UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers

Hendriksen, M.; Kuiper, E.; Nauts, P.; Schelter, S.; de Rijke, M.

Publication date
2020

Document Version
Final published version

Published in
The 2020 SIGIR Workshop On eCommerce

License
CC BY-NC-ND

Link to publication

Citation for published version (APA):
Hendriksen, M., Kuiper, E., Nauts, P., Schelter, S., & de Rijke, M. (2020). Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers. In *The 2020 SIGIR Workshop On eCommerce: July 30 : accepted papers* Article 23 SIGIR eCom'20. https://sigir-ecom.github.io/ecom20Papers/paper23.pdf

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:10 Oct 2026

Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers

Mariya Hendriksen[1] Ernst Kuiper[2] Pim Nauts[3] Sebastian Schelter[4,5] Maarten de Rijke[4,5]
[1]AIRLab, University of Amsterdam [2]Bol.com [3]Albert Heijn [4]University of Amsterdam [5]Ahold Delhaize
m.hendriksen@uva.nl,ekuiper@bol.com,pim.nauts@ah.nl,s.schelter@uva.nl,m.derijke@uva.nl

ABSTRACT

The popularity of e-commerce platforms continues to grow. Being able to understand, model, and predict customer behaviour is essential for customizing the user experience through personalized result presentations, recommendations, and special offers. Previous work has considered a broad range of prediction models as well as features inferred from clickstream data to record session characteristics, and features inferred from user data to record customer characteristics. So far, most previous work in the area of purchase prediction has focused on known customers, largely ignoring anonymous sessions, i.e., sessions initiated by a non-logged-in or unrecognized customer. However, in the de-identified data from a large European e-commerce platform available to us, more than 50% of the sessions start as anonymous sessions.

In this paper, we focus on purchase prediction for both anonymous and identified sessions on an e-commerce platform. We start with a descriptive analysis of purchase vs. non-purchase sessions. This analysis informs the definition of a feature-based model for purchase prediction for anonymous sessions and identified sessions; our models consider a range of session-based features for anonymous sessions, such as the channel type, the number of visited pages, and the device type. For identified user sessions, our analysis points to customer history data as a valuable discriminator between purchase and non-purchase sessions. Based on our analysis, we build two types of predictors: (1) a predictor for anonymous sessions that can accurately predict purchase intent in anonymous sessions, beating a production-ready predictor by over 17.54% F_1; and (2) a predictor for identified customers that uses session data as well as customer history and achieves an F_1 of 96.20% on held-out data collected from a real-world retail platform. Finally, we discuss the broader practical implications of our findings.

ACM Reference Format:
Mariya Hendriksen, Ernst Kuiper, Pim Nauts, Sebastian Schelter and Maarten de Rijke. 2020. Analyzing and Predicting Purchase Intent in E-commerce: Anonymous vs. Identified Customers. In *Proceedings of the 2020 SIGIR Workshop on eCommerce (SIGIR eCom'20), July 30, 2020, Virtual Event, China.* ACM, New York, NY, USA, 10 pages.

1 INTRODUCTION

Information retrieval (IR) technology is at the heart of today's e-commerce platforms, in the form of search engines, recommenders, and conversational assistants that connect users to the products they may be interested in [24]. To help improve the effectiveness of IR technology in an e-commerce context, the problem of analyzing, modeling, and, ultimately, predicting customers' purchase intent has been studied extensively in academia and industry [1, 2, 18]

Purchase intent prediction. Here, purchase intent is defined as a predictive measure of subsequent purchasing behavior [21].

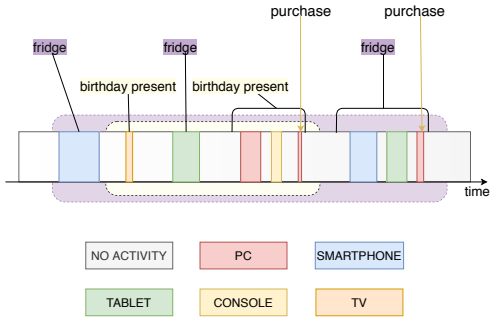

Figure 1: Customer journeys across sessions, with multiple interests and devices; the colors indicate different devices.

Figure 1 illustrates the complexities of customer behavior during a sequence of sessions, when multiple tasks, interests, and devices may play a role. Areas in the back of the figure are meant to signify different user journeys across time, purple for one that is focused on fridges, yellow for one that is focused on a birthday present. Colored rectangular blocks in the front indicate different devices used by the user. Initial exploration of a relatively expensive item (a fridge) starts on a smartphone and continues on a tablet, while the journey ends with a purchase of a fridge on a PC. The purchase of a fridge is interleaved with the purchase of a (lower-priced) birthday present, with initial exploration on a PC, followed by further exploration on a TV and PC, and, ultimately, a purchase on a PC.

Online search behavior that targets transactions has been analyzed at scale at least since the work by Broder [5], who identified a class of so-called *transactional* queries, where the user is seeking to reach a page with more interaction opportunities, e.g., to conduct a purchase, download or sign-up. In particular, factors influencing online purchases have been described as early as in 2002 [9], and work on predicting purchases goes back to at least to the work of [3], where the task was to predict whether a given customer is going to purchase within a given session.

Challenges. Despite the many advances, purchase intent prediction still has many challenges [34]. In particular, previous work on purchase intent prediction has focused mostly on customers of an e-commerce platform who are identified or recognized by the platform. A diverse range of models has been considered, from traditional feature-based models such as boosted decision trees

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGIR eCom'20, July 30, 2020, Virtual Event, China
© 2020 Copyright held by the owner/author(s).

to sequence-based neural models such as RNNs. However, based on the analysis of de-identified data from an e-commerce website available to us, more than 50% of traffic comes from anonymous users. Purchase intent detection for anonymous users is particularly challenging because it cannot rely on historical information about the user on which many of the existing models rely.

Features for purchase intent prediction. In this paper, we focus on identifying signals that suggest purchase intent in an anonymous and identified setting. We do this by analyzing purchase vs. non-purchase sessions sampled from a large European e-commerce website and testing the features based on our observations on a production-ready model. We further test the obtained feature sets on five other classifiers to explore the generalizability of our findings. In particular, we include features derived from session-based data such as page dwell time and customer-specific data such as the number of days since the last purchase. Session-based features have the advantage that they are available both during sessions when a user is identified (i.e., the customer has logged-in or is recognized through cookies) and anonymous sessions (when the customer is not known). Customer-related features are only available during identified sessions. Interestingly, many of the features proposed previously [26] apply only to identified sessions: *purchase intent prediction for anonymous sessions has been studied very little.*

To fill this gap, we analyze a dataset of more than 95 million sessions, sampled from four weeks of anonymized user interaction data in a European e-commerce platform. We answer the following research questions:

RQ1: How do purchase sessions differ from non-purchase sessions? In Section 4 we compare purchase vs. non-purchase sessions in such aspects as session length, temporal variations, device and channel type, queries. Among others, we find out that purchase sessions tend to be longer than non-purchase ones, customers are more likely to purchase in the evening and during a weekday, and more likely to own more than 1 device.

RQ2: What are the important session-based features that allow us to tell purchase sessions apart from non-purchase sessions? What are the important historical features that should inform predictors for identified sessions? How does the importance of features change across the session? Based on the experiments described in Section 5, we conclude that historical features related to previous purchasing behavior are highly important for detecting purchases in the identified setting. For the anonymous setting, however, dynamic features related to page dwell time and sequence of pages are most important. Besides, the importance of dynamic features increases as the session continues, while the importance of static features decreases.

RQ3: How effective are models used for purchase intent prediction for anonymous vs. identified sessions? Furthermore, to which degree do the proposed features help improve performance for anonymous sessions? In Section 5, we show that in the anonymous setting, tree-based and neural classifiers demonstrate the best performance, and adding extra features to models improves F_1 by about 17%. In contrast, for identified setting all models demonstrate high performance and adding extra features do not provide a significant gain.

The principal contributions of our research are the following:

- We conduct an in-depth analysis of a real-world customer interaction dataset with more than 95 million sessions, sampled from a large European e-commerce platform. We identify session features such as device type and conversion rate, weekday, channel type, and features based on historic customer data such as number of previous orders and number of devices to distinguish between purchase and non-purchase sessions (see Section 4).
- We define two feature sets for purchase prediction, tailored towards anonymous sessions and identified sessions (see Section 5).
- We evaluate our proposed features by extending an existing production-ready model and run additional experiments with classifiers generally used for this task. We find F_1 improvements of up to 17% in purchase intent prediction for anonymous sessions and reach an F_1 of 96% for identified sessions on held-out data collected from a real-world retail platform (see Section 5).

2 BACKGROUND AND DEFINITIONS

In our study, we operate with the following definitions.

A *session* is a sequence of requests made by a single end-user during a visit to a particular site. A session ends if the user is idle for more than 30 minutes. We define two types of sessions: *purchase sessions*, during which the customer buys an item, and *non-purchase sessions*, during which the customer does not buy anything. In connection to this, we define *purchasers* as customers who had at least one purchase session, whereas *non-purchasers* are customers who were identified but have never purchased anything. We furthermore distinguish between *identified sessions*, where a customer is logged in or recognized with a browser cookie, and *anonymous* sessions where this is not the case. Additionally, we denote the number of actions taken during a given session as the *session length*, where an *action* corresponds to opening a new web page, submitting a search query, or adding/removing an item to/from the shopping basket.

Device switch is the act of changing the type of browsing device between two consecutive sessions that belong to the same journey. For instance, if a customer first explores the platform on a smartphone and afterward accesses the platform on a PC, she switches from a smartphone to a PC.

A *channel* indicates the way through which a customer enters the platform. For example, if the customer comes to the platform via an advertisement, she uses a paid channel.

The *conversion rate* denotes the fraction of visits during which a purchase was made [19]. We use this metric to compare device popularity in a purchasing context. We calculate the conversion rate by dividing the number of purchasing sessions by the overall number of sessions. In order to protect sensitive information, we only report *standardized conversion rates* for each device; since we are interested in differences across devices types, this suffices for our purposes. The standardized conversion rate is computed by subtracting the mean conversion rate per device type from the desired conversion rate and dividing the result by the standard deviation of the device-specific conversion rate. For instance, if our device specific conversion rates are $\textit{Conversion Rates} = \{0.5, 0.2, 0.3\}$, the mean of device-specific conversion rates is $\overline{\textit{Conversion Rate}} = 0.33$ and the standard deviation of conversion rates is $\sigma = 0.12$. Therefore, the resulting standardized conversion rates are $\textit{Standardized Conversion Rates} = \{1.34, -1.07, -0.27\}$

3 DATASET DESCRIPTION

In this section, we describe how we extract a dataset consisting of anonymized user interaction data from the search logs of an e-commerce platform, and summarize dataset statistics.

Data Collection. Our dataset comprises four weeks (28 days) of anonymized visits sampled from a European e-commerce platform in October 2019. The original sample of the log entries includes a unique non-personal customer identifier (for identified users), the type of browsing device used during the session, as well as a timestamp for every query, and a URL of each clicked page. We convert all the timestamps to the Central European Time Zone (CET). We additionally recorded the price of every product the customers have seen and the prices of the items they ended up buying. In cases where a customer starts a session without logging in and ends up logging in at a later point in the session, we assign the session to the customer.

To filter out bot traffic, we apply several measures related to location and device type [4]. First, we filter out sessions based on location, to only include entries from the European countries from which the majority of the customers come; bots come mostly from non-European IPs, especially North-America. Second, we specify the set of device types we are interested in and remove all the entries from other devices, leaving us with *PC*, *Smartphone*, *Tablet*, *Game Console*, and *TV*; bots often do not specify a device type.

Dataset Statistics. Table 1 provides descriptive statistics of the resulting dataset. Overall, the dataset contains 95,757,177 sessions, out of which 54,144,152 (about 56.5%) are anonymous. In total, the dataset contains 9,663,509 identified users. We additionally keep track of the device types used for browsing and distinguish between five such device types: PC, smartphone, tablet, game console, and TV. The table also lists the number of search queries; these are the queries submitted during the sessions captured in the log.

Table 1: Dataset statistics.

Description	Total
Sessions	94,402,590
Anonymous	55,305,709
Logged-in or recognized	39,096,881
Logged-in or recognized customers	6,125,781
Queries	31,185,176
Device types	PC, Smartphone, Tablet, Game Console, TV

4 CHARACTERIZING PURCHASE INTENT

We explore customer behavior and, in particular, the difference in the behavior of purchasing and non-purchasing users. These explorations aim to identify characteristics that may help us improve the effectiveness of purchase intent predictors. We analyze several aspects of sessions, such as the length of purchase and non-purchase sessions, the temporal characteristics of sessions, and device information. Furthermore, we investigate the channels from which customers start sessions and issue queries during purchase sessions and non-purchase sessions.

4.1 Session Length

First, we examine the overall session length for purchase sessions and non-purchase sessions. Figure 2 plots the complementary cumulative distribution function (CCDF) of the session lengths of purchase sessions and non-purchase sessions per device type.

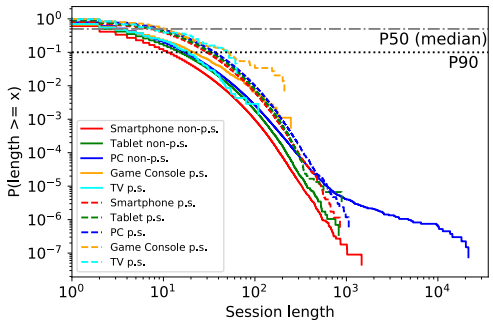

Figure 2: CCDF of the session length per device type for purchase sessions (p.s.) and non-purchase sessions (non-p.s.).

As can be seen in the area between the P50 and P90 percentiles in Figure 2, purchase sessions are in general longer than non-purchase sessions. Moreover, the purchase session length per device varies less than the non-purchase session length per device. It can be explained by the fact that non-purchase sessions can be both very short or rather long, depending on the underlying user intents. For instance, a user could quickly look something up or spend some time exploring the catalog. On the other hand, in the case of purchase sessions, user intentions are less ambiguous. Usually, users look for a specific product that they have in mind and, upon finding it, proceed to purchase.

From a device perspective, the shortest sessions take place on smartphones, whereas sessions on tablets are generally longer. The longest sessions occur on the PC, TV, and game console. This finding holds for both purchase sessions and non-purchase sessions. However, in the tail of the distributions, the distinction between purchase session length and non-purchase session length is not as clear as between the P50 and P90 percentiles. The non-purchase session length distribution on the PC has an exceptionally long tail. Overall, we can attribute these findings to the fact that smartphones have a smaller screen and are therefore less convenient for longer sessions. Tablet screens are bigger than smartphone screens; hence, the sessions can last longer. The PC screen is the biggest one, and therefore PC users exhibit event longer sessions.

4.2 Temporal Variations

Next, we look into the temporal characteristics of purchase sessions and non-purchase sessions, such as their distribution across days of the week and the sessions' starting hours.

First, we want to understand customer activity during the days of the week. Figure 3 shows how the number of purchase and non-purchase sessions varies across days of the week. The three most popular days for purchase sessions are Thursday, Tuesday, and Wednesday. In total, the purchase sessions of these three days amount to 48.55% of all purchase sessions. On the other hand, the least popular purchase days are Sunday, Saturday, Monday, and

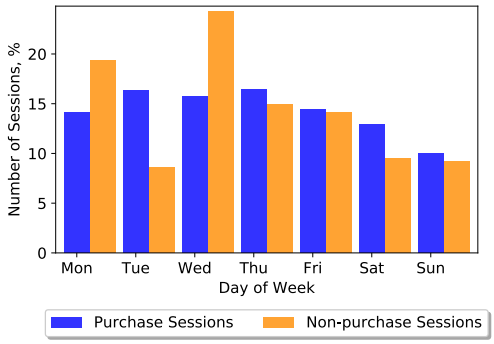

Figure 3: The fraction of purchase sessions and non-purchase sessions across days of the week w.r.t. total amount of purchase and non-purchase sessions. Most activity occurs on weekdays.

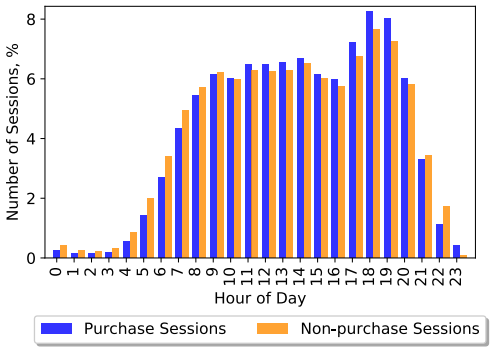

Figure 4: The fraction of purchase session and non-purchase sessions across the hours of the day w.r.t. total amount of purchase and non-purchase sessions. Most purchase sessions start in the evening.

Friday. They contribute to 51.45% of purchase sessions. The observed pattern of purchase behavior hints at the fact that customers prefer to buy during weekdays, which aligns with their workweek. Besides, we conclude that the lower purchase activity on Monday and Friday attributes to their proximity to weekends.

In the case of non-purchase sessions, the most active session days are Wednesday, Monday, and Thursday. Altogether, these days contribute to 58.55% of non-purchase sessions. The least active days are Tuesday, Sunday, Saturday, and Friday. All the sessions of these days amount to 41.44%. Just like for purchase sessions, the activity for non-purchase sessions also centers around weekdays. However, the difference between the three most active days and the four least active days for non-purchase sessions is bigger than the corresponding difference for purchase sessions. For purchase sessions, the difference is only 2.9%, whereas, for non-purchase sessions, the difference is 17.11%. Moreover, Tuesday, the 2-nd most popular day for purchase sessions is the least popular day for non-purchase sessions. On the other hand, Monday, the 2-nd most popular day for non-purchase sessions is the 3-rd least popular day for purchases. The observation indicates that people need time to consider a purchase before making the buying decision. Hence, they spend Monday, the first day of the new week on considering the purchase, and the purchase itself happens on Tuesday or later in the week. In general, the most active day of the week is Wednesday, whereas the least active day is Sunday. These findings strongly suggest that user behavior depends on the day of the week. In general, people are most active on weekdays, during their workweek, their activity peaks in the middle of the week. On the other hand, at the beginning and end of the workweek, user activity is generally lower.

Next, we look at user behavior on the level of the hour during which a session starts. As mentioned in Section 3, all the hours are represented in CET. Figure 4 shows how purchase sessions and non-purchase sessions spread across the hours of the day.

As expected, the least active hours are in the early morning, in the period from 1 am to 3 am. That can be explained by the fact that most people sleep during the night. (Note that the majority of the e-commerce platform customers come from Europe; hence, time does not vary that much.) Moreover, the activity on the platform during the period from 10 am till 5 pm is stable both for purchase and non-purchase sessions, whereas the most active hours are in the evening, i.e., from 6 pm till 8 pm. In general, our observations correspond to the established rhythm of the daily life of the majority of people, who sleep during the night, browse e-commerce platforms both during work hours and in the evening after work.

4.3 Channel Types

Next, we look at whether channel types distributions change across purchase and non-purchase sessions. We define the following channel types: *direct* where a user enters the platform directly; *paid* where a user enters the platform through search engine advertisement, and *organic* where a user enters the platform through a web search engine and unpaid results. Table 2 displays the channel distribution across purchase and non-purchase sessions.

Table 2: Channel types for purchase and non-purchase sessions.

	Sessions		Stand
Channel	**Purchase (%)**	**Non-purchase (%)**	**conv. rate**
Direct	**71.07**	**77.30**	-0.56
Paid	16.74	12.92	0.54
Organic	11.78	7.83	**0.94**
Other	0.31	1.05	-1.33

Both for purchase and non-purchase sessions, the direct channel is the most used channel to enter the platform. However, for purchase sessions, the percentage of sessions which start with the direct channel is 8.06% less than the fraction of non-purchase sessions, which started with the direct channel. The second most popular channel for purchase and non-purchase sessions is a paid channel. However, in the case of this channel, the fraction of purchase sessions is 12.92% bigger than the corresponding channel type fraction for non-purchase sessions. The organic channel is the third channel in terms of popularity for both session groups. The organic channel fraction for purchase sessions is 50.40% bigger for purchase sessions when compared with non-purchase sessions.

Overall, during purchase sessions, users are more likely to enter the platform through paid or organic channels, whereas for non-purchase sessions the direct channel is more common. It can be explained by the fact that purchasers decide to converge after being offered an advertisement or a search result that matches their

interest, whereas non-purchasers may enter the platform directly to explore the catalog.

4.4 Devices

In this subsection, we investigate purchase intent from the perspective of device types. In particular, we look at the device types used by purchasers and non-purchasers and analyze device switches.

4.4.1 Device type. First, we want to understand how many users are using multiple devices and which devices customers use for purchase and non-purchase sessions.

Table 3: User device statistics per session.

Device(s)	Purchasers (%)	Non-purchasers (%)
> 1 device	24.05	16.22
1 device	75.95	83.78
2 devices	22.23	15.39
3 devices	1.82	0.82
4 devices	≈ 0	≈ 0
5 devices	0	0

Table 3 shows how many devices purchasers and non-purchasers own. The majority of users from both groups are single-device users. However, the fraction of single-device purchasers is 9.35% smaller than the corresponding fraction of non-purchasers. On the other hand, the fraction of multi-device users for purchasers is 45.28% bigger than the corresponding fraction for non-purchasers. In general, multi-device users represent almost a quarter of the purchasers. As the number of devices increases, the difference between purchasers and non-purchasers grows. Our observations support the statement that multi-device users tend to be more engaged [20].

Table 4: Purchase and non-purchase sessions per device type and standardized conversion rates.

Device	Purchase sessions (%)	Non-purchase sessions (%)	Stand. conv. rate
Smartphone	47.00	58.09	−0.56
PC	44.97	34.40	1.61
Tablet	8.03	7.50	0.61
Game Console	0.004	0.004	−0.40
TV	0.001	0.002	−1.25

Next, we examine the distributions of purchase and non-purchase sessions across device types and device-specific standardized conversion rates; see Table 4. The PC is the device with the highest conversion rate. Indeed, the fraction of purchase sessions is 30.70% bigger than the fraction of non-purchase sessions. The device with the second-highest conversion rate is a tablet. For this device, purchase sessions are 7.12% more frequent than non-purchase sessions. The Smartphone is the device with the second-lowest conversion rate. For this device, the number of purchase sessions is 19.10% less frequent than the number of non-purchase sessions.

Game consoles and TVs are relatively new devices in e-commerce; hence, sessions with these devices are relatively less frequent. Nevertheless, based on our observations, we find that the game console is a device with the third-highest conversion rate. Interestingly, its conversion rate is close to that of the smartphone. It can be explained by the fact that device functionalities of smartphones and tablets in e-commerce context blur due to the similarity of their interfaces and screen sized. The number of purchase sessions on a game console is 15.47% less than the number of non-purchase sessions. The TV is the least common device, with the lowest conversion rate. The number of purchase sessions on this device is 34.01 less than the number of non-purchase sessions.

We can explain our findings by the fact that customers use different devices for different purposes. For example, PCs and tablets seem to be used for the purchase, whereas smartphones, game consoles, and TVs for exploration.

4.4.2 Device switches. Next, we analyze how users switch between devices before a purchase session.

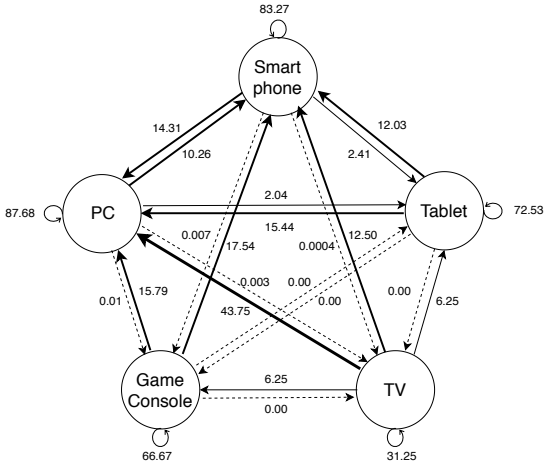

Figure 5: Device transition probability before a purchase session, including self-transitions. The thickness of an arrow indicates the connection strength; the dashed line is the weakest connection.

Figure 5 shows device transition probability, including self-transitions. Generally, the situation when a user remains on the same device is the most likely outcome for all devices, except TV. There, the self-transition probability is lower than the probability of switching from TV to PC, a device with the highest self-transition probability. A probability of remaining on a smartphone is 5.03% lower than a self-transition probability for PC, whereas a probability to remain on a tablet is 17.28% lower than the probability to remain on PC. The game console has the second-lowest self-transition probability.

Next, we consider connections between two different devices. We characterize those interconnections based on how likely a user is to switch from one device to another one and vice versa.

Strong interconnections. Some pairs of devices have high probability interconnections. The strongest connection is between a smartphone and a PC, the two most popular devices. The second strongest connection is between PC and tablet. There is a bigger discrepancy between probability rates, with the probability of switching to PC being 656.86% higher than of switching to tablet. The third strongest interconnection is between a smartphone and a tablet with a stronger connection switch to a smartphone, a more popular device. The probability of switching to a smartphone is 399.17% higher than switching to a tablet. Overall, the three interconnections form a triangle that includes the three most popular

devices: PC, smartphone, and tablet.

One-sided interconnections. For a one-sided interconnection there is a high probability of switching from one device to another, but a close to zero probability of switching back. There are six cases of this type in Figure 5. TV is the device with the largest number of one-sided interconnections, with PC, smartphone, tablet, and game console. In all cases, the transition probability is low when TV is a target device, which can be explained by relative difficulty to purchase on TV. The strongest one-sided interconnection is between TV and PC. The probability of switching from TV to PC is 43.75%, the highest transition probability for TV, and the highest probability to transition to another device. We explain this by the fact that PC is one of the most popular devices for purchase. The second most likely device people switch to from TV is a smartphone, whereas a probability to switch to a tablet or game console is 6.25%.

Another device with a significant number of one-sided interconnections is the game console. Apart from the connection with TV discussed above, the device also has this connection type with smartphone and PC. The transition probability is close to zero when a game console is a target device. Unlike the situation with TV, came console has a higher probability of switching to a smartphone, whereas the probability of switching to PC is 1.75% less.

In general, all one-sided interconnection cases include switching from a less common device type such as game console or TV to a more conventional device, such as PC, smartphone, or tablet.

Weak interconnections. In some cases, the switch between two devices rarely happens, i.e., the transition probability is close to zero. As can be seen in Figure 5, there is only one case of this type. It is a connection between a game console and a tablet.

Overall, the analysis of device switches before a purchase session supports the conclusion that users tend to switch from less popular devices such as TV and game console to more popular ones such as PC, smartphone, and tablet.

4.5 Queries

The next aspect of purchase intent that we examine is queries. We look at the number of queries in purchase and non-purchase sessions and per device type. In total, the dataset contains 31,185,176 queries, 1,302,195 or 4.17% of which are unique. Given the number of sessions in the dataset, we can conclude that queries are infrequent.

Table 5: Queries per session for purchase and non-purchase sessions per device type, percentages are computed w.r.t. total number of queries per purchase or non-purchase session.

Device	Purchase sessions		Non-purchase sessions	
	query/session	%	query/session	%
Smartphone	4	**52.92**	0.05	42.40
PC	2	36.38	0.09	**57.54**
Tablet	4	10.67	0.0003	0.049
Game Console	≈ 0	≈ 0	≈ 0	≈ 0
TV	≈ 0	≈ 0	≈ 0	≈ 0
Avg	3.16	100	0.06	100

Table 5 shows the query per session frequencies across five devices for both purchase and non-purchase sessions. Besides, it also demonstrates which devices are most popular for querying during purchase and non-purchase sessions. Overall, queries are more common in purchase sessions. This can be explained by the fact that querying is more likely to happen when customers are determined to buy something.

Naturally, queries are most common for smartphones, PCs and tablets, and uncommon for game consoles and TVs. Indeed, the current interface of game console and TV makes it difficult to type queries, especially when compared to a PC or a smartphone.

The PC has the highest query per session frequency for non-purchase sessions and second-highest frequency for purchase sessions. A smartphone has the second-highest query per session frequency for non-purchase sessions and the highest query frequency per non-purchase session. Tablet, on the contrary, has the third-highest frequency for non-purchase sessions and the highest frequency for purchase sessions.

When it comes to query distributions per device for purchase and non-purchase sessions, the ranking is somewhat consistent for both groups. During purchase sessions, most queries are issued on a smartphone, whereas during non-purchase sessions PC prevails. On the other hand, PC is the second most popular device for purchase sessions, whereas for non-purchase sessions smartphone takes the second place. Tablet is third for both groups. In general, the query distribution across devices correlates with the session distribution across devices (see Table 4).

Table 6: Unique query counts for purchase and non-purchase sessions per device type. The percentage is computed w.r.t. total number of queries per purchase or non-purchase session.

Device	Purchase sessions		Non-purchase sessions	
	count	%	count	%
Smartphone	180,542	36.89	321,640	39.57
PC	224,763	45.92	312,855	38.49
Tablet	83,881	17.14	175,909	21.64
Game Console	136	0.03	1,841	0.23
TV	46	0.01	582	0.07
Total	489,368	1.91	812,827	14.52

Next, we look at the number of unique queries for purchase and non-purchase sessions and per device. Table 6 shows unique queries count and their corresponding fractions. The fractions are computed w.r.t. the total number of queries per session type and device. Overall, during purchase sessions users issue less unique queries, it holds for every device class but a PC. This can be explained by the fact that during purchase sessions users may retype a previous query to revisit the results they have seen earlier, whereas non-purchasers want to explore and hence use more unique queries.

4.6 Purchase Intent Characteristics

What have we learned from the log analysis conducted in this section that might help us to devise better models for purchase intent prediction? We found out that purchase sessions tend to be longer what suggests that session length is an essential indicator of purchase intent. Besides, the difference in session length depends on the type of device customer use. Moreover, we discovered how

the day of week and hour of the day influence purchase behavior. In particular, customers are more likely to buy during the weekdays and in the evening. From the perspective of channels, there is a difference, too. In particular, for non-purchase sessions, the direct channel is more common, whereas purchase sessions are more likely to start with paid or organic channels. From the device perspective, we found out that multi-device users are more common among purchasers. Besides, we figured the probability of purchase for every device and characterized transitions between devices. After looking into queries in the dataset, we discovered that during purchase sessions, users issue more queries per session. Besides, during purchase session, there are less unique queries.

5 PREDICTING PURCHASE INTENT

Next, we turn to predict purchase intent when a user is anonymous ("anonymous setting") and when a user is logged-in or recognized ("identified setting"). The goal of our experiments is to evaluate how the features which we discovered during dataset exploration influence purchase predictor performance in both settings. To accomplish this, we derive a feature set for each setting, and evaluate the features by adding them to an existing production-ready model, based on a Random Forest. To showcase the generalizability of our findings, we additionally test the impact of our features on five additional popular classifiers. To investigate how the models' ability to predict purchase evolves throughout a session, we evaluate all models on 11 session steps (corresponding to the visits of 10 pages). We are interested in longer sessions because the outcome of such sessions is more difficult to predict. As we do not want to evaluate the model's performance on the very last step, (where the outcome is clear), we set up a buffer of 2 pages. Therefore, we filter out all the sessions which are shorter than 12 pages. We conclude the section by analyzing the features which contributed most to the model performance in both the anonymous and the identified setting, and explore how dynamic and static feature importance change as the session continues.

5.1 Experimental Setup

In this section, we discuss the feature sets which we use in the experiments for the anonymous and identified setting, the models on which we test the features, and the evaluation setup.

Feature sets. We start by designing a set of features for purchase prediction in identified and anonymous user settings. Since our initial analysis demonstrated that about 56% of all sessions are anonymous (see Table 1), it is worth to pay special attention to this category. Based on the findings obtained thus far and on an analysis of best-performing features available in the literature [13, 17, 22, 26], we compile a feature set presented in the Table 7.

We categorize features into two classes: *session features* and *customer history features*. We derive session features from the information of the given session and base customer history features on the information from previous sessions of the given customer.

Since we run experiments in the anonymous and the identified setting, we use different feature sets for each setting. In the anonymous setting, the information about the customer is not available and, therefore, we can only use session features. On the other hand, when a customer is identified, we can use both session and customer history features. The feature set contains both static and dynamic features. Dynamic features can change throughout the session, whereas static features remain constant.

Table 7: Complete feature set. "Dynamic" indicates that a feature may change during a session.

	Feature	Dynamic	Baseline
Session	current page dwell time, mean	✓	✓
	current page dwell time, σ	✓	✓
	page sequence score	✓	✓
	number of pages	✓	✓
	channel type		
	start hour		
	week day		
	device type		
	device conversion rate		
History	number of orders		✓
	days since last purchase		✓
	number of sessions		
	number of devices		
	device sequence score		
	switch probability		

Models. Next, we select models on which we evaluate the features discovered during the dataset analysis. As our primary model, we use a production-ready classifier. This is a random forest (RF) with a baseline feature set as described in Table 7. Additionally, to showcase the general utility of our feature set, we experiment on additional models. After reviewing previous work in the domain of purchase prediction (see Section 6), we choose the following models for our experiments: logistic regression (LR), K-nearest neighbors (KNN), support vector machines (SVM), neural classifier, and gradient boosted decision tree (GBDT). Each model is trained on the baseline and extended feature set in both settings.

Prediction setup. Since we want to explore how models' performances change across sessions, we select points of a session for which we predict the probability of purchase.

We define a point by the number of pages opened in the session up until the point of prediction. Overall, we select 11 points of measurement. The first point is at the very beginning of the session when the user did not open any pages yet. At this point, the classifier makes a prediction based solely on static features. The following point of measurement is right after the user opened the first page. The subsequent nine points happen after the next nine pages. To make the evaluation possible and to ensure that we do not predict for the very last session page, we filter out sessions with fewer than 12 pages, with 2 pages as a buffer. The buffer is there to avoid the situation when the model predicts at the very end of a session when the outcome is clear. Therefore, we filter out all the sessions which are less than 12 pages long. For example, in step 2 we only have a session with at least 12 actions, which is a hard setting.

Evaluation setup. For both settings, we evaluate model performance with 10-fold cross-validation. To account for class imbalance, we set class weights to be inversely proportional to class frequencies and use F_1 score as a primary evaluation metric.

5.2 Prediction for Anonymous Users

First, we evaluate how the added features influence model performance in the anonymous setting, where the user is not known.

Setup. For the anonymous setting, we sampled 22,982 sessions. We use the data to create a feature set for the baseline model and our model. In the anonymous setting, there is no available information about customer history, therefore, we only use session features (see Table 7. As can be seen from the table, the baseline feature set comprises four dynamic features, whereas the extended feature set offers five extra features. Since we predict purchase for different points in the session, we compute all dynamic features for a particular session point on which we evaluate. For each session point, we train the baseline and extended model on the obtained feature sets.

Results. The results in Figure 7 show that the additional features boost model performance across all session steps. The performance boost is especially significant at step 0 when a customer has not opened any pages yet. In general, tree-based models (RF and GBDT) and the neural classifier demonstrate the highest scores across all steps. The models are followed by SVM, LR, and KNN classifiers.

The performance of all the models with the baseline feature set improved on step 1. The gain can be explained by the introduction of dynamic session features (step 0 means that the user did not open any pages yet, hence, no dynamic session features). Conversely, for models with the extended feature set, the introduction of dynamic features on step 1 does not significantly increase the performance. After step 1, models' performances reach a plateau.

5.3 Prediction of Identified Users

Second, we test models' performances with baseline and extended feature sets in identified setting, when the user is known.

Setup. For the identified setting, we sampled 6,319 sessions. The feature set for this setting includes session and customer history features (see Table 7). During our experiments, we found out that information about the previous session device (such as device type and conversion rate) decrease model performance, so we excluded those features from the training and evaluation sets. This can be explained by the fact that the information about what kind of device users previously used and what was the probability of purchase on that device is not relevant for predicting purchase on the current device. In analogy with the anonymous setting, we prepare feature sets for each of the eleven session points and train and evaluate the models with the baseline and extended feature sets.

Results. Figure 7 shows the performance of the models. Overall, the performance of all models for both baseline and extended feature sets and across all steps stays around 96%. The only exception is the k-nearest neighbors classifier where adding extra features on step 0 increases the model's performance by 6.74%. On step 1, however, the gain from the extended feature set is not present. This can be explained by the introduction of the dynamic session features.

5.4 Feature Importance Analysis

The experimental results raise a natural question that is 'Which features contribute most to model performance in both settings?' To answer this question, we look at the feature importance scores

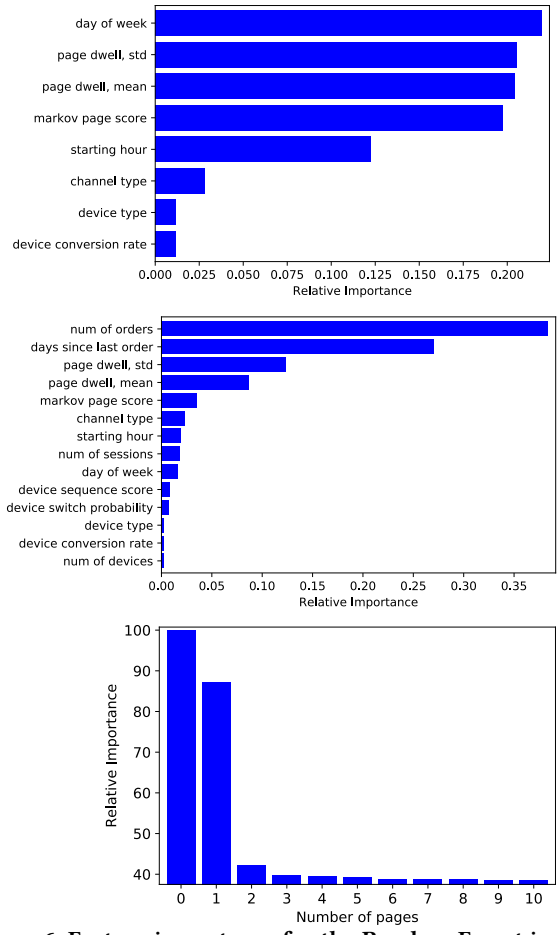

Figure 6: Feature importance for the Random Forest in the anonymous setting (top) and identified setting (center), as well as summed for static features in the anonymous setting (bottom).

of a production-ready classifier which shows one of the best performances in both settings, random forest.

Figure 6 (top) demonstrates that in anonymous setting, day of the week is the feature with the highest importance. It is followed by three dynamic features (standard deviation and mean of page dwell time, and Markov page sequence score), and four static features (starting hour, channel type, device type and conversion rate).

Figure 6 (center) shows that in the identified user setting, number of previous orders, and number of days since last order are the features with the highest relative importance. Both features describe user historical purchasing behavior what can explain their high relative importance. The features are followed by three dynamic features (standard deviation and mean of page dwelling time, and Markov page sequence score), which also have relatively high importance in the anonymous setting. The high relative importance of the dynamic session features (standard deviation and mean of page dwelling time, and Markov page sequence score) in both settings explain the gain all models with baseline feature set got on step 1 in the anonymous setting (see Figure 7).

Next, we determine how static feature importance changes across

sessions. We consider the importance in the anonymous setting because the introduction of dynamic features in this setting showed an improvement. Figure 6 (bottom) shows that static session feature importance decrease as the session evolves, which entails that the importance of dynamic features increases. On step 0 the cumulative importance of static features is 100% because there are no dynamic features introduced. However, from step 1 the relative importance starts to drop. The figure supports the hypothesis that as the session progresses dynamic features become more important.

6 RELATED WORK

E-commerce user purchase behavior analysis. Research on understanding online users' purchasing behavior has been ongoing since the very beginning of e-commerce [2]. Studies have investigated user motivation [2], factors that influence e-commerce adoption [23], as well as purchasing behavior [6, 14], with a focus on perceived security [9, 25], the decision-making process [27], and purchaser profiles [12, 32]. Besides, there has also been work on user behavior on content discovery platforms and its relationship to subsequent purchases [18] as well as work dedicated to the identification of a taxonomy of product search intents and the prediction of user satisfaction [29].

Unlike previous work, our study focuses on the exploration of user purchasing behavior by comparing purchase vs. non-purchase sessions. Besides, we analyze the data from the perspective of device types, and explore aspects such as session length, price of the seen products from the perspective of different devices. On top of that, we also look into the way customer switches between devices.

Purchase prediction in e-commerce. The problem of e-commerce user behavior modeling has been studied from various angles, such as building multiple classifiers based on genetic algorithms [15], mining purchase patterns with association rules and using those patterns for purchase prediction [31]. Research has been focused on creating models robust to noise in session data [1], and using a recurrent neural network to predict customer behaviour [16].

Sismeiro and Bucklin [28] predict purchasing task completion for a given user who completed at least one task earlier, whereas Cheng et al. [7] explore user behavior on a content discovery platform to determine intent specificity and time in the future when a purchase is estimated to take place. Some work in the field focuses on using queries for purchasing behavior modeling. For instance, Dai et al. [8] predict purchase based on input query. Besides using general session data, there has been work that incorporates demographic data and perceived attributes [36], scrolling and mouse movements [11], payment data [35], log-trace data [33], and phone touch actions [10]. There has been work on analyzing behavioral patterns and the exploration of different model architectures. In particular, support vector machines, K-nearest neighbor approach, random forest, and logistic regression were used [17, 22, 30].

Unlike previous work in this domain, our study focuses on purchase prediction with two types of users, identified and anonymous. Therefore, we develop two models, run them in two settings, and evaluate their results. The possibility to experiment with identified users also allows us to leverage information from previous user sessions, such as user purchasing history and the number of devices a user owns. In contrast, anonymous users contribute to a higher share of traffic, which makes it important to understand their behavior too. Additionally, we explore how the relevance of dynamic and static features changes as a session progresses.

7 DISCUSSION & CONCLUSION

In this paper, we have carried out an analysis of user purchase intent in e-commerce. We have analyzed four weeks of session logs from a European e-commerce platform to identify signals in user behavior that can imply purchase intent. We have considered aspects such as session length, day of the week, and session start hour, as well as information about device, channel, and queries.

In the second part of our study, we have analyzed the relevance of the discovered signals by running a series of experiments aimed at purchase intent prediction in the anonymous and identified settings. We tested the features on random forest, the model which fits production requirements. Additionally, we tested the features on five other models. The experiments demonstrated the value of the features that we engineered based on our insights into the data. We explored which features contribute to performance improvement.

One of the implications of our study is enhanced understanding of purchasing user behavior in e-commerce. Understanding the behavior is the first step towards modeling it, as we demonstrated in the second part of the paper. Modeling user behavior can contribute towards reducing friction in the customer journey and, therefore, to better customer experience. Besides, we explored the topic of detecting the purchase intent of anonymous users. We showed that, while anonymous users contribute to more than half of the traffic, their user intent is harder to detect because all the predictions have to be made without knowledge about the prior behavior.

Our research has several limitations; one of them is limited generalizability. Even though the data we use in our study comes from a dominant e-commerce platform, it is still only one platform. Hence, it would be interesting to verify the findings against other e-commerce platforms and explore the differences. Moreover, we sampled four weeks of data, thereby introducing a sample bias that could make our findings sensitive to unknown temporal or seasonal patterns. Therefore, it would be interesting to explore if expanding our dataset will lead to new insights. For example, if we had several months of data, we could explore how user purchase intent changes across different months or seasons. Furthermore, we evaluated our purchase intent prediction models in an offline setting. The next logical step is to evaluate them in an online setting.

Future research on the topic includes several directions. First, there is an opportunity to continue research into general purchase behavior analysis and modeling in e-commerce. It would be interesting to explore more aspects of purchasing behavior and try out more models. Another direction for further research concerns predicting purchase intent for anonymous users. Another exciting direction for further research includes modeling device-specific purchase behavior. It can include both relatively common devices such as PC, smartphone, and tablet, and relatively less popular and studied devices such as TV or game console.

8 REPRODUCIBILITY

All plots for our paper, as well as the code to regenerate them, can be found in our Git repository:

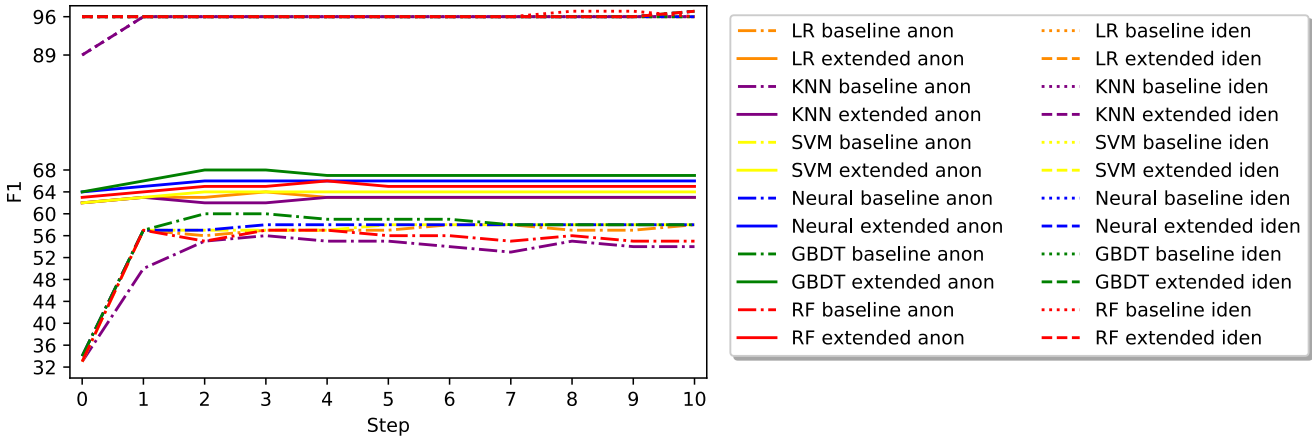

Figure 7: Experimental results in anonymous (anon) and identified (iden) setting, across different session steps, F_1.

https://github.com/mariyahendriksen/purchase_intent.

ACKNOWLEDGMENTS

This research was supported by Ahold Delhaize. We thank the three anonymous reviewers whose comments helped to improve and clarify the manuscript.

REFERENCES

[1] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. 2006. Learning User Interaction Models for Predicting Web Search Result Preferences. *SIGIR*.
[2] Steven Bellman, Gerald Lohse, and Eric J Johnson. 1999. Predictors of Online Buying Behavior. *Commun. ACM* 42 (1999), 32–48.
[3] David Ben-Shimon, Alexander Tsikinovsky, Michael Friedmann, Bracha Shapira, Lior Rokach, and Johannes Hoerle. 2015. RecSys Challenge 2015 and the YOOCHOOSE Dataset. In *RecSys*. 357–358.
[4] Christian Bomhardt, Wolfgang Gaul, and Lars Schmidt-Thieme. 2005. Web Robot Detection-preprocessing Web Logfiles for Robot Detection. In *CLADAG*. 113–124.
[5] Andrei Broder. 2002. A Taxonomy of Web Search. *SIGIR Forum* 36, 2, 3–10.
[6] Mark Brown, Nigel Pope, and Kevin Voges. 2003. Buying or Browsing? An Exploration of Shopping Orientations and Online Purchase Intention. *EJM* 37, 11/12 (2003), 1666–1684.
[7] Justin Cheng, Caroline Lo, and Jure Leskovec. 2017. Predicting Intent Using Activity Logs: How Goal Specificity and Temporal Range Affect User Behavior. In *WWW*.
[8] Honghua Kathy Dai, Lingzhi Zhao, Zaiqing Nie, Ji-Rong Wen, Lee Wang, and Ying Li. 2006. Detecting Online Commercial Intention. In *WWW*. ACM, 829–837.
[9] Joey F George. 2002. Influences on the Intent to Make Internet Purchases. *Internet Research* 12, 2 (2002), 165–180.
[10] Long Guo, Lifeng Hua, Rongfei Jia, Binqiang Zhao, Xiaobo Wang, and Bin Cui. 2019. Buying or Browsing?: Predicting Real-time Purchasing Intent using Attention-based Deep Network with Multiple Behavior. *KDD*. 1984–1992.
[11] Qi Guo and Eugene Agichtein. 2010. Ready to Buy or Just Browsing?: Detecting Web Searcher Goals from Interaction Data. In *SIGIR*. 130–137.
[12] Blanca Hernández, Julio Jiménez, and M José Martín. 2011. Age, Gender and Income: Do They Really Moderate Online Shopping Behaviour? *Online information review* 35, 1 113–133.
[13] Walter Hop. 2013. *Web-shop Order Prediction Using Machine Learning*. Master's thesis. Erasmus University Rotterdam.
[14] Meng-Hsiang Hsu, Chia-Hui Yen, Chao-Min Chiu, and Chun-Ming Chang. 2006. A Longitudinal Investigation of Continued Online Shopping Behavior: An Extension of the Theory of Planned Behavior. *International Journal of Human-Computer Studies* 64, 9 (2006), 889–904.
[15] Eunju Kim, Wooju Kim, and Yillbyung Lee. 2003. Combination of Multiple Classifiers for the Customer's Purchase Behavior Prediction. *Decision Support Systems* 34, 2 (2003), 167–175.
[16] Tobias Lang and Matthias Rettenmeier. 2017. Understanding Consumer Behavior with Recurrent Neural Networks. In *MLRec*.
[17] Munyoung Lee, Taehoon Ha, Jinyoung Han, Jong-Youn Rha, and Ted Taekyoung Kwon. 2015. Online footsteps to purchase: Exploring consumer behaviors on online shopping sites. In *Proc. WebSci*. 1–10.
[18] Caroline Lo, Dan Frankowski, and Jure Leskovec. 2016. Understanding Behaviors that Lead to Purchasing: A Case Study of Pinterest. *KDD*. 531–540.
[19] Wendy W Moe and Peter S Fader. 2004. Dynamic Conversion Behavior at E-commerce Sites. *Management Science* 50, 3 (2004), 326–335.
[20] George Montanez, Ryen White, and Xiao Huang. 2014. Cross-device Search. *CIKM*. 1669–1678.
[21] Vicki G Morwitz and David Schmittlein. 1992. Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which "Intenders" Actually Buy? *Journal of marketing research* 29, 4 (1992), 391–405.
[22] Xi Niu, Chuqin Li, and Xing Yu. 2017. Predictive Analytics of E-commerce Search Behavior for Conversion. (2017).
[23] Aron O'cass and Tino Fenech. 2003. Web Retailing Adoption: Exploring the Nature of Internet Users Web Retailing Behaviour. *Journal of Retailing and Consumer services* 10, 2 (2003), 81–94.
[24] Jennifer Rowley. 2000. Product Search in E-shopping: A Review and Research Propositions. *Journal of Consumer Marketing* 17, 1 (2000), 20–35.
[25] W David Salisbury, Rodney A Pearson, Allison W Pearson, and David W Miller. 2001. Perceived Security and World Wide Web Purchase Intention. *Industrial Management & Data Systems* 101, 4 (2001), 165–177.
[26] Hannah Sophia Seippel. 2018. *Customer Purchase Prediction through Machine Learning*. Master's thesis. University of Twente.
[27] Sylvain Senecal, Pawel J Kalczynski, and Jacques Nantel. 2005. Consumers' Decision-making Process and their Online Shopping Behavior: A Clickstream Analysis. *Journal of Business Research* 58, 11 (2005), 1599–1608.
[28] Catarina Sismeiro and Randolph E Bucklin. 2004. Modeling Purchase Behavior at an E-commerce Web Site: A Task-completion Approach. *Journal of marketing research* 41, 3 (2004), 306–323.
[29] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping Ma. 2018. User Intent, Behaviour, and Perceived Satisfaction in Product Search. *WSDM*. 547–555.
[30] Grażyna Suchacka, Magdalena Skolimowska-Kulig, and Aneta Potempa. 2015. A k-Nearest Neighbors Method for Classifying User Sessions in E-commerce Scenario. *Journal of Telecommunications and Information Technology* (2015).
[31] Euiho Suh, Seungjae Lim, Hyunseok Hwang, and Suyeon Kim. 2004. A Prediction Model for the Purchase Probability of Anonymous Customers to Support Real Time Web Marketing. *Expert Systems with Applications* 27, 2 245–255.
[32] William R Swinyard and Scott M Smith. 2004. Activities, Interests, and Opinions of Online Shoppers and Non-shoppers. *IBER* 3, 4 (2004).
[33] Zhiqiang Tao, Sheng Li, Zhaowen Wang, Chen Fang, Longqi Yang, Handong Zhao, and Yun Fu. 2019. Log2Intent: Towards Interpretable User Modeling via Recurrent Semantics Memory Unit. *KDD*. 1055–1063.
[34] Manos Tsagkias, Tracy Holloway King, Surya Kallumadi, Vanessa Murdock, and Maarten de Rijke. 2020. Challenges and Research Opportunities in eCommerce Search and Recommendations. *SIGIR Forum* 54, 1 (June 2020).
[35] Yu-Ting Wen, Pei-Wen Yeh, Tzu-Hao Tsai, Wen-Chih Peng, and Hong-Han Shuai. 2018. Customer Purchase Behavior Prediction from Payment Datasets. In *WSDM*. 628–636.
[36] Eun Young Kim and Youn-Kyung Kim. 2004. Predicting Online Purchase Intentions for Clothing Products. *EJM* 38, 7 (2004), 883–897.