A Deep Markov Model for Clickstream Analytics in Online Shopping

Yilmazcan Ozyurt
ETH Zurich
Zurich, Switzerland
yozyurt@ethz.ch

Tobias Hatt
ETH Zurich
Zurich, Switzerland
thatt@ethz.ch

Ce Zhang
ETH Zurich
Zurich, Switzerland
ce.zhang@inf.ethz.ch

Stefan Feuerriegel
LMU Munich
Munich, Germany
feuerriegel@lmu.de

ABSTRACT

Machine learning is widely used in e-commerce to analyze clickstream sessions and then to allocate marketing resources. Traditional neural learning can model long-term dependencies in clickstream data, yet it ignores the different shopping phases (i. e., goal-directed search vs. browsing) in user behavior as theorized by marketing research. In this paper, we develop a novel, theory-informed machine learning model to account for different shopping phases as defined in marketing theory. Specifically, we formalize a tailored attentive deep Markov model called *Clickstream*DMM for predicting the risk of user exits without purchase in e-commerce web sessions. Our *Clickstream*DMM combines (1) an attention network to learn long-term dependencies in clickstream data and (2) a latent variable model to capture different shopping phases (i. e., goal-directed search vs. browsing). Due to the interpretable structure, our *Clickstream*DMM allows marketers to generate new insights on how shopping phases relate to actual purchase behavior. We evaluate our model using real-world clickstream data from a leading e-commerce platform consisting of 26,279 sessions with 250,287 page clicks. Thereby, we demonstrate that our model is effective in predicting user exits without purchase: compared to existing baselines, it achieves an improvement by 11.5 % in AUROC and 12.7 % in AUPRC. Overall, our model enables e-commerce platforms to detect users at the risk of exiting without purchase. Based on it, e-commerce platforms can then intervene with marketing resources to steer users toward purchasing.

CCS CONCEPTS

• **Mathematics of computing** → **Markov processes**; *Time series analysis*; • **Applied computing** → **Online shopping**; • **Information systems** → **Web log analysis**; • **Computing methodologies** → **Neural networks**; *Supervised learning by classification.*

KEYWORDS

clickstream data; online user behavior; attention network; latent states; Markov model; deep Markov model

ACM Reference Format:
Yilmazcan Ozyurt, Tobias Hatt, Ce Zhang, and Stefan Feuerriegel. 2022. A Deep Markov Model for Clickstream Analytics in Online Shopping. In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3485447.3512027

WWW '22, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9096-5/22/04…$15.00
https://doi.org/10.1145/3485447.3512027

1 INTRODUCTION

In 2020, over 97 % of global online shopping users exited the e-commerce websites with no purchase [34]. For this reason, e-commerce websites use machine learning to detect when users are about to exit, which helps them to trigger potential interventions that steer users toward purchase. For instance, some e-commerce websites leverage coupons or dynamic price promotions [8, 21]. These have the goal to steer users that otherwise would have exited with no purchase towards conversion (i. e., turning a "user" into a "buyer"). The effectiveness of such interventions depends on the accuracy with which users at risk of exiting with no purchase are identified [5]. Hence, accurate predictions of when users are about to exit without a purchase are needed.

Modeling clickstream data for predicting user exits without purchase is difficult for two main reasons. First, clickstreams exhibit long-term dependencies of page sequences. For example, whether a user visits *first* a product and *then* the shopping cart has a different meaning for the risk of user exits than whether the same user *first* visits the shopping cart and *then* moves on to a new product. Oftentimes, users also go back to previous product pages at the end of a session until eventually making a purchase. Second, clickstreams are highly variable over time as users undergo different (but unobservable) shopping phases [5, 24]. This has been previously formalized by marketing theory, according to which shopping phases are characterized by "goal-directed search" vs. " browsing" [23, 27]. The former, goal-directed search, is characterized by a high propensity towards purchasing, while the latter, browsing, is characterized by experiential activities such as information collection (e.g., when users visit overview pages).

There are several machine learning models for the purpose of predicting user exits without purchase from clickstream data. These can be grouped into two literature streams. (1) On the one hand,

recurrent neural networks and variants thereof have been applied to make inferences from clickstreams. These models have been powerful in modeling long-term dependencies within clickstream data. Specific examples are long short-term memory (LSTM) [16], bi-directional LSTM [39], and mixture LSTM [31, 36]. (2) On the other hand, latent variable models have been proposed, which model different shopping phases through latent variables. Specific models developed for clickstream analytics are hidden Markov models (HMMs) [5, 24] and Markov-modulated marked point processes (M3PP) [11]. However, to the best of our knowledge, there is no prior work that combines the strengths of both neural networks and latent variable models in predicting user exits without purchase.

Proposed model[1]: In this paper, we propose a novel deep probabilistic model called *Clickstream*DMM for predicting whether online users are at the risk of exiting with no purchase. Our *Clickstream*DMM is tailored to model (1) long-term dependencies of page sequences via an attention network. Furthermore, our *Clickstream*DMM models (2) different shopping phases of the online users via a latent variable model. Thereby, we propose the first combination of attention network and latent variable model tailored to clickstream modeling. Moreover, we propose an approach for interpreting *Clickstream*DMM via clustering the latent variable space. This provides relevant marketing insights and enables marketers to tailor their interventions specifically to a user's shopping phase.

Results: We conducted computational experiments based on a real-world clickstream dataset, comprising 26,279 online sessions. The dataset was provided by *Digitec Galaxus*, the largest online retailer in Switzerland offering more than a million products. The results show that our *Clickstream*DMM consistently outperforms state-of-the-art algorithms. First, we evaluated the task of identifying whether the current page will be an exit without purchase, which is analogous to the prior literature [5, 11, 16, 18, 24, 31, 36, 39]. In this task, *Clickstream*DMM achieves a performance improvement over the state-of-the-art baseline in the area under the receiver operating characteristic curve (AUROC) by 1.4 % and in area under the precision-recall curve (AUPRC) by 0.8 %. Second, and more importantly, we evaluated the performance in the task when making multi-step ahead predictions (i.e., will there be an exit without purchase within the next n pages). The latter is particularly relevant for real-world applications as it gives additional time for e-commerce websites to trigger marketing interventions. In this task, our *Clickstream*DMM yields a substantial performance improvement in AUROC by 11.5 % and in AUPRC by 12.7 %.

Contributions: In summary, our main contributions are the following:

(1) We propose a novel deep probabilistic model called *Clickstream*DMM to analyze user behavior in online shopping and predict users at the risk of exiting without purchase. Our *Clickstream*DMM models long-term dependencies via an attention network and further captures different shopping phases via a latent variable model.
(2) We demonstrate that our *Clickstream*DMM achieves a robust performance, even at early stages of a session (i. e., several pages before the actual exit occurs). In particular, our model outperforms state-of-the-art baselines by substantial margin.
(3) We suggest an approach for model interpretability. Specifically, we examine the latent space of our *Clickstream*DMM to generate marketing-relevant insights and thereby characterize clickstreams based on marketing theory.

[1]The code is available from https://github.com/oezyurty/ClickstreamDMM

2 RELATED WORK

Clickstream modeling has been the subject of several works in the literature. These works exploited the sequential structure of the page clicks in clickstream to answer different research questions regarding online user behavior. Examples are predicting the next page that a user will visit [2, 3, 10, 13, 15] and recommending the most suitable item for a user [1, 9, 19, 30, 35, 42–44]. On the other hand, there has been a particular focus on predicting purchases based on the clickstream behavior from entire user history (i.e., including previous sessions) [26, 38, 41]. In contrast, this paper will focus on predicting whether user will exit a web session with vs. without purchase using *only* the current session.

Marketing theory: Online shopping users exhibit varying clickstream patterns depending on their shopping phases [5, 24]. Specifically, a so-called *shopping phase* refers to whether the user engagement with the website is "goal-directed" or "browsing". The goal-directed shopping phase is characterized by page clicks towards a planned purchase. In contrast, browsing phase is associated with exploratory activities in online shopping without a specific purpose. Further, users in browsing phase may experience "flow" [14], which refers to prolonged duration in user activity with a distorted sense of time. A user may undergo different shopping phases in a single web session [14, 20]. Since such phases are unobservable to the online marketer, they must be modeled by latent variables.

Predicting exits in clickstreams: Several approaches have been proposed for predicting the user exit without purchase in clickstream data (cf. [22] for an overview). The approaches can be loosely grouped into two categories (see Table 1): recurrent neural networks and latent variable models, as follows.

(1) Recurrent neural networks: Recent works have approached the prediction of users at the risk of exiting with no purchase through the use of recurrent neural networks. For this purpose, several adaptations of recurrent neural networks have been implemented such as long short-term memory (LSTM) [16, 32], bi-directional LSTM [39], and mixture LSTM [31, 36]. These models have been tailored to capture long-term dependencies among the sequence of page clicks in clickstream data. However, these do not model different shopping phases (i.e., latent states) of users throughout web sessions.

(2) Latent variable models: Latent variable models have been proven useful for capturing unobserved variables [e. g., 12, 17, 25, 29]. Some studies have modeled the clickstream via such latent variable models. For the purpose of predicting user exit without purchase, these are hidden Markov models [5, 24] and Markov-modulated marked point processes (M3PP) [11]. Here, the idea is that latent variables capture different yet unknown states (i. e., shopping phases). This allows latent variable models to model states representing goal-directed and browsing behavior that can eventually be recovered [5, 11, 24]. Traditionally, the states are modeled as being discrete and, in practice, there is an upper limit for the

number of states to prevent overfitting [4]. Moreover, latent variable models have a key limitation: long-term dependencies within clickstream are typically ignored.

Table 1: Overview of key literature for predicting user exits without purchase from clickstreams.

Model	Long-term dependence	Latent states
LSTM [16, 32]	✓	
BiLSTM [39]	✓	
Mixture LSTM [31, 36]	✓	
HMM [5, 24]		✓
M3PP [11]		✓
***Clickstream*DMM** (ours)	✓	✓

Research gap: To the best of our knowledge, there is no prior work that combines the strengths of neural networks and latent variable models for predicting user exits without purchase. Therefore, we propose a novel deep probabilistic model called *Clickstream*DMM which captures both (1) long-term dependencies and (2) latent states with varying shopping phases in clickstream data.

3 PROPOSED *Clickstream*DMM

In this section, we describe our prediction task, as well as the model components of our *Clickstream*DMM.

3.1 Prediction Task

We predict whether a user is at the risk of exiting without purchase. The task is analogous to earlier research [5, 11, 16, 18, 24, 31, 36, 39].

Input: The input to the prediction is given by clickstream data $\mathcal{D}$. Clickstream data comprises a set of sessions, $d = 1, \ldots, D$. Each session is a sequence of pages that have been visited by a user together with timestamps of when they were visited. Note that sessions can have variable lengths (i. e., number of pages visited).

Output: The prediction should output y^d for a given session whether the user will end the session with vs. without purchase.

Specifically, we denote a clickstream dataset by $\mathcal{D}$. It comprises D sessions, i. e., $\mathcal{D} = \{\mathcal{S}^d\}_{d=1}^{D}$, where $\mathcal{S}^d$ is the d-th session. The d-th session is defined as $\mathcal{S}^d = (\{t_m^d, x_m^d, \}_{m=1}^{M_d}, s^d, T^d, y^d)$, with s^d being some static user features (e. g., gender and age) and $x_m^d \in \mathcal{X}$ being the m-th page visited at time t_m^d, and $\mathcal{X}$ is the set of available pages on the website. The total number of pages observed during this session is given by M_d, and the duration of the session (in seconds) is T^d. We further compute the time spent on page x_m^d, i. e., $t_{m+1}^d - t_m^d$, which we refer to by τ_m^d. The time spent on a page is an important quantity in clickstream analytics [11], and we thus make later deliberate choices of how it enters our model. The outcome of the session $\mathcal{S}^d$ is denoted by y^d, which is either a purchase ($y^d = 1$) or an exit with no purchase ($y^d = 0$).

3.2 Specification of *Clickstream*DMM

Our *Clickstream*DMM is a deep probabilistic model tailored to predict the users at the risk of exiting with no purchase. Aligned with marketing theory, our model captures both (1) long-term dependencies in the clickstream data and (2) different shopping phases during the online session.

Our *Clickstream*DMM has 5 components (see Fig. 1). These are: a **(1) transition network** models the transition between latent variables z_{m-1}^d and z_m^d for the d-th session. These latent variables model a user's latent shopping phases. Based on the set of latent variables $\{z_m^d\}_{m=1}^{M_d}$, the page clicks $\{x_m^d\}_{m=1}^{M_d}$ and the time spent on each page $\{\tau_m^d\}_{m=1}^{M_d}$ are modeled conditionally independent from each other. This follows the motivation in earlier research on clickstream analytics [6, 33, 40]. A **(2) page-emission network** models the probability of the visited page x_m^d, and a **(3) time-emission network** models the probability of the time spent on the page τ_m^d given the latent variable. The previous three components of our model constitute the latent variable part of *Clickstream*DMM. An **(4) attention network** aggregates the information from the set of latent variables $\{z_m^d\}_{m=1}^{M_d}$ by further considering their corresponding timestamps $\{t_m^d\}_{m=1}^{M_d}$. In the end, the aggregated latent variable $\tilde{z}^d$ is processed by a **(5) predictor network** to infer the risk of exiting with purchase or with no purchase for a given session.

Notation. For ease of readability, we remove the superscript d and change the set notation for the mathematical equations below.[2] Further, let ReLU denote the rectified linear unit, $\odot$ denote the element-wise multiplication, and $[\cdot\,;\,\cdot]$ denote the concatenation of two vectors.

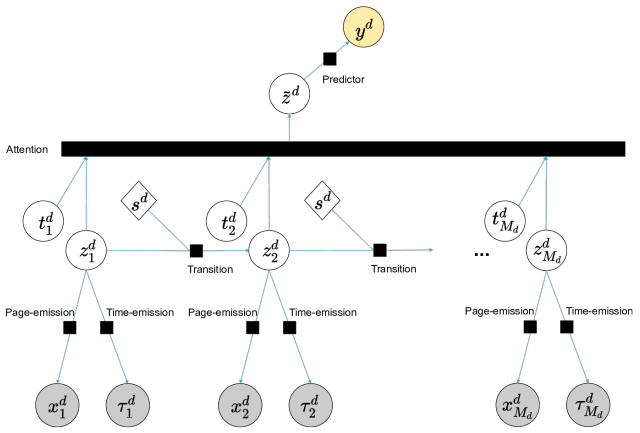


Figure 1: *Clickstream*DMM. Black squares denote neural networks.

3.3 Model Components

(1) Transition network: This component specifies the transition probability among the consecutive latent variables. The latent variables model the latent shopping phases of individual users. For this, we further accommodate static user features (e.g., gender, age) to consider heterogeneity among users.

Formally, the transition network learns the probability distribution of the latent variable z_m given z_{m-1} and the static user feature s, denoted by $p(z_m \mid z_{m-1}, s)$. This distribution is modeled as multivariate Gaussian distribution with a mean μ_{z_m} and

[2] For instance, set of page clicks $\{x_m^d\}_{m=1}^{M_d}$ is represented as $x_{1:M}$.

a diagonal covariance Σ_{z_m}, denoted as $\mathcal{N}(\mu_{z_m}, \Sigma_{z_m})$, where the diagonal entries of Σ_{z_m} are represented by the vector σ_{z_m}.[3] Then, we formalize the transition network as follows:

$$g'_{z_m} = \text{ReLU}(W^{tr}_{g'} [z_{m-1}; s] + b^{tr}_{g'}), \quad (1)$$
$$g_{z_m} = \text{sigmoid}(W^{tr}_{g} g'_{z_m} + b^{tr}_{g}), \quad (2)$$
$$\tilde{\mu}'_{z_m} = \text{ReLU}(W^{tr}_{\tilde{\mu}'} [z_{m-1}; s] + b^{tr}_{\tilde{\mu}'}), \quad (3)$$
$$\tilde{\mu}_{z_m} = W^{tr}_{\tilde{\mu}} \tilde{\mu}'_{z_m} + b^{tr}_{\tilde{\mu}}, \quad (4)$$
$$\bar{\mu}_{z_m} = W^{tr}_{\bar{\mu}} [z_{m-1}; s] + b^{tr}_{\bar{\mu}}, \quad (5)$$
$$\mu_{z_m} = g_{z_m} \odot \tilde{\mu}_{z_m} + (1 - g_{z_m}) \odot \bar{\mu}_{z_m}, \quad (6)$$
$$\sigma_{z_m} = \text{softplus}(W^{tr}_{\sigma} \text{ReLU}(\tilde{\mu}_{z_m}) + b^{tr}_{\sigma}) + \text{const.}, \quad (7)$$
$$z_m \sim p(z_m \mid z_{m-1}, s) = \mathcal{N}(\mu_{z_m}, \Sigma_{z_m}) \quad (8)$$

with matrices $W^{tr}_{g'}$, W^{tr}_{g}, $W^{tr}_{\tilde{\mu}'}$, $W^{tr}_{\tilde{\mu}}$, $W^{tr}_{\bar{\mu}}$, and W^{tr}_{σ} and bias vectors $b^{t}_{g'}$, b^{tr}_{g}, $b^{tr}_{\tilde{\mu}'}$, $b^{tr}_{\tilde{\mu}}$, $b^{tr}_{\bar{\mu}}$, and b^{tr}_{σ}.

(2) Page-emission network: This component models the probability of observing a specific page by the user given the latent variable.

Formally, the page-emission network learns the probability distribution of observing a page x_m given the latent variable z_m, which is denoted by $p(x_m \mid z_m)$. This distribution is modeled as a categorical distribution, denoted as Categorical(α_{x_m}), where the vector α_{x_m} contains the probabilities of $|\mathcal{X}|$ possible page types. This yields

$$h_{x_m} = \text{ReLU}(W^{pe}_{h} z_m + b^{pe}_{h}), \quad (9)$$
$$h'_{x_m} = \text{ReLU}(W^{pe}_{h'} h_{x_m} + b^{pe}_{h'}), \quad (10)$$
$$\alpha_{x_m} = \text{softmax}(W^{pe}_{\alpha} h'_{x_m} + b^{pe}_{\alpha}), \quad (11)$$
$$x_m \sim p(x_m \mid z_m) = \text{Categorical}(\alpha_{x_m}) \quad (12)$$

with matrices W^{pe}_{h}, $W^{pe}_{h'}$, and W^{pe}_{α} and bias vectors b^{pe}_{h}, $b^{pe}_{h'}$, and b^{pe}_{α}.

(3) Time-emission network: This component models the probability of the time spent on a page by the user given the latent variable.

Formally, this network learns the probability distribution of the time spent on a page τ_m given the latent variable z_m, denoted by $p(\tau_m \mid z_m)$.[4] This distribution is modeled as Gaussian distribution with a mean μ_{τ_m} and a variance $\sigma^2_{\tau_m}$, denoted as $\mathcal{N}(\mu_{\tau_m}, \sigma^2_{\tau_m})$. The mathematical formulation of this network is the following:

$$g'_{\tau_m} = \text{ReLU}(W^{et}_{g'} z_m + b^{et}_{g'}), \quad (13)$$
$$g_{\tau_m} = \text{sigmoid}(W^{et}_{g} g'_{\tau_m} + b^{et}_{g}), \quad (14)$$
$$\tilde{\mu}'_{\tau_m} = \text{ReLU}(W^{et}_{\tilde{\mu}'} z_m + b^{et}_{\tilde{\mu}'}), \quad (15)$$
$$\tilde{\mu}_{\tau_m} = W^{et}_{\tilde{\mu}} \tilde{\mu}'_{\tau_m} + b^{et}_{\tilde{\mu}}, \quad (16)$$
$$\bar{\mu}_{\tau_m} = W^{et}_{\bar{\mu}} z_m + b^{et}_{\bar{\mu}}, \quad (17)$$
$$\mu_{\tau_m} = g_{\tau_m} \odot \tilde{\mu}_{\tau_m} + (1 - g_{\tau_m}) \odot \bar{\mu}_{\tau_m}, \quad (18)$$
$$\sigma_{\tau_m} = \text{softplus}(W^{et}_{\sigma} \text{ReLU}(\tilde{\mu}_{\tau_m}) + b^{et}_{\sigma}) + \text{const.}, \quad (19)$$
$$\tau_m \sim p(\tau_m \mid z_m) = \mathcal{N}(\mu_{\tau_m}, \sigma^2_{\tau_m}) \quad (20)$$

with matrices $W^{te}_{g'}$, W^{te}_{g}, $W^{te}_{\tilde{\mu}'}$, $W^{te}_{\tilde{\mu}}$, $W^{te}_{\bar{\mu}}$, and W^{te}_{σ} and bias vectors $b^{e}_{g'}$, b^{te}_{g}, $b^{te}_{\tilde{\mu}'}$, $b^{te}_{\tilde{\mu}}$, $b^{te}_{\bar{\mu}}$, and b^{te}_{σ}.

(4) Attention network: This component aggregates the information carried by the latent shopping phases modeled by set of latent variables $z_{1:M}$. The attention network is carefully tailored to give more importance to the most recent shopping phases towards the prediction task.

Formally, the set of latent variables $z_{1:M}$ constructs the aggregated representation of the latent variables, denoted as $\tilde{z}$. Here, different from most attention mechanisms, the attention weights are subject to a decay. As such, more weight is given to recent shopping phases. For this, we use the inverse scaling based on the time difference between t_M and t_m to the power of β, which can be learned during the training process. The attention network is formalized via

$$z'_m = \tanh(W_{z'} z_m + b_{z'}), \quad (21)$$
$$\gamma_m = \frac{\exp(K(v, z'_m))/(t_M + 1 - t_m)^{\beta}}{\sum_{m'=1}^{M} \exp(K(v, z'_{m'}))/(t_M + 1 - t_{m'})^{\beta}}, \quad (22)$$
$$\tilde{z} = \sum_{m=1}^{M} \gamma_m z_m \quad (23)$$

with matrix $W_{z'}$, bias vector $b_{z'}$, query vector v, and scalar β, where $K(\cdot, \cdot)$ defines the dot product scaled by the inverse square root of the dimension v, i. e.,

$$K(v, z'_m) = \frac{v \cdot z'_m}{\sqrt{d_v}}. \quad (24)$$

(5) Predictor network: This component is the final step of inferring the risk of user exit with/without purchase. For this, the predictor network takes the aggregated representation of the latent variables $\tilde{z}$ as an input, and it yields the prediction $\hat{y}$, i.e.,

$$\hat{y} = \text{sigmoid}(U_y \text{ReLU}(W_y \tilde{z} + b_y) + c_y) \quad (25)$$

with matrices W_y and U_y and bias vectors b_y and c_y.

3.4 Posterior approximation

In order to estimate the parameters of our *Clickstream*DMM, we compute the posterior distribution of the latent variables $p(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s)$. To compute the posterior distribution efficiently despite the complexity of our model, we develop a posterior approximation of the latent variables (see Fig. 2), which is updated via stochastic variational inference.

[3] A constant is added to the variance terms to ensure the stability in the ELBO computation.

[4] For the numerical stability of the training process, we applied log transformation to τ_m, as motivated in [28]

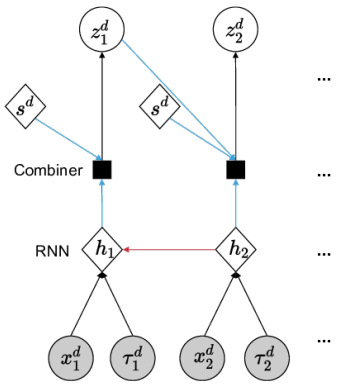


Figure 2: Posterior approximation of *Clickstream*DMM. Black squares denote neural networks.

Our *Clickstream*DMM approximates the posterior of the latent variables based on page clicks, time spent on each page, and the static user features. Formally, the posterior approximation computes $q(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s)$. For this, we first utilize the conditional independence of the latent variables, i. e.,

$$q(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s,) = \prod_{m=1}^{M} q(z_m \mid z_{m-1}, x_{1:M}, \tau_{1:M}, s). \tag{26}$$

We further hypothesize that the latent variable z_{m-1} contains the relevant past information. Therefore, it suffices to condition z_m only on *future* page clicks (i. e., $x_{m:M}$) and *future* time spent on them (i. e., $\tau_{m:M}$) instead of the entire sequence (i. e., $x_{1:M}$ and $\tau_{1:M}$). This leads to

$$q(z_m \mid z_{m-1}, x_{1:M}, \tau_{1:M}, s) = q(z_m \mid z_{m-1}, x_{m:M}, \tau_{m:M}, s). \tag{27}$$

Here, we use a recurrent neural network (RNN) to summarize the information contained by the sequence of $x_{m:M}$ and $\tau_{m:M}$ into the hidden state h_m, i.e.,

$$h_m = \text{ReLU}(W_h\,[x_m;\, \tau_m] + U_h\, h_{m+1} + b_h) \tag{28}$$

with matrices W_h, U_h and bias vector b_h. Next, we use another neural network component called **combiner network** (see Fig. 3.4) to aggregate all the relevant information to approximate the latent variable z_m. Formally, our combiner network takes z_{m-1}, h_m, and s as input to model the posterior approximation of z_m. In this setting, z_{m-1} and h_m contain information the past and future observations of the clickstream data, and s contains the relevant information to capture the heterogeneity between the users. Similar to the transition network, $q(z_m \mid z_{m-1}, x_{m:M}, \tau_{m:M}, s)$ is modeled as multivariate Gaussian distribution $\mathcal{N}(\mu_{z_m}, \Sigma_{z_m})$, where Σ_{z_m} is the diagonal covariance matrix with $\sigma^2_{z_m}$ on the diagonal, and 0 otherwise. The mathematical formulation of the combiner networks is as follows:

$$c_m = W_c\,[z_{m-1};\, s] + b_c, \tag{29}$$

$$\mu_{z_m} = W_\mu\,[h_m;\, c_m] + b_\mu, \tag{30}$$

$$\sigma_{z_m} = \text{softplus}(W_\sigma\,[h_m;\, c_m] + b_\sigma) + \text{const.}, \tag{31}$$

$$z_m \sim q(z_m \mid z_{m-1}, s, x_{m:M}, \tau_{m:M}) = \mathcal{N}(\mu_{z_m}, \Sigma_{z_m}) \tag{32}$$

with matrices W_c, W_μ, and W_σ and bias vectors b_c, b_μ, and b_σ. The above posterior approximation is leveraged to optimize the evidence lower bound (ELBO) via stochastic variational inference. For the details of the estimation procedure, see the Appendix A.

4 EXPERIMENTAL SETUP

4.1 Dataset

We evaluate the performance of our *Clickstream*DMM based on a real-world clickstream dataset provided by our partner company *Digitec Galaxus*, the leading online e-commerce platform in Switzerland. Its website is visited millions of times every month and, thereby, comprises a diversity of user profiles and shopping patterns.

In our clickstream dataset, each session contains a sequence with page clicks (Home, Account, Overview, Product, Marketing Content, Community, and Checkout) and their corresponding timestamps, i. e., when these pages were visited. Furthermore, the sessions include static user features such as the gender and age of the user, as well as the type of the user (i. e., B2B or B2C). The outcome of each session (exit with or without purchase) is further provided by our partner company. Note that the label becomes available only after a session is completed.

Overall, our clickstream dataset includes 26,279 sessions with a total of 250,287 page clicks. Each session contains between 5 and 50 page clicks. The duration of sessions has a mean of 14.58 minutes and 13.28 pages. This makes our dataset highly representative to the proprietary datasets in earlier research [11].

4.2 Baselines

We compare our *Clickstream*DMM against an extensive set of state-of-the-art baselines[5] that have been proposed in prior literature on clickstream analytics [5, 11, 16, 18, 24, 31, 32, 36, 39]. These models have been tailored for predicting the user exit probability with no purchase from the clickstream data.

On the one hand, we have implemented baselines that capture long-term dependencies within web sessions via recurrent neural networks. These are: long short-term memory (**LSTM**) [16, 32] and bi-directional LSTM (**BiLSTM**) [39]. We also implemented a mixture LSTM [31, 36] but it was on par or inferior to the above LSTMs and, hence, omitted for brevity. Further, we use a combination of LSTM and gradient boosting machine (**LSTM+GBM**) [16]. The previous networks are fed with full session information (e. g., static user features, sequence of page clicks and timestamps) and thus receive input identical to our model.

On the other hand, we use latent variable models as baselines and thus capture latent states (i. e., shopping phases) during web sessions. Here, we implemented the hidden Markov model (**HMM**)

[5]We also implemented some of the best practices from the recommender systems [19, 35] but omitted due to their inferior results. See Appendix E for details.

from [5, 24], where the discrete latent variables correspond to the latent states in web sessions. In addition, we use a Markov-modulated point process (**M3PP**) [11]. The latter incorporates the time spent on each page when constructing the latent variables. For better comparability, we also report Markov chains with varying orders from 1 to 3 (**MC-1**, **MC-2**, **MC-3**) [18]. These model page transitions based on last T pages.

5 RESULTS

In this section, we present the result of our experiments. In particular, we demonstrate the improvement of our *Clickstream*DMM over existing baselines on two prediction tasks. Moreover, we present the contribution to the improvement of every component of our model in an ablation study. Finally, we propose an approach for model interpretation via clustering the latent variable space, which reveals important insights for practitioners.

5.1 Prediction performance (overall)

In Table 2, we present the performance of predicting the risk of exiting with vs. without purchase when all page clicks of the online sessions are available. That is, the prediction is here evaluated at the last page before a session ends. This prediction setting is analogous to earlier research [5, 16, 24]; however, it does not allow for "early warnings". We nevertheless report the results for comparability but focus on a multi-step ahead prediction in the next section.

Among the state-of-the-art baselines, the best performance obtained by a BiLSTM with an AUROC of 0.913. In contrast, our *Clickstream*DMM achieves the AUROC of 0.926, which improves the performance by 1.4 %. Further, our *Clickstream*DMM achieves the best AUPRC with 0.690, improving over the best baseline performance (GBM and LSTM+GBM) by 0.9 %.

Table 2: Performance in prediction task in which the risk of exiting with no purchase is estimated based on full online session.

Model	AUROC	AUPRC
LSTM [16, 32]	0.899 ± 0.002	0.613 ± 0.006
BiLSTM [39]	0.913 ± 0.003	0.647 ± 0.009
GBM	0.897 ± 0.003	0.684 ± 0.010
LSTM+GBM [16]	0.909 ± 0.004	0.684 ± 0.013
HMM [5, 24]	0.756 ± 0.003	0.375 ± 0.002
M3PP [11]	0.765 ± 0.010	0.528 ± 0.012
MC-1 [18]	0.880 ± 0.005	0.634 ± 0.012
MC-2 [18]	0.874 ± 0.007	0.644 ± 0.011
MC-3 [18]	0.846 ± 0.010	0.606 ± 0.019
***Clickstream*DMM** (ours)	**0.926 ± 0.004**	**0.690 ± 0.013**

Higher is better. Best value in bold.

5.2 Prediction performance for early warnings

While the above evaluation setting was consistent with prior research, it is not representative of real-world deployments, where early warnings are needed so that e-commerce websites can trigger marketing interventions. Hence, to provide "early warnings" of exits without purchase, we make predictions as follows. We now evaluate the performance in predicting exits with vs. without purchase in a multi-step ahead setting. Specifically, we predict n steps ahead and further vary the time window n.

In Table 3, we present the average performance when making early warning predictions. The best baseline performance is achieved by BiLSTM with an AUROC of 0.733. In comparison, our *Clickstream*DMM outperforms the best baseline with an AUROC of 0.817. This is an improvement by 11.5 %. Further, our *Clickstream*DMM achieves the best performance in AUPRC with 0.569. This improves over the best baseline performance LSTM+GBM with an AUPRC of 0.505 by 12.7 %.

Table 3: Performance in prediction task in which the risk of exiting with no purchase is estimated starting from $n = 6$ pages ahead to exit.

Model	AUROC	AUPRC
LSTM [16, 32]	0.726 ± 0.014	0.472 ± 0.012
BiLSTM [39]	0.733 ± 0.017	0.497 ± 0.015
GBM	0.705 ± 0.006	0.500 ± 0.011
LSTM+GBM [16]	0.725 ± 0.011	0.505 ± 0.012
HMM [5, 24]	0.572 ± 0.005	0.259 ± 0.003
M3PP [11]	0.701 ± 0.012	0.447 ± 0.012
MC-1 [18]	0.732 ± 0.006	0.484 ± 0.009
MC-2 [18]	0.697 ± 0.007	0.456 ± 0.010
MC-3 [18]	0.655 ± 0.009	0.407 ± 0.011
***Clickstream*DMM** (ours)	**0.817 ± 0.007**	**0.569 ± 0.013**

Higher is better. Best value in bold.

This performance improvement of our *Clickstream*DMM facilitates earlier marketing interventions in real-world deployment. To illustrate this, Fig. 3 shows the performance of each model at each of the n pages ahead of exit. Across both metrics AUROC and AUPRC, our *Clickstream*DMM demonstrates considerable performance improvements over the existing methods, especially when "early" predictions are needed. For instance, the baseline models need to process ∼3 more pages than *Clickstream*DMM to yield the same performance.

5.3 Ablation study

We conduct an ablation study to investigate the contribution of the different components of *Clickstream*DMM. For this, we compare two variants of our *Clickstream*DMM in the following. (1) We examine the importance of attention and predictor networks for our model. Removing both, we essentially yield a mixture deep Markov model (DMM). Here, two DMMs are separately fitted to web sessions with different labels. We then perform the prediction task by comparing the likelihoods of observations under two different hypothesis of which model is more likely (i. e., the DMM trained for purchase or the DMM trained for no purchase). Thereby, we show the importance of directly incorporating the label information into the latent variable model. (2) Time spent on each page (TSP) has been earlier found decisive for online behavior [11]. Motivated by this, we now exclude the information TSP, and, hence, we remove the time-emission network and exclude TSP from posterior approximation (with suffix "w/o TSP"). Thereby, we show the importance of modeling the page sequence *together* with their TSPs.

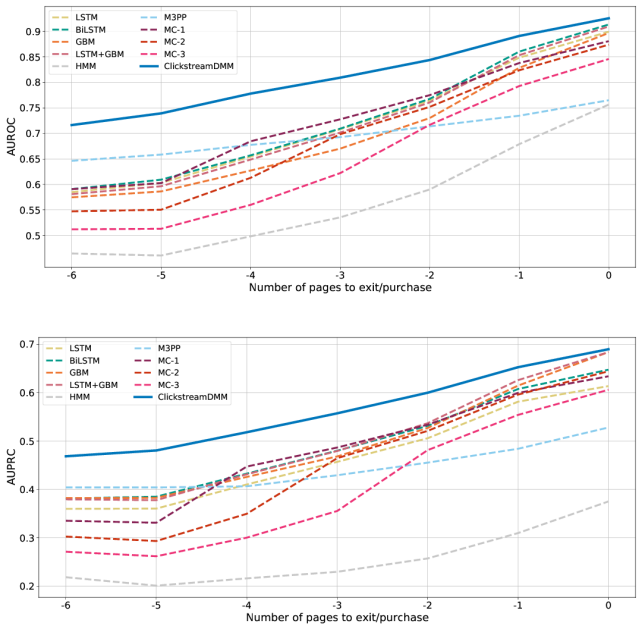


Figure 3: Prediction performance across varying time windows, that is, when making predictions n pages ahead of a user exit. Shown are two performance metrics: AUROC (top) and AUPRC (bottom).

Table 4 lists the prediction performance when the online sessions are fully available. In addition, Table 5 lists the average performance of the models starting from 6 pages ahead to exit. We further provide Fig. 8 in Appendix for the performance of each model for different pages ahead of exit. Overall, the mixture DMM shows an inferior performance. This demonstrates the importance of modeling predictions via a combination of attention network and predictor network. Further, our *Clickstream*DMM has a better prediction performance when processing TSP. For instance, in Table 5, our *Clickstream*DMM improves over the *Clickstream*DMM w/o TSP in terms of AUROC by 2 % (0.817 vs. 0.801) and in terms of AUPRC by 9.2 % (0.569 vs. 0.521). This justifies our choice behind building our *Clickstream*DMM based on two emission networks for pages and TSP, respectively. (We also tested other variants to fuse TSP into our model but none outperformed the above *Clickstream*DMM).

Table 4: Ablation study comparing different variants of *Clickstream*DMM in predicting risk of existing with vs. without purchase (here: data with full online sessions).

Model	AUROC	AUPRC
Mixture DMM	0.862 ± 0.009	0.577 ± 0.013
Mixture DMM w/o TSP	0.864 ± 0.010	0.575 ± 0.014
*Clickstream*DMM w/o TSP	0.904 ± 0.009	0.611 ± 0.009
***Clickstream*DMM**	$\mathbf{0.926 \pm 0.004}$	$\mathbf{0.690 \pm 0.013}$

Higher is better. Best value in bold.

5.4 Model interpretability for marketing insights

Consistent with earlier marketing research [5, 11, 24], we examine the latent variables produced by our model. Thereby, we provide additional marketing insights, which characterizes user behavior in different latent shopping phases. Specifically, we show that e-commerce websites may tailor marketing interventions to the latent shopping phases in order to target users based on specific behavior.

Table 5: Ablation study predicting the risk of exiting with vs without purchase starting from $n = 6$ pages ahead to exit.

Model	AUROC	AUPRC
Mixture DMM	0.754 ± 0.010	0.483 ± 0.012
Mixture DMM w/o TSP	0.756 ± 0.008	0.485 ± 0.011
*Clickstream*DMM w/o TSP	0.801 ± 0.007	0.521 ± 0.007
***Clickstream*DMM**	$\mathbf{0.817 \pm 0.007}$	$\mathbf{0.569 \pm 0.013}$

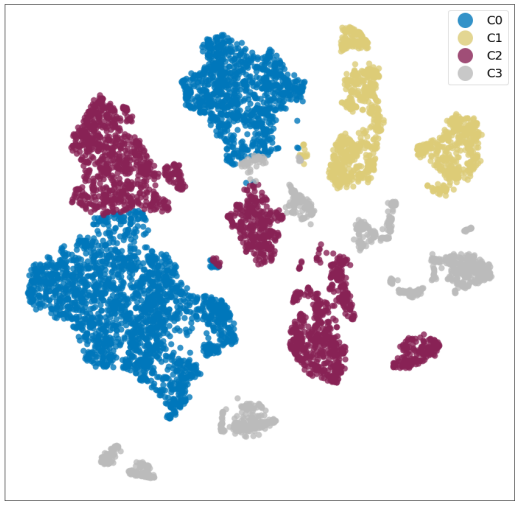
Higher is better. Best value in bold.

To extract the latent shopping phases of the users, we cluster the estimated latent variables via a Gaussian mixture model. The reason for choosing this clustering method is that our *Clickstream*DMM models the distribution of latent variables as a Gaussian distribution. To decide upon the number of clusters, we employ two metrics, namely the Bayesian information criterion (BIC) and the Silhouette score. Both scoring methods suggest to use $k = 4$ clusters. For terminology, we refer to the clusters as C0, C1, C2, and C3. We further visualize the latent variable clusters via a t-SNE plot [37] of latent variables (see Fig. 4a). The t-SNE plot reveals that the clusters have little to no overlap and are thus disjoint. This implies that the clusters are characterized by different behavior but with large within-cluster similarity. We discuss the different characteristics of the clusters (i. e., shopping phases) from three perspectives: risk of exit with no purchase, emission and transition.

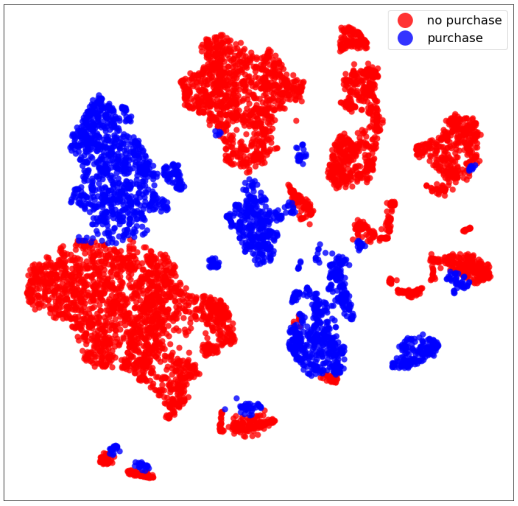
Risk of exit with no purchase. The clusters vary in the risk that users exit without purchase. Fig. 4b shows the same clusters, but with each data point colored by whether the sessions is concluded with purchase vs. no purchase. Hence, the clusters can be interpreted as follows: (1) C0 and C1 are the latent shopping phases where users typically exit the website without purchase. (2) C2 reveals a high tendency towards purchase before exiting the website. (3) C3 is of a mixed behavior.

Emissions. We further examine the characteristics of the clusters and their association with the shopping phases in marketing literature. For this, we report the relative frequency of visiting specific pages by each cluster (see Fig. 5). We map the different pages onto 7 categories. Here, we make the following observations: (1) C0 mostly comprises the visits of Overview and Product pages, meaning that the user visits the website for the exploratory purposes but without tendency toward purchasing. Such behavior has been termed "experiential browsing" in marketing [23, 27]. C1 are mostly visits of Home and and Marketing pages, meaning that the user again engages in browsing. (2) C2 has primarily Product and Checkout pages, suggesting that users are directly navigating to specific shopping items of interests without extensive search. Such behavior reflects marketing where it has termed "goal-directed search" in marketing [23, 27]. (3) C3 has clicks to different pages, including Home and Product along with Overview and Marketing. It thereby corresponds to mixed behavior of previous clusters.

Transitions. A user may go through different latent shopping phases in a single web session [14, 20]. Fig. 6 shows the transition



(a) Coloring shows the $k = 4$ different clusters.



(b) Coloring shows sessions with vs. without purchase.

Figure 4: t-SNE plot of latent variables.

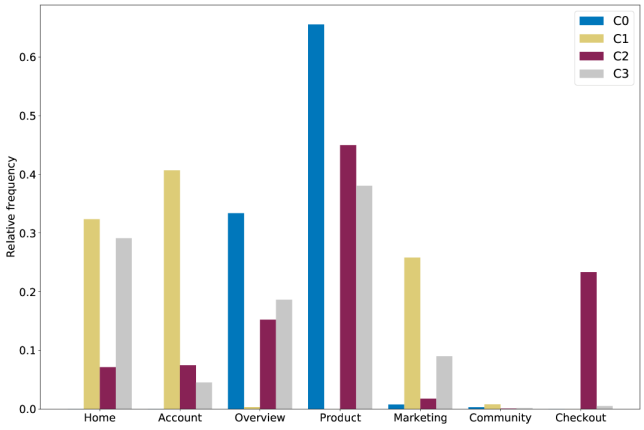
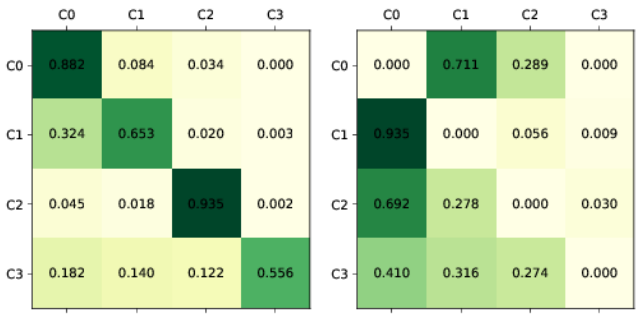


Figure 5: Relative frequency of page clicks by cluster (in %).



(a) with self-transition (b) without self-transition

Figure 6: Transition matrix between clusters (left: current, top: next).

between the clusters of the latent variables (with and without self-transition at each page click). All clusters are relatively sticky, i.e., the latent variable from the next page belongs to the previous cluster. This further differentiates users in C0 from users in C1: browsing in C0 is characterized by "flow", where users are likely to continue with the same shopping behavior, whereas C1 is "without flow", where users are likely to transition to a different cluster. This is further supported when comparing the average time spent on page (TSP) by cluster: the average TSP is larger for C0 (mean: 36.75 s) than for C1 (mean: 22.91 s), consistent with our flow hypothesis.

Overall, we summarize the characteristics of each cluster as follows: C0 is mostly "browsing in flow", C1 is "browsing without flow", C2 is "goal-directed search", and C3 is of mixed nature.

6 DISCUSSION

Theory-informed model: We develop a tailored attentive deep Markov model called *Clickstream*DMM for predicting users at the risk of exiting without purchase in e-commerce web sessions. Specifically, our *Clickstream*DMM jointly models both (1) long-term dependencies of page sequences and (2) latent shopping phases of the online users. This is aligned with marketing theory [23, 27], as we model (1) long-term dependencies via an attention network and (2) latent shopping phases via latent variables, which allow to distinguish between "goal-directed search" and "browings". This explains the superior performance of our model.

Performance: Our *Clickstream*DMM is a powerful tool for making accurate predictions of whether a user is about to exit the e-commerce website without a purchase. Our model is particularly effective when predictions are made multiple pages ahead. This is needed upon deployment, so that early warnings are generated to trigger marketing interventions that steer users toward purchase.

For this task, our *Clickstream*DMM yields a substantial performance improvement in AUROC by 11.5 % and in AUPRC by 12.7 %.

Marketing insights: Our *Clickstream*DMM successfully models the latent dynamics behind the page clicks of the user. In this way, we identify four clusters within the latent shopping phases of the online users, which are characterized by flow and goal-directed search vs. browsing. Hence, e-commerce websites may use the interpretation of the latent variables to tailor marketing interventions accordingly. For instance, they may trigger different interventions for "browsing with flow" and "browsing without flow". Users in the former group should not be interrupted by interventions as long as they are in flow, while users in the latter group may need more profound interventions (e. g., coupons, price discounts).

Future research: We proposed *Clickstream*DMM to predict the user exits without purchase based on the clickstream data. *Clickstream*DMM may facilitate other settings, where the sequences of events have both long-term dependencies and latent dynamics. Examples may be churn prediction and online fraud prediction.

REFERENCES

[1] David Adedayo Adeniyi, Zhaoqiang Wei, and Y Yongquan. 2016. Automated web usage data mining and recommendation system using k-Nearest Neighbor (kNN) classification method. *Applied Computing and Informatics* 12, 1 (2016), 90–108.

[2] Shelby D Bernhard, Carson K Leung, Vanessa J Reimer, and Joshua Westlake. 2016. Clickstream prediction using sequential stream mining techniques with Markov chains. In *International Database Engineering & Applications Symposium*.

[3] Veronika Bogina and Tsvi Kuflik. 2017. Incorporating dwell time in session-based recommendations with recurrent neural networks.. In *RecTemp@RecSys*.

[4] Mukund Deshpande and George Karypis. 2004. Selective Markov models for predicting web page accesses. *ACM Transactions on Internet Technology* 4, 2 (2004), 163–184.

[5] Amy Wenxuan Ding, Shibo Li, and Patrali Chatterjee. 2015. Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research* 26, 2 (2015), 339–359.

[6] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *KDD*.

[7] Magdalini Eirinaki, Michalis Vazirgiannis, and Dimitris Kapogiannis. 2005. Web path recommendations based on page ranking and Markov models. In *WIDM*.

[8] Alex Gofman, Howard R. Moskowitz, and Tõnis Mets. 2009. Integrating science into web design: Consumer-driven web site optimization. *Journal of Consumer Marketing* 26, 4 (2009), 286–298.

[9] Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, and Dawei Yin. 2020. Hierarchical user profiling for e-commerce recommender systems. In *WSDM*.

[10] Şule Gündüz and M Tamer Özsu. 2003. A web page prediction model based on click-stream tree representation of user behavior. In *KDD*.

[11] Tobias Hatt and Stefan Feuerriegel. 2020. Early detection of user exits from clickstream data: A Markov modulated marked point process model. In *WWW*.

[12] Tobias Hatt and Stefan Feuerriegel. 2021. Sequential Deconfounding for Causal Inference with Unobserved Confounders. *arXiv preprint arXiv:2104.09323* (2021).

[13] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based recommendations with recurrent neural networks. (2016).

[14] Donna L Hoffman and Thomas P Novak. 1996. Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing* 60, 3 (1996), 50–68.

[15] Porter Jenkins. 2019. ClickGraph: Web page embedding using clickstream data for multitask learning. In *WWW*.

[16] Dennis Koehn, Stefan Lessmann, and Markus Schaal. 2020. Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications* 150 (2020), 113342.

[17] Milan Kuzmanovic, Tobias Hatt, and Stefan Feuerriegel. 2021. Deconfounding Temporal Autoencoder: estimating treatment effects over time using noisy proxies. In *Machine Learning for Health*. PMLR.

[18] Choudur Lakshminarayan, Ram Kosuru, and Meichun Hsu. 2016. Modeling complex clickstream data by stochastic models. In *WWW Companion*. 879–884.

[19] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural attentive session-based recommendation. In *CIKM*.

[20] Naomi Mandel and Eric J Johnson. 2002. When web pages influence choice: Effects of visual primes on experts and novices. *Journal of Consumer Research* 29, 2 (2002), 235–245.

[21] William C McDowell, Rachel C Wilson, and Charles Owen Kile Jr. 2016. An examination of retail website design and conversion rate. *Journal of Business Research* 69, 11 (2016), 4837–4842.

[22] Bamshad Mobasher. 2007. Data mining for web personalization. In *The Adaptive Web*. Springer, 90–135.

[23] Wendy W Moe. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology* 13, 1-2 (2003), 29–39.

[24] Alan L Montgomery, Shibo Li, Kannan Srinivasan, and John C Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing Science* 23, 4 (2004), 579–595.

[25] Christof Naumzik, Stefan Feuerriegel, and Markus Weinmann. 2021. I Will Survive: Predicting Business Failures from Customer Ratings. *Marketing Science* (2021).

[26] Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, and Jiro Iwanaga. 2018. A latent-class model for estimating product-choice probabilities from clickstream data. *Information Sciences* 429 (2018), 406–420.

[27] Thomas P Novak, Donna L Hoffman, and Adam Duhachek. 2003. The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology* 13, 1-2 (2003), 3–16.

[28] Takahiro Omi, Naonori Ueda, and Kazuyuki Aihara. 2019. Fully neural network based model for general temporal point processes. In *NeurIPS*.

[29] Yilmazcan Ozyurt, Mathias Kraus, Tobias Hatt, and Stefan Feuerriegel. 2021. AttDMM: An Attentive Deep Markov Model for Risk Scoring in Intensive Care Units. In *KDD*.

[30] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *RecSyS*.

[31] Borja Requena, Giovanni Cassani, Jacopo Tagliabue, Ciro Greco, and Lucas Lacasa. 2020. Shopper intent prediction from clickstream e-commerce data with minimal browsing information. *Scientific Reports* 10, 1 (2020).

[32] Cemal Okan Sakar, Suleyman Olcay Polat, Mete Katircioglu, and Yomi Kastro. 2019. Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications* 31 (2019), 6893–6908.

[33] Oleksandr Shchur, Marin Biloš, and Stephan Günnemann. 2020. Intensity-Free learning of temporal point processes. In *ICLR*.

[34] Statista. 2021. Conversion rate of online shoppers worldwide as of 3rd quarter 2020.

[35] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *CIKM*.

[36] Arthur Toth, Louis Tan, Giuseppe Di Fabbrizio, and Ankur Datta. 2017. Predicting shopping behavior with mixture of RNNs. In *CEUR Workshop Proceedings*, Vol. 2311.

[37] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008), 2579–2605.

[38] Armando Vieira. 2015. Predicting online user behaviour using deep learning algorithms. *arXiv preprint arXiv:1511.06247* (2015).

[39] Zhenzhou Wu, Bao Hong Tan, Rubing Duan, Yong Liu, and Rick Siow Mong Goh. 2015. Neural modeling of buying behaviour for e-commerce from clicking patterns. In *RecSyS Challenge*.

[40] Shuai Xiao, Junchi Yan, Xiaokang Yang, Hongyuan Zha, and Stephen Chu. 2017. Modeling the intensity function of point process via recurrent neural networks. In *AAAI*.

[41] Jinyoung Yeo, Seung-won Hwang, Eunyee Koh, Nedim Lipka, et al. 2018. Conversion prediction from clickstream: Modeling market prediction and customer predictability. *IEEE Transactions on Knowledge and Data Engineering* 32, 2 (2018), 246–259.

[42] Jiaxuan You, Yichen Wang, Aditya Pal, Pong Eksombatchai, Chuck Rosenburg, and Jure Leskovec. 2019. Hierarchical temporal convolutional networks for dynamic recommender systems. In *WWW*.

[43] Meizi Zhou, Zhuoye Ding, Jiliang Tang, and Dawei Yin. 2018. Micro behaviors: A new perspective in e-commerce recommender systems. In *WSDM*.

[44] Yu Zhu, Hao Li, Yikang Liao, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. 2017. What to do next: Modeling user behaviors by time-LSTM.. In *IJCAI*.

A ESTIMATION DETAILS

Loss function: *Clickstream*DMM minimizes the following loss function during training:

$$\mathcal{L}(y, \hat{y}, \mathcal{S}) = \ell(y, \hat{y}) - \alpha \, \mathrm{ELBO}(\mathcal{S}), \tag{33}$$

where the two terms $\ell(y, \hat{y})$ and the evidence lower bound (ELBO) of a session $\mathcal{S}$ are described below.

Cross-entropy loss: The term $\ell(y, \hat{y})$ denotes weighted cross-entropy loss between the observed label y and the corresponding prediction of exiting without purchase $\hat{y}$. It is given by

$$\ell(y, \hat{y}) = -\rho\, y\, \log(\hat{y}) - (1 - y)\, \log(1 - \hat{y}) \tag{34}$$

with a weight $\rho = \frac{|\{y \in Y : y = \text{"exit with no purchase"}\}|}{|\{y \in Y : y = \text{"exit with purchase"}\}|}$ denoting the ratio of users who exited the online website without purchase vs. who made a purchase. Such weight in the cross-entropy loss helps our *Clickstream*DMM in discriminating the minority class (i. e., purchase from the website) in the imbalanced dataset.

ELBO: We denote the evidence lower bound via ELBO. It regulates the generative part of our model and serves as a regularization term in the loss function. The weight of ELBO is parametrized by α, which can be tuned as the other hyperparameters of the model. The ELBO formulation is given by

$$\begin{aligned}\text{ELBO}(\mathcal{S}) = \mathbb{E}_{q(z_{1:M} \mid x_{1:M},\, \tau_{1:M},\, s)}[\log p(x_{1:T},\, \tau_{1:T} \mid z_{1:T})] \\ - \text{KL}(q(z_{1:M} \mid x_{1:M},\, \tau_{1:M},\, s) \mid\mid p(z_{1:M})).\end{aligned} \tag{35}$$

In the above formulation, the expectation term denotes the expected log-likelihood of $x_{1:M}$ and $\tau_{1:M}$ given the latent variables $z_{1:M}$. The KL-divergence measures the distance between the posterior approximation and the prior formulation of the latent variables $z_{1:M}$, which is minimized by *Clickstream*DMM during the training.

In this ELBO formulation, $q(z_{1:M} \mid x_{1:M},\, \tau_{1:M},\, s)$ is decomposed into further terms, as explained in Sec. 3.4. Hence, the posterior approximation computes the following:

$$q(z_{1:M} \mid x_{1:M}, \tau_{1:M}, s) = \prod_{m=1}^{M} q(z_m \mid z_{m-1}, x_{m:M}, \tau_{m:M}, s). \tag{36}$$

Similarly, $p(x_{1:M},\, \tau_{1:M} \mid z_{1:M})$ is decomposed into individual terms of x_m and τ_m, which are produced by the two emission networks of *Clickstream*DMM. Specifically, given the latent variables, we model the emission probabilities of pages and time spent on pages as conditionally independent from each other. Thereby, we write

$$p(x_{1:M},\, \tau_{1:M} \mid z_{1:M}) = p(x_{1:M} \mid z_{1:M})\, p(\tau_{1:M} \mid z_{1:M}). \tag{37}$$

Further, based on the Markov property of our *Clickstream*DMM, we model that the observation of one time-step is conditionally independent from the other observations given the latent variables. Thereby, we further decompose the probability distributions of pages and time spent on pages as follows:

$$p(x_{1:M} \mid z_{1:M}) = \prod_{m=1}^{M} p(x_m \mid z_m), \tag{38}$$

$$p(\tau_{1:M} \mid z_{1:M}) = \prod_{m=1}^{M} p(\tau_m \mid z_m). \tag{39}$$

Sampling: The continuous distribution of the latent variables and the non-linearity encoded in components of *Clickstream*DMM introduce complexity for the computation of ELBO. To alleviate such complexity, we use Monte Carlo sampling for the posterior of latent variables and minimize ELBO via stochastic variational inference.

B PREDICTION OF EXIT WITHOUT PURCHASE IN REAL TIME

*Clickstream*DMM estimates the probability of the user exit without purchase in real time, and updates its estimate with every new page click.

At any given time t_M, *Clickstream*DMM takes the sequence of page clicks $x_{1:M}$, the sequence of time spent on each page $\tau_{1:M}$ and the user's static user feature s. Given these inputs, *Clickstream*DMM samples the sequence of latent variables $z_{1:M}$ from the posterior approximation. The sampling procedure is repeated N times, yielding the set of sampled latent variables $\{z^n_{1:M}\}_{n=1}^N$. After sampling, the latent variables $\{z^n_{1:M}\}_{n=1}^N$ are coupled with their corresponding timestamps $t_{1:M}$ and fed into the attention network. The output is the set of aggregated representation of latent variables, denoted by $\{\tilde{z}^n\}_{n=1}^N$. As the final step, the predictor network takes $\{\tilde{z}^n\}_{n=1}^N$ as input and yields the set of probabilities for user exit with no purchase, denoted as $\{\hat{y}^n\}_{n=1}^N$. The mean of $\{\hat{y}^n\}_{n=1}^N$ is used as the estimate for the probability of user exit with no purchase.

C DATA PREPARATION

Aligned with the prior research [7, 11, 18, 24], we applied the standard preprocessing before feeding the clickstream data into the models. We categorized the pages into seven distinct page types as Home, Account, Overview, Product, Marketing Content, Community, and Checkout (incl. Shopping Cart). With the built-in tool provided by our partner company, we discarded sessions not originated by a human (e. g., generated by the web crawlers). To capture long-term dependencies among page clicks, we discarded the sessions having less than 5 page clicks. To eliminate outliers, we excluded the sessions having more than 50 page clicks. Overall, the resulting clickstream dataset comprises 26,279 sessions.

D IMPLEMENTATION DETAILS

We split the dataset into 5 folds of the same size. Each fold has the same ratio of *purchase*-to-*no purchase* for user exits. We run all models 5 times such that each fold has been the test set once and the other 4 folds has been used for training (3 folds) and validation (1 fold for early stopping). Thereby, in all experiments, we report the results averaged over 5 folds. Further, we include the standard deviation of the results across folds.

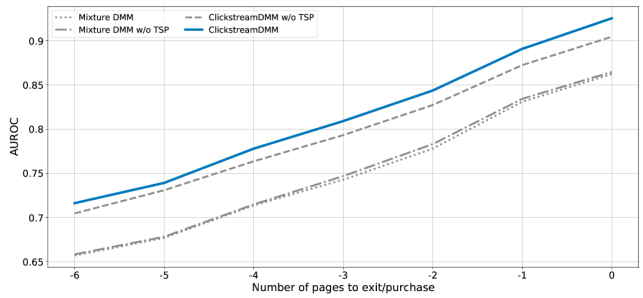
We implemented all models (except M3PP) in Python. Overall, we used the following libraries for each model: pyro for *Clickstream*DMM, PyTorch for LSTM and BiLSTM, scikit-learn for GBM, and hmmlearn for HMM. We used the original implementation of M3PP, which is provided in Stan. We trained and evaluated all models using an NVIDIA TITAN V GPU with 12 GB of memory.

E PERFORMANCE OF RECOMMENDER SYSTEMS

We experimented with some of the best practices from the recommender systems such as Neural Attentive Recommendation Machine (NARM) [19] and BERT4Rec [35]. However, we omitted them in the main results due to their inferior performance and their *different* task specification. For completeness, we show their prediction performance for early warning in Figure 7.

F PREDICTION OF MINORITY CLASS

Table 6 and Table 7 show how our *Clickstream*DMM succeeded to predict the minority class (i.e., purchase) in imbalanced dataset.



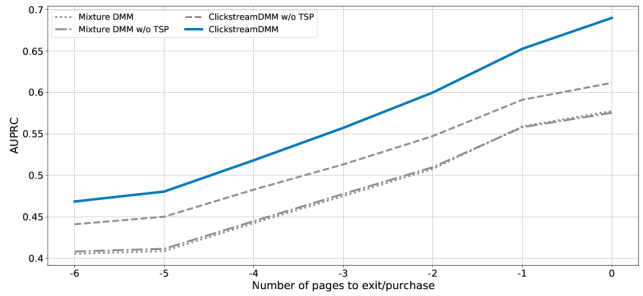


Figure 8: Ablation study comparing different variants of our *Clickstream*DMM when making predictions n pages ahead of an user exit with/without purchase. Shown are two performance metrics: AUROC (top) and AUPRC (bottom).

Table 8: Hyperparameters used for tuning the baselines.

Model	Tuning parameters	Tuning range
LSTM, BiLSTM	Dimension of cell state	2, 4, 8, 16, 24
	Number of layers	1, 2, 3, 4
	Dropout rate	0, 0.1, 0.2, 0.5
	Learning rate	0.00002, 0.0002, 0.001, 0.01
	Batch size	32, 64, 128, 256
GBM	Number of trees	50, 100, 500, 1000
	Maximum depth	3, 5, 7, 9, 11
	Class weight	none, balanced
LSTM+GBM*	Model weights	25-75, 50-50, 75-25
HMM, M3PP	Number of states	2, 3, ..., 10

* Other parameters of LSTM+GBM are tuned in the same way as in LSTM and GBM.

Table 9: Hyperparameters for tuning *Clickstream*DMM.

Tuning parameters	Tuning range
Dimension of latent variable	2, 4, 8, 12, 16
Dimension of transition hidden layer	4, 8, 12, 16
Dimension of page-emission hidden layer	4, 8, 12, 16
Dimension of time-emission hidden layer	4, 8, 12, 16
Dimension of attention mechanism	8, 12, 16, 24
Dimension of MLP hidden layer	2, 4, 8, 12
Dimension of RNN cell	4, 8, 16, 24
Regularization strength of ELBO (α)	0.001, 0.01, 0.1, 1
Number of Monte Carlo samples (N)	1, 2, 5, 10, 20
Learning rate	0.00002, 0.0002, 0.001, 0.01
Batch size	32, 64, 128, 256

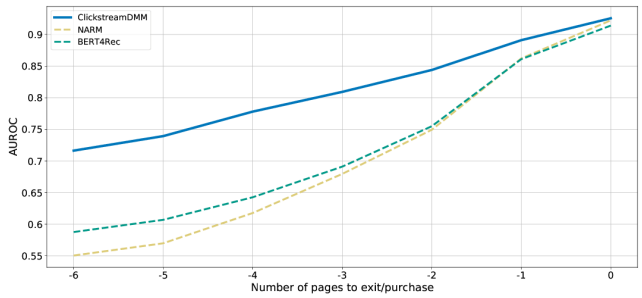


Figure 7: Early warning performance of recommender systems against our *Clickstream*DMM.

Table 6: Confusion matrix of *Clickstream*DMM at Section 5.1

		Ground Truth: No Purchase	Ground Truth: Purchase
Prediction	**No Purchase**	3406	662
	Purchase	30	1158

Table 7: Confusion matrix of *Clickstream*DMM at Section 5.2

		Ground Truth: No Purchase	Ground Truth: Purchase
Prediction	**No Purchase**	24809	2969
	Purchase	3269	4948

G PERFORMANCE OF *Clickstream*DMM VARIANTS

Fig. 8 compares the performance of each *Clickstream*DMM variant at different pages ahead of exit.

H TRAINING DETAILS

The large parameter space of the models makes grid search computationally expensive. For this reason, we employed a *ceteris paribus* strategy, meaning that we tuned each model parameter individually while keeping the others fixed. We iterated this procedure for a couple of loops, until reaching convergence in the validation score. For the experiment results, we used the hyperparamer set providing the best result for each model.

Table 8 and Table 9 list the tuning parameters of the baseline models and of our *Clickstream*DMM, respectively.