EARLY BIRD - PREDICT HEALTHY PRODUCT CHOICES IN VIRTUAL COMMERCE

Completed Research Paper

Tobias Weiss, Justus Liebig University Gießen, Gießen, Germany, tobias.weiss@wi.jlug.de

Jella Pfeiffer, University Stuttgart, Stuttgart, Germany, jella.pfeiffer@bwi.uni-stuttgart.de

Thies Pfeiffer, University of Applied Sciences Emden/Leer, Emden/Leer, Germany, this.pfeiffer@hs-emden-leer.de

Abstract

Due to advances in extended reality technology, an increasing number of head-mounted displays are equipped with eye trackers. These sensors allow to predict customers' preferences on-the-fly. Such information can serve as features for recommender systems. We propose to treat eye tracking data as time series and utilize a deep time series classifier for inference. Our evaluation investigates possibly early predictions about customer preferences for healthy products in a virtual reality environment. The results, that are based on data from a large-scale laboratory experiment, demonstrate superior performance of the time series classifier, compared to a shallow gradient boosting baseline. They indicate a trade-off between prediction quality and how early this prediction is made. Overall, our study suggests that eye tracking and time series classification are valuable avenues for research and practice. Adaptive (shopping) assistants and recommendations based on artificial intelligence and bio sensors seem to be in close vicinity.

Keywords: Extended Reality, Eye Tracking, Healthy Consumption, Time Series Classification, Virtual Commerce

1 Introduction

Healthy food choices are a highly relevant topic for making predictions and recommendations in retail context (Cho et al. 2014; Naruetharadhol et al. 2023), as food choices are an important determinant of physical health and well-being (Wahl et al. 2017; Block et al. 2011; Bublitz et al. 2013). After the disruptive retail transformation from physical warehouses to e-commerce, a slower but continuous development towards virtual commerce is taking place (Evans and Wurster 1999; Bourlakis et al. 2009; Gadalla et al. 2013; Kovacova et al. 2022). Extended Reality (XR), an umbrella term for Augmented Reality (AR) and Virtual Reality (VR), found its way into Western society. Through interaction and high realism, this new technology offers unprecedented opportunities that may encourage consumers to make healthier choices. Research on the topic is needed that investigates new challenges and opportunities. Thus, we think that retailers should seize the opportunity and adjust their user assistance capabilities in order to meet the eminent needs of consumers who visit their future (at least partly virtual) commerce environments (Regt and Barnes 2019). Examples are adaptive head-up displays (HUDs that display customized product information and comparison options), personalized side-by-side recommendations, contextual advertising, and cross-platform nudges based on individual characteristics and preferences (Mariotti et al. 2023).

While acknowledging that research should advocate for rigid privacy measures within any XR environment, technological developments will most likely lead consumers to wear XR headsets, equipped with various bio sensors, for prolonged periods. Today, the first consumer-grade XR devices offer bio sensor based features, such as foveated rendering (Patney et al. 2016) and gaze-based interactions (Piumsomboon et al., 2017). One reason for our anticipated proliferation of biosensors is the privacy-personalization paradox, which describes the fact that people readily give personal information away, if they expect utility while misjudging the real value of their personal information (Hoang et al. 2023).

Especially eye tracking (ET) based applications are a unique selling point in the current XR adoption phase. Eventually, ET could become a quality-of-life feature which consumers take for granted, like the camera in smartphones. ET can help to achieve a high degree of personalization and serve as an additional source of information for recommender systems (Meißner et al. 2019). In XR recommendation scenarios, ET may eventually replace click streams and historical data to a large extent. This is because ET allows close investigation of the user's decision process and at the same time is available in the early phase of a purchase situation (Pfeiffer et al. 2020; Meißner et al. 2019).

Regarding the consumer preference (the dependent variable), we focus on healthy consumption because in different societies around the world, an increased attention on a healthy lifestyle is noticeable (Parashar et al. 2023). Policy makers are introducing healthiness indicators like the Nutri-score label and are actively fostering a healthy consumption (Hercberg et al. 2021), which is even included in the United Nations Sustainable Development Goals (Fernandez 2019). Therefore, a valuable customer insight is whether a person is open to suggestions that support healthy product choices or not (Tran et al. 2018). In the light of these developments, we pose following research question:

Can we identify customers who buy healthy products possibly early during their decision process in a virtual commerce scenario?

Shallow machine learning approaches have already been successfully applied in previous studies that predicted other aspects of the customer journey, for example the customers' search motives (Pfeiffer et al. 2020) or the duration of intermediate decision stages (Weiß et al. 2023). A logic next step is to leverage deep learning to make predictions. An increasing amount of data and architectural improvements are likely to allow training of highly generalizing (or very precise, specialized) models. We treat the ET data as a discrete time series and, as further contrast to previously mentioned works, compare InceptionTime, one of the most promising deep learning approaches for time series classification, with the shallow gradient boosting method XGBoost which uses cross-sectional features.

With this paper, we contribute to the information systems literature in theoretical and practical manner. (i) As theoretical contribution, we show the superiority of using the complete time series of ET data in contrast to treating the ET data as cross-sectional data (by aggregating the number of fixations and other attributes). (ii) On the practical side, we show a promising way to personalize assistance systems in future metaverse applications based on the inobtrusive collection of ET data. Our paper describes a machine learning approach based on ET data which can be used to personalize XR experiences. The resulting features are of particular interest for new products or, more generally, in cases where user data is absent. (iii) Moreover, we investigate the trade-off between prediction quality and timing. Overall, our results inform the reader about interesting time windows during the decision process in our experimental purchase situation. From a broader research perspective, we show a promising way to personalize assistance systems in future metaverse applications.

2 Related Work

Already several *Second Life* studies pioneered connected 3D environments in virtual retail platforms (Bourlakis et al. 2009; Gadalla et al. 2013; Papagiannidis and Bourlakis 2010). The authors have depicted a transformation of traditional retail and outlined evolving marketing opportunities in the virtual space. Their conclusions emphasize the need for highly personalized and precisely timed customer service. Today, such connected virtual environments are thought of as the *Metaverse*, which are accessible via various XR devices. Recent comprehensive literature reviews about Metaverse shopping (Kliestik et al. 2022; Alcañiz et al. 2019; Shen et al. 2021) and AR shopping (Popescu et al. 2022) show how earlier claims, that were made for desktop environments, remain valid in XR. Virtual commerce research has diversified while recommendations and personalization remain highly relevant. A further recent review by Xi and Hamari (2021) categorizes 83 XR shopping studies along different axes (theories, in- and output devices, tracking technology, products, cognitive reactions, behavioral outcomes) and suggests a number of avenues for future research. Among these suggestions is an effective and efficient design of XR shopping, which is the area this work contributes to. The Metaverse is steadily taking shape (Peukert et al. 2022; Sriram 2022), head-mounted displays (HMDs) technology is advancing (Spagnolo et al. 2023), and HMD prices are deteriorating (Jensen and Konradsen 2018).

Various experiments have shown the significant impact of recommendations on the shopping behavior of customers, such as Li et al. (2022). Particularly in advertisement driven environments, recommender systems are very important business components. For instance, Google[1] accounts 40% of the Play Store app installations and 60% of the YouTube watch time to recommendations made by their recommender system. Collecting implicit information which reflects user preferences, like ET data, is an unobtrusive approach. This is important, as finding similarities between individuals should happen without any disruption of the consumer. Working with ET data in the context of recommender systems is nothing new (Castagnos et al. 2010; Xu et al. 2008; Zhao et al. 2016), but previous studies focused on desktop based e-commerce websites. Moreover, these studies do not aim for an early prediction of user preferences.

Generally speaking, gaze patterns have potential to improve various aspects of digital and virtual commerce. Takahashi et al. (2022) presented a work in which they utilized ET to optimize a desktop-based 3D store layout. With the goal to support customers' decision-making processes, the experiment software used gaze information to rearrange the displayed products. Another step towards gaze-pattern utilization in shopping context was made by David-John et al. (2021). Their experimental design consisted of selection tasks of food items listed on recipes in a VR scene. The authors predicted the participants' intent to interact using logistic regression on gaze patterns. They treated the data as time series but only for a relatively short prediction horizon of 0.17 to 1 second. The results suggest that the used model can predict the users' interaction timing in real-time with above-chance accuracy.

Further ET studies have examined healthy food choices (Fenko et al. 2018; Kim et al. 2018) but the prediction horizon of these studies covered the whole decision-making process until the very end. Typical research using ET in the field of consumer behavior focuses on understanding and modelling the entire decision process up to the final purchase. For example, ET research has found the gaze cascade effect which describes a pre-decisional focus of attention on the chosen product (Shimojo et al. 2003; Krajbich and Rangel 2011). Regarding our research gap, none of these studies predicted customers' preferences early in the decision process.

In a hybrid field study, Pfeiffer et al. (2020) investigated grocery shopping behavior, especially the differences between a real and virtual supermarket. The authors did not predict consumers' preferences but two different shopping patterns, namely goal directed and exploratory search behavior. To predict shopping patterns, they analyzed the collected ET data of 29 participants in VR (a room-sized CAVE environment) and 20 in a real supermarket. Their evaluation covered increasing time windows on a per second basis. These windows were calculated using the intervals from the start of each trial to $[5; 100]$ seconds into the decision-making process, increasing by one second. Due to the experimental setup, the classes were balanced, which is different compared to data presented in our study. They used shallow machine-learning approaches for point-in-time related features and not for time series. We call these features cross-sectional, as they are single values which are aggregated over the whole predefined period. This work identified the total number of fixated products and the variance of the average fixation duration among the most important predictor variables.

Millecamp et al. (2021) reported gaze pattern classification results for personality traits in the context of a browser-based music recommender system. The authors conducted a study with 30 participants in which eye movements were recorded using a desktop-based tracker. Their goal was to acquire predictions about the participants' openness, need for cognition, and musical sophistication. The authors considered 30%, 60%, and 90% of the data as time windows for their predictions. These time windows were less than the whole task duration but 60% and 90% of the decision-making process cannot be considered as particularly early stages. In general, their work showed the potential of using ET for

[1] https://developers.google.com/machine-learning/recommendation/overview

adaptation of recommendations and explanations. However, in the conclusion they outlined improvement potential for the model's performance and called for further research on different tasks and interfaces.

Our search for related work indicates a research gap that previous authors did not particularly focus on early prediction of consumer preferences based on gaze patterns. So far, no proposal has been made to leverage ET data to generate features for recommender systems in VR which are generated possibly *early* in customer decision-making processes. Furthermore, to the best of our knowledge, no previous study used ET data with a state-of-the-art time series classification model to predict customer choices for healthy products. Using time series can improve performance because of leveraging information retrieved from behavior over time.

3 Method

Experimental Design

As dependent variable, we are interested in the healthiness of different muesli (cereal) purchase decisions. To categorize all products as healthy or unhealthy, the package label serves as a discriminative criterion. Representatives of the healthy and unhealthy classes are illustrated in Figure 1 (slightly distorted due to copyright reasons), where the left package is the healthy and the right package is the unhealthy alternative. The highlighted healthy label reads "without added sugar, wholegrain". We categorized a product as healthy if the packaging indicated at least reduced (or no) sugar or fat. According to this definition, seven out of the total 40 available products in the experiment were marked as healthy products. In total, out of 1040 product choices, 158 (15.2%) were for healthy products. The imbalanced class ratio leads to methodological challenges, which we discuss in the section on the treatment of class imbalances.

Figure 1. Criterion for healthy (left) and unhealthy (right) is the packaging label.

Our observations of retail purchase decisions in VR were collected in a controlled environment in a European University laboratory. The experimental design allowed our research group to answer several questions. Thus, the data is used in further studies which investigate the impact of low versus high immersion on system adoption (Peukert et al. 2019) and the impact of virtual reality in a conjoint-based choice analysis (Meißner et al. 2020). The VR scene was created using Unity 5.5.3f1 game engine. Participants were situated in a plain virtual room with a shelf of product packages and a shopping cart, as shown in Figure 2. We used an HTC Vive HMD with a dual display with 2160×1200 pixels resolution, an integrated SMI eye tracker, and HTC hand-held controllers.

Figure 2. Virtual environment with a muesli package shelf and a shopping cart.

After signing an informed consent form, all participants made multiple product choices in front of a product shelf. Participation compensation amounted to 14 Euro in total. To provide an economic incentive, participants received one of their product choices at random as part of their compensation. We instructed the participants to choose according to their natural preference and subtracted the cost of the chosen product from the monetary payout. Each experimental session was preceded by a training phase to familiarize the participants with the virtual environment. For this training, the shelf was filled with baking mixtures. In the subsequent experimental trials, the virtual shelf contained muesli products. In total, it held 24 different options which were selected from a product pool of 40 mueslis. Their arrangement followed a design which was suited for a conjoint-based choice analysis (Chrzan and Orme 2000). At any time, the product positioning ruled out centrality effects (Atalay et al. 2012). Furthermore, we positioned mueslis of the same brand close to each other. For each trial, one out of 171 product arrangements were displayed on the shelf. On average, the shelf contained 4.27 (SD 1.09) healthy products.

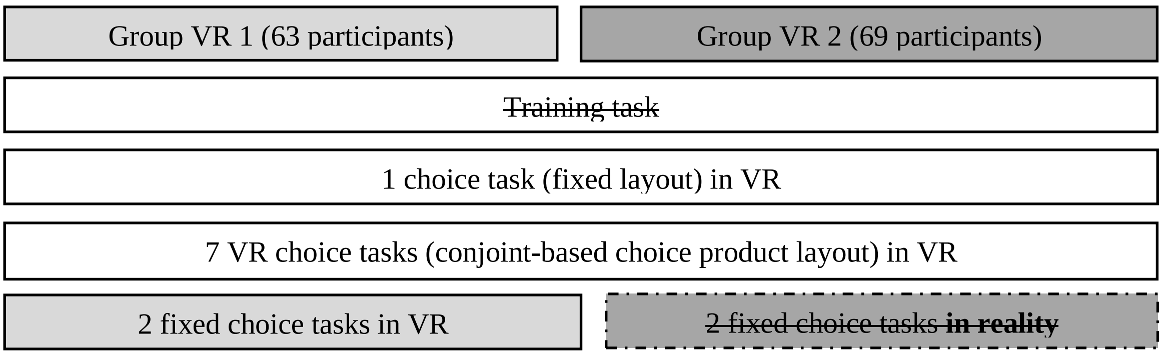

Figure 3. The experimental setup. We exclude the training task and the real-world decisions.

Our sample consists of 132 student recordings, of which 45 were females and 87 males, with an average age of 22.13 (SD 1.98). The experiment followed a between-subjects design in which one treatment group was asked to make their last two purchase decisions in front of a real shelf (with real products). We had to exclude these real-world tasks because the ET equipment differed substantially between the VR and real-world setup. Thus, each participant made a total of either eight or ten product choices in VR, depending on the treatment group (see Figure 3). In other words, for the present study, we only used purchase decisions that were made in VR. After excluding the training task and erroneous recordings, the VR trials yield 1040 product choices, with an average decision duration of 54.91 seconds (SD 33.49). However, we further reduced the number of evaluated trials in the preprocessing because many of the respective decision-making processes were too short (less than 45 seconds) to separate them meaningfully into sub-phases (like orientation and evaluation). We chose 45 seconds as cutoff duration because of logic considerations about a decision process: a participant would need approximately 15 seconds to get an overview over the assortment and another 30 seconds to decide between the items in their consideration set (Hauser 2014).

Preprocessing

First, we determined fixations from the raw ET data and calculated the subject's gaze target for each fixation, which we tracked by means of ray casting (Pietroszek 2019). We did not consider blinks, pupil dilation and saccades. However, we emphasize that additional features could further improve predictive performance. In this paper we deliberately chose to focus on visual attention, which is best described by fixations (Holmqvist et al. 2011). In general, fixations last between 0.2 and 0.4 seconds. Fixations of less than 0.1 seconds were excluded, as they are too short for conscious information processing (Duchowski 2017). Fixations lasting longer than 10 seconds were also excluded, as they most likely indicate unnatural behavior or faulty sensor information. Predefined areas of interest comprised different parts of the individual product packages and their related price tags. This enabled us to discriminate fixations on different product groups (healthy and unhealthy products). Furthermore, fixations on each individual product and individual product's nutrition table were treated separately.

Transforming the gaze data into a discrete multivariate time series is the next preprocessing step. To aggregate the fixations into discrete bins, it was necessary to choose different step sizes for the cut-off points of the bins. We evaluated the step values $(0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 9)$ seconds for the time series generation process. These values are based on reasoning about the average and maximum duration of a single fixation as described above. Shorter steps would often contain no fixation at all, and longer periods would cover too many fixations and be too coarse. We applied a sliding window technique (Hota et al. 2017) such that all bins overlapped with the previous one by 50%. The purpose of applying a sliding window is to capture interesting patterns that might be hidden by disjoint intervals. For each step size, we calculated the number of fixations, mean, variance, and skewness of the fixation duration (overall and for each of the areas of interest separately).

Our goal is to provide recommendations as early as possible during the evaluation phase of the respective decision. Therefore, we aimed to partially cut off the orientation and validation phase of the decision process as described in the on-the-fly-detection decision phase model by Peukert et al. (2020). In the orientation phase, consumers scan their environment, get an overview of the assortment, and do not compare different product choices in detail. For our data, the average transition from orientation to evaluation occurred in second 8 and the second transition from evaluation to verification occurred in second 47. Accordingly, we considered all integers in the interval $[0;\ 15]$ seconds as start values for our time series and all integers in the interval $[20;\ 45]$ seconds as stop values. Using these intervals logically entailed to exclude decisions which lasted less than 45 seconds. Therefore, keeping shorter decisions would have confounded the input time series because trials shorter than 45 seconds would have to be filled with default values. After excluding all purchase processes shorter than 45 seconds, 516 relevant product choices remained for evaluation, with 78 (15.1%) healthy choices. To train and evaluate the classification models, a random split of training (60%), validation (20%), and test (20%) was used. We also allowed for recurring customers, i.e., we did not assign all trials of one participant to a single set. This means we assume that customers can return to the store, which is typical for grocery shopping.

Time Series Classifier

The deep learning approach InceptionTime (Ismail Fawaz et al. 2020) is a time series specific successor to the image classification model Inception, also referred to as GoogLeNet (Szegedy et al. 2015). InceptionTime is one of the current state-of-the-art deep learning approaches for time series classification (Middlehurst et al. 2021). The InceptionTime building blocks mainly consist of convolutional layers and pooling layers (Aggarwal 2018). The reference implementation proposes to stack six InceptionTime modules sequentially. As shown on the left in Figure 4, each module consists of several stages. A bottleneck layer (stage 1a) reduces the input dimensionality. The main components are three convolutional layers of different kernel sizes (stage 2a). Additionally, a parallel MaxPooling layer (stage 1b) makes the model invariant to small perturbations. This is followed by another bottleneck layer (stage 2b) to reduce dimensionality. At the end of each module (stage 3), the output of the convolutions and the max pooling operation are concatenated and serve as input to the next layer. As shown on the right in Figure 4, InceptionTime uses shortcut connections between every third InceptionTime module. These shortcuts help to overcome the vanishing gradient problem (Hochreiter 1998) and overfitting (Goodfellow et al. 2016). Finally, a dense classification head (a fully connected softmax layer) outputs the predicted probabilities for each class.

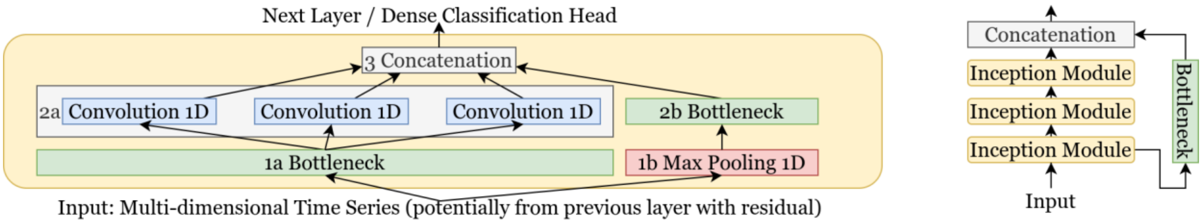

Figure 4. An InceptionTime module on the left and a shortcut connection on the right. Adapted from Ismail Fawaz et al. (2020).

The Inception architecture is based on two main ideas: First, reducing the dimensionality (via bottleneck layers) keeps the computational complexity low and mitigates overfitting for small datasets. Second, convolutional components with different receptive fields capture different aspects of the time series (Luo et al. 2016). For temporal data, the receptive field can be thought of as the maximum field of view of a neuron. The larger the receptive field is, the longer the patterns that can be detected by the neuron. The model uses multiple parallel, densely connected convolutional layers with different kernel sizes (see Figure 4, stage 2a) that allow to capture different aspects of the time series. During the trials, an asynchronous hyperband scheduler (Li et al. 2020) facilitated the exploration of 50 different combinations. Table 1 shows the complete hyperparameter space.

Name	**Values**	**Description**
Activation function	ReLU (Agarap 2019), eLU (Clevert et al. 2015)	
Alpha	[0.1; 0.3] Uniform	Focal loss
Bottleneck size	(32, 64, 128)	Inception Module 1a, 2b
Gamma	[0.1; 0.3] Uniform	Focal loss
Kernel Multiplier	(4, 6, 8, 18)	Inception Module 2a
Learning Rate	[1e-1, 1e-6] Log uniform	Optimizer
Num Filters	(8, 16, 32)	Inception Module 2a
Num Modules	(3, 6)	InceptionTime

Table 1. The hyperparameter space which we used in the InceptionTime tuning process.

In total, the different start, stop, and step size values resulted in 4990 possible combinations. A high-performance cluster was used to compute all respective trials. The Ray Tune framework (Liaw et al. 2018), combined with the slurm task scheduler (Yoo et al.), allowed us to partially parallelize the optimization of the InceptionTime instances, which all ran for a maximum of 100 epochs, using up to 75 compute nodes equipped with 24 CPU cores.

Class Imbalance Treatment

In our data, only 15.2% of choices were for healthy products. The applied methods need to take this class imbalance into account. Otherwise, classifiers tend to always predict the majority class. Different paradigms to treat imbalanced data exist, namely data-level, algorithm-level, and hybrid methods (Krawczyk 2016). We used an α balanced focal loss function (Lin et al. 2017) for the neural network optimizer to discount the majority classes. It is a hybrid approach that combines cost modifying and algorithmic adjustments. Focal loss is a modification of the widely used cross-entropy loss function (Goodfellow et al. 2016). The main idea is to discount correctly classified samples of the majority class, i.e., the contribution to the total loss value is large for wrong predictions of the minority class. Focal loss and can be denoted as

$$FocalLoss(p_t) = -\alpha_t\,(1-p_t)^{\gamma}\log(p_t), \text{ with } p_t = \begin{cases} p, & if\ y=1 \\ 1-p, & otherwise \end{cases}. \tag{1}$$

Parameter α_t specifies the minority class proportion in the test data set, $p_t \in [0, 1]$ is the predicted class probability for the sample, and $y \in \{0, 1\}$ is the target label. The focusing intensity $\gamma \geq 0$ determines the rate for discounting easy samples. Note that, when $\gamma = 0$, focal loss equals cross-entropy.

A further algorithmic measure is the evaluation with a suited scoring metric. For the prediction of imbalanced data the accuracy metric is unexpressive (Bekkar et al. 2013). Accuracy would put too much attention on unhealthy product choices (precision) and too little on healthy ones (recall). The F_β metric allows to adjust the trade-off between recall and precision (Maratea et al. 2014). A value for parameter β greater than one emphasizes the importance of recall while a value less than one emphasizes the importance of precision. For this study, $\beta = 1.5$ is used because we focus more on recall than on precision. Choosing $F_{\beta=1.5}$ means we deliberately expose some of the purchasers of unhealthy mueslis to recommendations for healthy products as trade-off for a higher classification rate of intended healthy product choices (which may be interpreted as a form of nudging). The F_β score (Maratea et al. 2014) can be denoted as

$$F_\beta = (1+\beta^2)\cdot\frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall}. \tag{2}$$

Gradient Boosting Trees Baseline

This gradient boosting baseline considers aggregated, one-dimensional, features, which is the current standard. Utilizing multi-dimensional features in form of time series is more promising because it allows considering the complete decision-making process in form of a vector, from the start of the purchase situation until the point-of-time when a recommendation should be made.

Gradient boosting served as a baseline for this study, as it has shown good results in similar setups (Millecamp et al. 2021; Pfeiffer et al. 2020). We implemented it using the XGBoost (Chen and Guestrin 2016) and scikit-learn (Pedregosa et al. 2011) packages. This model did not require a distinct validation dataset for training. Instead a 10-fold cross validation (Refaeilzadeh et al. 2009) ensured generalizability on the data set, permuting the combined training and validation subsets. The features for the gradient boosting model consist of the same underlying information (e.g., number of fixations) but aggregate it with respect to the total interval length. Analogous to the time series, we used $F_{\beta=1.5}$ as scoring metric and chose the intervals $[0;\ 15]$ for start timestamps and $[20;\ 45]$ for stop timestamps. To find a good set of hyperparameters (colsample_bytree, gamma, learning_rate, max_depth, min_child_weight, n_estimators, scale_pos_weight, subsample) a randomized search was performed for 100 trials on all possible start-stop combinations.

4 Results

Figure 5 shows two different prediction horizons (i) the first 25 seconds and (ii) the first 45 seconds of the decision process. The 25-second horizon is based on the idea of making recommendations early in the product evaluation process. A recommender system would have enough time to generate content after a feature extraction phase of 25 seconds at the beginning of the decision process. On average, the 45-second horizon covers the entire evaluation phase and can be seen as the upper limit for a recommender system to make suggestions.

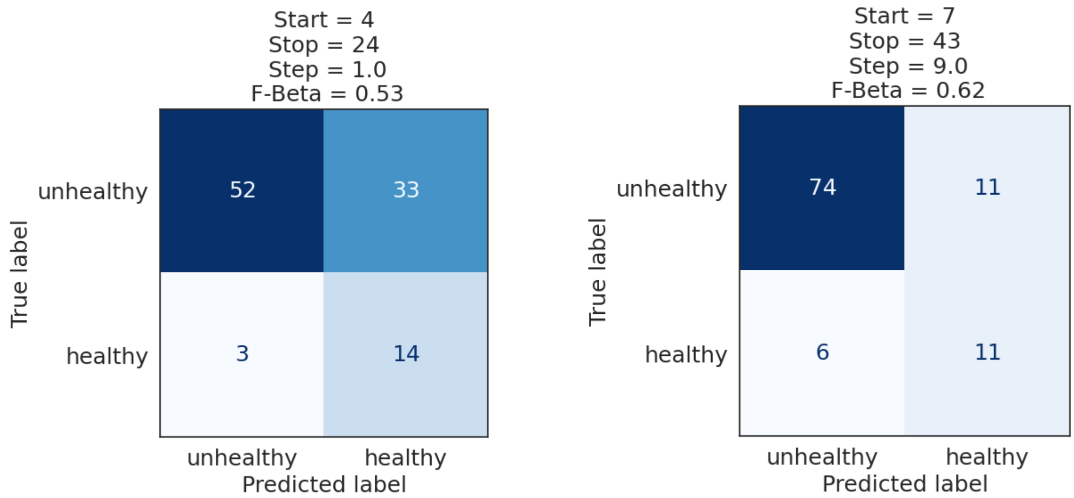

Figure 5. The confusion matrices represent the best InceptionTime models for healthiness preference predictions within the first 25 (left) and 45 (right) seconds.

The best model for the entire prediction horizon of 45 seconds ($F_{\beta=1.5} = 0.62$) did not use the full 45 seconds. It performed best when considering the time series from second 7 to 43, with a step size of 9.0 seconds. This model correctly classified 87.05% of the unhealthy choices and 64.70% of the healthy choices.

The model for the shorter prediction horizon of 25 seconds does not perform much worse overall (61.17% correct unhealthy classification and 82.35% correct healthy classification). It achieved an $F_{\beta=1.5}$ score of 0.53. We remind the reader that with a beta of 1.5, we value recall higher than precision, i.e., finding most of the healthy choices has priority. This model considered the period from second 4 to 24 as a time series, using a start-stop interval of [4; 24], and a step size of 1.0 second. It even correctly classified more healthy choices correctly compared to the best model for the 45-second prediction horizon.

In contrast, the best performing XGBoost model achieved an $F_{\beta=1.5}$ score of 0.48, using a start-stop interval of [0; 38]. It correctly classified 90.5% of the unhealthy choices but only 47.1% of the healthy choices. With respect to the prediction horizon of 25 seconds, the best XGBoost model performed slightly worse with an $F_{\beta=1.5}$ score of 0.42, using a start-stop interval of [5; 21].

In Figure 6, we provide information about the effect of different start and stop values on the maximum F_β classification performance. The left plot shows the average effect of different start values. For our data, starting in second 5 results in the best average F_β value. As expected, a decrease in performance occurs when a long onset duration is omitted before *feeding* the model. The right plot shows the average impact of different stop values with a peak at second 30. The positive trend for later stop values is also plausible, as more information becomes available over time.

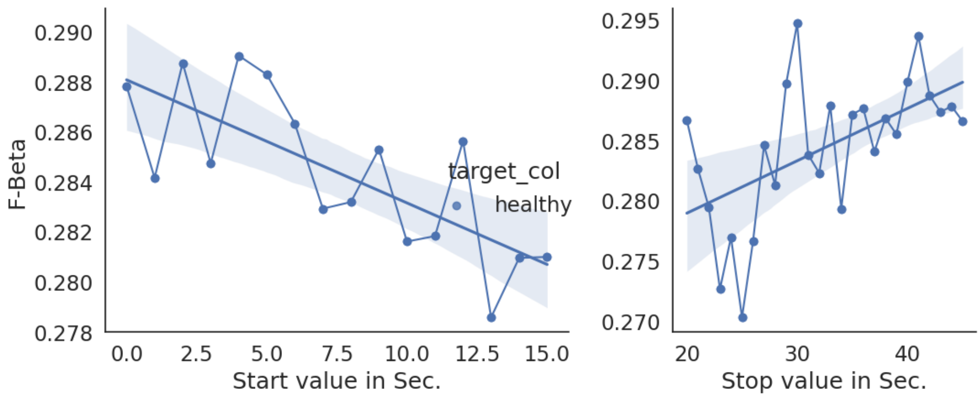

Figure 6. A timeline showing the average F-Beta value for healthiness preferences predictions regarding all evaluated start and stop values.

5 Discussion

Although the amount of training data we used was limited and the class distribution imbalanced, our work demonstrates a way to use gaze behavior, in our case the extracted fixations and gaze targets, as input for recommender systems. With the algorithmic adjustments regarding the misclassification of the majority class, we can clearly answer our research question. Yes, with reasonable performance in relation to the limited amount of data, we can identify customers who buy healthy products early during their decision-making process in a virtual commerce scenario using the InceptionTime deep learning approach. However, we acknowledge that current classification rates are not production ready and continuous model improvement and data collection are required to eventually allow for accurate predictions.

Our main aim was to correctly classify as many samples of the minority class (healthy choices) as possible during the evaluation phase of the decision processes. For the given data set, our results suggest that a time series based approach like InceptionTime is a more appropriate classifier compared to the shallow XGBoost method. The InceptionTime model with a 1.0 second step size and a start-stop interval of [4; 24] is a promising predictor for healthy and unhealthy product choices early in the decision process. This model showed that focal loss and the F_β metric are effective measures to cope with the class imbalance inherent to the data set. It achieved the highest $F_{\beta=1.5}$ score of 0.53 in our evaluation and correctly classified most of the healthy choices (14 out of 17) while generating nudges candidates (a fraction of customers with unhealthy choices, 33 out of 87). The extent of candidate-generation could be adjusted by the β parameter for the evaluation metric (in our case we chose β=1.5 and argue that it was a good choice because the amount of nudge candidate seems to be appropriate).

The best XGBoost model achieved an $F_{\beta=1.5}$ score of 0.48 in our evaluation. It correctly classified less than half of the healthy choices correctly (8 out of 17) and generated only a small number of nudge candidates (8 out of 87). One reason for the lower performance could be the fact that the scikit-learn implementation of XGBoost does offer a focal loss function. However, recently Wang et al. (2020) implemented a focal loss function for the XGBoost algorithm that may serve as drop-in replacement in scikit-learn. A comparison between InceptionTime and an XGBoost model with implemented focal losses might offer a more equitable benchmark and could potentially alter the results' significance. The XGBoost model could also benefit from advanced sampling techniques, such as creating synthetic samples with small deviations from the real observations (Chawla et al. 2002), but this is beyond the scope of this research.

Regarding start values, the best InceptionTime models started early for both the 45 and the 25 second prediction horizon (second 4 and 7, respectively). Thus, it seems advisable to start the time series possibly early. Another viable option may be to decide on an individual case basis when the orientation phase ends, e.g., by detecting the gaze pattern which represents the first comparison of two products

(Peukert et al. 2020). In terms of stop values, the results unsurprisingly exhibit a positive linear trend for the maximum F_β score, i.e., an increase of performance with the duration of the prediction horizon. However, the graph shows a lot of variances around second 25, 30, and 40 and it might be counterintuitive that for the stop values after second 40, the maximum F_β values mainly decrease. For earlier stop values, our results show that the prediction quality can remain relatively good, e.g., when stopping after 24 seconds. The corresponding InceptionTime model correctly classified only one healthy customer less (out of 17) than the best InceptionTime model which had access to additional 16 seconds of ET data of the decision processes. This further supports the importance of the early decision phase for the correct classification of healthy customers.

An open question remains the choice of a possibly ideal step size. We evaluated many different step values, which cost a lot of (computation) time and energy. Finding and validating a better theoretical foundation, to explain for what reasons a certain overlapping technique should be applied, could prove very helpful.

As theoretical contribution, our study confirms that leveraging a complete time series of ET data and feed it into a convolutional network can be superior to treating the ET data as cross-sectional data. However, the performance gain in comparison to a basic XGBoost model is only a first proof of concept and both the baseline and the classification model can further improve.

Before closing, we reflect on ethical considerations, particularly with regard the use of our classification model as input for recommender and other context-aware AI systems. We used gaze information of our participants to infer their willingness to buy healthy food and prioritized healthy purchases. In the design of our model, we accepted a bias towards healthy classification, what may lead to a nudge for a certain fraction of customers who would not necessarily appreciate suggestions for a healthy product. We argue that such a nudge would be ethically valid, as it fosters socially desired behavior. However, there is an ongoing debate in what situations nudging is desirable and when it should be avoided altogether (Hausman and Welch 2010). In any case, “[c]hoice architecture, both good and bad, is pervasive and unavoidable, and it greatly affects our decisions.” (Thaler and Sunstein 2021, p. 252).

From a technical standpoint, our study suggests that time series classification enables real-time feature generation for recommender systems using gaze patterns. Our results indicate that the longitudinal point of view offers more relevant information than aggregations to statistical moments that span over the whole decision period. We acknowledge that further research and validation are needed to improve the reliability and generalizability of our findings. Nonetheless, we hope that the presented approach encourages practitioners to integrate recommender systems in virtual commerce environments. From our point of view, it is only a question of time until we experience various (most likely artificial intelligence assisted) tools which support and improve healthy food choices based on individual sensor data. Overall, the use of suitable deep learning models, such as InceptionTime, could potentially change the state-of-the-art for developing personalized interventions. In combination with large language models, time series classification and cutting-edge deep learning methods are likely to transform user assistance as we know it today. Researchers and practitioners might think about further contexts beyond classic collaborative filtering, such as personal trainers and instructors, medical advisors, psychotherapeutic treatments, and more. The presented approach could be applied everywhere where learning about users' preferences or their decision processes in general can be helpful. Therefore, it seems advisable to continue with data acquisition, model evaluation, and workflow integration.

6 Summary and Outlook

We proposed to use an InceptionTime classifier to infer customer preferences during the evaluation phase of customer decision processes using gaze patterns. Our focus was on classifying customers who buy healthy products in a VR setup. The results show that InceptionTime, in combination with class imbalance measures, can outperform a shallow gradient boosting model in classifying healthy purchase decisions while generating candidates for healthy food nudges.

The main limitation of this study is the fact that our sample consists of only 516 purchase decisions, of which only 15.1% were made for healthy products. Deep learning models are typically trained on much larger datasets (Szegedy et al. 2016), and we believe that the full potential of deep time series classification approaches will remain unexplored until such a large dataset becomes (publicly) available. However, in order to collect such a dataset, the legal consensus regarding privacy concerns for ET data needs to be solidified. Another limitation of this study is that we only considered product labels (the most visually salient information) to classify products as unhealthy or healthy when defining the ground truth. Future research could use more fine-grained information, such as ingredient lists and nutritional tables. With detailed information about a product's composition, recommendations could take additional aspects into account. A highly relevant example is the detection of allergies, e.g., many people are allergic to nuts. Consumers could decide whether to hide such products altogether or receive a multi-sensory warning when they focus on a critical product.

Overall, we see several avenues for future virtual commerce focused research. One prominent concern is the treatment of privacy issues. Deliberate actions, such as body movements or use of voice, can be controlled by the customers. In contrast, the gaze as such is less under consumer control and fundamental to decision making. ET data can identify individuals and might reveal unwanted personal aspects (Cantoni et al. 2018). Thus, research should invent, evaluate, and reflect on different suitable (pseudo) anonymization techniques (Steil et al. 2019). Privacy research enables device vendors and digital commerce providers to avoid pitfalls and fosters trust among customers. The nudge aspect of this work is another route to follow. Healthiness is only one aspect of socially desirable behavior but there are further areas, such as sustainable consumption, which could be investigated by further research.

Regarding data collection, upcoming studies should include a broader variety of available information. Pupillometry and additional bio sensors seem to be a promising source for additional input features (Halbig and Latoschik 2021). Furthermore, time series classification evolves quickly and new classifiers emerge frequently, e.g., InceptionFCN (Usmankhujaev et al. 2021) or TapNet (Zhang et al. 2020). These models may have the potential to yield better classification rates and should be compared with the presented results.

Future research should predict further dependent variables and showcase a real recommendation pipeline. In addition to healthy products, we argue that brand and flavor preferences are particularly interesting. Such a follow-up study should rethink the large-scale hyperparameter searches. These searches do not necessarily enumerate all presented start and stop value combinations as presented in this study. Instead, it should benchmark different algorithmic design aspects, like predicting preferences for new customers only or limiting the feature set, which would provide further managerial insights. Next, a follow up should introduce a better baseline, e.g., by comparing InceptionTime with previously mentioned deep learning time series classification methods. Overall, we suggest iterative improvements by means of ongoing experiments with the latest sensor technology available, such as electroencephalography (event related potentials), facial features, body posture, pupil dilation, and maybe functional near-infrared spectroscopy (fNIRS). With all measures combined, we expect the predictive performance and validity to improve significantly (unfortunately, the same is true for complexity). From our perspective, a long-term goal should be to hone a publicly available machine learning pipeline, similar to the presented one, and ultimately showcase it as real-time feature generator for a recommender system in *real* virtual commerce setups.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD2School – Designing Adaptive Systems for Economic Decisions

References

Agarap, Abien Fred (2019): Deep Learning using Rectified Linear Units (ReLU): arXiv.

Aggarwal, Charu C. (2018): Neural networks and deep learning. In *Springer* 10, 978-3.

Alcañiz, Mariano; Bigné, Enrique; Guixeres, Jaime (2019): Virtual reality in marketing: a framework, review, and research agenda. In *Frontiers in psychology* 10, p. 1530.
Atalay, A. Selin; Bodur, H. Onur; Rasolofoarison, Dina (2012): Shining in the center: Central gaze cascade effect on product choice. In *Journal of consumer research* 39 (4), pp. 848–866.
Bekkar, Mohamed; Djemaa, Hassiba Kheliouane; Alitouche, Taklit Akrouf (2013): Evaluation measures for models assessment over imbalanced data sets. In *Journal of Information Engineering and Applications* 3 (10).
Block, Lauren G.; Grier, Sonya A.; Childers, Terry L.; Davis, Brennan; Ebert, Jane E.J.; Kumanyika, Shiriki et al. (2011): From Nutrients to Nurturance: A Conceptual Introduction to Food Well-Being. In *Journal of Public Policy & Marketing* 30 (1), pp. 5–13. DOI: 10.1509/jppm.30.1.5.
Bourlakis, Michael; Papagiannidis, Savvas; Li, Feng (2009): Retail spatial evolution: paving the way from traditional to metaverse retailing. In *Electronic Commerce Research* 9 (1), pp. 135–148.
Bublitz, Melissa G.; Peracchio, Laura A.; Andreasen, Alan R.; Kees, Jeremy; Kidwell, Blair; Miller, Elizabeth Gelfand et al. (2013): Promoting positive change: Advancing the food well-being paradigm. In *Journal of Business Research* 66 (8), pp. 1211–1218. DOI: 10.1016/j.jbusres.2012.08.014.
Cantoni, Virginio; Musci, Mirto; Nugrahaningsih, Nahumi; Porta, Marco (2018): Gaze-based biometrics: An introduction to forensic applications. In *Pattern Recognition Letters* 113, pp. 54–57.
Castagnos, Sylvain; Jones, Nicolas; Pu, Pearl (2010): Eye-tracking product recommenders' usage. In : Proceedings of the fourth ACM conference on Recommender systems, pp. 29–36.
Chawla, Nitesh V.; Bowyer, Kevin W.; Hall, Lawrence O.; Kegelmeyer, W. Philip (2002): SMOTE: synthetic minority over-sampling technique. In *Journal of artificial intelligence research* 16, pp. 321–357.
Chen, Tianqi; Guestrin, Carlos (2016): Xgboost: A scalable tree boosting system. In : Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.
Cho, Jaehee; Park, Dongjin; Lee, H. Erin; others (2014): Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. In *Journal of medical Internet research* 16 (5), e3283.
Chrzan, Keith; Orme, Bryan (2000): An overview and comparison of design strategies for choice-based conjoint analysis. In *Sawtooth software research paper series* 98382, p. 360.
Clevert, Djork-Arné; Unterthiner, Thomas; Hochreiter, Sepp (2015): Fast and accurate deep network learning by exponential linear units (elus). In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics* 20, pp. 924–932.
David-John, Brendan; Peacock, Candace; Zhang, Ting; Murdison, T. Scott; Benko, Hrvoje; Jonker, Tanya R. (2021): Towards gaze-based prediction of the intent to interact in virtual reality. In : ACM Symposium on Eye Tracking Research and Applications, pp. 1–7.
Duchowski, Andrew T. (2017): Eye tracking methodology: Theory and practice: Springer.
Evans, Philip; Wurster, Thomas S. (1999): Getting real about virtual commerce. In *Harvard Business Review* 77, pp. 84–98.
Fenko, Anna; Nicolaas, Iris; Galetzka, Mirjam (2018): Does attention to health labels predict a healthy food choice? An eye-tracking study. In *Food quality and preference* 69, pp. 57–65.
Fernandez, Rosa Maria (2019): SDG3 Good Health and Well-Being: Integration and Connection with Other SDGs. In : Encyclopedia of the UN Sustainable Development Goals. Cham: Springer International Publishing, pp. 629–636.
Gadalla, Eman; Keeling, Kathy; Abosag, Ibrahim (2013): Metaverse-retail service quality: A future framework for retail service quality in the 3D internet. In *Journal of Marketing Management* 29 (13-14), pp. 1493–1517.
Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016): Deep learning: MIT press.
Halbig, Andreas; Latoschik, Marc Erich (2021): A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. In *Front. Virtual Real.* 2, Article 694567, p. 89. Available online at https://www.frontiersin.org/articles/10.3389/frvir.2021.694567/full.

Hauser, John R. (2014): Consideration-set heuristics. In *Journal of Business Research* 67 (8), pp. 1688–1699.
Hausman, Daniel M.; Welch, Brynn (2010): Debate: To nudge or not to nudge. In *Journal of Political Philosophy* 18 (1), pp. 123–136.
Hercberg, Serge; Touvier, Mathilde; Salas-Salvado, Jordi (2021): The Nutri-score nutrition label: Hogrefe AG.
Hoang, Hanh Thi Hong; Nguyen, Lam Son; Nguyen, Chinh Hong; Le, Nga Viet; Nguyen, Nhung Thi; Dinh, Len Thi (2023): Personalisation-Privacy Paradox from Marketing Perspectives: Literature Review and Future Research Directions. In *Business Innovation for the Post-pandemic Era in Vietnam*, pp. 87–105.
Hochreiter, Sepp (1998): The vanishing gradient problem during learning recurrent neural nets. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6 (02), pp. 107–116.
Holmqvist, Kenneth; Nyström, Marcus; Andersson, Richard; Dewhurst, Richard; Jarodzka, Halszka; van de Weijer, Joost (2011): Eye tracking: A comprehensive guide to methods and measures: OUP Oxford.
Hota, H. S.; Handa, Richa; Shrivas, Akhilesh Kumar (2017): Time series data prediction using sliding window based RBF neural network. In *International Journal of Computational Intelligence Research* 13 (5), pp. 1145–1156.
Ismail Fawaz, Hassan; Lucas, Benjamin; Forestier, Germain; Pelletier, Charlotte; Schmidt, Daniel F.; Weber, Jonathan et al. (2020): Inceptiontime: Finding alexnet for time series classification. In *Data Mining and Knowledge Discovery* 34 (6), pp. 1936–1962.
Jensen, Lasse; Konradsen, Flemming (2018): A review of the use of virtual reality head-mounted displays in education and training. In *Education and Information Technologies* 23 (4), pp. 1515–1529.
Kim, Eojina; Tang, Liang Rebecca; Meusel, Chase; Gupta, Manjul (2018): Optimization of menu-labeling formats to drive healthy dining: An eye tracking study. In *International Journal of Hospitality Management* 70, pp. 37–48.
Kliestik, Tomas; Novak, Andrej; Lazaroiu, George (2022): Live Shopping in the Metaverse: Visual and Spatial Analytics, Cognitive Artificial Intelligence Techniques and Algorithms, and Immersive Digital Simulations. In *Linguistic & Philosophical Investigations*.
Kovacova, Maria; Machova, Veronika; Bennett, Daniel (2022): Immersive extended reality technologies, data visualization tools, and customer behavior analytics in the metaverse commerce. In *Journal of Self-Governance and Management Economics* 10 (2), pp. 7–21.
Krajbich, Ian; Rangel, Antonio (2011): Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (33), pp. 13852–13857. DOI: 10.1073/pnas.1101328108.
Krawczyk, Bartosz (2016): Learning from imbalanced data: open challenges and future directions. In *Progress in Artificial Intelligence* 5 (4), pp. 221–232.
Li, Liam; Jamieson, Kevin; Rostamizadeh, Afshin; Gonina, Ekaterina; Hardt, Moritz; Recht, Benjamin; Talwalkar, Ameet (2020): A system for massively parallel hyperparameter tuning. In *Proceedings of Machine Learning and Systems* 2, pp. 230–246.
Li, Xitong; Grahl, Jörn; Hinz, Oliver (2022): How Do Recommender Systems Lead to Consumer Purchases? A Causal Mediation Analysis of a Field Experiment. In *Information Systems Research* 33 (2), pp. 620–637.
Liaw, Richard; Liang, Eric; Nishihara, Robert; Moritz, Philipp; Gonzalez, Joseph E.; Stoica, Ion (2018): Tune: A Research Platform for Distributed Model Selection and Training: arXiv.
Lin, Tsung-Yi; Goyal, Priya; Girshick, Ross; He, Kaiming; Dollár, Piotr (2017): Focal loss for dense object detection. In : Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.
Luo, Wenjie; Li, Yujia; Urtasun, Raquel; Zemel, Richard (2016): Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems* 29.

Maratea, Antonio; Petrosino, Alfredo; Manzo, Mario (2014): Adjusted F-measure and kernel scaling for imbalanced data learning. In *Information Sciences* 257, pp. 331–341.

Mariotti, Thomas; Schweizer, Nikolaus; Szech, Nora; Wangenheim, Jonas von (2023): Information nudges and self-control. In *Management Science* 69 (4), pp. 2182–2197.

Meißner, Martin; Pfeiffer, Jella; Peukert, Christian; Dietrich, Holger; Pfeiffer, Thies (2020): How virtual reality affects consumer choice. In *Journal of Business Research* 117, pp. 219–231.

Meißner, Martin; Pfeiffer, Jella; Pfeiffer, Thies; Oppewal, Harmen (2019): Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. In *Journal of Business Research* 100, pp. 445–458.

Middlehurst, Matthew; Large, James; Flynn, Michael; Lines, Jason; Bostrom, Aaron; Bagnall, Anthony (2021): HIVE-COTE 2.0: a new meta ensemble for time series classification. In *Machine Learning* 110 (11), pp. 3211–3243.

Millecamp, Martijn; Conati, Cristina; Verbert, Katrien (2021): Classifeye: classification of personal characteristics based on eye tracking data in a recommender system interface. In : Joint Proceedings of the ACM IUI 2021 Workshops co-located with 26th ACM Conference on Intelligent User Interfaces, vol. 2903. CEUR Workshop Proceedings.

Naruetharadhol, Phaninee; Wongsaichia, Sasichakorn; Pienwisetkaew, Teerapong; Schrank, Johannes; Chaiwongjarat, Kullanan; Thippawong, Peeranut et al. (2023): Consumer Intention to Utilize an E-Commerce Platform for Imperfect Vegetables Based on Health-Consciousness. In *Foods* 12 (6), p. 1166.

Papagiannidis, Savvas; Bourlakis, Michael A. (2010): Staging the New Retail Drama: at a Metaverse near you! In *Journal of Virtual Worlds Research* 2 (5), pp. 425–446.

Parashar, Sapna; Singh, Supriya; Sood, Gunjan (2023): Examining the role of health consciousness, environmental awareness and intention on purchase of organic food: A moderated model of attitude. In *Journal of Cleaner Production* 386, p. 135553.

Patney, Anjul; Salvi, Marco; Kim, Joohwan; Kaplanyan, Anton; Wyman, Chris; Benty, Nir et al. (2016): Towards foveated rendering for gaze-tracked virtual reality. In *ACM Transactions on Graphics (TOG)* 35 (6), pp. 1–12.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O. et al. (2011): Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research* 12, pp. 2825–2830.

Peukert, Christian; Lechner, Jessica; Pfeiffer, Jella; Weinhardt, Christof (2020): Intelligent invocation: Towards designing context-aware user assistance systems based on real-time eye tracking data analysis. In : Information systems and neuroscience: Springer, pp. 73–82.

Peukert, Christian; Pfeiffer, Jella; Meißner, Martin; Pfeiffer, Thies; Weinhardt, Christof (2019): Shopping in virtual reality stores: the influence of immersion on system adoption. In *Journal of Management Information Systems* 36 (3), pp. 755–788.

Peukert, Christian; Weinhardt, Christof; Hinz, Oliver; van der Aalst, Wil M. P. (2022): Metaverse: How to Approach Its Challenges from a BISE Perspective: Springer. In *Business & Information Systems Engineering*.

Pfeiffer, Jella; Pfeiffer, Thies; Meißner, Martin; Weiß, Elisa (2020): Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. In *Information Systems Research* 31 (3), pp. 675–691.

Pietroszek, Krzysztof (2019): Raycasting in Virtual Reality: Springer International Publishing AG, part of Springer Nature 2018.

Popescu, Gheorghe H.; Valaskova, Katarina; Horak, Jakub (2022): Augmented Reality Shopping Experiences, Retail Business Analytics, and Machine Vision Algorithms in the Virtual Economy of the Metaverse. In *Journal of Self-Governance & Management Economics* 10 (2).

Refaeilzadeh, Payam; Tang, Lei; Liu, Huan (2009): Cross-validation. In *Encyclopedia of database systems* 5, pp. 532–538.

Regt, Anouk de; Barnes, Stuart J. (2019): V-Commerce in Retail: Nature and Potential Impact. In : Augmented Reality and Virtual Reality: Springer, Cham, pp. 17–25.

Shen, Bingqing; Tan, Weiming; Guo, Jingzhi; Zhao, Linshuang; Qin, Peng (2021): How to promote user purchase in metaverse? A systematic literature review on consumer behavior research and virtual commerce application design. In *Applied Sciences* 11 (23), p. 11087.

Shimojo, Shinsuke; Simion, Claudiu; Shimojo, Eiko; Scheier, Christian (2003): Gaze bias both reflects and influences preference. In *Nat Neurosci* 6 (12), pp. 1317–1322. Available online at https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/nn1150&casa_token=utehjyj_ytwaaaaa:rtettuxa6c3gj4lv9z9ee8kek0kd1j78i5ovbk6tsoh0t5nnjql7y8mpuljyizhgnw8d4h6-l2fvglda7q.

Spagnolo, Fanny; Corsonello, Pasquale; Frustaci, Fabio; Perri, Stefania (2023): Design of a low-power super-resolution architecture for virtual reality wearable devices. In *IEEE Sensors Journal* 23 (8), pp. 9009–9016.

Sriram, G. K. (2022): A Comprehensive Survey on Metaverse. In *International Research Journal of Modernization in Engineering Technology*.

Steil, Julian; Hagestedt, Inken; Huang, Michael Xuelin; Bulling, Andreas (2019): Privacy-aware eye tracking using differential privacy. In : Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA. New York, NY, USA: ACM.

Szegedy, Christian; Liu, Wei; Jia, Yangqing; Sermanet, Pierre; Reed, Scott; Anguelov, Dragomir et al. (2015): Going deeper with convolutions. In : Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Szegedy, Christian; Vanhoucke, Vincent; Ioffe, Sergey; Shlens, Jon; Wojna, Zbigniew (2016): Rethinking the inception architecture for computer vision. In : Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Takahashi, Shigeo; Uchita, Akane; Watanabe, Kazuho; Arikawa, Masatoshi (2022): Gaze-driven placement of items for proactive visual exploration. In *Journal of Visualization* 25 (3), pp. 613–633.

Thaler, Richard H.; Sunstein, Cass R. (2021): Nudge: The final edition: Yale University Press.

Tran, Thi Ngoc Trang; Atas, Müslüm; Felfernig, Alexander; Stettinger, Martin (2018): An overview of recommender systems in the healthy food domain. In *Journal of Intelligent Information Systems* 50 (3), pp. 501–526.

Usmankhujaev, Saidrasul; Ibrokhimov, Bunyodbek; Baydadaev, Shokhrukh; Kwon, Jangwoo (2021): Time Series Classification with InceptionFCN. In *Sensors* 22 (1), p. 157.

Wahl, Deborah R.; Villinger, Karoline; König, Laura M.; Ziesemer, Katrin; Schupp, Harald T.; Renner, Britta (2017): Healthy food choices are happy food choices: Evidence from a real life sample using smartphone based assessments. In *Scientific Reports* 7 (1), p. 17069.

Wang, Chen; Deng, Chengyuan; Wang, Suzhen (2020): Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. In *Pattern Recognition Letters* 136, pp. 190–197.

Weiß, Tobias; Merkl, Lukas; Pfeiffer, Jella (2023): Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study using a Hidden Markov Model. In : Davis, Riedl et al. (Hg.) 2023 – Information Systems and Neuroscience.

Xi, Nannan; Hamari, Juho (2021): Shopping in virtual reality: A literature review and future agenda. In *Journal of Business Research* 134, pp. 37–58.

Xu, Songhua; Jiang, Hao; Lau, Francis C. M. (2008): Personalized online document, image and video recommendation via commodity eye-tracking. In : Proceedings of the 2008 ACM conference on Recommender systems, pp. 83–90.

Yoo, Andy B.; Jette, Morris A.; Grondona, Mark: SLURM: Simple Linux Utility for Resource Management, pp. 44–60.

Zhang, Xuchao; Gao, Yifeng; Lin, Jessica; Lu, Chang-Tien (2020): TapNet: Multivariate Time Series Classification with Attentional Prototypical Network. In *AAAI* 34 (04), pp. 6845–6852.

Zhao, Qian; Chang, Shuo; Harper, F. Maxwell; Konstan, Joseph A. (2016): Gaze prediction for recommender systems. In : Proceedings of the 10th ACM Conference on Recommender Systems, pp. 131–138.