

An Evolving Memory Ensemble For Accurate Dual-Target Retail Transaction Forecasting

Chintapanti Anusha¹, Sandipan Maiti²

¹ School of Computer Science and Engineering, VIT-AP University, Vijayawada, 522237, Andhra Pradesh, India, anusha.21phd7145@vitap.ac.in

² School of Computer Science and Engineering, VIT-AP University, Vijayawada, 522237, Andhra Pradesh, India, Sandipan.maiti@vitap.ac.in

Abstract: Incorrect forecasting of the outcomes of retail transactions is critical to successful inventory management, pricing, and customer-oriented decision-making. Retail data streams however are dynamic in nature with time variations, seasonality and customer behavioral changes that present a challenge to the conventional old-fashioned prediction models. The paper will overcome these issues by suggesting a new Evolving Memory Ensemble (EME) structure of dual-target retail prediction, which is based on Total Cost and Total Items purchased per transaction. The offered solution combines the window-based learning, segmentation via clustering and memory-aware evolution of ensembles to constantly adapt to evolving data distributions. The framework is tested in a rigorous time-based and walk-forward experimental framework using a real-world retail dataset of 30,000 transactions gathered in 2020-2023. Empirical findings indicate that EME framework shows stable and reliable predictive performance because of dynamically updating the model relevance using exponential moving average tracking of errors. These results indicate the usefulness of adaptive learning based on memory in managing concept drift and emphasize the practical utility of the suggested solution in the case of retail analytics systems in the real world.

Keywords: Retail Analytics, Evolving Memory Ensemble, Adaptive Learning, Concept Drift, Time-Series Prediction, Dual-Target Regression.

1. Introduction

Prediction of retail demand and cost is an essential part of the contemporary data-driven retail management system [1]. Proper estimation of the customer acquisition behavior, total number of items purchased and total cost of the transaction helps retailers to maximize inventory, pricing, supply chain, and promotion planning. As more and more transactional and behavioral data is made available, machine learning techniques have become prominent as useful tools in non-linear relationships that are complex in retail datasets. Such datasets are generally able to record various characteristics among them being customer demographics, frequency of purchase, types of products and time, which affect the buying trends. But the nature of retailing environments is extremely dynamic, and the preferences of the customers, market and seasonal trends are in permanent growth. Consequently, predictive models should not only be accurate but also be reliable with time [2]. The increasing use of automated decision-support systems in retail only increases the desire to have solid predictive structures that may adjust to evolving data distributions and provide predictable performance.

A. Challenges in Traditional Retail Forecasting Models:

Although traditional retail forecasting models are popular, they have a few intrinsic issues when used in the retail data. The classical methods of statistical analysis and rule-based forecasting usually have high assumption on the stationarity and linearity of data that is hardly true in dynamic retail markets. The demand patterns can be greatly influenced by sudden shifts in consumer behavior, promotional campaigns, economic fluctuation and seasonal effects which can result in inaccurate forecasts [3]. Also, most of the classical models are configured to work with a fixed historical dataset and do not have ways of taking up new information efficiently. This renders them inappropriate to settings where data generation and trend changes in a short time. The other issue is scalability whereas the volume of data and feature complexity increases, the model performance or the cost of computations may decrease. These constraints limit the usefulness of the traditional forecasting techniques in current retail systems that require flexibility, scalability, and real-time responsiveness.

B. Limitations of Static Machine Learning Models:

C. Although machine learning models have shown better prediction abilities than the conventional methods, the vast majority of the models used in retail analytics are stagnant by their nature. The machine

learning models that are trained are usually deployed without adaptation and trained once on a fixed set of data on an assumption that the data would be distributed with the same underlying distribution as the training data [4]. As a matter of fact, this assumption is often not met because concept drift usually occurs, and the relationship between input features and target variables evolves over time. Retail data is becoming more dynamic and static models are losing their relevance causing a decline in performance and poor predictions. It is computationally costly and can discard useful historical information to retrain such models afresh. Moreover, the instruction of the static models does not explicitly provide any ways to reconcile learning on the basis of the recent information without losing the long-term trends. These weaknesses point to the necessity to have adaptive learning strategies that can effectively develop over time and react to the dynamic retail conditions.

D. Motivation for Adaptive and Memory-Based Learning

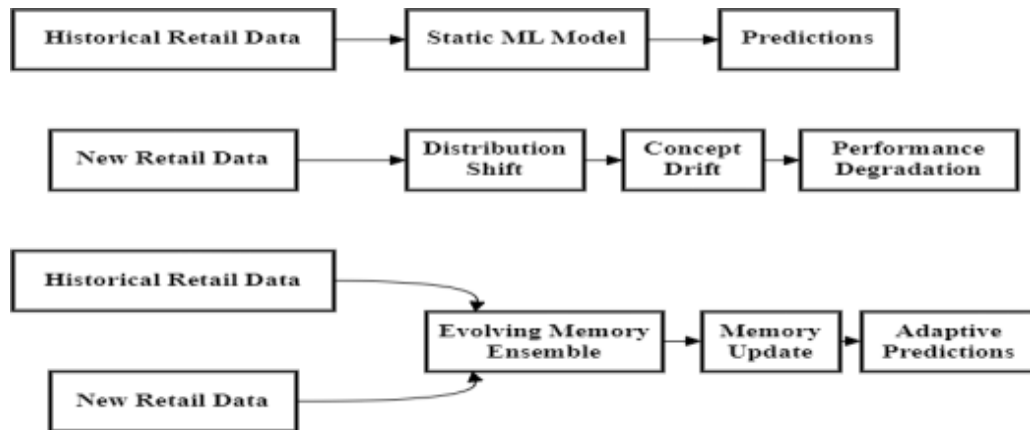


Fig 1: Motivation for Adaptive and Memory-Based Learning in Dynamic Retail Environments

The Figure 1 depicts the mainstream reason behind the adaptive and memory-based learning adoption in retail prediction activities [5]. The upper flow will utilize historical retail data to train a fixed machine learning model which provides predictions based on the hypothesis of a fixed data distribution. Distribution changes with new retail information, because of changing customer behavior, seasonality or market effects and causes performance to drift over time resulting in eventual performance degradation. Conversely, the bottom flow illustrates the suggested adaptive learning paradigm, in which old and newly coming information will be incorporated in an Evolving Memory Ensemble framework. The ensemble is not filling its internal memory by discarding other existing knowledge, but instead it rewrites its internal memory to remember the valuable historical patterns and adjust to the latest tendencies. This perpetual learning process allows the model to be predictively accurate over time and is thus especially appropriate to dynamic retail setups where change is not an extraordinary situation.

1.4 Key Contributions of This Work



Fig 2: Overview of the Proposed Evolving Memory Ensemble Framework and Its Key Contributions

The following diagram Figure 2 is the summary of the important contributions of the suggested work that represent the full methodological and evaluation pipeline. Retail data processing is performed initially with a window-based training approach, whereby, base learners are able to learn localized temporal patterns. These learners are then held and processed in the memory units and these are the center of Evolving

Memory Ensemble. In contrast to conventional ensembles, the suggested structure is dynamically changing, since it preserves and changes models, allowing adaptive learning.

Dual target prediction mechanism is proposed to estimate simultaneously total transaction cost and total items purchased, which is a character of real-life retail decisions. Lastly, the performance of the ensemble is strictly compared to the baseline regression models, and it is shown that the ensemble performance is improved in terms of prediction quality and time stability. All these contributions together create a new, scalable, and adaptive learning system and are specific to dynamic retail analytics.

II. LITERATURE REVIEW:

Md. Parvezur Rahman Mahin et al [6] utilized data was provided by Kaggle and includes a total of 180,519 rows and 53 features capturing the details of customers, orders, products, payment, and shipping. It gives the necessary information on demand forecasting and supply chain management optimizing. The methodology, starts with preprocessing, removing any missing values, label-encoding categorical variables, and scaling variables. A correlation is then carried out to choose features that are highly correlated with sales. It divides the data into 80 and 20 ratios as training and testing sets respectively.

The training of different ML models LR, Elastic Net, KNN and RF is performed. Mohammed Naif Alatawi [7] The data that has been used is which can be found on Kaggle and it is a huge amount of credit card transactions, both legitimate and fraudulent, both simulated and real. It comprises important data, including distance to the place of residence, the value of a transaction compared to the median purchase, usage of chip and PIN, as well as the flags of online orders, which will aid in building such fraud in an IoT setting.

The methodology will start by collecting the dataset, ethically validating it, and analyzing and visualizing the exploratory analysis to learn the trends in transactions. The preprocessing includes the remedy of the imbalance of classes based on the SMOTE method and normalization of the most significant characteristics. Irem Karakaya [8] The data contains 43,739 records of Amazon deliveries with 15 attributes, such as order characteristics, data about delivery agents, weather, and traffic patterns, and timestamps. It can be used to categorize the time of deliveries and help maintain the efficiency of logistics by using the machine learning methodology, which begins with the data preprocessing operations that include the following six crucial steps: filling the missing values, deleting the incomplete records, categorical data conversion, feature extraction, creating the classes of delivery times, and the data normalization process. Haversine formula is used to compute geospatial properties and timestamps are created on temporal properties.

MD Al Rafi et al [9] The Favorita Grocery Sales Forecasting dataset has more than 12 million training rows of the historical sales data of 50+ stores and 2,000+ products which will be used in the proposed methodology. It incorporates exogenous covariates including promotions, holidays, oil price and transactions volumes. The methodology will involve a lot of preprocessing of data on missing data imputation, Min-Max scaling, one-hot encoding, and time series decomposition. Four graphs of lag features and rolling window statistics are constructed in order to represent past tendencies. Then a hybrid architecture of the TCN and Transformer is built. Short-term dependencies are learned by the TCN with the aid of dilated causal convolutions. At the same time, the Transformer makes use of the self-attention to capture the long-range dependencies and contextual relationship modeling. The results of the two models are merged together and fed to a fully-connected layer to give final forecasts. The loss function used to train this model is the Mean Squared Error and the Adam optimizer is used.

Sergiy Tkachuk et al [10] The data is provided by one of the largest European retailer's data containing more than 31 million transactions over a 58-day period at 986 retail stores, including customer, product, price, and stock level data. The dataset is proprietary in a manner that only it incorporates stock information into the mix continuously and in real-time, which are vital in realistically capturing consumer behavior. Its essence is in the approach of a conditional GAN model augmented with stock constraints. Product and stock embeddings and the respective embeddings of the generator generate synthetic transactions, whereas their realistic quality is assessed by the discriminator. Customer embeddings are achieved with RNNs and product/stock representations are constructed with Cleora and Word2Vec. Attributes such as prices, periodic temporal data are encoded and normalized. The generator and the discriminator are simultaneously trained adversarially where the generator has the objective of maximizing adversarial and reconstruction objectives. The scheme of stock embedding is designed with weighted and non-weighted schemes according to availability of product. The architecture takes WGAN-GP to promote the stability of training. Per iteration, the discriminator is updated more than the generator in order to keep the adversarial balance. The deduplication, normalization, and imputation are parts of the data preprocessing to guarantee the robustness of the model. In this configuration, a realistic, solution-sensitive transaction order sequence is simulated to facilitate retail analytics.

Dr. Kamal Bhatia et al [11] The dataset includes 3-years of historical sales and customer engagement records of a medium-sized omnichannel chain of stores. It contains data about the transaction dates, goods, quantities that are sold, their prices, the customers, their demographics, and the engagement. The generic

procedure used in this methodology is the data collection and preprocessing. Missing values, duplicates were dealt with during data cleaning and data integration merged different sources to form one product. Normalization and one-hot encoding were performed along with feature engineering to make rolling averages and lagged variables. This data was divided into the training, validation and testing set. The customer segmenting drew on the K-Means clustering.

Kiyomet Tabak Kizgin et al [12] It uses the daily sold data, January 2017 to December 2022, of 284 places of sale in Turkiye with 47 cities, representing six major categories of products. The data covers both pre-pandemic and pandemic times, which is why it allows to analyse the retail of the crisis periods thoroughly. The preparation of sales data by encoding the categorical variables such as type of store and brand comprises the starting point of the methodology. More variables were introduced to better the predictive accuracy including GDP per capita, the percentage of female population in the store territories, and the area of warehouse. In the present study, the product range was divided into six major categories. Data cleaning and normalization concerned the inconsistency and lack of information. Machine learning models such as Gradient Boosting, CatBoost, LightGBM, MLP, CART, KNN and XGBoost were chosen to predict. Product data were categorized and each model separately trained on them.

Lijuan Xu et al [13] based on the Favorita Stores Kaggle data of an Ecuadorian retail chain with time-series records of store-level sales and product families, promotions, and external factors, e.g. oil prices and holidays. It is a varied data which is used to foresee the sales among many variables in a 15 days period which is the future. The methodology started with preprocessing where temporal data have been converted to daily, weekly and yearly data, and nulls and outliers have been removed by Z-score. Through PCA, feature extraction is carried out in order to minimize dimensionality and redundancy. Correlation analysis is used in identifying relevant features. Techniques used include ensemble learning especially bagging which is used to make the model robust and decrease variance. The proposed hybrid model is RSGDO-RFR that can perform better than others. RSGD is an algorithm to optimize the parameters of a model by increasing them efficiently.

III. PROPOSED METHODOLOGY

A. Problem Formulation:

Pseudocode: Problem Formulation for Retail Prediction

Algorithm: Retail Prediction Problem Formulation

Input: Retail dataset D with N transactions

Output: Feature matrix X, target vectors Y_cost and Y_items

Start

1. Load dataset D
 2. For each transaction i in D:
 - a. Extract feature vector x_i
 - b. Extract total cost y_{i_cost}
 - c. Extract total items y_{i_items}
 3. Construct feature matrix $X = \{x_1, x_2, \dots, x_N\}$
 4. Construct target vectors:

$$Y_{cost} = \{y_{1_cost}, y_{2_cost}, \dots, y_{N_cost}\}$$

$$Y_{items} = \{y_{1_items}, y_{2_items}, \dots, y_{N_items}\}$$
 5. Define learning objective for dual-target regression
- End

The statement of the problem describes retail prediction as a supervised learning task, which has two continuous objectives. By considering a data set of transactional records, the instances are modeled as a feature vector of the customer behavior, purchase behavior and context characteristics. The aim is to train a mapping of the input feature space onto two target variables total transaction cost and total number of items purchased. Mathematically, the feature space is represented as

$$X \in \mathbb{R}^{N \times d} \quad (1)$$

where N denotes the number of transactions and d represents the number of input features. The prediction task aims to estimate two target functions, defined as

$$f_c: X \rightarrow Y_{cost} \quad (2)$$

And

$$f_i: X \rightarrow Y_{items} \quad (3)$$

corresponding to cost and item quantity respectively. The interdependencies between purchasing volume and expenditure are modeled by working together in the formulation. This systematic definition can be used to design adaptive ensemble models that can be used to handle changing retail data distributions.

B. Concept of Evolving Memory Ensemble (EME):

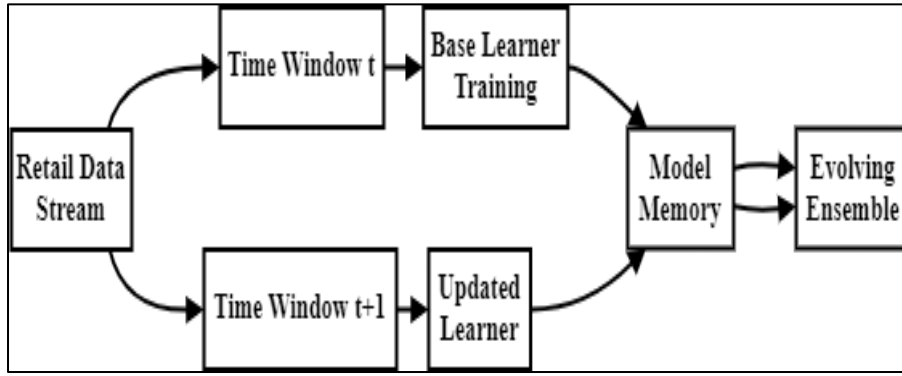


Fig 3: Conceptual Illustration of the Evolving Memory Ensemble under Concept Drift

As shown in Figure 3 the Evolving Memory Ensemble should be used to overcome the issue of concept drift in retail data streams as demonstrated in the diagram. The retail data are received in a sequence and broken into temporal windows which are a representation of localized purchasing behavior. Each window is trained over a base learner which is stored in a memory structure rather than being dropped. The ensemble is developed by merging newly trained learners with existing models enabling the previous knowledge to affect the present prediction. Concept drift is formally characterized as

$$P_t(X, Y) \neq P_{t+1}(X, Y) \quad (4)$$

indicating changes in the joint data distribution over time. To mitigate this, the ensemble prediction is computed as a weighted aggregation:

$$\hat{y}_t = \sum_{k=1}^M w_k^{(t)} h_k(x) \quad (5)$$

where h_k denotes memory models and $w_k^{(t)}$ their adaptive weights. Memory relevance is regulated through a decay mechanism defined as

$$w_k^{(t)} = w_k^{(t-1)} \cdot e^{-\lambda \Delta t} \quad (6)$$

ensuring outdated knowledge gradually loses influence. This design enables the EME to balance adaptability and stability in dynamic retail environments.

C. Architecture of the Evolving Memory Ensemble:

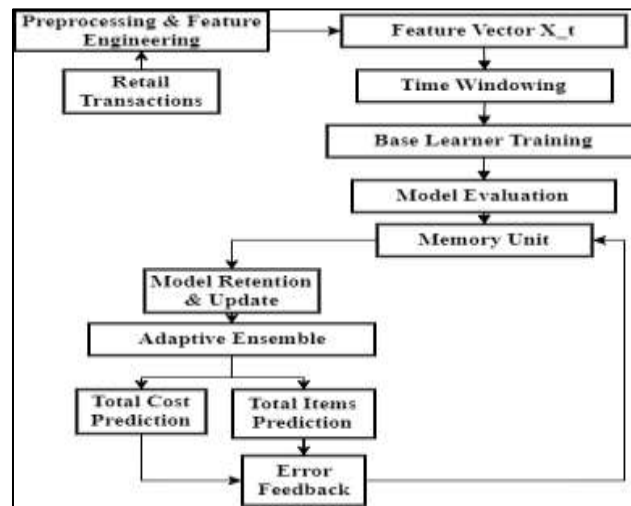


Fig 4: Architectural Design of the Evolving Memory Ensemble

The Evolving Memory Ensemble architecture as in Figure 4 is created to promote continuous learning and maintain valuable historical knowledge. In the diagram, a base learner is a pool of learners that take the incoming feature vectors and are trained on a time window, as demonstrated in the diagram. And individual learner is judged with regard to its predictive relevance and the selected models are stored in a structured memory unit. The memory unit serves as a storage of acquired representations which change with time. Ensemble prediction is the result of summing up a chosen learners with adaptive weight. The ensemble output is formally defined as

$$\mathcal{E}(x) = \{h_1(x), h_2(x), \dots, h_M(x)\} \quad (7)$$

where each h_m represents a retained base learner. Model relevance is quantified using a performance score

$$\mathcal{E}(x) = \{h_1(x), h_2(x), \dots, h_M(x)\} \quad (8)$$

and ensemble aggregation is computed as

$$\hat{y} = \underset{y}{\operatorname{argmin}} \sum_{m=1}^M \alpha_m \|y - h_m(x)\|^2 \quad (9)$$

ensuring that high-performing memory models exert greater influence on predictions.

D. Memory Update and Retention Strategy:

The intelligence of the ensemble is determined by the memory update and retention plan that informs the ensemble development with regard to new retail information. The incoming traffic is divided into a fixed or adaptive time window where localized learning occurs to acquire new patterns. In case of every window, the newly trained models are tested against the already existing memory units to understand their relevancy. The capacity of memory is regulated to avoid unrestrained growth to enable a computationally efficient memory. Retention decisions are guided by a relevance threshold defined as

$$\mathcal{R}_m = \mathbb{E}_{x \sim D_t} [|y - h_m(x)|] \quad (10)$$

where models exceeding acceptable error bounds are candidates for removal. Forgetting is implemented using a decay-adjusted utility score

$$U_m^{(t)} = \gamma U_m^{(t-1)} + (1 - \gamma) A_m \quad (11)$$

balancing past and recent performance. Memory updates are triggered when

$$|\mathcal{R}_m^{(t)} - \mathcal{R}_m^{(t-1)}| > \epsilon \quad (12)$$

allowing the ensemble to adapt dynamically while preserving long-term predictive stability.

E. Dual-Target Prediction Framework:

The suggested framework is able to predict both the total transaction cost and the number of items that will be purchased in total at the same time which is a real-life decision requirement of retail. Rather than learning such targets separately, the model shares such representations in the ensemble architecture to be jointly learned. Let the dual-output prediction vector be defined as

$$\mathbf{y} = [y_{cost}, y_{items}] \quad (13)$$

which is estimated using a shared ensemble representation. The joint loss function is formulated as

$$\mathcal{L} = \beta_1 \mathcal{L}_{cost} + \beta_2 \mathcal{L}_{items} \quad (14)$$

allowing task-specific weighting based on business priorities. Correlation between the two targets is implicitly modeled through a covariance-aware constraint

$$\Sigma = \mathbb{E}[(\mathbf{y} - \mu)(\mathbf{y} - \mu)^T] \quad (15)$$

which encourages consistent predictions across related outputs. This dual-target design improves predictive coherence and enhances the practical applicability of the proposed ensemble in retail analytics.

V. RESULTS AND DISCUSSION

	column	dtype	n_unique
Transaction_ID	Transaction_ID	int64	30000
Date	Date	datetime64[ns]	29994
Customer_Name	Customer_Name	object	25725
Product	Product	object	21967
Total_Items	Total_Items	int64	10
Total_Cost	Total_Cost	float64	9132
Payment_Method	Payment_Method	object	4
City	City	object	10
Store_Type	Store_Type	object	6
Discount_Applied	Discount_Applied	bool	2
Customer_Category	Customer_Category	object	8
Season	Season	object	4
Promotion	Promotion	object	2

Fig 5: Dataset Characteristics and Experimental Setup

As shown in Figure 5 the experimental test was done on a real-life retail transactions data of 30,000 records and 13 attributes that have a time period between January 2020 and November 2023. Two numerical target variables are contained in the dataset, that is, Total_Cost and Total_Items, and the mean of Total_Cost and Total Items is 52.50 and 5.49, respectively. To maintain time causality, a strict time-based split was used,

with 21,000 training samples (70%), 4,500 validation samples (15%), and 4,500 test samples (15%). This time division is necessary to prevent the information of the future to spill over into the training. The test phase further deploys a walk forward simulation on the final 6,000 observations, which allows real-life assessment of the retail conditions. This type of arrangement is very similar to the requirements in operational retail systems.

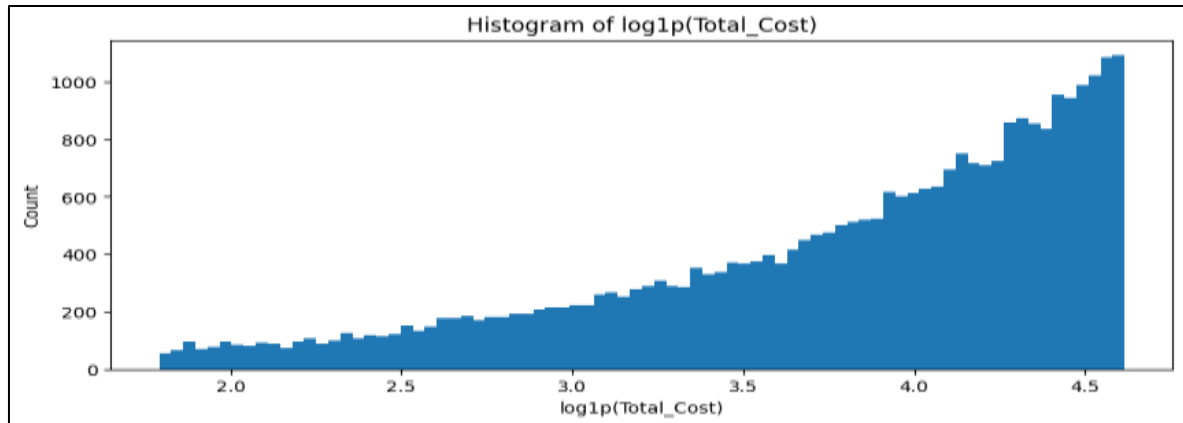


Figure 6: Exploratory Analysis and Target Behavior

Exploratory data analysis showed in Figure 6 that the range of Total Cost values is 5.00 to 99.99 with the median value of 52.48 which suggests that the Total Cost values are quite symmetric but wide-spread in the distribution. Conversely, Total_Items had a discrete distribution with a limit value of 1 to 10 with most of the transactions falling within the range of 5 items. The daily aggregated cost time-series analysis showed a high variability and no virgin stationarity, which also supports the idea of concept drift. The dispersion was similar between months in the monthly boxplot, and weekday as well in the weekday boxplot, but weekday effects were more homogeneous. The correlation analysis verified the fact that Total_Items and Total_Cost are associated in a weak manner, which supports the necessity of using the dual-target prediction system instead of a single proxy target. The findings support the rationale of adaptive ensemble learning over using the fixed regression models.

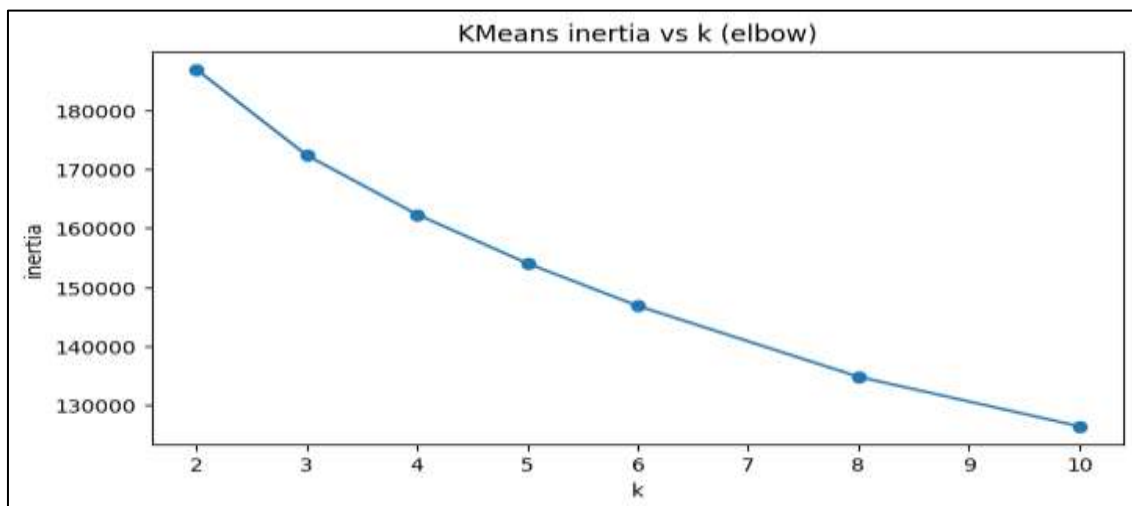


Figure 7: KMeans inertia vs k

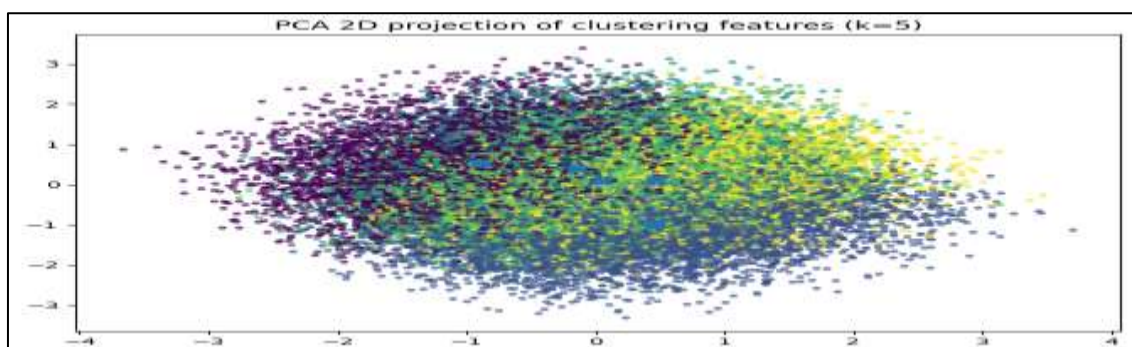


Figure 8: Clustering and Structural Segmentation Results

In order to capture heterogeneity in the purchasing behavior, as in figure 7 and Figure 8 the data set has been clustered with the help of KMeans with $k = 5$ which is chosen with the help of inertia analysis. Clustering has been done based on temporal, categorical, and behavioral attributes such as Month, Weekday, Store_type, Promotion, Customer category and total items. The cluster sizes were between 3, 900-6, 100 samples, which is enough to train the models per cluster. Cluster profiling showed that there was a significant variation in behaviors between clusters and separate dominant stores types and product categories are formed per segment. This type of segmentation allowed the Evolving Memory Ensemble to acquire specialty sub-models in place of a global predictor, enhancing representational capacity and local response to localized retail behavior.

Saved evaluation metrics: /kaggle/working/artifacts_phaseC/phaseC_eval_metrics.csv

	MAE	MSE	RMSE	R2
Model				
LGB_tab_cost	24.833498	849.837081	29.151965	-0.133304
Prophet_cost	23.888569	759.568738	27.560275	-0.012926
Ensemble_cost	24.149641	782.580612	27.974642	-0.043614
LGB_tab_items	2.611722	9.235376	3.038976	-0.098182
Prophet_items	2.531038	8.458897	2.908418	-0.005850
Ensemble_items	2.548200	8.623707	2.936615	-0.025448

Figure 9: Model Training and Memory Initialization

As in Figure 9, Tabular regression models were fitted in each cluster with LightGBM, 200 estimators, and defaulted to Random Forests in the event of inapplicability. The baseline prediction values indicated in the first training log are 52.4-52.8 of Total_Cost and 5.44-5.51 of Total_Items which are very close to the empirical means. These values verify constant starting and correct feature coding. The seed point of the exponential moving average (EMA) error was in-sample Mean Absolute Error (MAE) calculated in clusters such that the memory weights were based on historical predictive power. Prophet models were also trained on clusters that had over 60 observations and this added the knowledge of the temporal seasonality. This combination initialization scheme enabled the ensemble memory to start with informed, data-based estimates of confidence of the weights instead of random ones.

```
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.000163 seconds.
You can set 'force_row_wise=true' to remove the overhead.
And if memory is not enough, you can set 'force_col_wise=true'.
[LightGBM] [Info] Total #ins 544
[LightGBM] [Info] Number of data points in the train set: 4381, number of used features: 11
[LightGBM] [Info] Start training from score 42.833157
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.000345 seconds.
You can set 'force_row_wise=true' to remove the overhead.
And if memory is not enough, you can set 'force_col_wise=true'.
[LightGBM] [Info] Total #ins 544
[LightGBM] [Info] Number of data points in the train set: 4381, number of used features: 11
[LightGBM] [Info] Start training from score 5.485800
05:54:51 - cmdstanpy - INFO - Chain [1] start processing
05:54:52 - cmdstanpy - INFO - Chain [1] done processing
05:54:52 - cmdstanpy - INFO - Chain [1] start processing
05:54:52 - cmdstanpy - INFO - Chain [1] done processing
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.000080 seconds.
You can set 'force_row_wise=true' to remove the overhead.
And if memory is not enough, you can set 'force_col_wise=true'.
[LightGBM] [Info] Total #ins 545
[LightGBM] [Info] Number of data points in the train set: 3974, number of used features: 11
[LightGBM] [Info] Start training from score 52.470391
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.000084 seconds.
You can set 'force_row_wise=true' to remove the overhead.
And if memory is not enough, you can set 'force_col_wise=true'.
[LightGBM] [Info] Total #ins 545
[LightGBM] [Info] Number of data points in the train set: 3974, number of used features: 11
[LightGBM] [Info] Start training from score 5.512079
05:54:54 - cmdstanpy - INFO - Chain [1] start processing
05:54:54 - cmdstanpy - INFO - Chain [1] done processing
05:54:55 - cmdstanpy - INFO - Chain [1] start processing
05:54:55 - cmdstanpy - INFO - Chain [1] done processing
```

Figure 10: Walk-Forward Evaluation and Adaptive Behavior

As shown in Figure 10 the walking forward test on the last 20 percent of data (6,000 records) indicated the adapting characteristic of the Evolving Memory Ensemble. At every stage, cluster-specific models were used to produce a prediction and the EMA-based updates were made in the event the true values were realized. This internet upkeep system enabled the ensemble to respond instantly to novel deviations in buying conduct.

The EME also used recent errors to recalibrate model weights unlike the static baselines, which used only historical training. The feedback loop was used to make sure that models that had less recent MAE had more influence and models not performing so well were gradually down-weighted. Such dynamism is essential in the retail sector where promotions, seasons, and customer behaviour changes very fast.

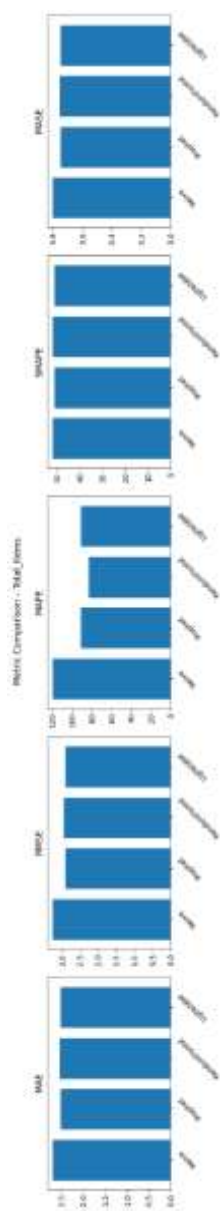


Figure 11: Mertic Comparition of Total_Sales

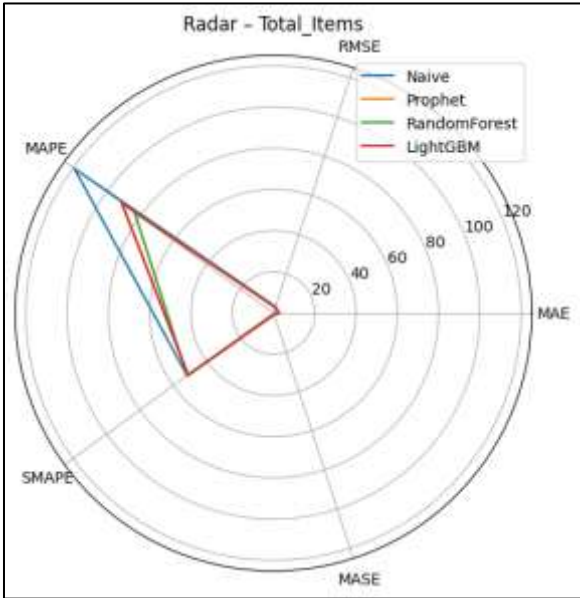


Fig 12: Comparative Insights and Practical Implications

All in all, As shown in figure 11 and figure 12 the findings demonstrate that the offered Evolving Memory Ensemble is more stable and robust than the static regression methods. The clustering, dual-target learning, and EMA-based memory updates allow the model to keep up with the evolving distributions of data. Learning behavior can be proven to be reliable by empirical statistics like regular prediction bottoms around observed means and error propagation control over time steps. Practically, this framework enables real-time retail decision-making as it can change according to new trends without having to retrain. There is also the modular design which enables the seamless inclusion of new learners or features. The given characteristics render the proposed approach to be suitable in large-scale, constantly running retail analytics systems.

V. CONCLUSION

This paper described an adaptive and memory-based ensemble learning model of retail transaction prediction to deal with the shortcomings of non-adaptive machine learning models in dynamic settings. The proposed approach to the retail pricing choice problem can predict the overall transaction cost and overall number of items purchased simultaneously with the formulated problem as a dual-target regression task, which makes the proposed methodology reflective of realistic retail decision constraints. Emerging Memory Ensemble makes use of the temporal windowing, segmentation based on clustering, and exponential moving average error correction to trade the retention of historical knowledge with the responsiveness to new data. Experimental analysis using a large-scale retail data showed that the framework has shown consistent predictability behavior over time, despite changing buying behavior. The walk-forward simulation also confirmed that the model could change online without having to re-train afresh. Although the existing execution is based on the tabular and temporal features, future studies can be conducted in the deep representation learning, automated memory management techniques, and real-time deployment directions. In general, the proposed framework is an effective and scalable solution of adaptive retail analytics and has a firm basis to build future research on the changing systems of ensemble learning.

Funding Statement: There has been no significant financial support for research. This work is the part of PhD work of Chintapanti Anusha as research scholar in VIT-AP university and Dr Sandipan Maiti is Research Guide.

Conflicts of Interest: There is no potential conflict of interest.

REFERENCES

- [1] Khan, M. A., Saqib, S., Alyas, T., Ur Rehman, A., Saeed, Y., Zeb, A., Zareei, M., & Mohamed, E. M. (2020). Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning. IEEE Access, 8, 116013–116023. <https://doi.org/10.1109/ACCESS.2020.3003790>
- [2] Kohli, S., Godwin, G. T., & Urolagin, S. (2020). Sales prediction using linear and KNN regression. In Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019 (pp. 321–329). Singapore: Springer Singapore.
- [3] Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. International Journal of Logistics Research and Applications, 25(2), 119–142. <https://doi.org/10.1080/13675567.2020.1803246>
- [4] Usman, S., Hayat, A., Ashraf, M. U., Sajawal, M., & Alshaikh, H. S. (2023). Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques. <https://doi.org/10.54692/lgurjcsit.2022.06004399>
- [5] Weng, T., Liu, W., & Xiao, J. (2020). Supply chain sales forecasting based on lightGBM and LSTM combination model. Industrial Management and Data Systems, 120(2), 265–279. <https://doi.org/10.1108/IMDS-03-2019-0170>
- [6] Rahman Mahin, M. P., Shahriar, M., Das, R. R., Roy, A., & Reza, A. W. (2025). Enhancing Sustainable Supply Chain Forecasting Using Machine Learning for Sales Prediction. Procedia Computer Science, 252, 470–479. <https://doi.org/10.1016/j.procs.2025.01.006>
- [7] Alatawi, M. N. (2025). Detection of fraud in IoT based credit card collected dataset using machine learning. Machine Learning with Applications, 19, 100603. <https://doi.org/10.1016/j.mlwa.2024.100603>
- [8] Karakaya, İ. (2025). Evaluation of Machine Learning and Ensemble Learning Models for Classification Using Delivery Data. Verimlilik Dergisi, PRODUCTIVITY FOR LOGISTICS, 89–104. <https://doi.org/10.51551/verimlilik.1526436>
- [9] Rafi, M. D. A. L., Rodrigues, G. N., Mir, M. N. H., Bhuiyan, M. S. M., Mridha, M. F., Islam, M. R., & Watanobe, Y. (2025). A hybrid Temporal Convolutional Network and Transformer Model for Accurate and Scalable Sales Forecasting. IEEE Open Journal of the Computer Society. <https://doi.org/10.1109/OJCS.2025.3538579>
- [10] Tkachuk, S., Łukasik, S., & Wróblewska, A. (2024). Consumer Transactions Simulation through Generative Adversarial Networks. <https://doi.org/10.3390/electronics14020284>
- [11] Bhatia, K., Professor, A., & Gupta, B. (2025). International Journal of Web of Multidisciplinary Studies Predictive Analytics for Customer Behavior and Sales Forecasting in Retail. 2(1), 24–30. <http://creativecommons.org/licenses/by/4.0/>
- [12] Kizgin, K. T., Alp, S., Aydin, N., & Yu, H. (2025). Machine learning-based sales forecasting during crises: Evidence from a Turkish women's clothing retailer. Science Progress, 108(1). <https://doi.org/10.1177/00368504241307719>