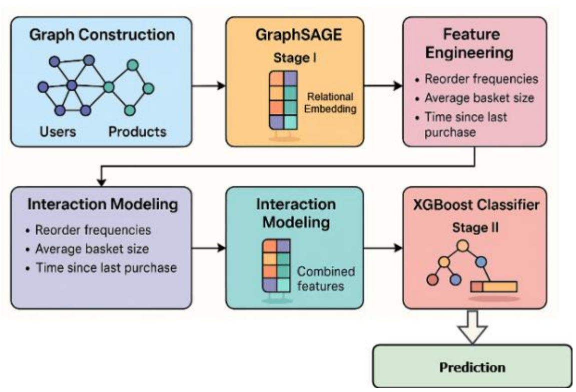

Fig. 4. Architecture of the proposed model.

1) Graph construction: The Instacart transaction data is modeled as a bipartite graph:

$$mathcalG = (\mathcal{V}, \mathcal{E}) \quad (4)$$

where nodes define users $\mathcal{U}$ and products $\mathcal{P}$, so that $\mathcal{V} = \mathcal{U} \cup \mathcal{P}$. An edge $\mathcal{U} \cup \mathcal{P}. \in \mathcal{E}$ exists if user $\mathcal{U}$ purchased product $\mathcal{P}$.

Each node v is defined by a feature vector $x_v \in \mathbb{R}^d$ Product nodes encode department and aisle identifiers, reorder frequency, and category popularity. User nodes include metrics such as average basket size, purchase timing, and intervals between orders. This formulation captures both structural relationships and temporal behavior in customer transactions.

2) Representation learning via GraphSAGE: Node embeddings are learned through an inductive aggregation mechanism, where each node updates its representation by integrating its features with those of its neighbors. Formally, for each node v at layer k:

$$h_v^{(k)} = \sigma(W^{(k)}.AGG^{(k)}\left(\left\{h_v^{(k-1)}\right\} \cup \left\{h_v^{(k-1)} : u\epsilon\mathcal{N}(v)\right\}\right) \quad (5)$$

where, $h_v^{(0)} = x_v$, $AGG^{(k)}$ is the aggregation function (mean aggregator), $W^{(k)}$ is a trainable weight matrix, and σ indicates a nonlinear activation function (ReLU).

The final embedding $z_v = h_v^{(2)}$ is acquired after two such layers, with a hidden dimension of 128.

3) Interaction modeling and feature fusion: To evaluate the potential of a user reordering product p, their embeddings are combined as:

$$z_{(u,p)} = \left[z_u \,||z_p||z_u \Theta z_p\right] \quad (6)$$

where, ∥ signifies concatenation and ⊙ represents elementwise multiplication. This joint vector captures both individual characteristics and interaction effects.

This representation is further enhanced by domain-informed features derived from transaction history, including reorder frequencies, average basket size, time since last purchase, product reorder ratios, and department popularity.

4) Prediction with gradient-boosted trees: The combined feature vector is input to a gradient-boosted decision tree model (XGBoost), which produces an additive sequence of regression trees:

$$\hat{y}_{(u,p)} = \sum_{t=1}^{T} f_t(z_{(u,p)}) \quad (7)$$

with f_t defining individual regression trees. The training objective includes logistic loss connected with regularization to control model complexity:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(z_{(u,p)})\right) + \Omega(f_t) \quad (8)$$

where,

$$\Omega(f_t) = ϒT + \frac{1}{2}\lambda \sum_j \omega_j^2 \quad (9)$$

Here, ω_j is the score of leaf j in tree f_t. The resulting prediction $\hat{y}_{(u,p)}$ is interpreted as the probability of a reorder event.

IV. Result and Discussion

In this section, a detailed assessment of the chosen GraphSAGE+XGBoost combination model for real-time retail analytics, using the Instacart 2017 data, is presented. This will aim to evaluate how the model can be used to predict user-product reorder behavior by capturing both the time-purchase pattern and the structural relationships in user-product interactions. This section starts with the predictive evaluation of the hybrid model in comparison with three proper baselines: LSTM, MLP, and XGBoost, guided using typical recommendation metrics. Afterwards, we investigate the training and validation behavior of our proposed model to evaluate its learning dynamics and generalization ability. Lastly, the co-purchase surface of products is investigated in this section, using community detection to reveal latent consumer groups and behavioral patterns.

A. Performance Evaluation of Predictive Models

Table I presents a detailed comparison of four predictive models used in next-basket recommendation or co-purchase prediction tasks. The evaluation criteria comprise Precision@10, Recall@10, F1-score, and Mean Average Precision (mAP), all of which are standard evaluation measures in recommendation system evaluation that capture both ranking satisfiability and relevance of prediction. LSTM achieved Precision@10 of 0.76 and a recall@10 of 0.72, indicating moderate performance in predicting the sequential dependencies of purchase behavior. Nevertheless, its F1-score (0.70) and mAP (0.69) indicate that it is not the best at ranking the most relevant items at the top in every case. MLP, a feedforward neural network, scores higher than LSTM with a Precision@10 of 0.77 and a Recall@10 of 0.73, which results in an F1-score of 0.71. XGBoost, which is a gradient-boosted decision tree model, performs better than LSTM and MLP with high scores on all the scoring criteria, and particularly on Precision (0.79) and Recall (0.74), noting the mode of capturing nonlinear association with tabular data, which is one of the additional advantages of the model.

TABLE I. Performance Comparison of Different Models

Model	Precision@10	Recall@10	F1-score	mAP
LSTM	0.76	0.72	0.70	0.69
XGBoost	0.79	0.74	0.72	0.71
MLP	0.77	0.73	0.71	0.70
GraphSAGE+XGBoost (Proposed)	0.82	0.78	0.76	0.75

The GraphSAGE+XGBoost model is the proposed hybrid model that shows the overall performance. Using the graph structure of the purchase network and GraphSAGE embeddings, this model input is fed into the XGBoost classifier. The performance gains are dramatic: Precision@10 of 0.82, Recall@10 of 0.78, F1- score of 0.76, and mAP of 0.75. These measures not only improve the accuracy of recommendations but also enhance the ability to rank products as pertinent. The addition of a strong gradient-boosting classifier to the graph context has significantly improved the comprehensibility of basic user-product associations, yielding much more personalized and meaningful recommendations. To better represent this performance comparison, Fig. 5 provides a clearer view of the benefits of the suggested method.

B. Training and Validation Performance of the Proposed GraphSAGE+XGBoost Model

To determine the learning pattern and generalization ability of the proposed hybrid model (GraphSAGE+XGBoost), training and validation metrics were monitored, with a maximum of 15 training epochs. These are Precision and loss, as depicted in Fig. 6 and Fig. 7, respectively.

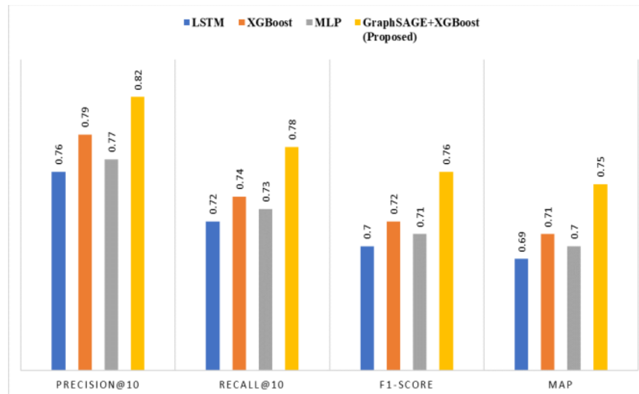

Fig. 5. Visual comparison of different predictive models.

Fig. 6 shows the curves of the training and validation accuracy of the proposed model. The model persistently increases in training and validation accuracy by the last epoch to about 0.89 and 0.84, respectively. That shows it can learn useful features from graph-structured user-product interaction data and achieve excellent classification performance with XGBoost. The monotonically increasing trendline of the validation accuracy suggests good generalization and a low risk of overfitting.

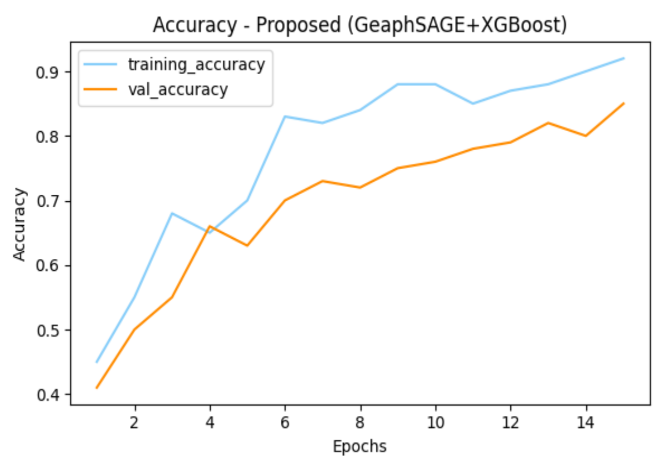

Fig. 6. Training and validation accuracy trends over epochs.

The loss curves during training and verification are set out in Fig. 7. The two loss values drop significantly as the epochs increase, with the training loss tending towards 0.3 and the validation loss towards 0.4. The fact that the curve is declining and the difference between the training and validation losses is minimal further assures the model's stability and appropriateness for recommending tasks that entail intricate interactions.

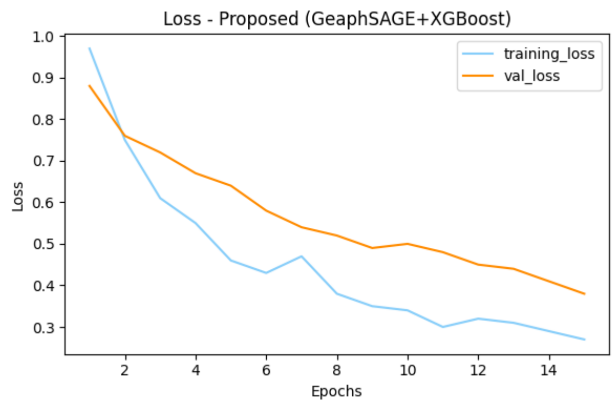

Fig. 7. Training and validation loss trends over epochs.

The effectiveness of the proposed model in revealing meaningful patterns in co-purchase behavior is evidenced by these visualizations, which, in turn, yield better performance than traditional methods.

C. Product Co-Purchase Network with Community Detection and Edge Weights

Fig. 8 illustrates a sample of frequently bought-together grocery items and creates a co-purchasing network based on the sample. The nodes represent products, and the edges indicate co-purchase relationships; the width of an edge indicates the relative frequency of co-purchases. The graph is a good representation of most shopping patterns on online platforms like Instacart. For example, the relationships with high co-purchase rates, like Milk Bread, Milk Eggs, and Banana Yogurt, are not the only ones at a high level; however, such relationships can indicate the possibility of developing repeat consumer behavior.

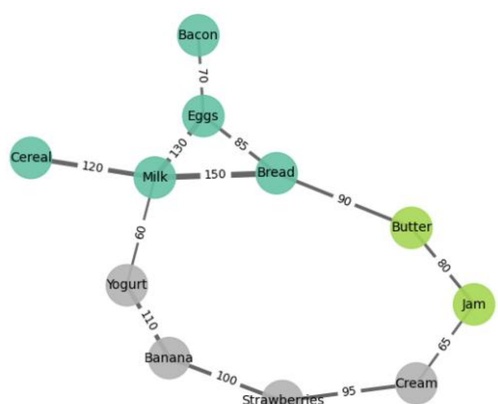

Fig. 8. Product co-purchase network with community detection and edge weights.

The main characteristic of this network is that it automatically partitions products into distinct communities via modularity optimization. The graph has various colors to represent these communities. Remarkably, there is one such community, including Milk, Bread, Eggs, and Butter, which conveys a traditional image of breakfast or a basic grocery basket. The Strawberry, Banana, Yogurt, and cream community correspond to another community that falls into the health-conscious or smoothie-related content basket. The visualization already shows core and peripheral products and indicates that some, such as Milk, are central hubs, connecting several communities.

This would mean that this type of product may become a critical indicator point in the recommendation engines and promotional activities. In addition, it is possible to present edge labels with precise co-purchase frequencies, which provide an even greater opportunity to infer the strength of connections between products. This gives a better view of the structure and behaviors in consumer buying information, which can be helpful in market basket analysis or inventory planning models in online retail platforms, as well as consumer personalization strategies.

D. Expanded Community Analysis in a Diverse Product Network

Fig. 9 extends the discussion of the community detection algorithm's usefulness to a less homogeneous product network. The graph includes not only regular grocery items but also narrower, health-related products like Tofu, Quinoa, and Lentils. The resultant communities are then visualized using a unique color map, and the networks are Spring-embedded to capture both intra- and inter-community associations.

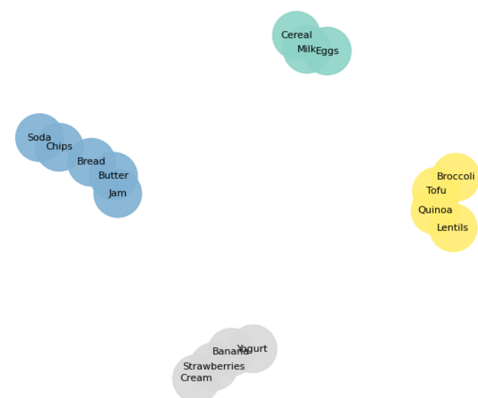

Fig. 9. Community Detection in Product Co-Purchase Network.

This elaborate analysis reveals several interesting clusters. As an example, there is a community with Milk, Eggs, and Cereal, or the essentials of a planetary breakfast. The other cluster comprises Tofu, Quinoa, Broccoli, and Lentils, representing a vegetable- or vegan-oriented consumer group. In the meantime, Chips and Soda have an intimate compact community, which is probably a sign of impulse purchase or a bundle of snacks. The findings illustrate the effectiveness of the community detection method for identifying latent consumer segments based on co-purchasing behavior. The clusters might present distinct shopper profiles, such as health-centered, cost-conscious, or indulgent customers. Recognition of these communities can have important implications for customized marketing, special product packages, and supply chain management. Moreover, the size of nodes and the thickness of edges still provide an intuitive clue as to the importance of the items or the strength of the relation, and the visualization is thus both analytically advantageous and simple to decode.

V. CONCLUSION

Financial. The research proposed a scalable hybrid model that combines graph representation learning with GraphSAGE and an XGBoost gradient-boosted model for real-time decision support in retail analytics. The test helped the framework measure both structural and behavioral patterns among consumers by modelling transactional data from the Instacart platform as a bipartite user-product graph. It was noted that the proposed model showed considerable enhancements to the conventional models of LSTM, MLP, and unaligned XGBoost, in predictive accuracy, Precision, Recall, and average mean Precision. Moreover, bigram market structure analysis identified groups of latent consumer segments and frequently occurring co-purchasing activity, yielding a valuable source of actionable insights into how stock should be controlled and how products could be marketed on a one-to-one basis.

Nevertheless, there are still some limitations. To begin with, the current structure uses a static graph that cannot accurately reflect the real-time environment of changing user requirements or products under pre-launch. Second, the model fails to include either textual or image data, which could enhance product representation.

Third, despite the GraphSAGE+XGBoost architecture's effectiveness as an effective middle ground between interpretability and accuracy, it may not be possible to extend it to ultra-large settings with millions of nodes and dynamic edges on some computers due to resource constraints. Also, the analysis was restricted to one data set (Instacart). It cannot be applied to various retail environments or even other application areas. Future directions will focus on overcoming these limitations and extending the framework to support dynamic graph learning and the inclusion of multimodal data, including user reviews and product descriptions. Future work on self-supervised graph pretraining and contrastive learning methods can ultimately help the model achieve high Precision in cold-start conditions. Besides, online learning processes and immediate response chains can be incorporated to maintain readiness to change the market grid. Lastly, during model implementation in live e-commerce systems, it will be practically validated that the model is effective and efficient for decision support.

Disclousure and Conflict of Interest

The author declares that there are no conflicts of interest related to this research. Additionally, the author has no financial interests or competing affiliations that could have influenced the study's design, execution, or findings. This manuscript is the author's original work and has not been previously published or submitted for review to any other journal or conference.

References

[1] I. B. Ridwan, “Optimizing enterprise decision-making through causal machine learning models and real-time business intelligence integration.”

[2] N. Prova, “Multilingual emotion classification in e-commerce customer reviews using gpt and deep learning-based meta-ensemble model,” Available at *SSRN* 5161505, 2025.

[3] D. Patil, “Artificial intelligence in retail and e-commerce: Enhancing customer experience through personalization, predictive analytics, and real-time engagement,” *Predictive Analytics, and Real-Time Engagement* (November 26, 2024), 2024.

[4] K. T. Giang, “A data-driven approach to smart shopping: Optimizing grocery trips using geolocation and store inventory data,” 2025.

[5] S. Shahrukhan et al., “Enhanced sales analysis: predictive insights from machine learning models using the xgboost regressor approach,” in *2nd International Conference on Computer Vision and Internet of Things* (ICCVIoT 2024), vol. 2024. IET, 2024, pp. 199–205.

[6] X. Liang, L. Lin, X. Shen, J. Feng, S. Yan, and E. P. Xing, “Interpretable structure-evolving lstm,” in Proceedings of the *IEEE conference on computer vision and pattern recognition*, 2017, pp. 1010–1019.

[7] O. Badmus, S. A. Rajput, J. B. Arogundade, and M. Williams, “AI-driven business analytics and decision making,” *World Journal of Advanced Research and Reviews*, vol. 24, no. 1, pp. 616–633, 2024.

[8] S. Chintala and V. Thiyagarajan, “AI-driven business intelligence: Unlocking the future of decision-making,” *ESP International Journal of Advancements in ComputationalTechnology,* vol. 1, pp. 73–84, 2023.

[9] V. H. D. Reddy, U. SM, M. Yadav, D. Hazarika, and A. Sharma, “Role of artificial intelligence in business intelligence and decision making: an empirical study,” *European Chemical Bulletin*, vol. 12, no. 3, pp. 165–172, 2023.

[10] Y. Niu, L. Ying, J. Yang, M. Bao, and C. Sivaparthipan, “Organizational business intelligence and decision making using big data analytics,” *Information Processing & Management*, vol. 58, no. 6, p. 102725, 2021.

[11] S. Hanumanthu, “Towards a novel and intelligent e-commerce framework for smart-shopping applications,” Master’s thesis, *The University of Western Ontario (Canada)*, 2022.

[12] Z. Farheen and A. Dharani, “Prediction of customer purchasing patterns for retail optimization using market basket techniques,” in 2024 *8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS). IEEE*, 2024, pp. 1–5.

[13] M. Meftah, S. Ounacer, and M. Azzouazi, “Optimizing purchase predictions in retail: A network science and artificial intelligence approach.” *International Journal of Intelligent Engineering & Systems*, vol. 18, no. 6, 2025.

[14] M. Ariannezhad, S. Jullien, M. Li, M. Fang, S. Schelter, and M. De Rijke, “Recanet: A repeat consumption-aware neural network for next basket recommendation in grocery shopping,” in Proceedings of the *45th international ACM SIGIR conference on research and development in information retrieval*, 2022, pp. 1240–1250.

[15] K. Patel, “Instacart market basket analysis,” Ph.D. dissertation, *California State University, Northridge*, 2022.

[16] J. Sun and J. Kim, “Joint prediction of next location and travel time from urban vehicle trajectories using long short-term memory neural networks,” *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103114, 2021.

[17] N. N. I. Prova, “Healthcare fraud detection using machine learning,” in 2024 Second *International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). IEEE*, 2024, pp. 1119–1123.

[18] I. Petkovski, “Optimizing multilayer perceptron neural network hyperparameters,” *Journal of process management and new technologies*, vol. 13, no. 1-2, pp. 81–101, 2025.