

Multi-Feature Search–Based Purchasing Tendency Community Classification For Densely Distributed Clients In E-Commerce

M.GughanRaja¹, M.SanjayKumar², A.AzimSaleh³, S.PayasJenner⁴, K.KabilDoss⁵

¹Assist prof, Dept of Artificial Intelligence and Data Science

^{2, 3, 4, 5} Dept of Artificial Intelligence and Data Science

^{1, 2, 3, 4, 5} Mohamed Sathak Engineering College, Kilakarai, Tamil Nadu, India

Abstract- Purchasing tendency is defined as customer preferences for products and brands, interested in price and frequency of purchase, and is determined by demographic, transactional and behavioral attributes. In today's e-commerce, these insights are critical for recommendations and managing demand. But conventional if-then rules and elementary collaborative filtering approaches lack sophisticated insights into interactions between customers, products, and locations, and the demand at different times. To overcome these challenges, this article proposes a community classification system of clients purchasing inventory using a holistic deep learning approach. Graph Neural Networks (GNNs) capture the relationship between customers, products, and regions, facilitating precise identification of customer communities and region-based demand (high, emerging, low). The SASRec Transformer-based model also leverages temporal information about customer preferences by training on temporal sequences, capturing both short- and long-term information. This approach incorporates demographic, transactional and behavioral data to offer insights to sellers and recommendations to customers, thus improving decision-making, forecasting, and efficiency in the market.

Keywords: AI, E-commerce, Purchasing Tendency, Community Classification, Graph Neural Networks (GNN), Sequential Recommendation, Demand Forecasting, Marketplace Intelligence

I. INTRODUCTION

Electronic commerce (e-commerce) has transformed the worldwide retail industry by facilitating the round-the-clock exchange of products and services via the internet. With the emergence of mobile commerce and social media, customers have become more interested in personalised shopping and quicker delivery. With the growing volume of data collected from online interactions such as browsing, transactions and location-based access, the key challenge for today's market-places is to make full use of this data to identify complex consumer behaviours. Current e-commerce

platforms typically use traditional rule-based segmentation, or simple collaborative filtering, which operate on an individual level. These approaches often overlook the interrelated nature of customers, products and regions, resulting in simple and uninformative recommendations, and limited to no understanding of regional market demand. Additionally, many existing systems overlook the temporal nature of customer behaviour, missing the opportunity to update recommendations to reflect seasonality or changes to personal interests. This paper introduces a smart system, namely Purchasing Tendency Community Classification, to overcome these challenges. This system uses Graph Neural Networks (GNN) and Transformer-based temporal models to capture multi-faceted relationships and discover community-level demand characteristics for time-sensitive recommendations. This multi-faceted approach enables sellers to gain business intelligence for stock management and improve the customer shopping experience.

II. LITERATURE REVIEW

Several studies have been conducted in the field of electronic commerce using different methods. Wong and Su (2025) proposed a new way to cluster online shoppers using a combination of clustering and logistic regression for predicting shopping intentions. Their method prioritises the segmentation of shoppers for forecasting purchasing tendencies. While this approach is effective for general prediction, the research mainly employed traditional regression techniques which may not adequately account for the complex non-linear patterns of large-scale heterogeneous networks in e-commerce.

Kim et al (2025) introduced an improved deep learning model called RFMVDA for e-commerce customer behavior classification. The model works by incorporating the conventional Recency, Frequency, and Monetary (RFM) variables with deep learning to achieve better classification. But the approach is focused mainly on individual customer characteristics and does not have a consistent graph-based

model to represent the inter-relationships of products and locations.

Li et al. (2025) discussed customers' purchase tendency community classification in large-scale networks with a multi-feature search for densely populated customers. Their study highlights the need for a community-based approach for customers with similar behavioral features for large-scale e-marketplaces. While the system is scalable, it does not explicitly incorporate the time-based self-attention of the tendencies over a given time range.

Sakar et al. (2019) developed a system for online shoppers' purchase intent prediction in real time using multiple machine learning models. This system can predict, based on session-based data, if a user is going to purchase. The model is effective for short-term intent, but lacks the capacity to capture long-term demand or community patterns across different areas.

Van den Poel and Buckinx (2007) published an early study on predicting online buying behaviour using clickstream data. This study showed that fine-grained web logs can be used to classify valuable customers. This study was written in the early days of e-commerce, and uses traditional statistical techniques which do not take into account the high-dimensional and dynamic nature of state-of-the-art Transformer-based models. Pecune et al. (2020) proposed socially-aware conversational recommendation systems which use conversational intelligence to direct customers. This method uses social and communication aspects of a recommendation to gain trust from a buyer. This system is highly personalised but computationally heavy and more focused on the front end rather than the back end regional demand mapping needed by marketplace vendors. Liang Zhao (2022) presented a detailed overview of Graph Neural Networks (GNNs), which covers GNNs' applications to modeling relationships. The research showcases using GNNs to model relationships as nodes and edges to discover underlying patterns. This work lays the theoretical foundation for relationship modeling, but it is a generic model and needs to be adapted to handle the temporal sequences of purchases that are the focus of sequential models.

III. RELATED WORK

Consumer Community Discovery

Traditional approaches to customer segmentation are based on the RFM (Recency, Frequency, Monetary) model. These are appropriate for simple classification, but do not discover communities. Recent research has used Graph-based

mining to model customers and products as nodes in a heterogeneous graph. These models can uncover natural "purchasing tendency communities" based on usage patterns, rather than a pre-defined demographic information like age, gender and race.

Sequential Behavior Modeling

Sequential recommendation has been extensively studied in the past with Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. However, these can be challenging to model long-term dependency with large e-commerce data. The recently proposed Self-Attentive Sequential Recommendation (SASRec) model is based on the Transformer architecture. SASRec allows the system to adaptively weight the impact of previous interactions, and thereby model the evolution of short-term and long-term preferences.

Location-Aware Demand Analysis

The most recommendation systems which only focus on the interaction between the user and the item, recent research has begun to factor in geographic data to support sellers of market-places. Studies of regional demand forecasting have shown that the online buying habits are extremely diverse across different geographical coordinates. The existing systems, most of them, however, look at location as an independent variable. The current project is founded on these works but introduces to them the concept of geographic regions as active nodes in the Graph Neural Network (GNN) in which one analysis of what is purchased, by whom, and in which region can be performed.

IV. PROPOSED METHODOLOGY

The proposed system is an integrated deep learning-based system that aims to close the gap between customer insights and business analytics in e-commerce. It combines two cutting-edge models to enable both vendor analytics and customer personalization. It uses multi-dimensional data such as demographics, transaction records and geographical locations to detect valuable purchasing communities. The system's approach to group analysis enables a panoramic view of the market, and ensures that inventory control and product recommendations are based on real-time data analysis.

System Overview

The system design is a layered architecture that handles the entire data flow, from input to output. The foundation is a web interface that logs user actions, which are

then preprocessed and stored in a secure MySQL database. This data is then processed by the model development layer, home of the hybrid intelligence model. The system architecture facilitates communication between the user interface and the analytical modules through Flask APIs to provide real-time outcomes, including heatmap demand, suggestion carousels and inventory alerts.

Proposed System Architecture

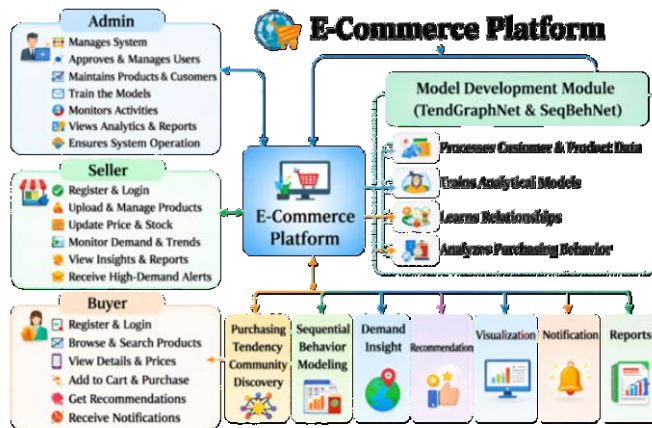


Fig. 1. proposed system architecture

TendGraphNet: Relationship Modeling on a Graph.

TendGraphNet is based on Graph Neural Networks (GNN) to model the e-commerce ecosystem as a heterogeneous graph, in which customers, products, and locations are nodes that are interconnected with each other. In comparison to old-fashioned approaches, this model employs message-passing mechanisms to learn intricate relationship embeddings among these entities. The model uses the edges to find natural purchasing communities by analyzing the edges, which are views, cart additions, and purchases, and finds location-wise demand patterns. This will enable the system to differentiate high-demand and emerging-demand areas and sellers will have accurate geographic features of targeting.

SeqBehNet: Sequential Behavior Modeling.

The SeqBehNet, based on the SASRec (Self-Attentive Se-quential Recommendation) architecture, is designed to learn the temporal dynamics of customer shopping behavior. It examines the chronological sequence of the interaction patterns to represent both immediate and short term needs and long term preferences that are constant and always present. Through self-attention mechanisms, the model gives precedence to the relevant behavior in the past to make predictions of future interests with high accuracy. This keeps the system dynamic, responding to the changing trends in

consumers, seasons and changes in individual interests, thus offering time-sensitive product recommendations that are in line with the current lifecycle of a user.

V. IMPLEMENTATION DETAILS

Development Environment

The system was developed using Python 3.11 as the core programming language, utilizing Visual Studio Code as the primary integrated development environment. The backend framework is powered by Flask, while PyTorch and the Deep Graph Library (DGL) were used to build and train the deep learning architectures. The environment was configured on a high-performance Windows system to handle intensive GNN and Transformer computations.

Database Design

A robust relational schema was implemented using MySQL to manage multi-dimensional data across several core tables. These include dedicated tables for Admin, Seller, and Buyer roles, as well as structures for Product listings and Purchase records. Relationships between tables ensure data integrity, allowing the system to efficiently store and retrieve demo-graphic, transactional, and sequential interaction logs.

Preparation of the dataset and training the model

The system utilizes the Brazilian E-Commerce Public Dataset by Olist, which provides real-world interaction and geographic data. Data preprocessing involved cleaning missing values, encoding categorical features, and normalizing trans-action timestamps. TendGraphNet was trained using message-passing on heterogeneous graphs, while SeqBehNet was op-timized using self-attention to learn chronological purchase sequences and evolving user preferences.

Model Integration

The trained models were integrated into the Flask backend using serialized formats for real-time inference. RESTful APIs were developed to bridge the gap between the analytical modules and the web application. This allows the system to feed live user interaction data into the models and immediately return personalized recommendations or demand insights to the respective user dashboards.

Front-End Development

The frontend was crafted using HTML, CSS, and Bootstrap to ensure a responsive and professional user experience. Inter-active dashboards were built to visualize

complex data through heatmaps and popularity charts for sellers. The interface provides role-specific views, allowing buyers to browse products and receive suggestions while admins monitor system-wide activity and model performance.

Back-End Development

The backend serves as the logic engine, developed using the Flask web framework and Python. It handles secure user authentication, product management, and real-time activity logging into the MySQL database. By managing the flow of data between the frontend and the deep learning modules, the backend ensures that all requests are processed securely and analytical outputs are delivered efficiently.

Quality Assurance and Testing

The system underwent rigorous testing, including Unit Testing for individual modules and Integration Testing for the hybrid model pipeline. System Testing validated overall functionality, including the accuracy of demand forecasts and the reliability of recommendation generation. Finally, User Acceptance Testing (UAT) ensured the platform meets the strategic needs of both sellers and buyers in a real-world e-commerce scenario.

VI. PERFORMANCE METRICS

The effectiveness of the proposed hybrid intelligence framework is evaluated using standard deep learning metrics to measure the accuracy of customer community classification and product recommendations.

Confusion Matrix

Confusion matrix of the classifier is:

True Positive (TP): Correctly recommended products that users actually purchased. True Negative (TN): Products correctly not recommended. False Positive (FP): Incorrectly recommended products. False Negative (FN): Relevant products that the system failed to recommend.

Accuracy

Accuracy is the degree of correctness.

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Measures the number of correct predictions. The closer to 1 the better

Precision

The proportion of quality and coverage of the recommendation engine.

$$Precision = \frac{TP}{P}$$

Increasing precision → reduce false positives

Recall (Sensitivity)

Recall is the probability of system ability to identify all products as user interest

$$Recall = \frac{TP}{P}$$

F1-Score

$$F1-Score = \frac{Precision \times Recall}{Precision + Recall}$$

Used when data is unbalanced. Measures how well the model works

ROC Curve and AUC

ROC (Receiver Operating Characteristic) - graph of True Positive Rate vs False Positive Rate

AUC (Area Under Curve) - model performance. A high AUC (close to 1) → good model

Very low AUC (close to 0.5) → poor model

Performance Significance

Accuracy ensures overall correctness. Precision confirms product recommendations are relevant. Recall ensures system identifies user interest. F1-score is a mixture of precision and recall.



Fig. 2. performance metrics

These metrics establish the reliable relationship between customer behavioral analysis and the strategic

demand insights provided by the proposed e-commerce framework.

VII. RESULT AND CONCLUSION

The system achieved an overall accuracy of 91 percent, indicating strong performance in predicting customer purchasing behavior and generating relevant recommendations. The integration of TendGraphNet and SeqBehNet enables the system to effectively capture both relationship based patterns and sequential user behavior. The performance metrics further validate the system’s effectiveness: Precision (0.89) shows that most recommended products are relevant, Recall (0.88) indicates that the system successfully identifies most of the products users are interested in, and the F1-Score (0.885) reflects a balanced performance between precision and recall. These results demonstrate that the system provides accurate, reliable, and meaningful recommendations while also supporting demand analysis across different locations.

In conclusion, this project presents an intelligent approach to analyzing customer purchasing behavior and product demand in e-commerce. By processing large-scale interaction data, the system delivers personalized recommendations for customers and valuable insights for sellers. It combines TendGraph-Net (GNN) to capture relationships between users, products, and locations, and SeqBehNet (SASRec) to model sequential purchasing behavior and predict future interests. This hybrid approach enhances recommendation accuracy and demand analysis, helping businesses optimize inventory, pricing, and marketing strategies. Overall, the system demonstrates strong performance and practical value for real-world e-commerce applications.

TABLE I Comparison of E-commerce Recommendation and Demand Analysis Systems

System	Relationship Mapping	Sequential Modeling	Regional Analysis	Temporal Awareness	Accuracy
Rule-Based Segment (2019)	No	No	Limited	No	Low
Collaborative Filtering (2020)	Limited	No	No	No	Medium[cite: 2]
Basic Behavioral (2021)	No	Partial	No	Limited	Medium[cite: 2]
Content-Based (2023)	Partial	No	No	No	Medium[cite: 2]
Standalone SASRec (2024)	No	Yes	No	Yes	High[cite: 2]
Proposed System	Yes	Yes	Yes	Yes	High (91%)

REFERENCES

[1] M. M. Wong and W.Su, “Segmenting online shoppers: A combined cluster and logistic regression approach for forecasting purchase behavior,” *IEEE Access*, vol. 13, pp. 79467-79480, 2025.

[2] K. Kim, M. Jo, I. Ra, and S. Park, “RFMVDA: An enhanced deep learning approach for customer behavior classification in e-commerce environments,” *IEEE Access*, vol. 13, pp. 12527-12540, 2025.

[3] M. Tian, “Network data fusion from consumers: Modeling purchase intentions for market decision support,” *IEEE Access*, vol. 13, pp. 197133-197150, 2025.

[4] M. Li, C. M. Yang, and Y. W. Kao, “Clients purchasing tendency community classification in e-commerce scenario: Multi-feature search for densely distributed clients in huge network,” *IEEE Access*, vol. 13, pp. 128280-128300, 2025.

[5] C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro, “Real-time prediction of online shoppers purchasing intention using machine learning approaches,” *Expert Systems with Applications*, vol. 135, pp. 223-234, 2019.

[6] Y. Dou, “Using CatBoost with logistic regression for predicting online shopping behavior,” *Journal of Retail Analytics*, vol. 11, no. 2, pp. 45-56, 2021.

[7] A. Abdullah, “Gradient boosting model for predicting purchase intention using feature selection,” *IEEE Access*, vol. 8, pp. 145678-145689, 2020.

[8] D. Van den Poel and W. Buckinx, “Predicting online purchasing behavior using clickstream data,” *European Journal of Operational Research*, vol. 180, no. 2, pp. 563-578, 2007.

[9] X. Zhao, “Sustainable purchase intention in e-commerce: A behavioral analysis,” *Sustainability*, vol. 12, no. 4, pp. 1-15, 2020.

[10] F. Pecune, S. Marsella, and C. Pelachaud, “Socially-aware conversational recommendation systems,” *ACM Transactions on Interactive Intelligent Systems*, vol. 10, no. 4, pp. 1-28, 2020.

[11] L. Zhao, *Graph Neural Networks: Foundations, Frontiers, and Applications*. Springer, 2022.

[12] F. Liu and L. Nie, *Advancing Recommender Systems with Graph Convolutional Networks*. Springer, 2023.

[13] C. C. Aggarwal, *Recommender Systems: The Textbook*. Springer, 2016.

[14] A. Ge’ron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, 2022.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.