An Interpretable Data-Driven Unsupervised Approach for the Prevention of Forgotten Items

Luca Corbucci [a,*], **Javier Alejandro Borges Legrottaglie** [a], **Francesco Spinnato** [a], **Anna Monreale** [a] and **Riccardo Guidotti** [a]

[a]University of Pisa

arXiv:2507.23303v1 [cs.LG] 31 Jul 2025

Abstract. Accurately identifying items forgotten during a supermarket visit and providing clear, interpretable explanations for recommending them remains an underexplored problem within the Next Basket Prediction (NBP) domain. Existing NBP approaches typically only focus on forecasting future purchases, without explicitly addressing the detection of unintentionally omitted items. This gap is partly due to the scarcity of real-world datasets that allow for the reliable estimation of forgotten items. Furthermore, most current NBP methods rely on black-box models, which lack transparency and limit the ability to justify recommendations to end users. In this paper, we formally introduce the forgotten item prediction task and propose two novel interpretable-by-design algorithms. These methods are tailored to identify forgotten items while offering intuitive, human-understandable explanations. Experiments on a real-world retail dataset show our algorithms outperform state-of-the-art NBP baselines by 10–15% across multiple evaluation metrics.

1 Introduction

Recommendation systems are evolving from simple suggestion tools into sophisticated, ubiquitous technologies that deliver personalized item selections across several domains, ranging from books and movies, to supermarket products and retail transactions. Especially within supermarkets and online marketplaces [10], recommendation systems play a crucial role in shaping consumer behavior, improving customer satisfaction, and driving business growth [18]. A significant portion of current research in retail analytics has focused on Next-Basket Prediction (NBP), the task of forecasting a customer's next purchase based on their historical transaction data [16]. Techniques developed for this problem range from models that leverage sequential shopping patterns to complex deep learning architectures. While these approaches have shown success in predicting general shopping behaviors, they fall short when it comes to addressing a more nuanced and practically important challenge: predicting forgotten items.

In the transactional retail setting, forgotten items are products that customers intended to purchase during a shopping trip but inadvertently left behind, often leading to a quick return trip to the store. For instance, we can imagine a customer shopping for their weekly groceries. Their cart includes jam and peanut butter, but it is missing bread, which they typically purchase along with the other breakfast items. A system that detects such omissions could prompt reminders, enhancing customer satisfaction and increasing timely sales. This simple scenario exemplifies the need for recommendation models that go beyond predicting what a user might like, but rather infer what they likely intended to buy but overlooked. Forgotten items create inefficiencies for customers, who must make additional store visits, and retailers, who miss opportunities for immediate sales. Mitigating this problem requires a different methodological approach than conventional NBP, as it involves identifying deviations from an individual's established shopping patterns, rather than predicting entirely new or evolving preferences. We empirically demonstrate that, although existing NBP methods can be applied to address the forgotten item problem, their performance on this specific task remains suboptimal.

Another key limitation of today's recommendation systems is that their increasing sophistication often comes at the expense of transparency. Indeed, many state-of-the-art models now operate as black boxes, achieving high predictive accuracy but offering little insight into their decision-making processes [11, 23]. This opacity presents a significant challenge in retail environments, where both businesses and customers benefit from understanding the rationale behind a recommendation [20]. The lack of interpretability can undermine user trust and diminish the practical utility of these systems [17]. Therefore, there is growing demand for models that not only perform effectively but also offer clear, easy-to-understand explanations, especially in contexts where recommendations influence consumer behavior.

We address these gaps through the following key contributions.

(i) We formalize the *Forgotten-Item Prediction* (FIP) problem, establishing a framework for systematically evaluating and comparing methods for this specific task. To support this, we define evaluation metrics that reflect the distinct nature of FIP, setting them apart from standard NBP criteria. ***(ii)*** We propose two FIP algorithms: `XMT`, which integrates historical purchase patterns, temporal dynamics, and contextual signals to identify likely forgotten items; and `TXMT`, an extension of `XMT` that incorporates recurrent purchasing behaviors using TARS [5, 7]. Both methods employ a multi-factor scoring system to enhance predictive accuracy. Furthermore, ***(iii)*** `XMT` and `TXMT` are interpretable-by-design, providing human-understandable justifications for each prediction. For instance, revisiting the earlier example, the model may suggest that *bread is a forgotten item because the current basket contains both peanut butter and jam, bread is typically purchased weekly, and the customer last bought it one week ago*. ***(iv)*** We evaluate `XMT` and `TXMT` on *real-world retail data*, demonstrating the effectiveness of our models, significantly outperforming both baseline NBP and state-of-the-art FIP methods by margins of 10–15% across different evaluation metrics.[1]

* Corresponding Author Email: luca.corbucci@phd.unipi.it

[1] Our code is available here: https://bit.ly/3H0r1C7

2 Related Works

Early recommender systems in retail focused on identifying item co-occurrence using collaborative filtering and association rule mining [18]. While effective at capturing static affinities, these approaches lacked temporal modeling to anticipate shopping intent. To incorporate sequential patterns, various time-aware and personalized strategies emerged. Simple yet strong baselines include predicting items from the user's most recent basket or selecting the most frequent items from a customer's purchase history [3]. More involved approaches use Markov Chains to introduce temporal dynamics as item transitions or binary classifiers on handcrafted temporal features to forecast item reoccurrence [3]. Matrix factorization methods brought further personalization. For example, Non-negative Matrix Factorization (NMF) [12] models latent item-user interactions. The Factorized Personalized Markov Chain (FPMC) [16] method combines matrix factorization with personalized Markov chains to account for user preferences and sequential transitions. Building on this, Hierarchical Representation Model (HRM) [21] learns unified representations of user and basket histories via hierarchical neural architectures. The advent of deep learning advanced NBP, enabling modelling of long-range dependencies and complex user behavior through recurrent networks [23], attention mechanisms [22], and hierarchical encoding [9].

However, most of these approaches function as black-box models, optimizing predictive performance at the cost of interpretability [6]. This opacity is particularly problematic in consumer-facing retail environments, where trust, transparency, and actionable insights are paramount. Indeed, as recommendation models become increasingly complex, explainability has become a crucial concern. In contemporary recommender systems, explaining why an item is suggested can be as important as the recommendation's accuracy for maintaining user trust and satisfaction [24]. This is especially true in retail, where customers may be wary of automated suggestions affecting their purchase decisions. Historically, straightforward techniques like association rules naturally offered simple explanations, e.g., *"customers who bought X also bought Y"*, based on co-purchase statistics [11]. However, modern NBP algorithms, from FPMC to deep neural networks, are largely opaque. Most next-basket recommendation methods have been optimized for predictive performance, with little attention to transparency [9]. The lack of interpretability can reduce user acceptance and complicate retailers' validation of recommendations.

Recently, research has started to bridge this gap by infusing explainability into NBP. One line of work integrates external knowledge to rationalize recommendations. In [9], the authors construct a basket-level knowledge graph using Reinforcement Learning to generate intuitive paths that justify a recommended item, providing explanations alongside the next-basket prediction. Attention-based models also offer a degree of interpretability by revealing which past items or baskets had the most influence on the prediction, albeit implicitly [22]. Nevertheless, providing direct and user-friendly explanations in a sequential recommendation setting remains challenging. Unlike single-item recommendations, next-basket suggestions may result from complex interactions among multiple prior purchases, which complicates the explanation task. [5, 7] introduces an interpretable Temporal Basket Prediction (TBP) model for NBP based on Temporal Annotated Recurring Sequences (TARS). Thanks to TARS, TBP can effectively capture multiple factors influencing customer decision-making, including co-occurrence, sequentiality, periodicity, and recurrence of purchased items.

Within this context, identifying products a user likely intended to purchase but accidentally forgot remains a relatively unexplored task. We define this task as Forgotten-Item Prediction (FIP). Prior work on basket completion touches on related themes, such as suggesting complementary or co-purchased items for an ongoing basket [4, 18], or identifying items that lost purchase relevance over time [14]. However, the task of predicting forgotten items in basket analysis is mostly absent from the literature. FIP requires modeling user intention and recall errors, rather than merely predicting what is likely to be bought next, making it a unique and under-investigated extension of the NBP problem. Indeed, to the best of our knowledge, at the time of writing, the only published work addressing the FIP task, albeit from a slightly different perspective, is [19]. In [19], the authors introduce the customer-centric Interval-Based Predictor (IBP), which addresses the FIP problem by offering customers a list of potentially forgotten items. This is achieved by considering numerical factors such as the average time between visits, the average interval between purchases of a given item, and the time elapsed since the item was last purchased, to estimate when the customer is likely to run out of stock. However, unlike most existing approaches to NBP, the score for each item is computed independently of the others, without considering item co-occurrence. Also, while the numerical estimation allows for some degree of interpretability, it remains notably limited.

In this paper, we address a research gap by formalizing the unexplored FIP problem and proposing an unsupervised, data-driven approach to prevent forgotten items. Our proposals simultaneously account for co-occurrence, sequentiality, periodicity, recurrence of purchased items, average time between visits, average intervals between purchases of specific items, and the time elapsed since an item was last bought. Furthermore, these factors are combined to provide customers with clear and intuitive explanations for why a particular item is being recommended.

3 Setting the Stage

This section outlines the key concepts behind our proposal.

Next-Basket Prediction (NBP): This task aims to forecast items a user is likely to purchase in their next shopping transaction based on historical shopping behaviors. Formally,

Definition 1 (Next-Basket Prediction)**.** *Let $U = \{u_1, u_2, \ldots, u_n\}$ denote the users and $I = \{i_1, i_2, \ldots, i_m\}$ the items available. Each user $u \in U$ has a shopping history $H_u = \{B_1, B_2, \ldots, B_{t-1}\}$ with baskets $B_j \subseteq I$. Given H_u, NBP predicts the probability distribution $P(i \in B_t \mid H_u)$ for each item $i \in I$.*

Thus, a NBP function g takes as input the shopping history of a customer H_u and, possibly, also a set of shopping history of a set of customers $\mathcal{H} = \{H_1, \ldots, H_n\}$, and returns the predicted next basket $\hat{B}_t = g(H_u, \mathcal{H})$. Key challenges in the algorithmic structure of g include modeling sequential item dependencies, co-occurrence patterns, and temporal dynamics of user behaviors [7]. Temporal Annotated Recurring Sequences (TARS), introduced in [5, 7], capture temporal and recurrent purchasing behavior. In simple terms, a sequence $X \rightarrow Y$, is composed of a set of items X called a head, followed later, in another shopping event, by a set of items Y named tail.

We say that a sequence is *minimal* when there is not a more smaller sequence using subsets of these items that occurs within the same time interval. TARS are mined from a customer's purchase history H_u by detecting frequent sequences and adapting temporal thresholds to individual behavior. Final item scores, Ω_i, assess the relevance of item i appearing in the tails Y of relevant TARS. They quantify how relevant each item is for future purchases, based on a customer's recent behavior and historical recurring patterns.

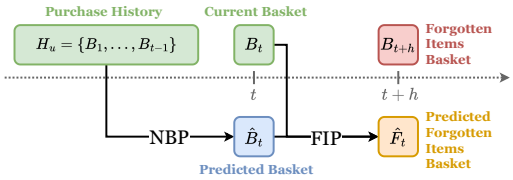


Figure 1. Depiction of the Next Basket Prediction (NBP), and Forgotten Items Prediction (FIP) problems. In NBP, the user's purchase history H_u is used to predict the current basket at time t, while FIP is used to predict which items the user is forgetting at the current time t.

Explainable AI. As NBP approaches grow increasingly complex, the need for Explainable AI (XAI) methods to interpret them becomes more pressing. XAI techniques can be classified as either *local* or *global*, depending on whether they explain individual predictions or the model's overall behavior [1]. They may also be *post-hoc*, applied after training to explain a black-box model, or *ante-hoc*, also known as *interpretable-by-design*, where the model is inherently transparent. We focus here on local, interpretable-by-design explainers, which provide direct, faithful insights into decision-making without relying on potentially misleading post-hoc approximations [17].

4 Problem Formulation

We define the Forgotten-Items Prediction (FIP) problem as follows.

Definition 2 (Forgotten-Items Basket)**.** *Let B_t represent a basket transaction purchased at time t. We name the basket B_{t+h}, occurring after a temporal horizon h relative to t, as a* forgotten-items basket *if the customer intended to purchase the items in B_{t+h} at the time t together with the items in B_t, i.e., the actual intended purchase should have been $B_t^* = B_t \cup B_{t+h}$.*

Thus, a FIP function g takes as input the current basket B_t, the shopping history of the customer H_u, and, depending on its nature, also a set of shopping history of a set of customers $\mathcal{H}$, and returns the predicted forgotten-items basket $\hat{F}_t = g(B_t, H_u, \mathcal{H})$. The temporal horizon h can be set based on observed shopping patterns and helps differentiate genuine forgotten-item purchases from regular shopping. For instance, it can be set in an interval between 0 days (i.e., re-purchasing the same day) and 2 days (i.e., repurchasing at most after 2 days). The intention of a customer to purchase two sequential baskets together can be inferred through various means, such as customer interviews or heuristics applied to their shopping history. Details specific to our case study are provided in Section 6.1.

We can now introduce the Forgotten-Item Prediction (FIP) problem, which has the objective to predicts items that are likely to be forgotten during the current transaction.

Definition 3 (Forgotten-Item Prediction)**.** *Let $U = \{u_1, u_2, \ldots, u_n\}$ denote the users, $I = \{i_1, i_2, \ldots, i_m\}$ the items available and $H_u = \{B_1, B_2, \ldots, B_{t-1}\}$ the shopping history of each user with baskets $B_j \subseteq I$. Given H_u, FIP consists in predicting the probability distribution $P(i \in B_{t+h} \mid H_u, B_t)$ for each item $i \in I$ where B_{t+h} is a forgotten-items basket of B_t.*

Hence, the objective of FIP is to predict the forgotten-items basket $\hat{F}_t \approx B_{t+h}$ given the customer history H_u and the current basket B_t. A depiction of the NBP and FIP tasks can be found in Figure 1. We report in Table 1 the list of symbols used in this paper to increase clarity and readability.

Table 1. Summary of notation.

Entities and Variables	
$\mathcal{H}, H$	set of user histories, single history
U, u	set of users, user
I, i	set of items, item
B_t, B_{t+h}	basket at time t, forgotten-items basket
$\hat{B}_t, \hat{F}_t$	predicted basket, predicted forgotten items
g	NBP or FIP function
Item Properties	
η_i	median inter-purchase interval
ϕ_i	days since last purchase
ρ_i	seasonal concentration ratio
$\chi(i, j)$	co-purchase count with item j
$\zeta_i(h)$	repurchases within h
λ_i	repurchase opportunities after large baskets
Item Scores	
f_i, τ_i, σ_i	frequency, temporal, seasonal scores
κ_i, ψ_i	context affinity, repurchase tendency
Ω_i	TARS-based score
MAP_i, $TMAP_i$	combined prediction scores
Hyperparameters	
ϵ, γ, G	min. freq., max. interval mult., discontinue threshold
α, υ, β	temporal boost, seasonality thresh., seasonal boost
Υ, Θ	co-purchase thresh. and boost
ν, Λ, Φ	repurchase window, boost, rate threshold

5 Methodology

We present here two interpretable unsupervised algorithms to tackle the FIP problem: the *eXplainable Missing items considering Time* (`XMT`), and *TARS-based* `XMT` (`TXMT`). For both algorithms, we face the FIP problem exploiting NBP as follows.

Definition 4 (Forgotten-Item Prediction via Next-Basket Prediction)**.** *Let B_t be the basket purchased at current time t, B_{t+h} the forgotten-items basket (unknown at time t), and $\hat{B}_t \subset I$ the prediction for the basket purchased at time t given by an NBP algorithm g. We solve FIP via NBP by approximating the prediction of the forgotten items B_{t+h} as the set of items predicted at time t but not purchased $\hat{F}_t = \hat{B}_t - B_t$.*

The objective of FIP via NBP is to define a function g that outperforms the user in remembering which items should be purchased, i.e., achieving higher recall than the user. Consequently, let $\hat{B}_t \cap B_t$ denote the set of items both purchased and correctly predicted to be purchased, $\hat{F}_t$ represent the set of items *potentially forgotten* by the user. Therefore, a core challenge in FIP via NBP is distinguishing genuinely forgotten items from intentionally omitted ones while maintaining strong predictive precision. In this setting, any NBP method can, in principle, be applied to address the FIP problem. However, in the experimental section, we demonstrate that standard NBP approaches, when evaluated by comparing B_{t+h} against $\hat{F}_t$ in cases where $B_{t+h} \neq \emptyset$, i.e., some items were indeed forgotten, are less effective than the two tailored methods we introduce, which are specifically designed to address FIP, while remaining within the NBP formulation. Finally, we highlight a latent yet critical issue that must be addressed alongside FIP: identifying real situations in which FIP is needed. Due to the highly imbalanced nature of real-world data, where users typically do not forget items, we leave this aspect for future research. In this study, we focus exclusively on scenarios where the next basket is known to be a forgotten-items basket. However, a simple yet practical strategy

Algorithm 1 XMT

Require: Customer purchase history H_u, current time t, minimum appearances ϵ, maximum interval multiplier γ, discontinuation threshold G, temporal score boost α, seasonality threshold υ, seasonal boost factor β, co-purchase threshold Υ, co-purchase boost Θ, repurchase window ν, repurchase boost Λ, repurchase threshold Φ, prediction length k.

Ensure: $\hat{F}_t$

1: $f \leftarrow$ calculateBaseFrequency(H_u, ϵ) ▷ Eq. 2
2: $\tau \leftarrow$ calculateTemporalProximity$(H_u, f, t, \gamma, G, \alpha)$ ▷ Eq. 3
3: $\sigma \leftarrow$ calculateSeasonalContext$(H_u, \tau, \upsilon, \beta)$ ▷ Eq. 4
4: $\hat{B}_t \leftarrow \{\text{top } |B_t| + k \text{ items according to } \sigma\}$
5: $\kappa \leftarrow$ calculateCoPurchase$(H_u, B_t, \Upsilon, \Theta)$ ▷ Eq. 5
6: $\psi \leftarrow$ calculateRepurchasePattern$(H_u, \theta, \nu, \Lambda, \Phi, O)$ ▷ Eq. 6
7: *MAP* $\leftarrow$ calculateMAPScore(σ, κ, ψ) ▷ Eq. 1
8: $\hat{F}_t \leftarrow \{\text{top } k \text{ items of } (\hat{B}_t \setminus B_t) \text{ according to } \textit{MAP}\}$
9: **return** $\hat{F}_t$

is to consistently present to customers at purchase time a small set $\hat{F}_t$ of k potentially forgotten items. Indeed, in the worst-case scenario, non-forgotten items will simply be ignored, incurring no harm.

5.1 XMT

The *eXplainable Missing items considering Time* (XMT) method, detailed in Algorithm 1, introduces a multi-factor scoring approach that combines historical purchase patterns, temporal relationships, and contextual evidence to identify potentially forgotten items. The forgotten item prediction process unfolds in two phases. In the first phase (lines 1-3 in Algorithm 1), XMT constructs the predicted basket $\hat{B}_t$ by selecting the top items ranked by the temporal and seasonal purchase likelihood, σ_i. This predicted basket $\hat{B}_t$ (line 4 in Algorithm 1) represents items that the customer is expected to purchase at time t, based purely on historical frequency, recency, and seasonality.

In the second phase (lines 5-7), the algorithm computes the Multi-factor Adjusted Prediction (MAP) score MAP_i for each item $i \in \hat{B}_t$:

$$MAP_i = \sigma_i + \kappa_i + \psi_i \tag{1}$$

which integrates three components: the previously computed seasonally adjusted purchase likelihood σ_i, the basket context affinity κ_i (line 5), and the post-basket repurchase tendency ψ_i (line 6). The final forgotten item predictions are obtained by selecting the top k items from $\hat{F}_t$ ranked by their MAP scores (line 8). This two-phase process ensures that the predictions are first guided by long-term behavioral signals and then refined by session-specific contextual information. In the following, we detail how to compute each component.

Temporal and Seasonal Purchase Likelihood. The first term σ_i, captures long-term purchasing patterns, short-term recency effects, and seasonal regularity. The relative purchase frequency is:

$$f_i = \frac{|H_u^{(i)}|}{|H_u|} \cdot \mathbb{1}\left[|H_u^{(i)}| \geq \epsilon\right], \tag{2}$$

where H_u is the customer's purchase history, $H_u^{(i)} = \{B_j \in H_u \mid i \in B_j\}$ is the set of baskets in the history containing item i, $\mathbb{1}$ is the indicator function that uses the threshold ϵ to filter out low-frequency items (Line 1). We treated ϵ as a hyperparameter, and we provide more details in Appendix B of Supplementary Material (SM) [2]. As shown by the effectiveness of the TOP method in [5], the relative purchase frequency provides an important estimate for an item's inclusion in any basket. Items with higher purchase frequencies are more likely to be regular purchases and thus more likely to be forgotten if missing from a large basket. Then, inspired by IBP [19], XMT incorporates temporal patterns into a scoring component τ_i that boosts items that are "due" for purchase based on their typical purchase cycle (line 2):

$$\tau_i = f_i \cdot (1 + (\alpha - 1) \cdot \mathbb{1}\left[\eta_i \leq \phi_i \leq \eta_i \cdot \gamma \wedge \phi \leq G\right]) \tag{3}$$

where ϕ_i is the number of days since the last purchase of item i, η_i is the median inter-purchase interval, and α, γ, and G are hyperparameters controlling the strength and eligibility of the temporal adjustment; more details on how we chose them are provided in Appendix B of the SM [2]. The temporal proximity computation scales the frequency score f_i by a factor α, capturing an increased likelihood of imminent repurchase. This boost prioritizes items that align with the customer's established purchase rhythm. On the contrary, the algorithm excludes items that have not been purchased for an extended period to avoid suggesting discontinued purchases. After that, XMT computes a seasonal adjustment σ_i that identifies and boosts items with strong seasonal tendencies (Line 3) computed as follows:

$$\sigma_i = \tau_i \cdot (1 + (\beta - 1) \cdot \mathbb{1}\left[\rho_i > \upsilon\right]) \tag{4}$$

where ρ_i is the fraction of annual purchases of item i that occur in the current season, and υ and β are hyperparameters defining the seasonality threshold and seasonal boost, respectively. More details about how we selected them are reported in Appendix B of the SM [2].

Basket Context Affinity. The second term κ_i measures the contextual relevance of item i given the content of the current basket B_t. This is computed by summing over all items $j \in B_t$ that have a significant history of co-purchase with i:

$$\kappa_i = \sum_{j \in B_t} \Theta \cdot \mathbb{1}\left[\chi(i,j) > \Upsilon\right] \tag{5}$$

where $\chi(i,j)$ is the number of past baskets in which items i and j co-occurred, and Υ and Θ control the minimum required co-occurrence frequency and the score boost per co-purchase, respectively (details about how we selected them are reported in Appendix B of the SM [2]). This component reflects associative strength and ensures that items frequently bought together are considered in tandem.

Post-Basket Repurchase Tendency. The final component ψ_i accounts for items that are commonly repurchased soon after large baskets, indicating a short-term replenishment behavior.

$$\psi_i = \Lambda \cdot \mathbb{1}\left[\frac{\sum_{h \leq \nu} \zeta_i(h)}{\lambda_i} > \Phi\right] \tag{6}$$

where $\zeta_i(h)$ is the number of times item i was repurchased within h days after the most recent basket, λ_i is the number of such opportunities, ν is the repurchase window, and Φ and Λ are the repurchase rate threshold and corresponding boost. More details on how we selected them are reported in Appendix B of the SM [2]. This term ensures that high-velocity repurchase patterns influence the final score even if the item's regular frequency is moderate.

5.2 TXMT

Building upon XMT, the *TARS-based eXplainable Missing items considering Time* (TXMT) extends its capabilities by integrating temporal patterns extracted through TARS [5]. The incorporation of TARS is crucial for capturing complex temporal purchasing behaviors, thereby enhancing the accuracy of the algorithm's predictions. However, this improvement comes at the cost of increased computational complexity due to the overhead introduced by the TARS extraction process.

To effectively integrate the TARS scores into XMT, TXMT modifies the MAP score by introducing an additional scoring component. First, the TARS scores Ω_i are computed using the customer's purchase history H_u as described in [5]. Subsequently, the TARS scores Ω_i are normalized by the maximum TARS score. These normalized scores are then added to the MAP scores creating the TARS-based MAP score $TMAP_i$ for item i as follows:

$$TMAP_i = MAP_i + \Omega_i \tag{7}$$

Similarly to XMT, given the $TMAP_i$ values, TXMT returns the top k items ranked by their TMAP scores.

5.3 Interpretability of XMT and TXMT

As discussed in Section 3, the increasing availability of black-box models has improved performance but often at the cost of interpretability, thereby diminishing user trust and limiting practical adoption. In contrast, both XMT and TXTM are interpretable-by-design, owing to the proposed MAP and $TMAP$ scoring algorithms, which rely solely on the combination of distinct scoring components. Unlike black-box approaches, each score type in our methods offers clear and actionable insights that can be directly communicated to users.

Indeed, through XMT, the different scores capture immediate contextual evidence and temporal relationships. A high base frequency f_i indicates the item is a regular part of the customer's purchase history. A temporal score boost from the Temporal Proximity Component $\tau_i > f_i$ suggests the item is "due" for repurchase based on their typical buying cycle. A boost from the Seasonal Context Component ($\sigma_i > \tau_i$) points to a higher likelihood of purchase due to the current season. A positive Co-Purchase score ($\kappa_i > 0$ indicates the item is frequently bought together with items currently in the basket. Finally, a positive Repurchase Pattern score ($\psi_i > 0$) suggests the item is often re-bought shortly after large shopping trips. The additive nature of the Multi-factor Adjusted Prediction (MAP) score ($MAP_i = \sigma_i + \kappa_i + \psi_i$) allows for a decomposition of the factors influencing the prediction for each item, enhancing the transparency of the model's reasoning. By presenting the contribution of each component, XMT offers interpretable predictions grounded in distinct aspects of the customer's purchasing behavior.

For TXMT, this combination used to compute $TMAP_i$ maintains the interpretability of both XMT and TARS while allowing each type of evidence to contribute independently to the final prediction. The TARS component contributes interpretability by highlighting items that are part of learned recurring purchase sequences, indicating established temporal dependencies in the customer's behavior.

6 Experiments

We outline here the experimental setup and present a selection of the obtained results. We begin by describing the dataset and the preprocessing steps used to identify forgotten-items baskets. Next, we define the evaluation metrics employed to assess model performance. We then introduce the baseline models used for comparison with XMT and TXMT. Finally, we report the results and provide a qualitative example illustrating the explainability of our proposed methods.

6.1 Experimental Setting

To evaluate XMT and TXMT, we relied on a proprietary dataset collected by an Italian supermarket chain. Table 2 provides a detailed overview of this dataset, which we call the Italian Supermarket dataset. It includes 10,867,976 individual transactions by 17,132 customers from 2007 to 2016, with a median basket size of 9.52 (Pre-processing details are in Appendix A of SM [2]). The products that are presented in this dataset are uniquely identified by an ID, and we mapped these IDS to a general category or product (524 unique items). Widely used supermarket datasets, such as the Ta Feng[2] dataset, are typically used to evaluate algorithms on the NBP or product assortment decisions task. However, these datasets are not well-suited for the FIP task, which requires a large amount of expenses and users to effectively identify and label forgotten expenses in a preprocessing phase. With over 10 million transactions, the Italian Supermarket dataset is a perfect candidate for evaluating XMT and TXMT on the FIP task. Moreover, it has been used in prior work [8], and a synthetic version is publicly available.[3]

Table 2. Dataset Statistics with different parameters to define forgotten-items baskets.

Customers	Unique Items	Baskets	Med. Hist. Len	Med. Days	Med. Basket Size
17,132	524	10.87M	452	3.50	9.52

θ_t	θ_{t+h}	h	# Forgotten Baskets	% Forgotten Basket
10	10	0	8,495	0.08%
		1	492,044	4.53%
		2	1,050,499	9.67%
10	1	0	49,314	0.45%
		1	1,207,712	11.11%
		2	2,043,673	18.80%

Given the absence of explicit forgotten purchase annotations in Italian Supermarket, i.e., we have no real ground truth for forgotten-items baskets B_{t+h}, it is necessary to define a set of criteria to identify baskets that likely represent forgotten items. In particular, we mark a basket as forgotten if:

- The forgotten-items basket B_{t+h} must be purchased after a *large basket* B_t containing at least θ_t items, i.e., $|B_t| > \theta_t$. In our experiments, we selected $\theta_t = 10$ to avoid small daily purchases, such as those containing only bread or milk, from being mistakenly flagged as forgotten expenses.
- The forgotten-items basket B_{t+h} must be purchased after a maximum of h DAYS after a large basket. We report the results of the experiments obtained with $0 \leq h \leq 2$, i.e., purchases made between the same day of B_t, i.e., $h = 0$, and two days after B_t, i.e., $h = 2$. The choice of these values is motivated by the observation that the most likely forgotten-items baskets tend to occur shortly after the *large basket* purchase while extending the horizon further would risk capturing unrelated expenses.
- The forgotten-items basket B_{t+h} cannot more than θ_{t+h} items. We consider experiments $\theta_{t+h} = 1$ and $\theta_{t+h} = 10$. However, to better capture scenarios where customers, the main results reported in the paper are those with $\theta_{t+h} = 10$. Indeed, upon returning to the store, a customer can take advantage of the trip to purchase additional items beyond just the forgotten ones.

Alongside the dataset details, Table 2 presents a summary of the forgotten items identified in the dataset following the filtering process. Specifically, we compare two different value for the parameter θ_{t+h}, i.e., 1 and 10, and three different values for h, i.e., 0, 1, and 2, to show how they impact the amount of baskets labeled as forgotten. As shown in Table 2, increasing h, leads to a higher number of forgotten-items baskets. In contrast, increasing the θ_{t+h} parameter results in fewer baskets labeled as forgotten-items baskets.

It is important to highlight that thanks to our user-centric approach,

[2] https://www.kaggle.com/datasets/chiranjivdas09/ta-feng-grocery-dataset
[3] Synthetic dataset available after free registration at http://bit.ly/4lLTVoE.

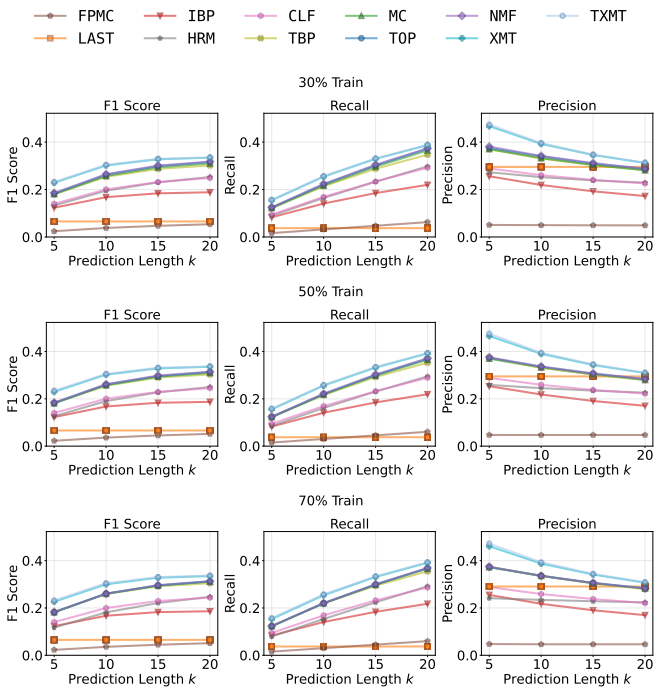


Figure 2. Comparison of XMT and TXMT against baselines and competitor using *30%*, *50%*, *70%* Train with $h = 2$. Prediction length k from 5 to 20.

each user can choose hyperparameters that best suit their needs for more tailored predictions. For example, the size θ_t of the *large basket* preceding a forgotten-item basket B_{t+h} can be smaller for users with typically small baskets and larger for those with bigger ones.

To evaluate the performance of the proposed methods, we implemented a *split-based* model validation strategy. For each customer $u \in U$, we partition their shopping history $H_u = \{B_1, B_2, \ldots, B_m\}$ into training and testing sets using different percentage splits. For instance, with the *30% Train* criteria, the model learns from the first 30% of a customer's transaction history. Then the models are evaluated to predict the first forgotten-items basket present in the test set (if at least one is present). The evaluation metrics reported in the paper are the average and the standard deviation of all the predictions computed on the test set. This lets us evaluate how the amount of historical data affects model performance. To assess the quality of a prediction of a model to solve the FIP problem, we employ the following evaluation metrics where $\hat{F}_t$ denotes the set of predicted forgotten items at time t, while B_{t+h} represents the set of items actually forgotten at time t and subsequently purchased at time $t + h$.

- *Precision*: the ratio of correctly predicted items to the total number of predicted items: $Precision = \frac{|\hat{F}_t \cap B_{t+h}|}{|\hat{F}_t|}$ where
- *Recall*: the ratio of correctly predicted items to the total number of actual items: $Recall = \frac{|\hat{F}_t \cap B_{t+h}|}{|B_{t+h}|}$.
- *F1-Score*: the harmonic mean of Precision and Recall: $F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2|\hat{F}_t \cap B_{t+h}|}{|\hat{F}_t| + |B_{t+h}|}$.

In the experiments, we compare XMT and TXMT, with eight solutions: four baseline methods introduced in [3] and four state-of-the-art approaches LAST [3]: predicts the same products contained in the customer's previous basket, TOP [3]: predicts the top most frequent items with respect to their appearance in the customer's history, MC [3]: makes the prediction based on the last customer purchase and on a Markov chain calculated on the customer's history, CLF [3]: uses a binary classifier based on temporal purchase history features to predict if an item will be in the next basket, Non-negative Matrix Factorization (NMF) [12]: uses non-negative matrix factorization on the customer-item matrix, Factorizing personalized Markov Chain (FPMC) [16]: combines personalized Markov chains with matrix factorization, Hierarchical Representation Model (HRM) [21]: combines personalized Markov chains with matrix factorization, and Interval Based Predictor (IBP) [19]. We remind the reader that, among the various baselines, all, except for IBP, are designed for NBP rather than FIP. Therefore, IBP serves as the only true competitor.

6.2 *Experimental Results*

This section reports results obtained using XMT and TXMT, and all baselines and competitors, on the Italian Supermarket dataset, with $h = 2$. This setting considers an item as forgotten if not purchased within two days of the previous expense. We evaluate performance using three train-test splits: *30% Train*, *50% Train*, and *70% Train* to analyze the training size impact on prediction performance. Due to space constraints, we report in Appendix C of the SM [2] further results with similar findings, including those with other split percentage values and with $h = 0$ and $h = 1$.

Figure 2 reports the *Recall*, *Precision* and *F1-score* obtained by the various methods while varying the prediction lengths k from 5 to 20 in the *30% Train*, *50% Train* and *70% Train* settings, respectively. All the metrics are averaged over the different customers on the test set with at least a forgotten-items basket. We can notice that XMT and TXMT always outperform all the competitors. In particular, TOP, TBP, NMF, and MC demonstrate relatively strong performance, while FPMC, LAST, and HRM perform poorly across all metrics, showing limited prediction power in solving the FIP problem. Notably, both XMT and TXMT outperform IBP [19], which is the state-of-the-art method for the FIP problem. From these results, we observe that *Recall* consistently increases, while *Precision* decreases. This behavior is expected: as prediction length k increases, the likelihood of capturing more correct items rises, but so does the risk of including incorrect ones.

Table 3 presents a detailed summary of the F1-scores obtained by each method for a subset of prediction length k equals to 5, 10, 15 and 20. As the table shows, XMT and TXMT demonstrate a substantial performance increase, outperforming all other competitor methods by approximately 5% to 85% in terms of F1-score across the different settings of prediction lengths and data splits. Notably, increasing the amount of training data does not consistently correspond to an improvement in predictive performance across all models. This lack of gain is particularly evident when examining Table 3, where all the models, including TXMT and XMT, show stable F1-score as the training split increases from 30% to 70% for a given prediction length k. This finding is particularly relevant for the design of personalized recommender systems, as having a method that achieves strong performance with a relatively small amount of training data is crucial for allowing the training of user-specific models that preserve privacy without sacrificing utility. Moreover, Table 3 reveals a trend of increasing performance measured with F1-score as the prediction length k increases. In particular, for the proposed methods XMT and TXMT, we can notice an increase in the F1 Score of 47% when transitioning from $k = 5$ to $k = 20$, showing how leveraging a longer prediction length could lead to more accurate predictions.

To compare XMT with TXMT and consider their real-world application, we present the training times (in seconds) for the evaluated methods in Table 4. We excluded the baseline methods TOP and

Table 3. F1-score for different prediction lengths k and data splits. For each row, the best performer is in bold, and the best performer runner-up is underlined.

k	split	CLF	FPMC	HRM	IBP	LAST	MC	NMF	TBP	TOP	TXMT	XMT
5	30	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
5	50	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
5	70	0.14 ± 0.10	0.02 ± 0.05	0.12 ± 0.09	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
10	30	0.20 ± 0.11	0.04 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	$\mathbf{0.30 \pm 0.13}$	$\mathbf{0.30 \pm 0.13}$
10	50	0.20 ± 0.11	0.04 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.11	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.31 \pm 0.13}$	0.30 ± 0.13
10	70	0.20 ± 0.11	0.04 ± 0.05	0.18 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.30 \pm 0.13}$	$\mathbf{0.30 \pm 0.13}$
15	30	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.13}$	$\mathbf{0.33 \pm 0.13}$
15	50	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.10	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.12}$	$\mathbf{0.33 \pm 0.12}$
15	70	0.23 ± 0.10	0.04 ± 0.05	0.22 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.12}$	$\mathbf{0.33 \pm 0.12}$
20	30	0.25 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.32 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	$\mathbf{0.34 \pm 0.12}$	0.33 ± 0.12
20	50	0.25 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.31 ± 0.11	$\mathbf{0.34 \pm 0.12}$	$\mathbf{0.34 \pm 0.12}$
20	70	0.24 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.31 ± 0.11	$\mathbf{0.34 \pm 0.12}$	0.33 ± 0.12

Table 4. Average Training Time (in seconds) for $h = 2$. For each row, the best performer is in bold, and the best performer runner-up is underlined.

split	CLF	FPMC	HRM	IBP	MC	NMF	TBP	TXMT	XMT
30	85.0 ± 12.5	2.4 ± 0.3	1595.3 ± 241.5	$\mathbf{1.3 \pm 0.1}$	1.5 ± 0.2	29.1 ± 26.9	4.3 ± 4.1	1524.2 ± 932.7	4.8 ± 0.3
50	147.9 ± 14.6	2.9 ± 0.6	2208.5 ± 203.3	$\mathbf{1.5 \pm 0.1}$	2.4 ± 0.8	42.0 ± 22.4	5.6 ± 5.8	2511.8 ± 1447.5	5.9 ± 0.4
70	188.8 ± 22.4	3.4 ± 0.5	2434.3 ± 224.0	$\mathbf{1.8 \pm 0.3}$	3.1 ± 0.3	47.5 ± 15.9	7.6 ± 8.3	3486.6 ± 1851.3	6.7 ± 0.5

Table 5. Example of explanations for the forgotten item *vegetables*.

Current Basket $B_t = \{rice, snacks, prepared\ food, bread, tomatoes, hazelnut\ fruit, wine, olive\ oil, cold\ cuts\}$
Forgotten Items $\hat{F}_t = \{drinks, dried\ fruit, rabbit, yogurt, vegetables\}$
Forgotten Item Analyzed: *vegetables*

XMT	Last purchased	3 days ago	Typically bought every 2.0 days
	Often bought with	bread, cold cuts, wine	Based on co-occurrence
	Often repurchased soon	71.3%	Of past opportunities
TXMT	TARS pattern likelihood	50.5%	Confidence in pattern match

LAST, which output the most frequent and most recent purchases, respectively. Regarding training time efficiency, IBP consistently outperforms all other methods. However, our proposed approach offers higher performance compared with IBP. This advantage is particularly noticeable when comparing IBP with XMT, which achieves higher performance with comparable training duration. Compared with XMT, TXMT requires significantly more computational resources, with training times up to 912x longer. Incorporating TARS patterns in TXMT enables more nuanced and explainable predictions, potentially justifying the longer training times through enhanced customer experience and potential revenue benefits. Conversely, XMT offers a competitive option for scenarios prioritizing rapid deployment, substantially reducing computation time for a minor trade-off in performance.

6.3 *Interpretability of XMT and TXMT*

Both XMT and TXMT provide interpretable justifications for predicted forgotten items, yet they differ in the type and richness of the information they offer. To illustrate this, we examine a FIP scenario with a prediction length of $k = 5$, in which the items *drinks*, *dried fruit*, *rabbit*, *yogurt*, and *vegetables* were predicted as forgotten, i.e., $\hat{F}_r = \{drinks, dried fruit, rabbit, yogurt, vegetables\}$ given the following current basket $B_t = \{rice, snacks, prepared\ food, bread, tomatoes, hazelnut\ fruit, wine, olive oil, cold\ cuts\}$.

Table 5 compares the explanations provided by XMT and TXMT for the item *vegetables*. The explanation generated by XMT consists of three elements: the recency of the last purchase of *vegetables* (3 days ago), its typical purchase frequency (every 2.0 days), and its co-occurrence with items such as *bread, cold cuts,* and *wine*. Additionally, XMT reports that *vegetables* are frequently repurchased shortly after being bought, occurring in 71.3% of relevant historical contexts. This explanation reflects aggregated behavioral patterns grounded in time and co-occurrence statistics. As another example of co-occurrence useful in FIP, we can consider the *dried fruit* item that is predicted as forgotten and a reason for suggesting it is the purchase of *hazelnut fruit* at time t. TXMT complements the aforementioned justifications with an additional layer, i.e., the likelihood of *vegetables* being needed again based on recurring temporal purchase patterns. Specifically, it reports a 50.5% confidence that *vegetables* will reoccur according to past behavior captured by mined sequences. This TARS-derived signal goes beyond surface-level frequency and highlights that the current context strongly aligns with a pattern where *vegetables* have consistently followed previous purchases. Thus, TARS likelihood provides a temporally structured rationale that improves interpretability by linking forgotten item to recurring behavioral sequences instead of isolated co-occurrence or recency metrics alone. Additional item explanations are in Appendix C of the SM [2], while Appendix D shows a dashboard visualization for user explanations.

7 Conclusion

In this paper, we introduced XMT and TXMT for tackling the Forgotten Item Prediction (FIP) problem. Evaluated on a complex real-world dataset, our models outperform nine existing methods, including those designed for both NBP and FIP tasks. Our user-centric approach lets individuals use their shopping history to train models that suggest forgotten items, ensuring highly personalized recommendations while preserving privacy, as no data is shared externally. A key advantage of XMT and TXMT is their interpretability, as they provide clear explanations for why specific items are predicted as forgotten, enhancing user trust and offering actionable insights. A limitation of our work is the use of a single dataset, due to scarce public data suited for FIP. Future work includes incorporating collaborative signals from similar users', helpful in cold-start scenarios. To reduce the privacy risks, we plan to explore privacy-preserving methods such as Federated Learning [13], enabling decentralized model training without centralizing user data.

Acknowledgements

This work has been partially supported by the Italian Project Fondo Italiano per la Scienza FIS00001966 "MIMOSA", by the PRIN 2022 framework project "PIANO" (Personalized Interventions Against Online Toxicity) under CUP B53D23013290006, by the European Community Horizon 2020 programme under the funding schemes ERC-2018-ADG G.A. 834756 "XAI", by the European Commission under the NextGeneration EU programme – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) Project: "SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics" – Prot. IR0000013 – Av. n. 3264 del 28/12/2021, M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR" - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", "FINDHR" that has received funding from the European Union's Horizon Europe research and innovation program under G.A. 101070212 and "TANGO" under G.A. 101120763. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

References

[1] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo. Benchmarking and survey of explanation methods for black box models. *Data Min. Knowl. Discov.*, 37(5):1719–1778, 2023.

[2] L. Corbucci, J. A. B. Legrottaglie, F. Spinnato, A. Monreale, and R. Guidotti. An interpretable data-driven unsupervised approach for the prevention of forgotten items, 2025. URL https://arxiv.org/abs/xxxx.xxxxx.

[3] C. M. Cumby, A. E. Fano, R. Ghani, and M. Krema. Predicting customer shopping lists from point-of-sale purchase data. In *KDD*, pages 402–409. ACM, 2004.

[4] R. Guidotti, A. Monreale, M. Nanni, F. Giannotti, and D. Pedreschi. Clustering individual transactional data for masses of users. In *KDD*, pages 195–204. ACM, 2017.

[5] R. Guidotti, G. Rossetti, L. Pappalardo, F. Giannotti, and D. Pedreschi. Market basket prediction using user-centric temporal annotated recurring sequences. In *ICDM*, pages 895–900. IEEE Computer Society, 2017.

[6] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5):93:1–93:42, 2019.

[7] R. Guidotti, G. Rossetti, L. Pappalardo, F. Giannotti, and D. Pedreschi. Personalized market basket prediction with temporal annotated recurring sequences. *IEEE Trans. Knowl. Data Eng.*, 31(11):2151–2163, 2019.

[8] R. Guidotti, M. Nanni, F. Giannotti, D. Pedreschi, S. Bertoli, B. Speciale, and H. Rapoport. Measuring immigrants adoption of natives shopping consumption with machine learning. In Y. Dong, G. Ifrim, D. Mladenić, C. Saunders, and S. Van Hoecke, editors, *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*, pages 369–385, Cham, 2021. Springer International Publishing. ISBN 978-3-030-67670-4.

[9] L. Huang, H. Zou, X. Huang, Y. Gao, Y. Kuang, and C. Wang. Knowledge-reinforced explainable next basket recommendation. *Neural Networks*, 180:106675, 2024.

[10] H. Ko, S. Lee, Y. Park, and A. Choi. A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics*, 11(1):141, 2022.

[11] D. Le, H. W. Lauw, and Y. Fang. Correlation-sensitive next-basket recommendation. In *IJCAI*, pages 2808–2814. ijcai.org, 2019.

[12] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562. MIT Press, 2000.

[13] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In A. Singh and J. Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 20–22 Apr 2017.

[14] F. Mourão, C. Fonseca, C. S. Araujo, and W. M. Jr. The oblivion problem: Exploiting forgotten items to improve recommendation diversity. In *DiveRS@RecSys*, volume 816 of *CEUR Workshop Proceedings*, pages 27–34. CEUR-WS.org, 2011.

[15] V. Nguyen. Bayesian optimization for accelerating hyper-parameter tuning. In *AIKE*, pages 302–305. IEEE, 2019.

[16] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *WWW*, pages 811–820. ACM, 2010.

[17] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.*, 1(5):206–215, 2019.

[18] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Analysis of recommendation algorithms for e-commerce. In *EC*, pages 158–167. ACM, 2000.

[19] A. Singh and P. Hosein. On the prediction of possibly forgotten shopping basket items. In *Artificial Intelligence and Applied Mathematics in Engineering Problems: Proceedings of the International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2019)*, pages 559–571. Springer, 2020.

[20] R. R. Sinha and K. Swearingen. The role of transparency in recommender systems. In *CHI Extended Abstracts*, pages 830–831. ACM, 2002.

[21] P. Wang, J. Guo, Y. Lan, J. Xu, S. Wan, and X. Cheng. Learning hierarchical representation model for nextbasket recommendation. In *SIGIR*, pages 403–412. ACM, 2015.

[22] S. Wang, L. Hu, L. Cao, X. Huang, D. Lian, and W. Liu. Attention-based transactional context embedding for next-item recommendation. In *AAAI*, pages 2532–2539. AAAI Press, 2018.

[23] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan. A dynamic recurrent model for next basket recommendation. In *SIGIR*, pages 729–732. ACM, 2016.

[24] Y. Zhang and X. Chen. Explainable recommendation: A survey and new perspectives. *Found. Trends Inf. Retr.*, 14(1):1–101, 2020.

A Pre-Processing

We applied a series of filtering rules to identify valid customers for our experiments in our dataset. These rules must be satisfied before applying the splitting algorithm to each customer's purchase history.

- **Minimum number of baskets per customer**: 10;
- **Minimum basket size**: $|B| \geq 1$;
- **Maximum basket size**: $|B| \leq \infty$;
- **Minimum item occurrences**: $f_i \geq 1$, where f_i represents the frequency of item i;

The splitting algorithm works in two phases. First, it analyzes the transaction history of each customer; those that do not satisfy the criteria are discarded. Second, it applies the split type, e.g., 30% Train, to divide them into training and test sets.

After applying the splitting criteria, we retained 17,131 (99.99%) customers from the `Italian Supermarket` datasets.

B Hyperparameters

The current implementation of `XMT` and `TXMT` relies on a series of hyperparameters whose values were chosen based on heuristic reasoning rather than systematic optimization. For the base frequency component, we set the minimum appearances threshold ϵ to 5, reasoning that fewer than five occurrences would provide insufficient data to establish reliable purchase patterns while being strict enough to filter out truly sporadic purchases.

In the temporal proximity component, we chose a temporal score boost α of 1.5 to provide a noticeable but not overwhelming increase in prediction likelihood. The maximum interval multiplier γ was set to 3.0, operating under the assumption that gaps in purchasing more than three times the usual interval likely indicate a change in buying pattern rather than forgetfulness. The 90-day discontinuation threshold G represents a conservative estimate of when an item should no longer be considered part of a customer's regular purchases.

For seasonal adjustments, we set the seasonality threshold υ to 0.4, considering that in a uniform distribution, each season would account for 25% of purchases, thus, 40% in a single season suggests meaningful seasonal preference. The seasonal boost factor β of 1.5 matches the temporal boost to maintain consistency in how we weigh different types of temporal evidence.

The co-purchase component uses a co-occurrence threshold Υ of 5 and a boost parameter Θ of 0.2, chosen to ensure that co-purchase evidence contributes meaningfully to predictions without dominating other factors. Similarly, the repurchase pattern component parameters ($\nu = 2$ days, $\Lambda = 0.5$, $\Phi = 0.3$, $O = 5$) were selected to identify patterns that seem intuitively likely to indicate forgotten items.

While these parameter choices have yielded promising results in our experiments, we acknowledge that they represent educated guesses rather than optimized values. Moreover, using fixed parameters across all customers may not be optimal given the highly personal nature of shopping behaviors. Future work should explore both systematic hyperparameter tuning through techniques such as grid search or Bayesian optimization [15], and the potential for dynamic, user-specific parameters that adapt to individual shopping patterns and purchase frequencies. Such personalization could increase performance, particularly for customers with unusual shopping cadences or those whose behaviour varies significantly from the norm.

C Other Results

In Section 6.2 we reported the results obtained using the `Italian Supermarket` with $h = 2$ and *30% Train*,*50% Train*, and *70% Train* i.e., using 30%, 50% and 70% of the dataset for training. In our experiments, we also tested other combinations of these parameters. In particular, we considered $h = 0$ and $h = 1$, i.e., the forgotten-item basket is purchased after a max of 0 or 1 day after the large basket. Moreover, we also considered different training dataset sizes ranging from 20% to 80%. Overall, the results obtained using these settings are coherent with the ones presented in Section 6.2. Figures 3, 4 and 5 presents the result with *20% Train* and $h = 0$, $h = 1$ and $h = 2$ respectively. Tables 6, 7 and 7 report respectively the precision, recall and F1 Score with *20% Train*. Even in this scenario, in which we use a small part of the dataset to train our model, `XMT` and `TXMT` offer performance comparable to the one we obtain when increasing the training dataset size. As highlighted in Section 6.2, this result is particularly important because it allows us to train models that offer good performance even with limited data, for example, when a new supermarket in the chain opens or when a new customer begins shopping and has no prior purchase history.

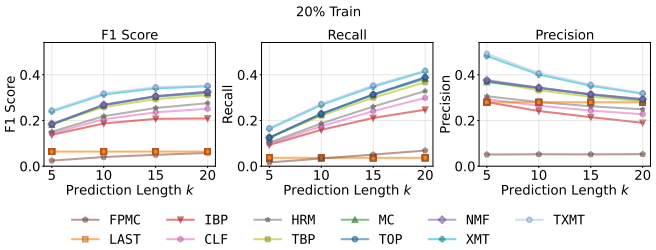


Figure 3. Comparison of `XMT` and `TXMT` with Baseline methods using a *20% Train* and h=0

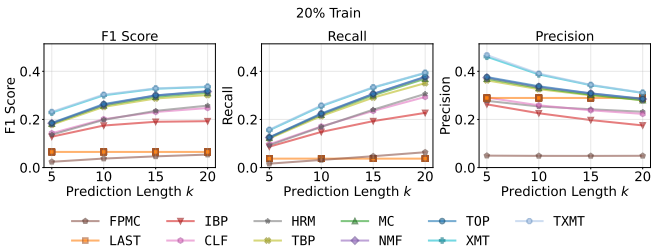


Figure 4. Comparison of `XMT` and `TXMT` with Baseline methods using a *20% Train* and h=1

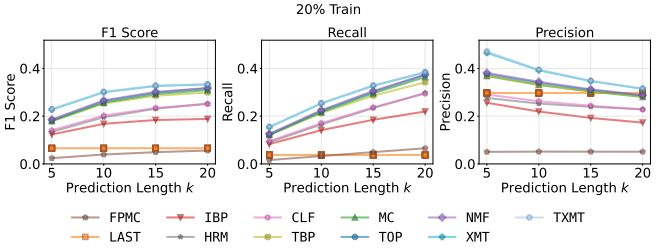


Figure 5. Comparison of `XMT` and `TXMT` with Baseline methods using a *20% Train* and h=2

Figures 6, and 7 presents the result with *30% Train* and $h = 0$ and $h = 1$ respectively. Tables 9, 10 and 10 report respectively the

Table 6. Precision comparison for different prediction lengths k and days h with *20% Train*

		20% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.29 ± 0.22	0.05 ± 0.10	0.31 ± 0.22	0.28 ± 0.22	0.28 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.49 ± 0.26**	0.48 ± 0.25
	1	0.29 ± 0.21	0.05 ± 0.10	0.28 ± 0.21	0.26 ± 0.22	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.36 ± 0.23	0.38 ± 0.23	**0.47 ± 0.25**	0.46 ± 0.25
	2	0.29 ± 0.21	0.05 ± 0.10	0.28 ± 0.	0.26 ± 0.21	0.30 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.37 ± 0.23	0.38 ± 0.23	**0.47 ± 0.25**	**0.47 ± 0.25**
10	0	0.26 ± 0.16	0.05 ± 0.07	0.28 ± 0.15	0.24 ± 0.16	0.28 ± 0.00	0.34 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.41 ± 0.19**	0.40 ± 0.19
	1	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.23 ± 0.16	0.29 ± 0.00	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.39 ± 0.19**	**0.39 ± 0.18**
	2	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.30 ± 0.00	0.33 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.40 ± 0.19**	0.39 ± 0.19
15	0	0.24 ± 0.13	0.05 ± 0.06	0.26 ± 0.13	0.21 ± 0.13	0.28 ± 0.00	0.31 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.36 ± 0.16**	0.35 ± 0.16
	1	0.24 ± 0.12	0.05 ± 0.06	0.24 ± 0.12	0.20 ± 0.13	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.34 ± 0.16**	**0.34 ± 0.15**
	2	0.24 ± 0.13	0.05 ± 0.06	0.24 ± 0.12	0.19 ± 0.12	0.30 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.35 ± 0.16**	**0.35 ± 0.16**
20	0	0.23 ± 0.11	0.05 ± 0.05	0.25 ± 0.11	0.19 ± 0.11	0.28 ± 0.00	0.29 ± 0.12	0.29 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	**0.32 ± 0.14**	**0.32 ± 0.14**
	1	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.11	0.18 ± 0.11	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	**0.31 ± 0.14**	**0.31 ± 0.14**
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.30 ± 0.00	0.28 ± 0.12	0.29 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	**0.32 ± 0.14**	0.31 ± 0.14

Table 7. Recall comparison for different prediction lengths (k) and days (d) with *20% Train*

		20% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.10 ± 0.07	0.02 ± 0.03	0.10 ± 0.08	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	**0.17 ± 0.09**	0.16 ± 0.09
	1	0.10 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.13 ± 0.08	**0.16 ± 0.09**	**0.16 ± 0.09**
	2	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.09
10	0	0.17 ± 0.10	0.03 ± 0.05	0.19 ± 0.10	0.16 ± 0.10	0.04 ± 0.00	0.23 ± 0.11	0.23 ± 0.11	0.22 ± 0.11	0.23 ± 0.11	**0.27 ± 0.12**	**0.27 ± 0.12**
	1	0.17 ± 0.10	0.03 ± 0.05	0.17 ± 0.09	0.15 ± 0.10	0.04 ± 0.00	0.22 ± 0.11	0.22 ± 0.10	0.21 ± 0.11	0.22 ± 0.11	**0.26 ± 0.12**	**0.26 ± 0.12**
	2	0.17 ± 0.09	0.03 ± 0.05	0.17 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.23 ± 0.11	0.21 ± 0.11	0.22 ± 0.11	**0.26 ± 0.12**	0.25 ± 0.12
15	0	0.24 ± 0.12	0.05 ± 0.06	0.26 ± 0.12	0.21 ± 0.12	0.04 ± 0.00	0.31 ± 0.13	0.31 ± 0.13	0.30 ± 0.13	0.31 ± 0.13	**0.35 ± 0.14**	**0.35 ± 0.14**
	1	0.24 ± 0.11	0.05 ± 0.06	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.31 ± 0.12	0.29 ± 0.13	0.31 ± 0.12	**0.33 ± 0.14**	**0.33 ± 0.14**
	2	0.24 ± 0.11	0.05 ± 0.06	0.23 ± 0.10	0.18 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.31 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	**0.33 ± 0.14**	**0.33 ± 0.14**
20	0	0.30 ± 0.13	0.07 ± 0.06	0.33 ± 0.13	0.25 ± 0.13	0.04 ± 0.00	0.38 ± 0.14	0.39 ± 0.14	0.37 ± 0.14	0.39 ± 0.14	**0.42 ± 0.15**	0.41 ± 0.15
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.38 ± 0.13	0.35 ± 0.14	0.38 ± 0.13	**0.39 ± 0.15**	**0.39 ± 0.15**
	2	0.29 ± 0.12	0.07 ± 0.06	0.30 ± 0.12	0.22 ± 0.12	0.04 ± 0.00	0.36 ± 0.13	0.37 ± 0.13	0.34 ± 0.14	0.37 ± 0.13	**0.38 ± 0.15**	**0.38 ± 0.15**

precision, recall and F1 Score with *30% Train*. This is one of the settings we considered in Section 6.2, here we offer a more in-depth view of the results, exploring how they change when we decrease the h. In particular, when $h = 0$, we can see a small increase in the performance of both XMT and TXMT. This is probably because using a low value for the h we better capture the forgotten baskets, and our methods can provide better results.

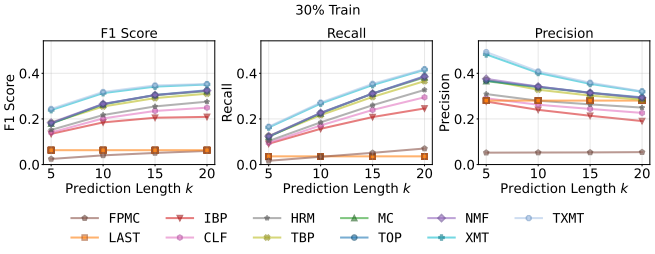


Figure 6. Comparison of XMT and TXMT with Baseline methods using a *30% Train* and h=0

Figures 8, 9and 10 presents the result with *40% Train* and $h = 0$, $h = 1$ and $h = 2$ respectively. Tables 12, 13 and 13 report respectively the precision, recall and F1 Score with *40% Train*. Increasing the size of the training dataset does not have an impact on the results, which are similar to the ones shown in previous tables and plots.

Figures 11, and 12 presents the result with *50% Train* and $h = 0$ and $h = 1$ respectively. Tables 15, 16 and 16 report respectively the precision, recall and F1 Score with *50% Train*. This is one of the

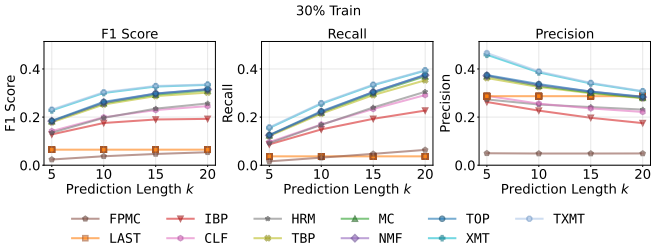


Figure 7. Comparison of XMT and TXMT with Baseline methods using a *30% Train* and h=1

settings we considered in Section 6.2, here we offer a more in-depth view of the results, exploring how they change when we decrease the h. In particular, when $h = 0$, we can see a small increase in the recall and F1 Score of both XMT and TXMT.

Figures 13, 14and 15 presents the result with *60% Train* and $h = 0$, $h = 1$ and $h = 2$ respectively. Tables 18, 19 and 19 report respectively the precision, recall and F1 Score with *60% Train*. Despite the increase of the training dataset, we notice only small improvements in terms of F1 Score, Recall, and Precision, which supports our thesis that our proposed methods offer good performance even when trained with a small amount of data.

Figures 16, and 17 presents the result with *30% Train* and $h = 0$ and $h = 1$ respectively. Tables 21, 22 and 22 report respectively the precision, recall and F1 Score with *20% Train*. This is one of the settings we considered in Section 6.2, here we offer a more in-depth

Table 8. F1 Score comparison for different prediction lengths (k) and days (d) with *20% Train*

		20% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.14 ± 0.10	0.02 ± 0.05	0.15 ± 0.11	0.14 ± 0.11	0.06 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.24 ± 0.13**	**0.24 ± 0.13**
	1	0.14 ± 0.10	0.02 ± 0.05	0.14 ± 0.10	0.13 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.19 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
	2	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
10	0	0.21 ± 0.11	0.04 ± 0.05	0.22 ± 0.11	0.19 ± 0.11	0.06 ± 0.00	0.27 ± 0.12	0.27 ± 0.12	0.26 ± 0.12	0.27 ± 0.12	**0.32 ± 0.14**	0.31 ± 0.13
	1	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.11	0.17 ± 0.11	0.06 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
	2	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.27 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
15	0	0.24 ± 0.11	0.05 ± 0.05	0.25 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.30 ± 0.12	0.31 ± 0.12	0.29 ± 0.12	0.31 ± 0.12	**0.35 ± 0.13**	0.34 ± 0.13
	1	0.23 ± 0.11	0.05 ± 0.05	0.24 ± 0.10	0.19 ± 0.11	0.06 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.13**	**0.33 ± 0.13**
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.13**	**0.33 ± 0.13**
20	0	0.25 ± 0.11	0.06 ± 0.05	0.28 ± 0.10	0.21 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.33 ± 0.11	0.31 ± 0.12	0.33 ± 0.11	**0.35 ± 0.12**	**0.35 ± 0.12**
	1	0.25 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.11	0.06 ± 0.00	0.31 ± 0.11	0.32 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	**0.34 ± 0.12**	**0.34 ± 0.12**
	2	0.25 ± 0.10	0.06 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.32 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**

Table 9. Precision comparison for different prediction lengths k and days h with *30% Train*

		30% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.29 ± 0.22	0.05 ± 0.10	0.31 ± 0.22	0.28 ± 0.22	0.28 ± 0.00	0.37 ± 0.22	0.38 ± 0.22	0.37 ± 0.23	0.37 ± 0.23	**0.49 ± 0.26**	0.48 ± 0.25
	1	0.29 ± 0.21	0.05 ± 0.10	0.27 ± 0.	0.26 ± 0.22	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.36 ± 0.23	0.37 ± 0.23	**0.47 ± 0.25**	0.46 ± 0.24
	2	0.29 ± 0.21	0.05 ± 0.10	0.27 ± 0.	0.26 ± 0.21	0.30 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.37 ± 0.23	0.38 ± 0.23	**0.47 ± 0.25**	**0.47 ± 0.25**
10	0	0.26 ± 0.16	0.05 ± 0.07	0.28 ± 0.15	0.24 ± 0.16	0.28 ± 0.00	0.34 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.41 ± 0.19**	0.40 ± 0.19
	1	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.23 ± 0.16	0.29 ± 0.00	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.39 ± 0.18**	**0.39 ± 0.18**
	2	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.30 ± 0.00	0.33 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.40 ± 0.18**	0.39 ± 0.18
15	0	0.24 ± 0.13	0.05 ± 0.06	0.26 ± 0.13	0.21 ± 0.13	0.28 ± 0.00	0.31 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.36 ± 0.16**	0.35 ± 0.16
	1	0.24 ± 0.12	0.05 ± 0.06	0.24 ± 0.12	0.20 ± 0.13	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.34 ± 0.15**	**0.34 ± 0.15**
	2	0.24 ± 0.12	0.05 ± 0.06	0.24 ± 0.12	0.19 ± 0.12	0.30 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.35 ± 0.15**	**0.35 ± 0.15**
20	0	0.23 ± 0.11	0.05 ± 0.05	0.25 ± 0.11	0.19 ± 0.11	0.28 ± 0.00	0.29 ± 0.12	0.29 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	**0.32 ± 0.13**	**0.32 ± 0.14**
	1	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.13	0.28 ± 0.12	**0.31 ± 0.13**	**0.31 ± 0.13**
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.30 ± 0.00	0.28 ± 0.12	0.29 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	**0.31 ± 0.14**	**0.31 ± 0.14**

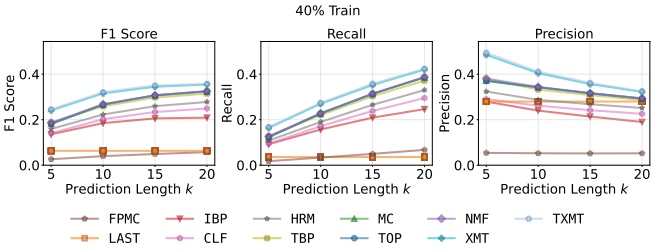


Figure 8. Comparison of XMT and TXMT with Baseline methods using a *40% Train* and h=0

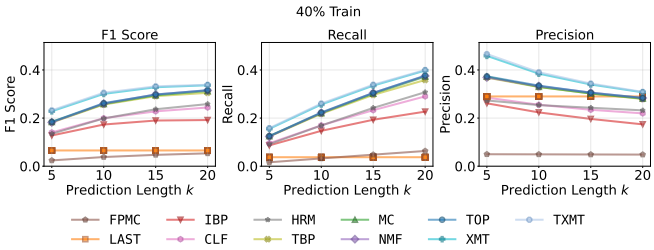


Figure 9. Comparison of XMT and TXMT with Baseline methods using a *40% Train* and h=1

view of the results, exploring how they change when we decrease the h.

Lastly, figures 18, 19 and 20 presents the result with *80% Train* and $h = 0$, $h = 1$ and $h = 2$ respectively. Tables 24, 25 and 25 report respectively the precision, recall and F1 Score with *80% Train*. This is the most extreme case that we considered in terms of training dataset size. However, the results are coherent and not dissimilar to the ones obtained when using smaller training datasets.

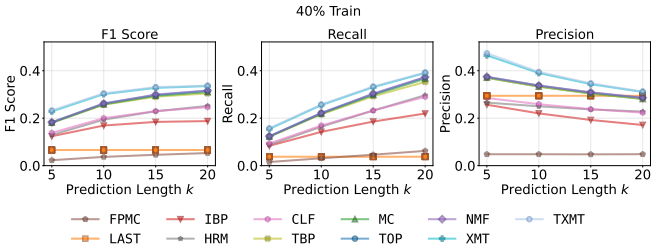


Figure 10. Comparison of XMT and TXMT with Baseline methods using a *40% Train* and h=2

Table 10. Recall comparison for different prediction lengths (k) and days (d) with *30% Train*

		30% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.10 ± 0.07	0.02 ± 0.03	0.10 ± 0.08	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.07	0.12 ± 0.08	0.12 ± 0.08	**0.17 ± 0.09**	0.16 ± 0.09
	1	0.10 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.13 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.09
	2	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.09
10	0	0.17 ± 0.10	0.03 ± 0.05	0.18 ± 0.10	0.16 ± 0.10	0.04 ± 0.00	0.22 ± 0.11	0.23 ± 0.10	0.22 ± 0.11	0.23 ± 0.11	**0.27 ± 0.12**	**0.27 ± 0.12**
	1	0.17 ± 0.09	0.03 ± 0.05	0.17 ± 0.09	0.15 ± 0.10	0.04 ± 0.00	0.22 ± 0.11	0.22 ± 0.10	0.21 ± 0.10	0.22 ± 0.11	**0.26 ± 0.12**	**0.26 ± 0.12**
	2	0.17 ± 0.09	0.03 ± 0.05	0.16 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.11	0.21 ± 0.10	0.22 ± 0.10	**0.26 ± 0.12**	0.25 ± 0.12
15	0	0.24 ± 0.11	0.05 ± 0.06	0.26 ± 0.11	0.21 ± 0.12	0.04 ± 0.00	0.31 ± 0.13	0.31 ± 0.12	0.30 ± 0.12	0.31 ± 0.12	**0.35 ± 0.14**	**0.35 ± 0.14**
	1	0.23 ± 0.11	0.05 ± 0.06	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	**0.34 ± 0.14**	0.33 ± 0.13
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	**0.33 ± 0.13**	**0.33 ± 0.13**
20	0	0.29 ± 0.12	0.07 ± 0.07	0.33 ± 0.12	0.25 ± 0.13	0.04 ± 0.00	0.38 ± 0.13	0.39 ± 0.13	0.37 ± 0.14	0.39 ± 0.13	**0.42 ± 0.15**	**0.42 ± 0.15**
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.35 ± 0.14	0.38 ± 0.13	**0.39 ± 0.15**	**0.39 ± 0.15**
	2	0.29 ± 0.12	0.06 ± 0.06	0.30 ± 0.12	0.22 ± 0.12	0.04 ± 0.00	0.36 ± 0.13	0.37 ± 0.13	0.35 ± 0.14	0.37 ± 0.13	**0.39 ± 0.15**	**0.39 ± 0.15**

Table 11. F1 Score comparison for different prediction lengths (k) and days (d) with *30% Train*

		30% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.14 ± 0.10	0.02 ± 0.05	0.15 ± 0.11	0.13 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.24 ± 0.13**	**0.24 ± 0.13**
	1	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.13 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
	2	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
10	0	0.20 ± 0.11	0.04 ± 0.05	0.22 ± 0.11	0.18 ± 0.11	0.06 ± 0.00	0.26 ± 0.12	0.27 ± 0.12	0.25 ± 0.12	0.27 ± 0.12	**0.32 ± 0.13**	0.31 ± 0.13
	1	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.10	0.18 ± 0.11	0.06 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
	2	0.20 ± 0.11	0.04 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
15	0	0.24 ± 0.11	0.05 ± 0.05	0.25 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.30 ± 0.12	0.30 ± 0.11	0.29 ± 0.12	0.31 ± 0.12	**0.35 ± 0.13**	0.34 ± 0.13
	1	0.23 ± 0.11	0.05 ± 0.05	0.24 ± 0.10	0.19 ± 0.11	0.06 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.13**	**0.33 ± 0.13**
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.13**	**0.33 ± 0.13**
20	0	0.25 ± 0.10	0.06 ± 0.05	0.28 ± 0.10	0.21 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.32 ± 0.11	0.31 ± 0.12	0.33 ± 0.11	**0.35 ± 0.12**	**0.35 ± 0.12**
	1	0.25 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.10	0.06 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	**0.34 ± 0.12**	0.33 ± 0.12
	2	0.25 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.32 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	**0.34 ± 0.12**	0.33 ± 0.12

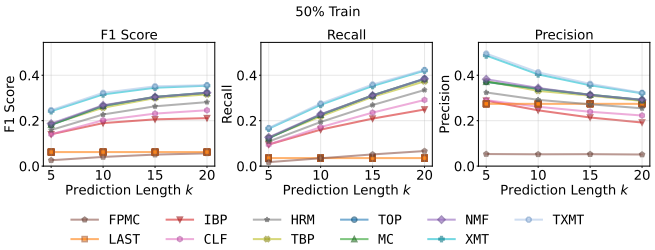


Figure 11. Comparison of XMT and TXMT with Baseline methods using a *50% Train* and h=0

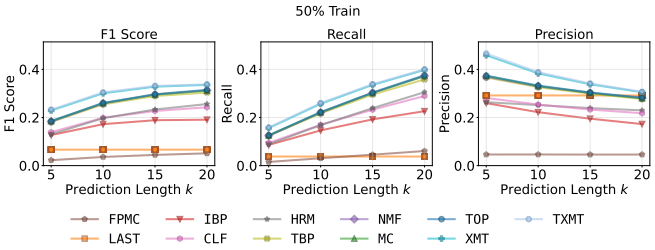


Figure 12. Comparison of XMT and TXMT with Baseline methods using a *50% Train* and h=1

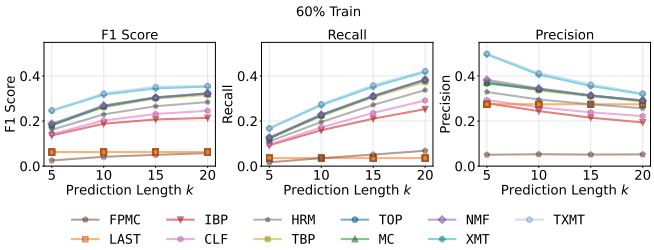


Figure 13. Comparison of XMT and TXMT with Baseline methods using a *60% Train* and h=0

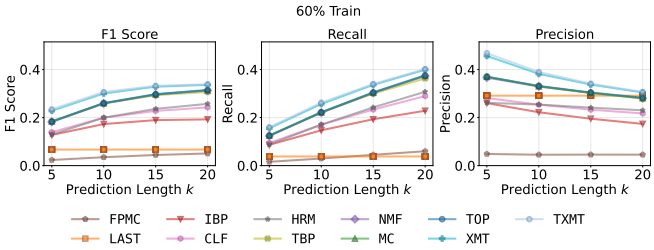


Figure 14. Comparison of XMT and TXMT with Baseline methods using a *60% Train* and h=1

Table 12. Precision comparison for different prediction lengths k and days h with *40% Train*

		40% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.29 ± 0.22	0.05 ± 0.10	0.32 ± 0.22	0.28 ± 0.21	0.28 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.38 ± 0.24	0.37 ± 0.23	**0.49 ± 0.25**	0.48 ± 0.25
	1	0.29 ± 0.21	0.05 ± 0.10	0.27 ± 0.	0.26 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.47 ± 0.24**	0.46 ± 0.24
	2	0.29 ± 0.21	0.05 ± 0.10	0.27 ± 0.	0.26 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.47 ± 0.25**	0.46 ± 0.24
10	0	0.26 ± 0.16	0.05 ± 0.07	0.29 ± 0.16	0.24 ± 0.16	0.28 ± 0.00	0.34 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.41 ± 0.19**	0.40 ± 0.18
	1	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	**0.39 ± 0.18**	0.38 ± 0.18
	2	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.39 ± 0.18**	**0.39 ± 0.18**
15	0	0.24 ± 0.13	0.05 ± 0.06	0.27 ± 0.13	0.21 ± 0.13	0.28 ± 0.00	0.32 ± 0.14	0.32 ± 0.14	0.31 ± 0.14	0.32 ± 0.14	**0.36 ± 0.15**	**0.36 ± 0.15**
	1	0.23 ± 0.12	0.05 ± 0.06	0.24 ± 0.12	0.20 ± 0.13	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	**0.34 ± 0.15**	**0.34 ± 0.15**
	2	0.24 ± 0.12	0.05 ± 0.06	0.24 ± 0.12	0.19 ± 0.12	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.13	**0.35 ± 0.15**	0.34 ± 0.15
20	0	0.23 ± 0.11	0.05 ± 0.05	0.25 ± 0.11	0.19 ± 0.11	0.28 ± 0.00	0.29 ± 0.12	0.29 ± 0.12	0.29 ± 0.12	0.29 ± 0.12	**0.32 ± 0.13**	**0.32 ± 0.13**
	1	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.13	0.28 ± 0.12	**0.31 ± 0.13**	**0.31 ± 0.13**
	2	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.29 ± 0.00	0.28 ± 0.12	0.29 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	**0.31 ± 0.13**	**0.31 ± 0.13**

Table 13. Recall comparison for different prediction lengths (k) and days (d) with *40% Train*

		40% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.09 ± 0.07	0.02 ± 0.03	0.11 ± 0.08	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	**0.17 ± 0.09**	0.16 ± 0.09
	1	0.10 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.13 ± 0.08	**0.16 ± 0.09**	**0.16 ± 0.09**
	2	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.08
10	0	0.17 ± 0.10	0.03 ± 0.05	0.19 ± 0.10	0.16 ± 0.10	0.04 ± 0.00	0.23 ± 0.10	0.23 ± 0.10	0.22 ± 0.11	0.23 ± 0.10	**0.27 ± 0.12**	**0.27 ± 0.12**
	1	0.17 ± 0.10	0.03 ± 0.05	0.17 ± 0.09	0.15 ± 0.10	0.04 ± 0.00	0.22 ± 0.11	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.11	**0.26 ± 0.12**	**0.26 ± 0.12**
	2	0.17 ± 0.09	0.03 ± 0.04	0.16 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	**0.26 ± 0.12**	0.25 ± 0.11
15	0	0.24 ± 0.11	0.05 ± 0.06	0.27 ± 0.11	0.21 ± 0.11	0.04 ± 0.00	0.31 ± 0.12	0.31 ± 0.12	0.30 ± 0.12	0.31 ± 0.12	**0.36 ± 0.13**	0.35 ± 0.13
	1	0.23 ± 0.11	0.05 ± 0.06	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	**0.34 ± 0.13**	0.33 ± 0.13
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	**0.33 ± 0.13**	**0.33 ± 0.13**
20	0	0.30 ± 0.12	0.07 ± 0.07	0.33 ± 0.12	0.25 ± 0.12	0.04 ± 0.00	0.38 ± 0.13	0.39 ± 0.13	0.37 ± 0.14	0.39 ± 0.13	**0.42 ± 0.15**	**0.42 ± 0.14**
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.38 ± 0.13	0.36 ± 0.14	0.38 ± 0.13	**0.40 ± 0.15**	**0.40 ± 0.15**
	2	0.29 ± 0.12	0.06 ± 0.06	0.30 ± 0.12	0.22 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.35 ± 0.14	0.37 ± 0.13	**0.39 ± 0.15**	**0.39 ± 0.15**

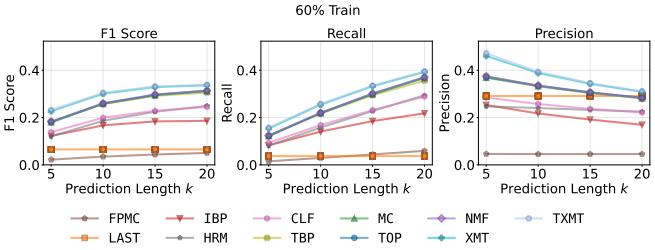


Figure 15. Comparison of XMT and TXMT with Baseline methods using a *60% Train* and h=2

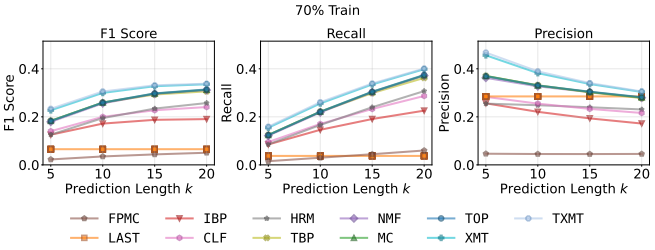


Figure 17. Comparison of XMT and TXMT with Baseline methods using a *70% Train* and h=1

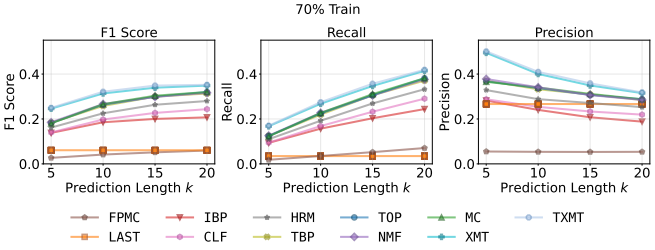


Figure 16. Comparison of XMT and TXMT with Baseline methods using a *70% Train* and h=0

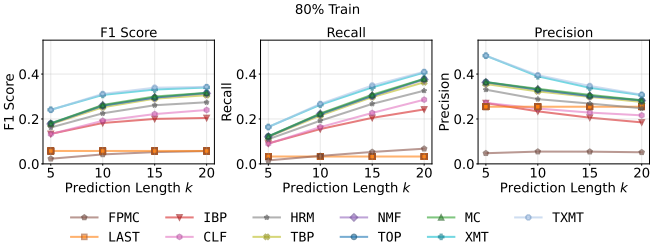


Figure 18. Comparison of XMT and TXMT with Baseline methods using a *80% Train* and h=0

Table 14. F1 Score comparison for different prediction lengths (k) and days (d) with *40% Train*

		40% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.14 ± 0.10	0.03 ± 0.05	0.16 ± 0.11	0.14 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.19 ± 0.12	0.18 ± 0.11	**0.25 ± 0.13**	0.24 ± 0.12
	1	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.13 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
	2	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
10	0	0.20 ± 0.11	0.04 ± 0.05	0.22 ± 0.11	0.18 ± 0.11	0.06 ± 0.00	0.27 ± 0.12	0.27 ± 0.12	0.26 ± 0.12	0.27 ± 0.12	**0.32 ± 0.13**	**0.32 ± 0.13**
	1	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.31 ± 0.13**	0.30 ± 0.13
	2	0.20 ± 0.11	0.04 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
15	0	0.23 ± 0.11	0.05 ± 0.05	0.26 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.31 ± 0.12	0.31 ± 0.11	0.30 ± 0.12	0.31 ± 0.12	**0.35 ± 0.13**	0.34 ± 0.13
	1	0.23 ± 0.10	0.05 ± 0.05	0.24 ± 0.10	0.19 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**
20	0	0.25 ± 0.10	0.06 ± 0.05	0.28 ± 0.10	0.21 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.33 ± 0.11	0.31 ± 0.11	0.33 ± 0.11	**0.36 ± 0.12**	0.35 ± 0.12
	1	0.24 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.32 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	**0.34 ± 0.12**	**0.34 ± 0.12**
	2	0.25 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.32 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	**0.34 ± 0.12**	**0.34 ± 0.12**

Table 15. Precision comparison for different prediction lengths k and days h with *50% Train*

		50% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.29 ± 0.22	0.05 ± 0.10	0.32 ± 0.22	0.29 ± 0.22	0.27 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.37 ± 0.24	0.37 ± 0.23	**0.50 ± 0.26**	0.49 ± 0.25
	1	0.28 ± 0.21	0.05 ± 0.10	0.26 ± 0.	0.26 ± 0.22	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.47 ± 0.24**	0.46 ± 0.24
	2	0.29 ± 0.21	0.05 ± 0.10	0.26 ± 0.	0.25 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.48 ± 0.24**	0.47 ± 0.24
10	0	0.26 ± 0.16	0.05 ± 0.07	0.29 ± 0.15	0.25 ± 0.15	0.27 ± 0.00	0.34 ± 0.17	0.34 ± 0.17	0.33 ± 0.17	0.34 ± 0.17	**0.41 ± 0.19**	0.40 ± 0.18
	1	0.25 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.33 ± 0.16	0.33 ± 0.17	0.33 ± 0.17	**0.39 ± 0.18**	0.38 ± 0.18
	2	0.26 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.34 ± 0.16	0.33 ± 0.17	0.34 ± 0.16	**0.39 ± 0.18**	**0.39 ± 0.18**
15	0	0.24 ± 0.13	0.05 ± 0.06	0.27 ± 0.13	0.21 ± 0.13	0.27 ± 0.00	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	**0.36 ± 0.15**	**0.36 ± 0.15**
	1	0.23 ± 0.12	0.05 ± 0.05	0.24 ± 0.12	0.20 ± 0.13	0.29 ± 0.00	0.30 ± 0.14	0.30 ± 0.14	0.30 ± 0.14	0.30 ± 0.14	**0.34 ± 0.15**	**0.34 ± 0.15**
	2	0.24 ± 0.12	0.05 ± 0.06	0.23 ± 0.12	0.19 ± 0.12	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.13	0.30 ± 0.14	0.31 ± 0.13	**0.35 ± 0.15**	0.34 ± 0.15
20	0	0.22 ± 0.11	0.05 ± 0.05	0.25 ± 0.11	0.19 ± 0.11	0.27 ± 0.00	0.29 ± 0.12	0.29 ± 0.12	0.29 ± 0.13	0.29 ± 0.12	**0.32 ± 0.13**	**0.32 ± 0.13**
	1	0.22 ± 0.10	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	**0.31 ± 0.13**	0.30 ± 0.13
	2	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.10	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	**0.31 ± 0.13**	**0.31 ± 0.13**

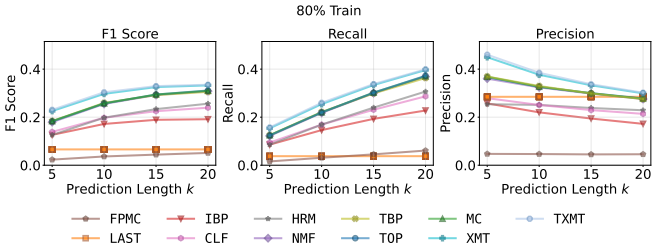


Figure 19. Comparison of XMT and TXMT with Baseline methods using a *80% Train* and h=1

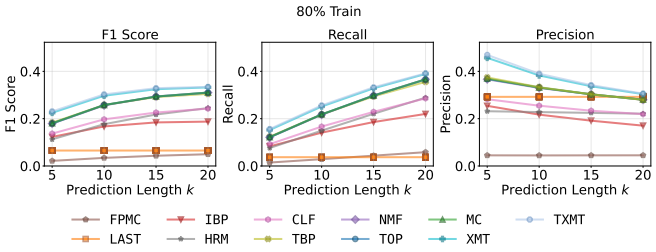


Figure 20. Comparison of XMT and TXMT with Baseline methods using a *80% Train* and h=2

Table 16. Recall comparison for different prediction lengths (k) and days (d) with *50% Train*

		50% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.09 ± 0.07	0.02 ± 0.03	0.11 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	$\mathbf{0.17 \pm 0.09}$	0.16 ± 0.09
	1	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.13 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.13 ± 0.08	$\mathbf{0.16 \pm 0.09}$	$\mathbf{0.16 \pm 0.09}$
	2	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	$\mathbf{0.16 \pm 0.09}$	$\mathbf{0.16 \pm 0.08}$
10	0	0.17 ± 0.10	0.03 ± 0.05	0.19 ± 0.10	0.16 ± 0.09	0.04 ± 0.00	0.23 ± 0.10	0.23 ± 0.10	0.22 ± 0.11	0.23 ± 0.11	$\mathbf{0.28 \pm 0.12}$	0.27 ± 0.12
	1	0.17 ± 0.10	0.03 ± 0.05	0.17 ± 0.09	0.15 ± 0.10	0.04 ± 0.00	0.22 ± 0.11	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	$\mathbf{0.26 \pm 0.12}$	$\mathbf{0.26 \pm 0.12}$
	2	0.17 ± 0.09	0.03 ± 0.04	0.16 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.10	$\mathbf{0.26 \pm 0.11}$	$\mathbf{0.26 \pm 0.11}$
15	0	0.24 ± 0.12	0.05 ± 0.06	0.27 ± 0.11	0.21 ± 0.11	0.04 ± 0.00	0.31 ± 0.12	0.31 ± 0.12	0.30 ± 0.12	0.31 ± 0.12	$\mathbf{0.36 \pm 0.14}$	0.35 ± 0.14
	1	0.23 ± 0.11	0.05 ± 0.06	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	$\mathbf{0.34 \pm 0.13}$	$\mathbf{0.34 \pm 0.13}$
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	$\mathbf{0.33 \pm 0.13}$	$\mathbf{0.33 \pm 0.13}$
20	0	0.29 ± 0.12	0.07 ± 0.06	0.33 ± 0.12	0.25 ± 0.12	0.04 ± 0.00	0.38 ± 0.13	0.38 ± 0.13	0.37 ± 0.14	0.38 ± 0.13	$\mathbf{0.42 \pm 0.15}$	$\mathbf{0.42 \pm 0.15}$
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.36 ± 0.14	0.38 ± 0.13	$\mathbf{0.40 \pm 0.14}$	$\mathbf{0.40 \pm 0.14}$
	2	0.29 ± 0.12	0.06 ± 0.06	0.30 ± 0.12	0.22 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.35 ± 0.13	0.37 ± 0.13	$\mathbf{0.39 \pm 0.14}$	$\mathbf{0.39 \pm 0.14}$

Table 17. F1 Score comparison for different prediction lengths (k) and days (d) with *50% Train*

		50% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.14 ± 0.10	0.03 ± 0.05	0.16 ± 0.10	0.14 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.18 ± 0.12	0.18 ± 0.11	$\mathbf{0.25 \pm 0.13}$	0.24 ± 0.12
	1	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.13 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.19 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
	2	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
10	0	0.20 ± 0.12	0.04 ± 0.05	0.23 ± 0.11	0.19 ± 0.11	0.06 ± 0.00	0.27 ± 0.12	0.27 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.32 \pm 0.13}$	$\mathbf{0.32 \pm 0.13}$
	1	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	$\mathbf{0.30 \pm 0.13}$	$\mathbf{0.30 \pm 0.13}$
	2	0.20 ± 0.11	0.04 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.11	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.31 \pm 0.13}$	0.30 ± 0.13
15	0	0.23 ± 0.11	0.05 ± 0.05	0.26 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.31 ± 0.12	0.30 ± 0.11	0.30 ± 0.12	0.30 ± 0.12	$\mathbf{0.35 \pm 0.13}$	$\mathbf{0.35 \pm 0.13}$
	1	0.23 ± 0.10	0.04 ± 0.05	0.23 ± 0.10	0.19 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.12}$	$\mathbf{0.33 \pm 0.12}$
	2	0.23 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.18 ± 0.10	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.12}$	$\mathbf{0.33 \pm 0.12}$
20	0	0.25 ± 0.11	0.06 ± 0.05	0.28 ± 0.10	0.21 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.32 ± 0.11	0.32 ± 0.12	0.32 ± 0.11	$\mathbf{0.36 \pm 0.12}$	0.35 ± 0.12
	1	0.24 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.32 ± 0.11	$\mathbf{0.34 \pm 0.12}$	$\mathbf{0.34 \pm 0.12}$
	2	0.25 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.31 ± 0.11	$\mathbf{0.34 \pm 0.12}$	$\mathbf{0.34 \pm 0.12}$

Table 18. Precision comparison for different prediction lengths k and days h with *60% Train*

		60% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.29 ± 0.22	0.05 ± 0.10	0.33 ± 0.22	0.28 ± 0.21	0.27 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.38 ± 0.24	0.37 ± 0.23	$\mathbf{0.50 \pm 0.26}$	$\mathbf{0.50 \pm 0.26}$
	1	0.28 ± 0.21	0.05 ± 0.10	$0.26 \pm 0.$	0.26 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	$\mathbf{0.47 \pm 0.25}$	0.46 ± 0.24
	2	0.28 ± 0.21	0.05 ± 0.09	$0.25 \pm 0.$	0.25 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	$\mathbf{0.47 \pm 0.24}$	0.46 ± 0.24
10	0	0.26 ± 0.16	0.05 ± 0.07	0.30 ± 0.16	0.24 ± 0.16	0.27 ± 0.00	0.34 ± 0.17	0.34 ± 0.17	0.34 ± 0.17	0.34 ± 0.17	$\mathbf{0.41 \pm 0.19}$	$\mathbf{0.41 \pm 0.19}$
	1	0.25 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	$\mathbf{0.39 \pm 0.18}$	0.38 ± 0.18
	2	0.26 ± 0.15	0.05 ± 0.07	0.24 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.34 ± 0.16	0.33 ± 0.17	0.33 ± 0.17	$\mathbf{0.39 \pm 0.18}$	$\mathbf{0.39 \pm 0.18}$
15	0	0.24 ± 0.13	0.05 ± 0.06	0.27 ± 0.13	0.21 ± 0.13	0.27 ± 0.00	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	$\mathbf{0.36 \pm 0.16}$	$\mathbf{0.36 \pm 0.15}$
	1	0.23 ± 0.12	0.05 ± 0.05	0.24 ± 0.12	0.20 ± 0.12	0.29 ± 0.00	0.30 ± 0.14	0.30 ± 0.14	0.30 ± 0.14	0.30 ± 0.14	$\mathbf{0.34 \pm 0.15}$	$\mathbf{0.34 \pm 0.15}$
	2	0.24 ± 0.12	0.05 ± 0.05	0.23 ± 0.12	0.19 ± 0.12	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.13	0.30 ± 0.14	0.31 ± 0.14	$\mathbf{0.35 \pm 0.15}$	0.34 ± 0.15
20	0	0.22 ± 0.11	0.05 ± 0.05	0.26 ± 0.12	0.19 ± 0.11	0.27 ± 0.00	0.29 ± 0.12	0.29 ± 0.12	0.29 ± 0.13	0.29 ± 0.12	$\mathbf{0.32 \pm 0.13}$	$\mathbf{0.32 \pm 0.13}$
	1	0.22 ± 0.10	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.10	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	$\mathbf{0.31 \pm 0.13}$	0.30 ± 0.13
	2	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.10	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	$\mathbf{0.31 \pm 0.13}$	$\mathbf{0.31 \pm 0.13}$

Table 19. Recall comparison for different prediction lengths (k) and days (d) with *60% Train*

		60% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.10 ± 0.07	0.02 ± 0.03	0.11 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	$\mathbf{0.17 \pm 0.09}$	$\mathbf{0.17 \pm 0.09}$
	1	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	$\mathbf{0.16 \pm 0.09}$	$\mathbf{0.16 \pm 0.09}$
	2	0.09 ± 0.07	0.01 ± 0.03	0.08 ± 0.07	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	$\mathbf{0.16 \pm 0.09}$	0.15 ± 0.08
10	0	0.17 ± 0.10	0.04 ± 0.05	0.20 ± 0.10	0.16 ± 0.10	0.04 ± 0.00	0.23 ± 0.10	0.23 ± 0.11	0.22 ± 0.11	0.23 ± 0.11	$\mathbf{0.28 \pm 0.12}$	0.27 ± 0.12
	1	0.17 ± 0.10	0.03 ± 0.04	0.17 ± 0.09	0.15 ± 0.09	0.04 ± 0.00	0.22 ± 0.11	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	$\mathbf{0.26 \pm 0.12}$	$\mathbf{0.26 \pm 0.12}$
	2	0.17 ± 0.09	0.03 ± 0.04	0.16 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	$\mathbf{0.26 \pm 0.11}$	0.25 ± 0.11
15	0	0.24 ± 0.12	0.05 ± 0.06	0.27 ± 0.12	0.21 ± 0.11	0.04 ± 0.00	0.31 ± 0.12	0.31 ± 0.12	0.31 ± 0.13	0.31 ± 0.12	$\mathbf{0.36 \pm 0.14}$	0.35 ± 0.14
	1	0.23 ± 0.11	0.04 ± 0.05	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.13	0.30 ± 0.12	$\mathbf{0.34 \pm 0.13}$	$\mathbf{0.34 \pm 0.13}$
	2	0.23 ± 0.11	0.04 ± 0.05	0.23 ± 0.10	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	$\mathbf{0.33 \pm 0.13}$	$\mathbf{0.33 \pm 0.13}$
20	0	0.29 ± 0.12	0.07 ± 0.07	0.34 ± 0.13	0.25 ± 0.13	0.04 ± 0.00	0.38 ± 0.13	0.38 ± 0.13	0.37 ± 0.14	0.38 ± 0.13	$\mathbf{0.42 \pm 0.15}$	$\mathbf{0.42 \pm 0.15}$
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.36 ± 0.13	0.38 ± 0.13	$\mathbf{0.40 \pm 0.14}$	$\mathbf{0.40 \pm 0.14}$
	2	0.29 ± 0.12	0.06 ± 0.06	0.29 ± 0.12	0.22 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.36 ± 0.13	0.37 ± 0.13	$\mathbf{0.40 \pm 0.14}$	0.39 ± 0.14

Table 20. F1 Score comparison for different prediction lengths (k) and days (d) with *60% Train*

		60% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.14 ± 0.10	0.03 ± 0.05	0.16 ± 0.10	0.14 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.19 ± 0.12	0.18 ± 0.11	$\mathbf{0.25 \pm 0.13}$	$\mathbf{0.25 \pm 0.13}$
	1	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.13 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
	2	0.14 ± 0.10	0.02 ± 0.05	0.12 ± 0.10	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	$\mathbf{0.23 \pm 0.12}$	$\mathbf{0.23 \pm 0.12}$
10	0	0.20 ± 0.11	0.04 ± 0.05	0.23 ± 0.11	0.19 ± 0.11	0.06 ± 0.00	0.27 ± 0.12	0.27 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.32 \pm 0.13}$	$\mathbf{0.32 \pm 0.13}$
	1	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.31 \pm 0.13}$	0.30 ± 0.13
	2	0.20 ± 0.11	0.03 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	$\mathbf{0.30 \pm 0.13}$	$\mathbf{0.30 \pm 0.13}$
15	0	0.23 ± 0.11	0.05 ± 0.05	0.27 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.31 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	$\mathbf{0.35 \pm 0.13}$	0.34 ± 0.13
	1	0.23 ± 0.10	0.04 ± 0.05	0.24 ± 0.10	0.19 ± 0.11	0.07 ± 0.00	0.30 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.12}$	$\mathbf{0.33 \pm 0.12}$
	2	0.23 ± 0.11	0.04 ± 0.05	0.22 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	$\mathbf{0.33 \pm 0.12}$	$\mathbf{0.33 \pm 0.12}$
20	0	0.25 ± 0.10	0.06 ± 0.05	0.28 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.32 ± 0.11	0.32 ± 0.12	0.32 ± 0.11	$\mathbf{0.36 \pm 0.12}$	0.35 ± 0.12
	1	0.24 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	0.32 ± 0.11	$\mathbf{0.34 \pm 0.12}$	$\mathbf{0.34 \pm 0.12}$
	2	0.24 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	$\mathbf{0.34 \pm 0.12}$	$\mathbf{0.34 \pm 0.12}$

Table 21. Precision comparison for different prediction lengths k and days h with *70% Train*

		70% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.29 ± 0.21	0.06 ± 0.10	0.33 ± 0.22	0.28 ± 0.21	0.27 ± 0.00	0.37 ± 0.23	0.38 ± 0.23	0.37 ± 0.24	0.37 ± 0.23	$\mathbf{0.50 \pm 0.26}$	0.49 ± 0.26
	1	0.28 ± 0.21	0.05 ± 0.09	$0.25 \pm 0.$	0.26 ± 0.21	0.28 ± 0.00	0.37 ± 0.23	0.36 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	$\mathbf{0.47 \pm 0.24}$	0.46 ± 0.24
	2	0.29 ± 0.21	0.05 ± 0.10	0.24 ± 0.19	0.26 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	$\mathbf{0.47 \pm 0.25}$	0.46 ± 0.24
10	0	0.25 ± 0.16	0.05 ± 0.07	0.29 ± 0.16	0.24 ± 0.16	0.27 ± 0.00	0.34 ± 0.17	0.34 ± 0.17	0.33 ± 0.18	0.34 ± 0.17	$\mathbf{0.41 \pm 0.19}$	0.40 ± 0.19
	1	0.25 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.28 ± 0.00	0.33 ± 0.17	0.33 ± 0.16	0.33 ± 0.17	0.33 ± 0.17	$\mathbf{0.39 \pm 0.18}$	0.38 ± 0.18
	2	0.26 ± 0.15	0.05 ± 0.07	0.23 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.34 ± 0.17	0.34 ± 0.16	0.34 ± 0.17	0.34 ± 0.17	$\mathbf{0.39 \pm 0.18}$	$\mathbf{0.39 \pm 0.18}$
15	0	0.23 ± 0.13	0.05 ± 0.06	0.27 ± 0.13	0.21 ± 0.13	0.27 ± 0.00	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	$\mathbf{0.36 \pm 0.15}$	0.35 ± 0.15
	1	0.23 ± 0.12	0.05 ± 0.05	0.24 ± 0.12	0.19 ± 0.12	0.28 ± 0.00	0.30 ± 0.14	0.30 ± 0.13	0.30 ± 0.14	0.30 ± 0.14	$\mathbf{0.34 \pm 0.15}$	$\mathbf{0.34 \pm 0.15}$
	2	0.24 ± 0.12	0.05 ± 0.06	0.23 ± 0.11	0.19 ± 0.12	0.29 ± 0.00	0.30 ± 0.14	0.31 ± 0.14	0.30 ± 0.14	0.31 ± 0.14	$\mathbf{0.34 \pm 0.15}$	$\mathbf{0.34 \pm 0.15}$
20	0	0.22 ± 0.11	0.05 ± 0.05	0.25 ± 0.12	0.19 ± 0.11	0.27 ± 0.00	0.29 ± 0.12	0.28 ± 0.12	0.28 ± 0.13	0.29 ± 0.12	$\mathbf{0.32 \pm 0.13}$	$\mathbf{0.32 \pm 0.13}$
	1	0.22 ± 0.11	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.10	0.28 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	$\mathbf{0.31 \pm 0.13}$	0.30 ± 0.13
	2	0.22 ± 0.11	0.05 ± 0.05	0.22 ± 0.10	0.17 ± 0.11	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	$\mathbf{0.31 \pm 0.13}$	$\mathbf{0.31 \pm 0.13}$

Table 22. Recall comparison for different prediction lengths (k) and days (d) with *70% Train*

		70% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.09 ± 0.07	0.02 ± 0.04	0.11 ± 0.08	0.09 ± 0.07	0.03 ± 0.00	0.12 ± 0.08	0.13 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.17 ± 0.10**	**0.17 ± 0.10**
	1	0.10 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.09
	2	0.09 ± 0.07	0.02 ± 0.03	0.08 ± 0.07	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.09
10	0	0.17 ± 0.10	0.04 ± 0.05	0.19 ± 0.10	0.16 ± 0.10	0.03 ± 0.00	0.23 ± 0.11	0.23 ± 0.11	0.22 ± 0.11	0.22 ± 0.11	**0.27 ± 0.13**	**0.27 ± 0.12**
	1	0.17 ± 0.10	0.03 ± 0.04	0.17 ± 0.09	0.15 ± 0.10	0.04 ± 0.00	0.22 ± 0.11	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	**0.26 ± 0.12**	**0.26 ± 0.12**
	2	0.17 ± 0.09	0.03 ± 0.04	0.15 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	**0.26 ± 0.11**	0.25 ± 0.11
15	0	0.23 ± 0.12	0.05 ± 0.06	0.27 ± 0.12	0.20 ± 0.12	0.03 ± 0.00	0.31 ± 0.13	0.30 ± 0.12	0.31 ± 0.13	0.31 ± 0.13	**0.36 ± 0.14**	0.35 ± 0.14
	1	0.23 ± 0.11	0.04 ± 0.05	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	**0.34 ± 0.13**	0.33 ± 0.13
	2	0.23 ± 0.11	0.04 ± 0.05	0.22 ± 0.10	0.18 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	**0.33 ± 0.13**	**0.33 ± 0.13**
20	0	0.29 ± 0.13	0.07 ± 0.07	0.33 ± 0.13	0.24 ± 0.13	0.03 ± 0.00	0.38 ± 0.14	0.38 ± 0.13	0.37 ± 0.14	0.38 ± 0.13	**0.42 ± 0.15**	0.41 ± 0.15
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.36 ± 0.13	0.37 ± 0.13	**0.40 ± 0.14**	**0.40 ± 0.14**
	2	0.29 ± 0.12	0.06 ± 0.06	0.29 ± 0.12	0.22 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.35 ± 0.13	0.37 ± 0.13	**0.39 ± 0.14**	**0.39 ± 0.14**

Table 23. F1 Score comparison for different prediction lengths (k) and days (d) with *70% Train*

		70% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.14 ± 0.10	0.03 ± 0.05	0.16 ± 0.11	0.14 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.19 ± 0.11	0.18 ± 0.12	0.18 ± 0.11	**0.25 ± 0.13**	**0.25 ± 0.13**
	1	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.13 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
	2	0.14 ± 0.10	0.02 ± 0.05	0.12 ± 0.09	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
10	0	0.20 ± 0.11	0.04 ± 0.06	0.23 ± 0.11	0.19 ± 0.11	0.06 ± 0.00	0.27 ± 0.12	0.27 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.32 ± 0.14**	0.31 ± 0.13
	1	0.20 ± 0.11	0.04 ± 0.05	0.19 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.31 ± 0.13**	0.30 ± 0.13
	2	0.20 ± 0.11	0.04 ± 0.05	0.18 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
15	0	0.23 ± 0.11	0.05 ± 0.06	0.26 ± 0.11	0.20 ± 0.11	0.06 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	**0.35 ± 0.13**	0.34 ± 0.13
	1	0.23 ± 0.10	0.04 ± 0.05	0.23 ± 0.10	0.19 ± 0.11	0.07 ± 0.00	0.30 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**
	2	0.23 ± 0.10	0.04 ± 0.05	0.22 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.30 ± 0.11	0.29 ± 0.12	0.30 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**
20	0	0.24 ± 0.10	0.06 ± 0.06	0.28 ± 0.11	0.21 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.32 ± 0.11	0.31 ± 0.12	0.32 ± 0.11	**0.35 ± 0.12**	**0.35 ± 0.12**
	1	0.24 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	**0.34 ± 0.12**	**0.34 ± 0.12**
	2	0.24 ± 0.10	0.05 ± 0.05	0.25 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.31 ± 0.11	**0.34 ± 0.12**	0.33 ± 0.12

Table 24. Precision comparison for different prediction lengths k and days h with *80% Train*

		80% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.27 ± 0.21	0.05 ± 0.09	0.33 ± 0.22	0.27 ± 0.21	0.25 ± 0.00	0.36 ± 0.23	0.36 ± 0.22	0.36 ± 0.22	0.36 ± 0.23	**0.48 ± 0.26**	**0.48 ± 0.26**
	1	0.28 ± 0.21	0.05 ± 0.09	0.25 ± 0.	0.26 ± 0.21	0.28 ± 0.00	0.37 ± 0.23	0.36 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.46 ± 0.24**	0.45 ± 0.24
	2	0.28 ± 0.21	0.05 ± 0.09	0.23 ± 0.19	0.25 ± 0.21	0.29 ± 0.00	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	0.37 ± 0.23	**0.47 ± 0.24**	0.46 ± 0.24
10	0	0.25 ± 0.15	0.05 ± 0.07	0.29 ± 0.16	0.23 ± 0.15	0.25 ± 0.00	0.33 ± 0.17	0.33 ± 0.16	0.32 ± 0.17	0.33 ± 0.17	**0.40 ± 0.18**	0.39 ± 0.18
	1	0.25 ± 0.15	0.05 ± 0.07	0.25 ± 0.14	0.22 ± 0.15	0.28 ± 0.00	0.33 ± 0.16	0.32 ± 0.16	0.33 ± 0.17	0.33 ± 0.17	**0.38 ± 0.18**	**0.38 ± 0.18**
	2	0.25 ± 0.15	0.04 ± 0.07	0.23 ± 0.14	0.22 ± 0.15	0.29 ± 0.00	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	0.33 ± 0.17	**0.39 ± 0.18**	0.38 ± 0.18
15	0	0.23 ± 0.12	0.05 ± 0.06	0.27 ± 0.13	0.21 ± 0.13	0.25 ± 0.00	0.31 ± 0.14	0.30 ± 0.14	0.30 ± 0.14	0.30 ± 0.14	**0.35 ± 0.15**	0.34 ± 0.14
	1	0.23 ± 0.12	0.05 ± 0.05	0.24 ± 0.12	0.19 ± 0.12	0.28 ± 0.00	0.30 ± 0.14	0.30 ± 0.13	0.30 ± 0.14	0.30 ± 0.14	**0.34 ± 0.15**	0.33 ± 0.15
	2	0.23 ± 0.12	0.05 ± 0.05	0.22 ± 0.11	0.19 ± 0.12	0.29 ± 0.00	0.30 ± 0.14	0.30 ± 0.13	0.30 ± 0.14	0.30 ± 0.14	**0.34 ± 0.15**	**0.34 ± 0.15**
20	0	0.22 ± 0.10	0.05 ± 0.05	0.25 ± 0.11	0.18 ± 0.11	0.25 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	**0.31 ± 0.12**	**0.31 ± 0.12**
	1	0.21 ± 0.10	0.05 ± 0.05	0.23 ± 0.10	0.17 ± 0.11	0.28 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
	2	0.22 ± 0.11	0.05 ± 0.05	0.22 ± 0.10	0.17 ± 0.10	0.29 ± 0.00	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	0.28 ± 0.12	**0.31 ± 0.13**	0.30 ± 0.13

Table 25. Recall comparison for different prediction lengths (k) and days (d) with *80% Train*

		80% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.09 ± 0.07	0.02 ± 0.03	0.11 ± 0.07	0.09 ± 0.07	0.03 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	**0.16 ± 0.09**
	1	0.09 ± 0.07	0.02 ± 0.03	0.09 ± 0.07	0.09 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.09**	0.15 ± 0.09
	2	0.09 ± 0.07	0.01 ± 0.03	0.08 ± 0.06	0.08 ± 0.07	0.04 ± 0.00	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	0.12 ± 0.08	**0.16 ± 0.08**	0.15 ± 0.08
10	0	0.16 ± 0.10	0.04 ± 0.05	0.19 ± 0.10	0.16 ± 0.10	0.03 ± 0.00	0.22 ± 0.11	0.22 ± 0.11	0.22 ± 0.10	0.22 ± 0.11	**0.27 ± 0.13**	0.26 ± 0.12
	1	0.17 ± 0.10	0.03 ± 0.05	0.17 ± 0.09	0.15 ± 0.10	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.11	0.22 ± 0.10	**0.26 ± 0.12**	0.25 ± 0.12
	2	0.17 ± 0.09	0.03 ± 0.04	0.15 ± 0.09	0.14 ± 0.09	0.04 ± 0.00	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.10	0.22 ± 0.10	**0.26 ± 0.11**	0.25 ± 0.11
15	0	0.23 ± 0.11	0.05 ± 0.06	0.27 ± 0.11	0.20 ± 0.12	0.03 ± 0.00	0.31 ± 0.13	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.13	**0.35 ± 0.14**	0.34 ± 0.14
	1	0.23 ± 0.11	0.05 ± 0.05	0.24 ± 0.11	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	0.30 ± 0.12	**0.34 ± 0.13**	0.33 ± 0.13
	2	0.23 ± 0.11	0.04 ± 0.05	0.22 ± 0.10	0.19 ± 0.11	0.04 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.12	0.30 ± 0.12	**0.33 ± 0.13**	**0.33 ± 0.13**
20	0	0.29 ± 0.12	0.07 ± 0.06	0.33 ± 0.13	0.24 ± 0.13	0.03 ± 0.00	0.38 ± 0.14	0.38 ± 0.13	0.36 ± 0.14	0.38 ± 0.14	**0.41 ± 0.15**	**0.41 ± 0.15**
	1	0.29 ± 0.12	0.06 ± 0.06	0.31 ± 0.12	0.23 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.36 ± 0.13	0.37 ± 0.13	**0.40 ± 0.14**	**0.40 ± 0.14**
	2	0.29 ± 0.12	0.06 ± 0.06	0.29 ± 0.11	0.22 ± 0.12	0.04 ± 0.00	0.37 ± 0.13	0.37 ± 0.13	0.36 ± 0.13	0.37 ± 0.13	**0.39 ± 0.14**	**0.39 ± 0.14**

Table 26. F1 Score comparison for different prediction lengths (k) and days (d) with *80% Train*

		80% Train										
k	h	**clf**	**fpmc**	**hrm**	**ibp**	**last**	**mc**	**nmf**	**tbp**	**top**	**txmt**	**xmt**
5	0	0.13 ± 0.10	0.02 ± 0.05	0.16 ± 0.11	0.13 ± 0.10	0.06 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.24 ± 0.13**	**0.24 ± 0.13**
	1	0.14 ± 0.10	0.02 ± 0.05	0.13 ± 0.10	0.13 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	**0.23 ± 0.12**
	2	0.14 ± 0.10	0.02 ± 0.05	0.11 ± 0.09	0.12 ± 0.10	0.07 ± 0.00	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	0.18 ± 0.11	**0.23 ± 0.12**	0.22 ± 0.12
10	0	0.19 ± 0.11	0.04 ± 0.05	0.23 ± 0.11	0.18 ± 0.11	0.06 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.25 ± 0.12	0.26 ± 0.12	**0.31 ± 0.14**	**0.31 ± 0.13**
	1	0.20 ± 0.11	0.04 ± 0.05	0.20 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
	2	0.20 ± 0.11	0.03 ± 0.05	0.18 ± 0.10	0.17 ± 0.11	0.07 ± 0.00	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	0.26 ± 0.12	**0.30 ± 0.13**	**0.30 ± 0.13**
15	0	0.22 ± 0.11	0.05 ± 0.05	0.26 ± 0.11	0.20 ± 0.11	0.06 ± 0.00	0.30 ± 0.12	0.30 ± 0.12	0.29 ± 0.11	0.30 ± 0.12	**0.34 ± 0.13**	0.33 ± 0.13
	1	0.22 ± 0.11	0.04 ± 0.05	0.23 ± 0.10	0.19 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.29 ± 0.11	0.29 ± 0.12	0.29 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**
	2	0.23 ± 0.11	0.04 ± 0.05	0.22 ± 0.10	0.18 ± 0.11	0.07 ± 0.00	0.29 ± 0.11	0.29 ± 0.11	0.29 ± 0.12	0.29 ± 0.11	**0.33 ± 0.12**	0.32 ± 0.12
20	0	0.24 ± 0.10	0.06 ± 0.05	0.27 ± 0.11	0.20 ± 0.11	0.06 ± 0.00	0.32 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	0.32 ± 0.11	**0.34 ± 0.12**	**0.34 ± 0.12**
	1	0.24 ± 0.10	0.05 ± 0.05	0.26 ± 0.10	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	0.31 ± 0.11	**0.34 ± 0.12**	0.33 ± 0.12
	2	0.24 ± 0.10	0.05 ± 0.05	0.24 ± 0.09	0.19 ± 0.10	0.07 ± 0.00	0.31 ± 0.11	0.31 ± 0.11	0.30 ± 0.11	0.31 ± 0.11	**0.33 ± 0.12**	**0.33 ± 0.12**

D Example of explanations

D.0.1 Example of XMT Explanations

To illustrate how XMT generates explanations, we examine here a real explanation for a prediction of the following sample taken from the Italian Supermarket dataset:

- **Customer ID:** 72764
- **Shopping Date:** September 9th, 2013
- **Current Basket (9 items):** rice, snacks, prepared food, bread, tomatoes, hazelnut fruit, wine, olive oil, cold cuts
- **Prediction Size:** 5 items
- **Predicted Forgotten Items:** drinks, dried fruit, rabbit, yogurt, vegetables

For each predicted item, XMT generated the following evidence-based explanations:

Item: vegetables
1. Last purchased 3 days ago (typically bought every 2.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. Often repurchased soon after large shopping trips (71.3% of opportunities)

Item: drinks
1. Last purchased 4 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. Often repurchased soon after large shopping trips (50.2% of opportunities)

Item: rabbit
1. Last purchased 6 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. Often repurchased soon after large shopping trips (40.3% of opportunities)

Item: dried fruit
1. Last purchased 4 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. Often repurchased soon after large shopping trips (34.1% of opportunities)

Item: yogurt
1. Last purchased 2 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. Often repurchased soon after large shopping trips (47.2% of opportunities)

These explanations demonstrate how XMTXMT combines temporal patterns, co-purchase relationships, and repurchase behaviors to provide transparent, evidence-based justifications for its predictions.

D.0.2 Example of TXMT Explanations

To demonstrate how TXMT enhances the explanation framework through TARS integration, we consider the prediction done on the same sample used for XMT in Appendix D.0.1:

- **Customer ID:** 72764
- **Shopping Date:** September 9th, 2013
- **Current Basket (9 items):** rice, snacks, prepared food, bread, tomatoes, hazelnut fruit, wine, olive oil, cold cuts
- **Prediction Size:** 5 items
- **Predicted Forgotten Items:** yogurt, rabbit, vegetables, dried fruit, drinks

Notice that while TXMT predicts the same items as XMT for this particular sample, its explanations incorporate additional TARS-based evidence:

Item: vegetables
1. Last purchased 3 days ago (typically bought every 2.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. TARS pattern analysis suggests this item is likely to be needed (confidence: 50.5%)
4. Often repurchased soon after large shopping trips (71.3% of opportunities)

Item: drinks
1. Last purchased 4 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. TARS pattern analysis suggests this item is likely to be needed (confidence: 39.0%)
4. Often repurchased soon after large shopping trips (50.2% of opportunities)

Item: rabbit
1. Last purchased 6 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. TARS pattern analysis suggests this item is likely to be needed (confidence: 29.5%)
4. Often repurchased soon after large shopping trips (40.3% of opportunities)

Item: dried_fruit
1. Last purchased 4 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. TARS pattern analysis suggests this item is likely to be needed (confidence: 28.6%)
4. Often repurchased soon after large shopping trips (34.1% of opportunities)

Item: yogurt
1. Last purchased 2 days ago (typically bought every 3.0 days)
2. Often bought with current basket items: bread, cold cuts, wine
3. TARS pattern analysis suggests this item is likely to be needed (confidence: 37.2%)
4. Often repurchased soon after large shopping trips (47.2% of opportunities)

The addition of TARS-based confidence scores provides users with a quantitative measure of how strongly each prediction is supported by recurring temporal patterns in their shopping history. This enhancement adds another layer to the explanations while maintaining the original, evidence-based structure of the original XMT algorithm.

E Visualization of the explanations

Alongside the development of the two proposed algorithms, XMT and TXMT, we developed a dashboard to visualize the predictions and the explanations. We report in Figure 21 a screenshot taken from the dashboard that shows the differences between the explanation obtained using XMT and TXMT. In particular, we highlight in green the TARS pattern extracted by the TXMT algorithm that increases the interpretability of the explanation.

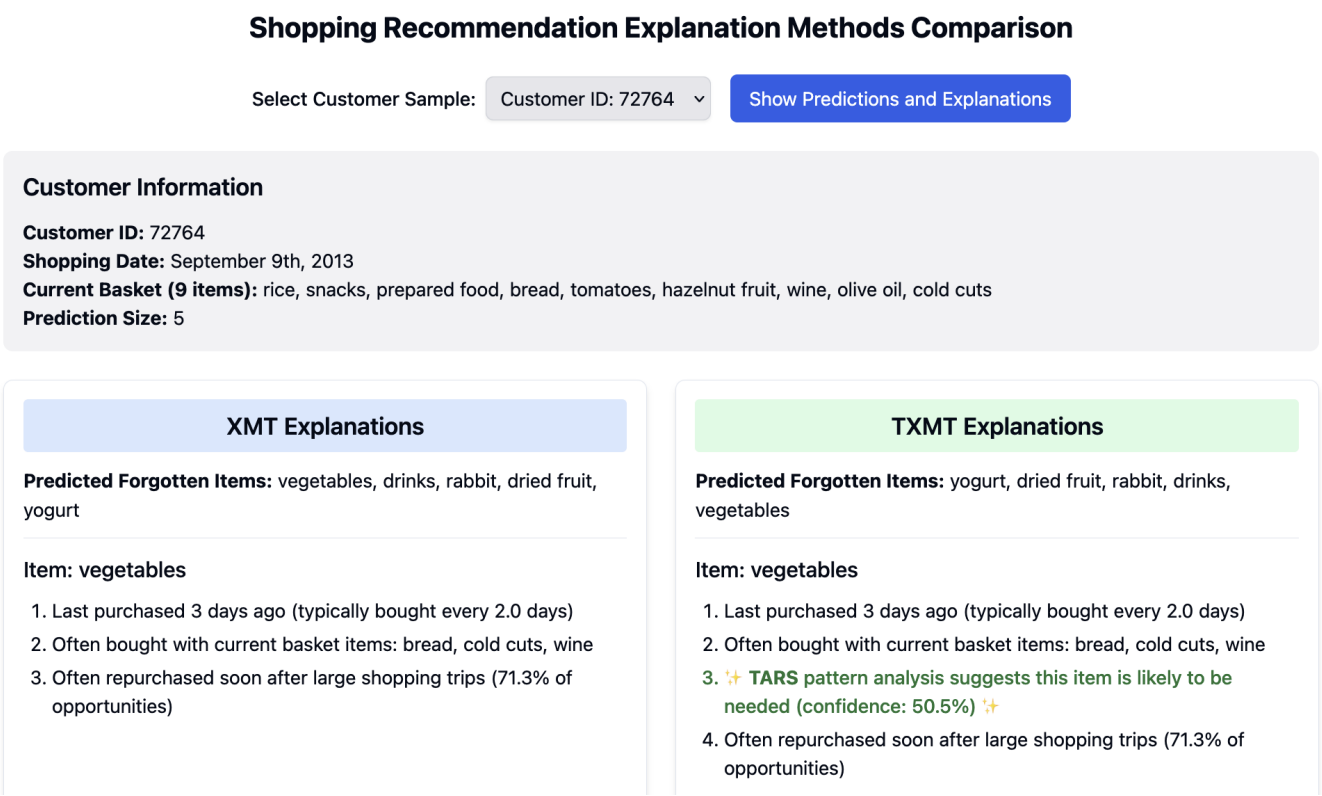


Figure 21. Visualization of an explanation of a prediction obtained with XMT and TXMT