



# Towards early purchase intention prediction in online session based retailing systems

Ramazan Esmeli<sup>1</sup> · Mohamed Bader-El-Den<sup>1</sup> · Hassana Abdullahi<sup>2</sup>

Received: 11 May 2020 / Accepted: 12 November 2020 / Published online: 19 December 2020  
© The Author(s) 2020

## Abstract

Purchase prediction has an important role for decision-makers in e-commerce to improve consumer experience, provide personalised recommendations and increase revenue. Many works investigated purchase prediction for session logs by analysing users' behaviour to predict purchase intention after a session has ended. In most cases, e-shoppers prefer to be anonymous while browsing the websites and after a session has ended, identifying users and offering discounts can be challenging. Therefore, after a session ends, predicting purchase intention may not be useful for the e-commerce strategists. In this work, we propose and develop an early purchase prediction framework using advanced machine learning models to investigate how early purchase intention in an ongoing session can be predicted. Since users could be anonymous, this could help to give real-time offers and discounts before the session ends. We use dynamically created session features after each interaction in a session, and propose a utility scoring method to evaluate how early machine learning models can predict the probability of purchase intention. The proposed framework is validated with a real-world dataset. Computational experiments show machine learning models can identify purchase intention early with good performance in terms of Area Under Curve (AUC) score which shows success rate of machine learning models on early purchase prediction.

**Keywords** Early purchase prediction · Session logs · Purchase prediction · Real-time offers · E-commerce · User behaviour analysing

**JEL Classification** M370

## Introduction

The market share of online/e-commerce sales has been rapidly increasing during the last three decades. In 2018,

for the first time, the amount of total online sales had exceeded in-store sales in the USA (Mu et al. 2020). Moreover, Google and Facebook generated 116.3 and 55.8 billion US dollars, respectively, from online advertising only (Corrigan et al. 2018). Unlike in-store sales, digital marketing, and online sales generate big and valuable data about products, consumers' intention and behaviour (where consumers are coming from, what devices are they using, what items do they buy or view and for how long, how do shoppers respond to digital marketing ads and emails, and much more.) which was not available for business before (Leeflang et al. 2014).

The e-commerce industry is moving rapidly towards targeted personalised adverts. Recent studies show that offering all potential consumers generic items recommendation has proven to be an ineffective strategy (Behera et al. 2020; de Pechpeyrou 2009; Stewart-Knox et al. 2016). One of the main issues in online sales is consumer conversion; the amount of online sessions (i.e. user visit to a website)

---

Responsible Editor: James E. Richard

✉ Ramazan Esmeli  
ramazan.esmeli@port.ac.uk

Mohamed Bader-El-Den  
mohamed.bader@port.ac.uk

Hassana Abdullahi  
hassana.abdullahi@port.ac.uk

<sup>1</sup> School of Computing, University of Portsmouth,  
Lion Terrace, Portsmouth PO1 3HE, UK

<sup>2</sup> School of Mathematics and Physics, University of Portsmouth,  
Lion Terrace, Portsmouth PO1 3HE, UK

that end with a purchase is negligible when compared to the total number of session/visits (Liu et al. 2019; Zhou et al. 2019; Behera et al. 2020). A substantial number of these abandoned sessions are due to lack of purchase intention from the consumers, which means that there is almost no chance for conversion; therefore, rendering marketing strategies ineffective. Conversely, a considerable number of abandoned sessions come from users exhibiting strong purchase intention. The lack of purchase, in this case, could be due to reasons such as lack of offers or inability to correctly interpret user behaviours. For consumers with strong purchase intention, personalised marketing strategies such as targeted discounts, personalised recommendation, targeted adverts and follow-up emails could be very effective. Moreover, in addition to the possibility of increasing conversion rate, correctly identifying and targeting consumers with strong purchase intention could lead to an increase in sales. There has been growing interest in developing different purchase prediction models. Not just online but also in physical environments (standard stores). Kim et al. (2020) developed a framework for real-time purchase behaviour prediction from the users' (shopper) physical movement in a store environment. They used camera sensors and object detection algorithms to recognise purchase action. However, deploying such systems is very expensive and challenging. Contrarily, online purchase prediction models are effective, easily deployed and integrated with the system (Mokryn et al. 2019).

Studies have proposed methods for purchase prediction in the literature in the last few years (Rust et al. 2011; Esmailian and Jalili 2015; Lo et al. 2016; Brodén et al. 2018; Mokryn et al. 2019; Martínez et al. 2020; Esmeli et al. 2020). However, most of these methods are offline and try to predict purchase from completed sessions (after the shopper has left the website) in order to define a follow-up action. This makes such methods ineffective for early purchase prediction while a session is on-going.

To the best of our knowledge, there is a lack of studies in the literature that propose methods for early purchase prediction. This lack could limit the application of personalised marketing strategies which in turn, could result in the loss of potential sales from users with high purchase intention. In this work, we close this gap by developing a novel framework for early purchase intention prediction (EPP). We design an EPP framework that could enable content personalising and the provision of real-time offers. The proposed EPP framework aids in the prediction of purchase intentions early by analysing characteristics of consumers' online shopping behaviours in an e-commerce website and extracting hidden features real-time. These features could facilitate the provision of smart personalised marketing strategies that could boost sales,

improve consumer experience and retention. The work is motivated by the following questions.

1. Given session data after a user's first interaction, how helpful can machine learning (ML) models be in predicting the likelihood of a purchase in an ongoing session.
2. What is the most critical session feature for early purchase prediction and how can it be identified?
3. How can ML models be evaluated to measure their performance on early purchase prediction?

In order to understand and answers these questions, we propose and develop an Early Purchase Prediction framework (EPP); we define 'early' in this framework as substantial purchase indication in a session between first item interaction, and last purchase action. We develop an early purchase prediction utility score where we aim to evaluate the performance of the ML models on early purchase prediction. This work can be seen as an extension of Mokryn et al. (2019) and Martínez et al. (2020), where the effect of temporal features and product trendiness on purchase prediction after sessions end is analysed, and Lo et al. (2016), which investigated how a registered user's actions can build up purchase intention in the long term. However, our work mainly focuses on real-time user purchase intention detection.

Our main contributions are summarised as follows:

1. We design a framework to predict users' purchase intention early that could enable content personalising and the provision of real-time offers and improve consumer retention.
2. We propose a utility scoring method in order to measure how ML models can predict purchase intention before the purchase happens.
3. The developed framework proposes a method to extract dynamic features from session logs.
4. A set of computational experiments that compares the ML models' performance on early purchase prediction is presented.
5. A detailed evaluation of the proposed EPP framework on a real-world e-commerce dataset is presented.

This paper is organised as follows. Section "[Related works](#)" provides an overview of previous works in purchase prediction. Section "[Dataset description](#)" describes the dataset used in this work. Section "[Early Purchase Prediction \(EPP\) framework](#)" introduces the proposed EPP framework. Section "[Experiments and results](#)" presents the results of the experiments. Section "[Discussion](#)" discusses the results, theoretical and practical implications of this research. Finally, Section "[Conclusion and future work](#)" provides a conclusion of this work. In addition, we provide key terminologies used in this work in [Appendix](#).

## Related works

This section gives a description of session logs, related works done in purchase prediction and an overview of ML models used in purchase prediction.

### Session logs

A session is described as a certain time duration that the user has been browsing on the website. The session interval time depends on company's policy. Session logs have been categorised as web usage logs (Zhuang et al. 2005) which could be utilised for analysing user behaviour for purchase prediction or product recommendations. Initially, these logs should be pre-processed to meaningful structured data such as session identification (Shu-Yue et al. 2011). Generally, session logs for an e-commerce data contain details of browsed products ID, products added to the cart or products purchased and the timestamp (Zeng et al. 2019). Since purchase prediction depends on extracted features from session logs, capturing relevant attributes from session logs is important in order to improve the accuracy level of the prediction models.

### Purchase prediction

Purchase prediction has been studied in several works in the literature. Recent studies have developed frameworks to investigate purchase prediction using users' previous session features and users' physical movement in a market environment in real-time (Martínez et al. 2020; Kim et al. 2020). Experiment results showed embedding these user behaviour/session features into ML models improves the performance of ML models in predicting users' purchase intention. In a physical environment, Zeng et al. (2019), proposed a purchase prediction model to analyse user behaviour during a festival in China. It was found that if a product is interesting to a user, the user was more likely to spend more time on it. In a virtual environment, Wu et al. (2015) proposed purchase behaviour prediction model focused on identifying click patterns rather than session features. Experiment results showed that using learned features from click patterns can improve purchase prediction as good as a conventional classification model trained using session features. When users are anonymous, users' web-logs can be extracted and used for purchase prediction (Suh et al. 2004; den Poel and Buckinx 2005). On the contrary, when users are registered, the performance of purchase prediction models can be improved by using extracted session features built up over time (Lo et al. 2016). In Lo et al. (2016), purchase prediction models for registered Pinterest users were proposed to analyse users' long and short term behaviours. The authors tested the

performance of prediction models on extracted features at different times before purchase action happened, and found that purchase intentions were built up by time. Results of the experiments indicated that prediction models produced better accuracy results when features are created based on the whole timeline until right before purchase action.

Some of the state-of-the-art ML models have been used for purchase prediction include Decision Tree (DT), Neural Networks (NN), Recurrent Neural Network (RNN) and logistic regression. Mokryn et al. (2019) and Li et al. (2016) investigated users' purchase prediction using various ML models, and evaluated the performance of these models. Li et al. (2016) proposed three ML models (Bagging, DT and Random Forest) and prediction results of the models were combined using a linear regression method. Their proposed method gained % 8 accuracy (Recall) on purchase prediction. In Mokryn et al. (2019), logistic regression, Bagging, DT algorithms were used to investigate the effect of temporal features (time, product trendiness, etc.) on purchase prediction performance, where it was found that Bagging performed best when temporal features were applied. Suh et al. (2004) created model attributes using Association Rules (AR) and applied a combination of different ML models; DT, NN, and Logistic Regression, and found that ML models performed better when these ML models are combined. When ML models are trained using anonymous sessions with extracted session features for purchase prediction for anonymous users, ML models performed well despite user anonymity (Yagci et al. 2015; Romov and Sokolov 2015; Esmailian and Jalili 2015; Pálovics et al. 2015).

### Limitations of previous approaches and proposed contributions

As discussed above and shown in Table 1, most of the current methods for purchase prediction are offline and try to predict purchase from completed sessions (i.e. after the user has left the website) in order to define a follow-up action (e.g. send a follow-up email). Moreover, existing methods could be ineffective in early purchase prediction while the shopper is still navigating the e-commerce website. Admittedly, there are methods for purchase prediction for registered users at the end of a session, however, there is a lack of early purchase prediction methods for anonymous users or registered users during an on-going session. This limits the use of personalised marketing strategies in real-time. Such marketing and business improvement strategies can boost sales, improve consumer experience and retain consumers if users are anonymous (no contact information) and cannot be contacted again after they leave the e-commerce website (Mokryn et al. 2019).

**Table 1** Recent works about purchase prediction in literature

Authors/years	Key findings	Anonymous	Registered	After session ends	Real-time
Zeng et al. (2019)	User behaviours were modelled and built firstly purchase prediction model then recommender system model to find out which product could be purchased.		X	X	
Wu et al. (2015)	Analysed purchase prediction from click patterns using a Neural Network Model.	X		X	
Li et al. (2016)	Designed a Linear Regression model to predict purchase action in a Chinese E-commerce website.		X	X	
Baumann et al. (2018)	User Navigational data was used for Graph modelling to build purchase prediction model.		X	X	
Lo et al. (2016)	User behaviours were modelled to investigate how the purchase intention is built up after in each website visit of the registered users.		X	X	
Mokryn et al. (2019)	Compared purchase prediction models when temporal attributes are considered.	X		X	
Park and Park (2016)	Analysed purchase conversion factors using a Bayesian probabilistic model.		X	X	
Martínez et al. (2020)	Designed a framework for predicting purchase behaviours in near future by analysing their behaviours.		X	X	
Kim et al. (2020)	Analysed purchase action on products in real time using only camera sensors in an unmanned product cabinet by object detection.	X			X

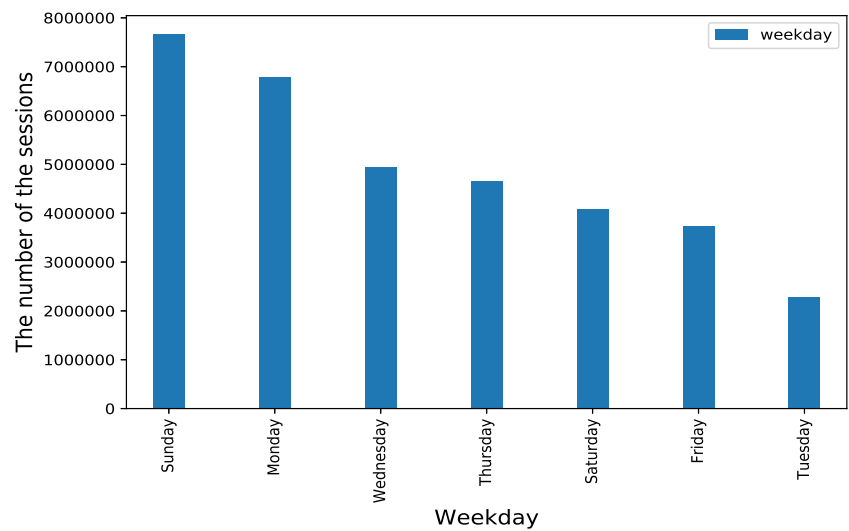
In this work, we present a novel approach for early purchase prediction in an active session in an e-commerce website in order to improve the provision of real-time discounts on products. This can be effective in convincing shoppers to purchase products. Also, we develop and implement an EPP framework to evaluate the performance of ML models for early purchase prediction.

## Dataset description

The dataset used in this work consists of 6 months of session logs (Ben-Shimon et al. 2015). The session logs are collected from a European e-commerce business (YooChoose<sup>1</sup>). YooChoose is a Germany-based company that offers a Software as a Service (SaaS) solution to help

online shops generate personalised shopping experience for their consumers via personalised product recommendation, search results and newsletters. Product categories in the dataset are not limited to clothes only, but also includes toys, electronics and garden tools. The dataset contains two files; one for click data that contains the click events on an item associated with the session id, item id, the category of the product and the timestamp (the time when the click occurred). The other file called purchase event logs contains purchase events from sessions that appear in the click events log and end with a purchase. Each purchase event is associated with a session id, item id, price and quantity of purchase. Sessions are diverse with their length (the number of browsed products) and timestamp (the time that session has started). Sessions last from a few minutes to a few hours, and the number of clicked products varies from one to hundreds depending on user activity. This dataset has been used widely for consumer intention prediction and

<sup>1</sup><https://www.yoochoose.com/>

**Fig. 1** Interaction frequency and weekday correlation

user behaviour modelling (Yeo et al. 2017; Wu et al. 2015; Brodén et al. 2018; Bogina and Kuflik 2017; Mokryn et al. 2019).

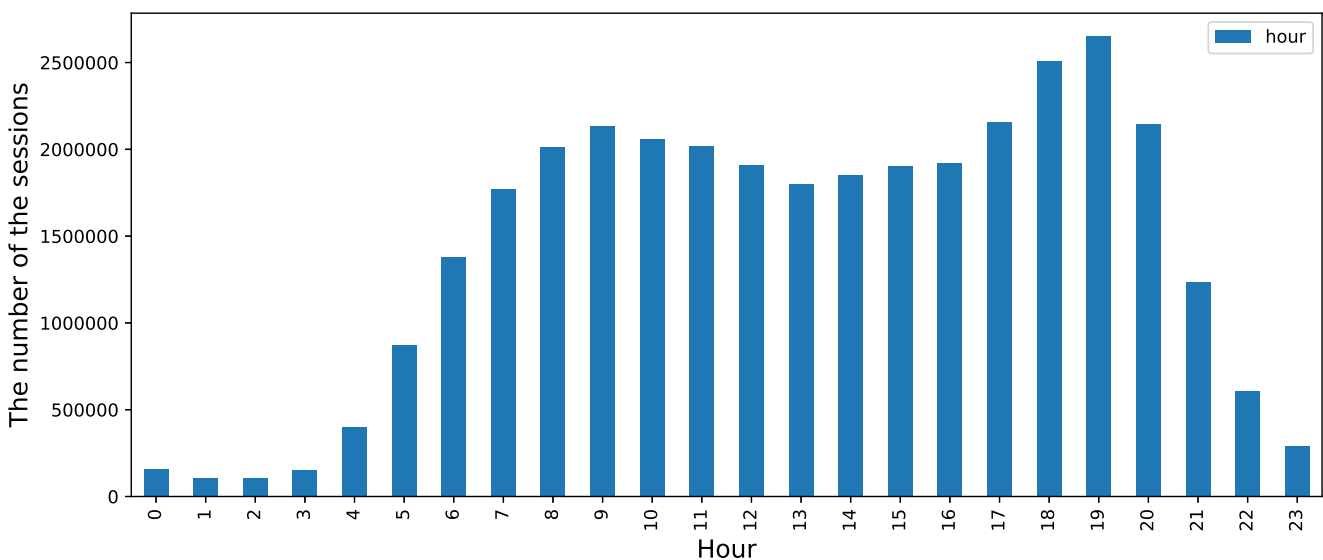
In the dataset, there are 52739 unique items, 9249729 sessions and 26637000 interactions. A majority of the sessions are only viewed session without any purchase, which means there is a massive class imbalance problem in the dataset. Users may have different habits; for example, they may prefer shopping on specific days of the week. We use the day of the week in the attributes. Consequently, we examine which days of the week do users mostly shop on the website. As seen in Fig. 1 users are more likely active on Sunday and Monday.

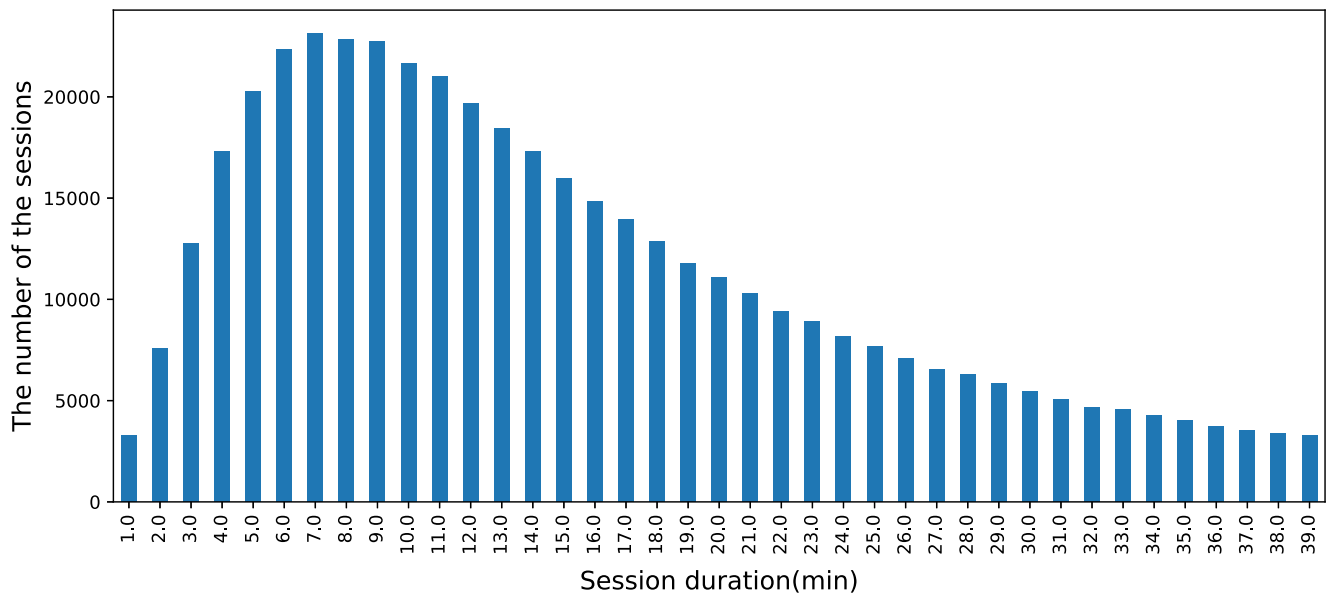
We analyse the dataset in terms of the hours users mostly visit the website. Figure 2, shows the distribution of the number of the user interactions of the website by hours.

Session duration is another indicator of purchase action. The longer a user spends browsing items in a session, the higher the probability of a purchase in that session. Users' spends more time on the e-commerce website, and there is a higher probability that the session will end with a purchase. We analyse the difference between the session duration distribution for purchase and non-purchase sessions. As shown in Fig. 3, the majority of purchase sessions last between 6 and 12 minutes while majority of the non-purchase sessions last less than 6 minutes (Fig. 4).

### Test dataset analysis

In this section, test dataset used to validate our framework is analysed. For the purpose of this work, we selected 44999 sessions randomly. Out of the selected sessions, Table 2

**Fig. 2** Session distribution by hour



**Fig. 3** Session distribution by session duration for purchase sessions

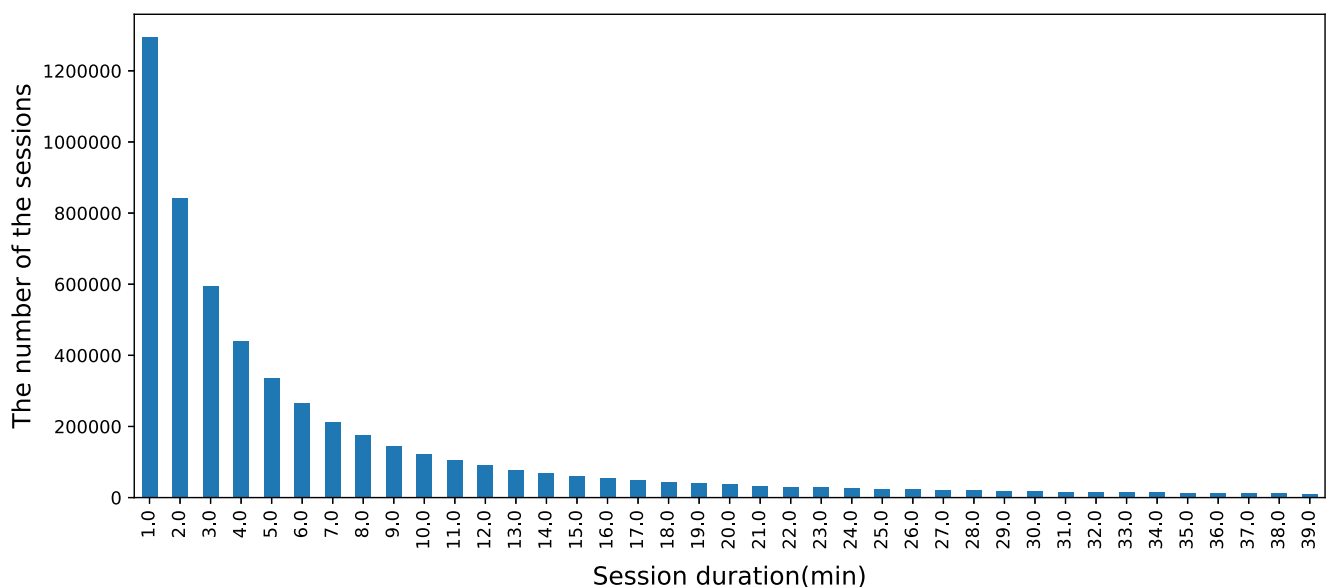
shows that only 5414 sessions end with purchase indicating that the test dataset is heavily imbalanced.

Two critical factors on test dataset that effect utility score is the first purchase location and last purchase location (if there is more than one purchase in a session) as will be explained later in the scoring function. Figure 5 shows that first purchase location is mainly after four viewed item, which means that shoppers' purchase actions focus between 2 and 5 product browsing. The last purchase location mainly concentrated around between five and ten viewed items. The line inside the box plot shows the mean of shoppers' first and last purchase locations.

### Early Purchase Prediction (EPP) framework

The EPP framework (Fig. 6) is designed in order to determine whether ML models can predict a purchase before purchase action happens in an ongoing session. Also, the EPP framework helps to evaluate the performance of ML models on early purchase prediction using the designed scoring function.

The proposed framework consists of three phases. In the first phase, the session logs from the e-commerce website are collected. Session logs store the data about which products are browsed in each session and start and end timestamps of

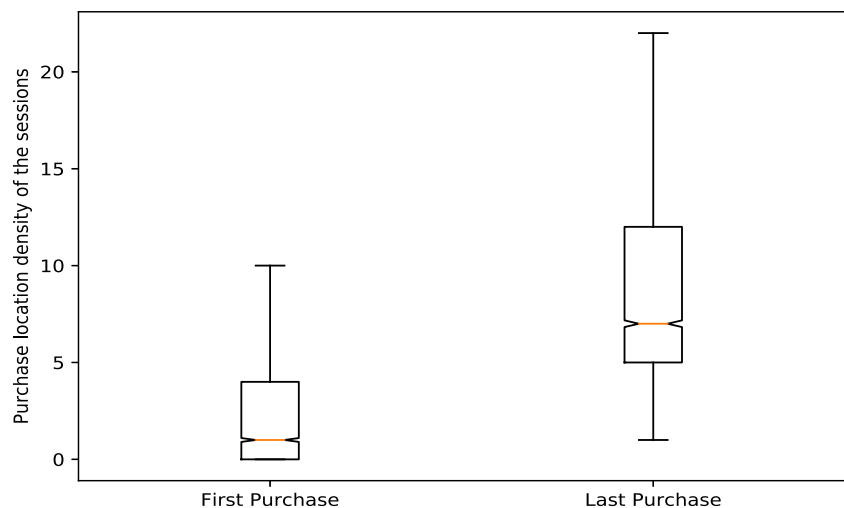


**Fig. 4** Session distribution by session duration for non-purchase sessions

**Table 2** Test session details

Session type	# of Sessions
Non-purchase	39585
Purchase	5414
Only one purchase	1401
More than one purchase	4013

the session. The collected session logs are unstructured for example, in some records, timestamp or session id can be missed. Therefore, to have a proper dataset to feed the ML models, these records are filtered out in the pre-processing stage of the EPP framework. In order to evaluate the ML models, their performance needs to be tested. In order to test ML models, the dataset is split as train and test datasets. Train dataset is used to train the ML models, while the test dataset is used to evaluate the models. The second phase solves the class imbalance problem, attribute selection, and features generation from selected attributes. Class imbalance can be defined as having more number of non-purchase sessions than purchase ended sessions. If models are trained with the imbalanced data, they will produce a bias on purchase prediction for the ongoing session; for instance, all the predictions will be non-purchase. To prevent this, we apply the class imbalance method as explained in Section “[Class imbalance](#)”. In order to build ML models, features about sessions need to be extracted. Choosing the right and relevant features is a critical step in achieving well-performing ML models. If the well-correlated features are selected, ML models perform better on predicting purchase action for an ongoing session. The details of selected features are given in Section “[Attribute selection and feature generation](#)”. In the last phase, trained ML models are evaluated using features created dynamically in a test session after each viewed product.

**Fig. 5** Distribution of first purchase and last purchase positions of sessions in test dataset

Since proposed EPP framework will provide real-time purchase prediction, the features for the sessions in the test dataset are generated dynamically, such as after each product viewing in the session (Fig. 7). The proposed scoring function detects if ML models can predict purchase action before actual purchase happens.

### Attribute selection and feature generation

In this work, we choose the **Timestamp** (see [Appendix](#)) attribute from session logs since this attribute will help to create other features such as session duration. The following features will be used to train the ML models. These features are generated from training dataset statistics based on users behaviour and timestamp.

1. **Total Viewed Items:** This feature shows how many items are viewed in the session and the length of the session. Clearly, an item can be clicked many times in a session and this provide an indication about users' intention on the item.
2. **Total Unique Items:** We take into account the unique seen items in the session to show how many of the clicked products are different products. If a user prefers to view the same item multiple times, it could be an indication of the user's interest in the item.
3. **Total Session Duration:** Shows the duration of a session. It is found that there is a positive correlation between dwell time and users interest on the items in a session (Bogina and Kuflik 2017).
4. **Click Rate:** Defines how many products are clicked within the duration of a session. This could show the user intention in the session. For example, if the user clicked many items within a short period in the duration of a session, this could be interpreted as the user having browsing intentions rather than buying an item in that session.

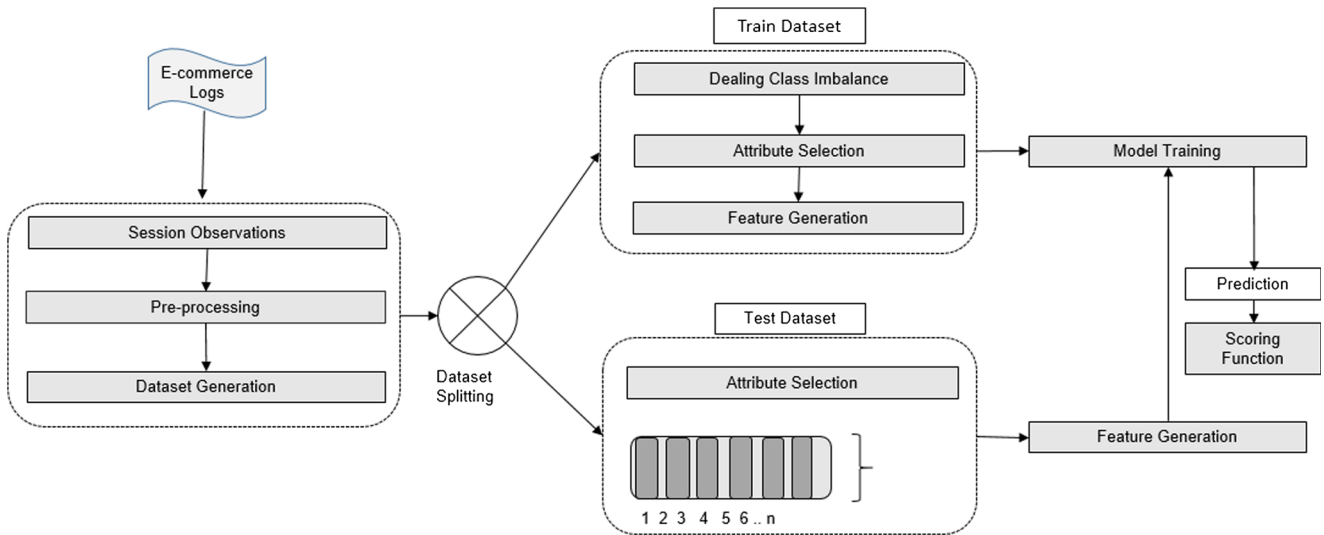


Fig. 6 Proposed early purchase prediction framework

5. **Max Popularity:** Shows highest popularity rate of a viewed item in a session. This feature shows the item trendiness in a session by looking at how popular items are browsed since popular products have higher probability of being purchased.
6. **Min Popularity:** Shows lowest popularity rate of the viewed item in a session. If the interacted products in a session have low popularity, this could give an indication of the users' purchase intention in the session.
7. **Duration Spent on a Product:** This shows the amount of time a user spends on a product in a session (minutes). It is found that the longer the time a user spends on a product, the higher the level of interest the user has on that product (Bogina and Kuflik 2017).
8. **Unique Categories:** Defines how many different categories have been browsed in the session. The number of unique categories is an important feature since it will indicate how the consumer is changing his/her category intention over the session.

We used temporal features to explore their effect on the performance of the ML models. The list of the created temporal features are described below.

1. **Hour:** This attribute shows the hour the session has starts. It is seen from the dataset analyses that users

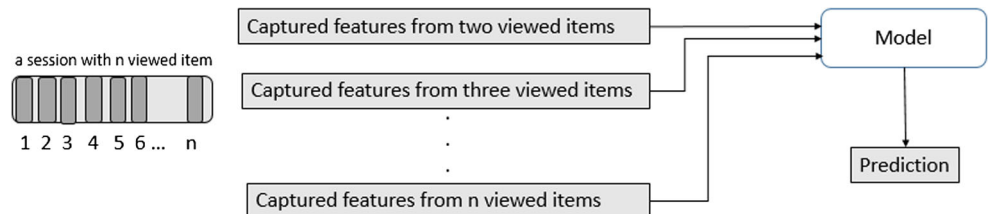
are more likely to proceed to check out after 5:00 pm (Fig. 2).

2. **Day of the week:** Since each day of the week has a different number of browse and purchases, we add to features which day of the week a session started (Fig. 1).
3. **Weekend:** Indicates whether the day of the week that the session started falls on the weekend or not. This feature can give an indication of a user's purchase intention. We found from our test dataset that Sunday is the busiest day with the highest number of visit to the e-commerce website (Fig. 1).
4. **Day of the Year:** Indicates the numerical day of the year that the session started. This attribute has an important role since user behaviours might change in seasons and special days such as Black Friday or festive periods such as Christmas.

### Model training

To test the proposed EPP framework, we train five ML models based on extracted features from the sessions. The ML models are Decision Tree (DT), Random Forests (RF), Bagging, K-Nearest Neighbour (KNN) and Naive Bayes (NB). DT (Berry and Linoff 2004) is a non-parametric supervised learning method used for classification and regression. The goal of DT is to create a model that predicts

Fig. 7 Dynamic feature creation process





the value of a target variable by learning simple decision rules inferred from the data features. RF (Breiman 2001) is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Bagging (Louppe and Geurts 2012) is an ensemble meta-estimator that trains base classifiers each on random subsets of the original dataset and then aggregates their individual predictions by voting or by averaging to form a final prediction. KNN classifier (Qian and Rasheed 2007) uses a simple majority vote of the nearest neighbours of each point for classification. A query point is assigned the data class which has the most representatives within the nearest neighbours of the point. Lastly, NB (Wang and Tseng 2015) is a classifier that can be extremely fast compared to other classifier methods and requires fewer data to have a well-performing model. All these models have their specific learning methods. We conduct experiments for each model to identify which model is performing better. We use the scikit-learn (Pedregosa et al. 2011) python ML library for the experiments.

### Dynamic feature building for the test dataset and purchase prediction

Feature extraction for the test dataset is done following a similar way with feature creation for training dataset. However, the features of the test sessions have been built after each viewed item (Fig. 7) since our purpose is to predict purchase intention in an on-going session. For each updated feature after each viewed product, we get a prediction if user will buy an item in the session or not in real-time.

After each prediction EPP utility score is calculated for the model. EPP utility score depends on the item position and session specification. For instance, if there is a purchase, the EPP utility score is calculated using the position detail of first purchased and last purchased items.

### EPP scoring function

EPP scoring system is an important component of the proposed EPP framework since we can evaluate ML models performance on how early they can determine purchase in the session based on the calculated utility score distribution. The frequency of higher utility score distribution for a model shows that the model is performing well in early purchase prediction. For example, after browsing the first item, based on the captured session features, the model can predict whether the session will end with a transaction or not. In this case, the proposed scoring system checks if there is actually purchase in the session. If there is a purchase, the scoring system will measure how early the model gives purchase prediction correctly by looking at the distance between the position of interacted item and position of purchase action.

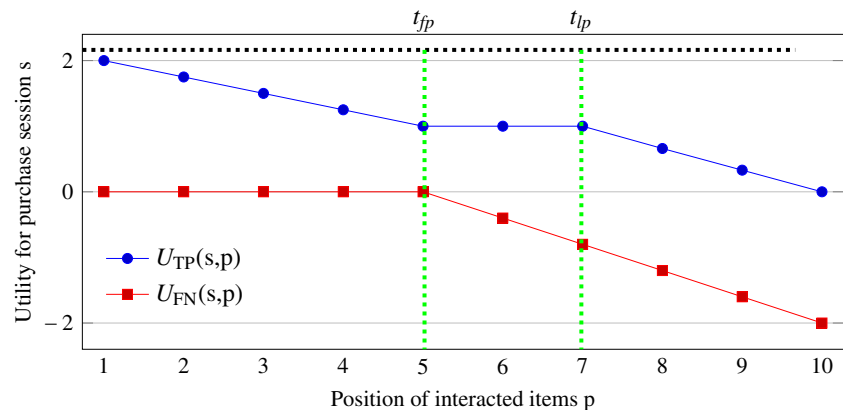
We define the proposed EPP utility scoring function as follows. Given a set of sessions  $S$  with session  $s \in S$ ,  $lp$  is the position of last purchased product in session  $s$ ,  $t$  is the number of the total items browsed in  $s$ ,  $p$  is the position of the last observed item  $i$ , and  $predictions$  use the item features to generate and hold a binary array of the prediction  $prediction$  results after each observed item  $i$  in  $s$ . If there is a purchase prediction in the  $predictions$  array,  $prediction$  is 1 otherwise,  $prediction$  is 0 (1).

$$prediction = \begin{cases} 1, & \text{if } purchaseprediction \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

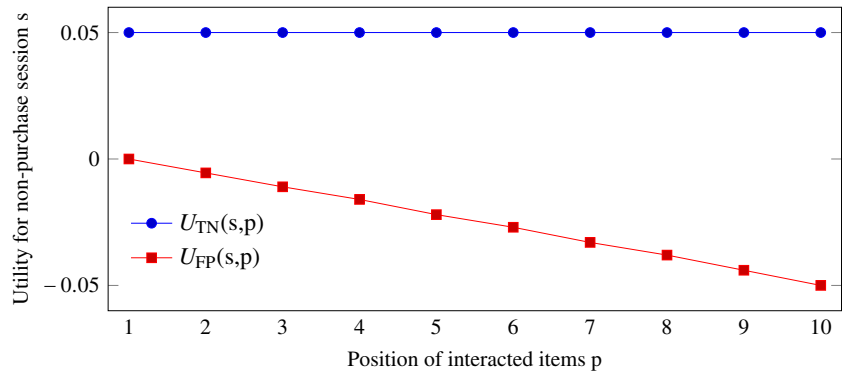
The utility scores are calculated based on the prediction location for the observed interacted items so far in the session. The classification prediction can be positive or negative (Figs. 8, 9). This means that if prediction is 1 we classify it as positive, if prediction is 0 we classify it as negative prediction.

The details of the different probable situations in the purchase prediction process is seen in Table 3. In this table,

**Fig. 8** Utility function for purchase session ( $p = 1$  means 2<sup>nd</sup> viewed item in the session)



**Fig. 9** Utility function for non-purchase session ( $p = 1$  means 2<sup>nd</sup> viewed item in the session)



$U(s,i)$  is the utility score function that gives a utility score for the observed features after each viewed item  $i$  in the session  $s$ .

Base on Table 3, there are 4 cases that need to be considered to calculate the EPP utility score;

- 1) In the case of  $U_{TP}(s,i)$ , Eq. 2 is used to calculate the utility score  $sc_1$  for item  $i$  in position  $p$  in the session  $s$ .

$$sc_1(p, fp, lp, t) = \begin{cases} 2, & \text{if } p == 1 \\ 1 + \frac{fp-p}{fp}, & \text{if } p > 1 \text{ and } p < fp \\ 1, & \text{if } p \geq fp \text{ and } p \leq lp \\ \frac{t-p}{t-lp}, & \text{if } p \geq lp \text{ and } p \leq t \text{ and } t \neq lp \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

- 2) In the case of  $U_{FN}(s,i)$ , we will give negative utility score based on the position  $p$  of last observed item  $i$  (3). In the utility score calculation ( $sc_2$ ), we do not apply negative values until  $p$  reaches the first purchase point ( $fp$ ) of the session  $s$ .

$$sc_2(p, fp, lp, t) = \begin{cases} 0, & \text{if } p \geq 1 \text{ and } p < fp \\ -1, & \text{if } p \geq fp \text{ and } p \leq lp \\ -1 - \frac{lp-p}{lp-t}, & \text{if } p \geq lp \text{ and } p \leq t \text{ and } t \neq lp \\ -2, & \text{otherwise} \end{cases} \tag{3}$$

- 3) If there is no purchase in the session but prediction model gives positive prediction ( $U_{FP}(s,i)$ ), we calculate

the utility score ( $sc_3$ ) as in Eq. 4. In this equation,  $-0.05$  indicates False Positive (FP) coefficient.

$$sc_3(p) = -0.05 \times \left(\frac{p}{t}\right) \tag{4}$$

- 4) If there is no purchase in the session, also the prediction model gives negative prediction ( $U_{TN}(s,i)$ ), the utility score is calculated in Eq. 5

$$sc_4(p) = 0.05 \tag{5}$$

We follow two methods to calculate overall utility score of the models' early prediction performance (Figs. 8 and 9):

1. If the model has a positive prediction for the sessions and ends up with a purchase, we will consider the utility score in the earliest position. However, if the model has a negative prediction for the session, but it ends up with a purchase, we will consider the utility score in the last position. For example, lets say  $fp = 5$ , purchase = 1, prediction = 1,  $lp = 5$ ,  $t = 10$ , and we have positive prediction in position  $p = 2$ . In this case, we will consider the calculated utility score in position  $p = 2$  neither negative utility scores calculated in previous positions nor positive or negative utility scores after position  $p = 2$ . Note that in this method, there must be at least one positive prediction.
2. If the model has negative predictions for all observed items in the session, this means that there is no positive prediction in this model; however, the session has a purchase, we will assign the lowest utility score to that model for that session.

Higher utility score indicates that the model is better at early purchase prediction. The details of the utility score function are shown in Algorithm 1.

**Table 3** Description of Confusion matrix

Value	Description
$U_{TP}(s,i)$	Positive prediction of item $i$ in purchase session $s$ .
$U_{FN}(s,i)$	Negative prediction of item $i$ in purchase session $s$ .
$U_{FP}(s,i)$	Positive prediction of item $i$ in non-purchase session $s$ .
$U_{TN}(s,i)$	negative prediction of item $i$ in non-purchase session $s$ .

**Algorithm 1** The Algorithm of Early Purchase Prediction Framework (EPPF).

---

```

T {consist of sessions in test dataset}
Model {classification model}
s {a session in T}
transaction = 0    ▷ holds if there is transaction in the
session
prediction = 0 {holds the model's purchase prediction
value}
predictions = [] {holds prediction values of the
classification model based on observed items.}
for each s in T do
  if hastransaction(s) then
    transaction = 1
    fp {holds positon of first transaction action}
    lp {holds position of last transaction action}
    totalItem {holds number of total browsed items}
  end if
  for each position in range(len(s)) do
    features = featureextract(s.head(position+
1))
    predictions[position] = model(features)
  end for
  if haspredicted(predictions) then
    prediction = 1
  end if
  for position in range(len(s)) do
    if transaction and prediction then
      if predictions[position] then
        sc(position) = sc1(position, fp, lp, t)
      end if
      if not predictions[position] then
        sc(position) = sc2(position, fp, lp, t)
      end if
    end if
    if not transaction then
      if predictions[position] then
        sc(position) = sc3(position, fp, lp, t)
      end if
      if not predictions[position] then
        sc(position) = sc4(position, fp, lp, t)
      end if
    end if
  end for
end for

```

---

## Experiments and results

This section shows the performance of the ML models on early purchase prediction on the dataset. As we discussed in the EPP scoring function in Section “[Dataset description](#)”, utility scores above 1 indicate that there is an early purchase

prediction for purchase ended session (Fig. 8). We conduct experiments on different ML models to see which model is better on early purchase predicting. The aim of this section is as follows.

1. Analyse the early purchase prediction performance of ML models in purchase ended sessions.
2. Analyse how accurate ML models can predict when sessions do not have a purchase.
3. Compare the effect of the two methods proposed to fix class imbalance problem in the dataset on early purchase prediction score.

### Class imbalance

Class imbalance (Berry and Linoff 2004) is a major problem in purchase prediction from session logs since the majority of the sessions end without any transaction. Techniques deal with imbalanced data including (i) oversampling, which is oversampling of the minority class, (ii) undersampling (Kubat and Matwin 1997) which is undersampling majority class and (iii) mixed methods (Batista et al. 2004) in which combination of undersampling and oversampling are applied. For oversampling, Synthetic Minority oversampling Technique (SMOTE) (Chawla et al. 2002) and the Adaptive Synthetic (ADASYN) (He et al. 2008) sampling are well-known methods. In this study, we apply the SMOTE oversampling and random undersampling class imbalance techniques. For each case, we train the classification algorithms. We use a Python library (Lemaître et al. 2017) to implement methods to deal with class imbalance problem.

### Evaluation metrics

We use the proposed EPP utility score designed in Section “[EPP scoring function](#)” to measure the performance of classification models on early purchase prediction. In addition, we use the Area Under Curve (AUC) (Martínez et al. 2020) score to evaluate which model performs better with highest EPP utility score. AUC show measurement of how the model is capable of differentiating between classes (Huang and Ling 2005). The classes in our work are purchasing and not purchasing. When a model has a high AUC score, it means the model is good at differentiating the purchase and not purchase intention. Additionally, we consider the confusion matrix (Martínez et al. 2020) in order to analyse the number of the correctly classified and unclassified purchase intention predictions.

### Experimental setup and analysis

In order to analyse how the ML classification models can predict early whether there will be a purchase in the session

**Table 4** Performed test categories and their details

Experiment	Detail
Category A	8 feature without temporal features.
Category B	8 feature filtered based on the number of interacted items in the session.
Category C	12 feature including temporal features added.
Category D	12 feature filtered based on number of interacted items in the session.

or not, a set of experiments are carried out (Table 4). We tested ML models on different class imbalance situations (SMOTE, undersample and without any class imbalance method) to understand the effect of class imbalance techniques (Section “[Class imbalance](#)”) on early purchase predictions. Also, we added additional features to our dataset to see the effect of temporal features on early purchase prediction. Moreover, we tested classification models on the filtered dataset, in which we eliminated sessions which have less than ten interacted items in the training dataset. NB, KNN, RF, Bagging and DT models were used to conduct the experiments and compare their performances. For the KNN model, we set  $n$  neighbour to 64, in RF, the number of estimators is set to 50, and the base estimator is chosen as DT.

All classification models were run on each dataset category. For each category, their performance was compared to find the best performing category and model.

### Utility score and evaluation of early purchase prediction for classification models

We used the proposed utility score to evaluate models in terms of how they can identify the earliest purchase intention. We created a decision condition that helps to labelling early purchase predicted sessions based on the utility score. The details of the utility score can be seen in Section “[Dataset description](#)”. As seen in Fig. 8, when the utility score is greater than one; it means classification model predicts correctly that there would be purchase in this session. On the other hand, when the utility score is greater than 0.05 for a session, this means that the session is labelled correctly as purchase ended session. However, in the case when utility score is between 0 and 1, this means purchase action is predicted after the transaction happened. Moreover, if utility score equals to 0.05 which means a non-purchase session classified correctly as non-purchase session (Fig. 9). Function 6 shows the possible early prediction results according to given session utility score series  $s$ .  $s$  is created

as a result of utility score calculation after predictions derived from classification models for each session.

$$label(s) = \begin{cases} early\ purchase, & \text{if } maxutilityscore(s) > 1 \\ correct\ prediction, & \text{if } 0 \leq maxutilityscore(s) \leq 1 \\ misclassified, & \text{otherwise} \end{cases} \quad (6)$$

In order to carry out the experiments, we pick the session features according to their utility score. For each interacted item, we create features, and we select the feature with the highest utility score to represent the session based on the utility score of the created feature. For example, let assume that a session ends with a purchase and prediction model gives the wrong prediction for the features derived from 2nd to the 4th interacted items in the session, and created session features after 5th interacted item leads to correct prediction. So, the utility score in the 5th location is picked as the early purchase prediction score for the session, and the session features created at the 5th viewed item are selected as the session representing features. On the other hand, if the model gives wrong purchase prediction for the session, we continue to check if we can get the right prediction until the session ends. If we do not have the correct prediction for sessions after the session ends, we use the features that are created after the last interacted item as representative features of the session.

The experiments aim to see which model is performing better on early purchase determination. After we find the best-performing model, we analyse how early the model predicted the purchase action by analysing utility score distribution (such as the number of steps after the session started). Having higher utility score shows earlier purchase prediction for the model. Also, we analysed the machine learning models with AUC score and confusion matrices. Our aim for using these assessment measures is to find out if there is any positive correlation between these evaluation metrics (AUC and confusion matrices) and calculated EPP utility score. Table 5 shows the performance comparisons of the models. For category A dataset, Bagging achieved the highest AUC score followed by KNN classifier. In other categories, DT classification models outperformed other classification models. While, NB shows consistency in terms of performance in different class imbalance methods and categories, with the worst AUC score. Undersampling method for the class imbalance problem outperformed SMOTE method in most classification models.

It can be seen from the results that the most affected model from class imbalance problem is KNN classifier (the lowest AUC score is 0.5124 in the imbalanced dataset)

**Table 5** AUC score comparison of early purchase prediction(EPP) Framework on NB, RF, Bagging, DT and KNN classifier models on different categories and class imbalance methods

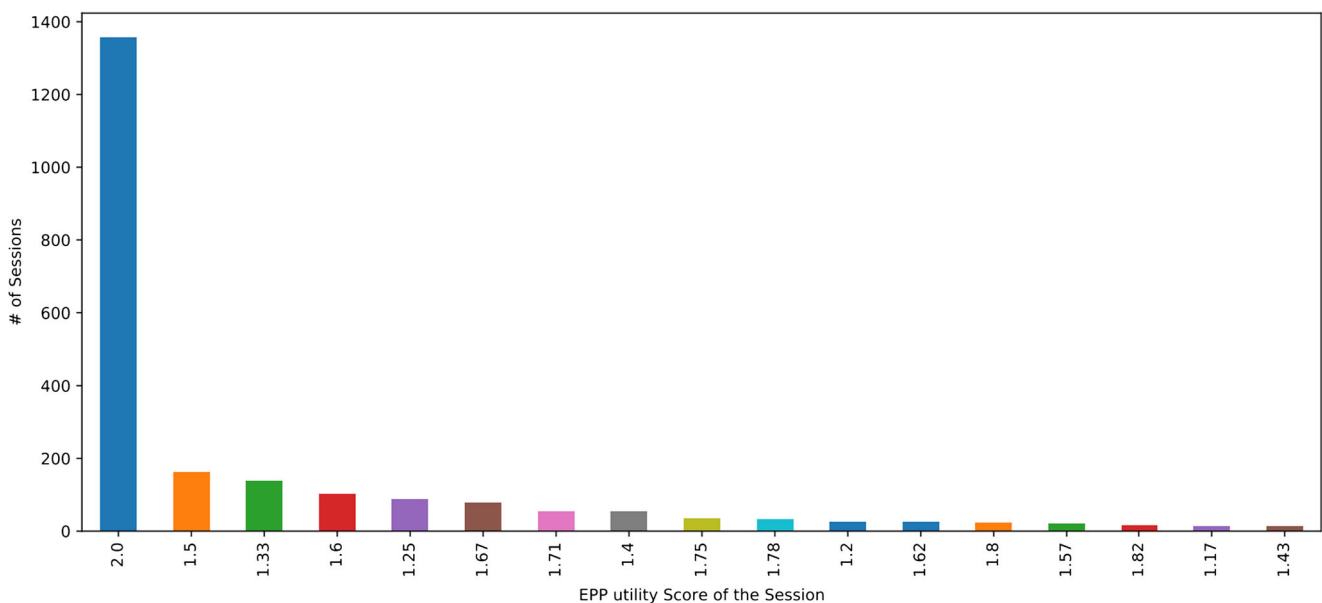
Categories	Experiments	NB	RF	Bagging	DT	KNN
Category A	Original	0.6626	0.5857	0.6401	0.8931	0.5345
	Undersampling	0.7518	0.9070	0.9448	0.9054	0.9363
	Smote	0.7524	0.6490	0.7328	0.9186	0.9121
Category B	Original	0.7874	0.5695	0.6477	0.9311	0.5389
	Undersampling	0.6987	0.9074	0.9176	0.9282	0.8166
	Smote	0.7464	0.6144	0.6794	0.9021	0.8188
Category C	Original	0.6460	0.5874	0.6597	0.9276	0.5124
	Undersampling	0.6391	0.8940	0.9153	0.9708	0.8541
	Smote	0.6322	0.7108	0.7783	0.9154	0.8602
Category D	Original	0.6638	0.7014	0.7491	0.9338	0.5389
	Undersampling	0.6400	0.8496	0.9075	0.9378	0.8404
	Smote	0.6454	0.7218	0.7837	0.8960	0.8350

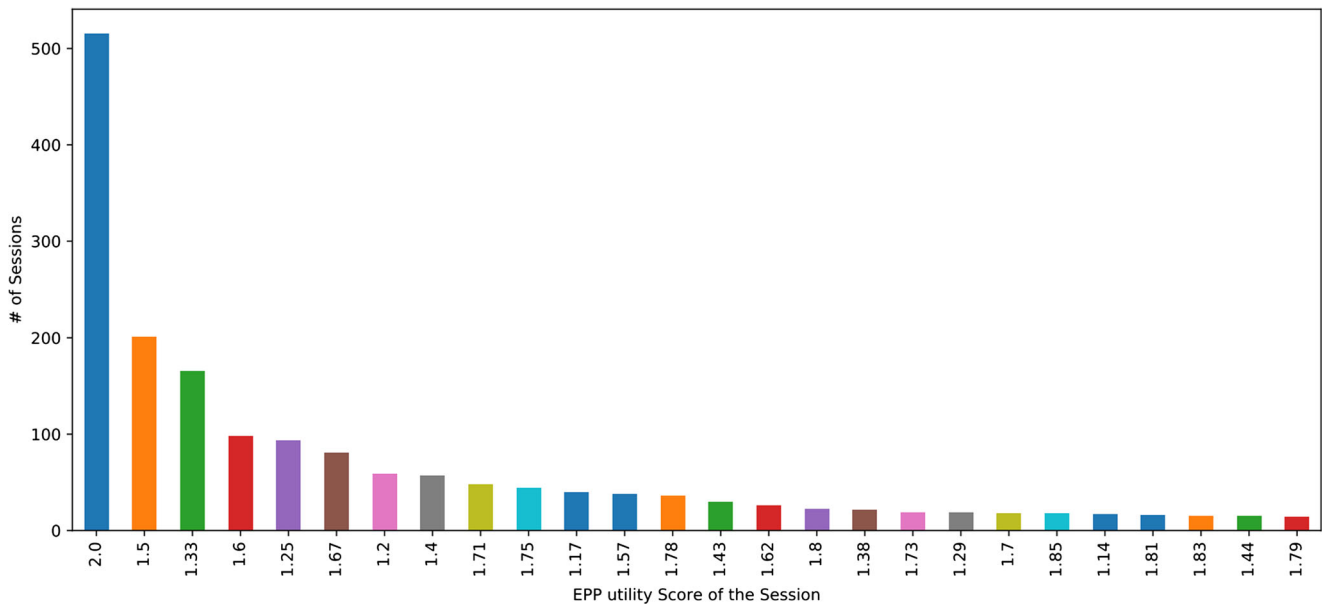
however when Undersampling is applied, the performance of KNN classifier AUC score improved to 93%. DT is the best robust classification model for imbalanced data case with the highest AUC score (the lowest AUC score is 0.89 %). In general, the highest AUC score for early purchase prediction is produced by DT classification model with 97 % of AUC score. Moreover, we analysed two highest performing models in terms of how early they are successful on purchase prediction. The first model is DT trained on Category C undersampled dataset, and the second dataset is Bagging trained on Category A undersampled dataset. In order to compare their differences, we look closer to distribution of EPP utility score of the sessions. As seen from the Figs. 10 and 11, DT has more number of 2.0 utility

score than Bagging which means DT is better than Bagging in terms of early purchase prediction after two viewed item.

We analysed the confusion matrices of three best-performed models trained on Category C undersampled dataset. It can be observed that (Table 6) DT classifier has superiority predicting purchase ended sessions comparing to Bagging classifier and KNN classifier.

The importance of data features during the ML model establishment process can be sorted and scored. We use DT classifier to conduct the feature importance analysing following the same method in Dutta et al. (2019) and Dou (2020) as we found DT classifier is the best performing model among all ML models we analysed for early purchase intention prediction. Based on feature importance analyses,

**Fig. 10** EPP score distribution for DT classifier trained on category C and undersampled dataset



**Fig. 11** EPP score distribution for Bagging classifier trained on Category C and undersampled dataset

we found that the most important feature for the purchase prediction is session duration. The reason might be that session duration is a good indicator for identifying users' purchase intention. For example, when a shopper spends more time on the e-commerce platform, it can show that the user has purchase intention, which increases the probability of purchase action. The importance of other features is shown in Fig. 12. It can be seen that using the minimum popularity value of browsed items in the session has very little effect on identifying purchase prediction.

## Discussion

This work explores how ML models perform on early purchase prediction. We train our ML models on anonymous

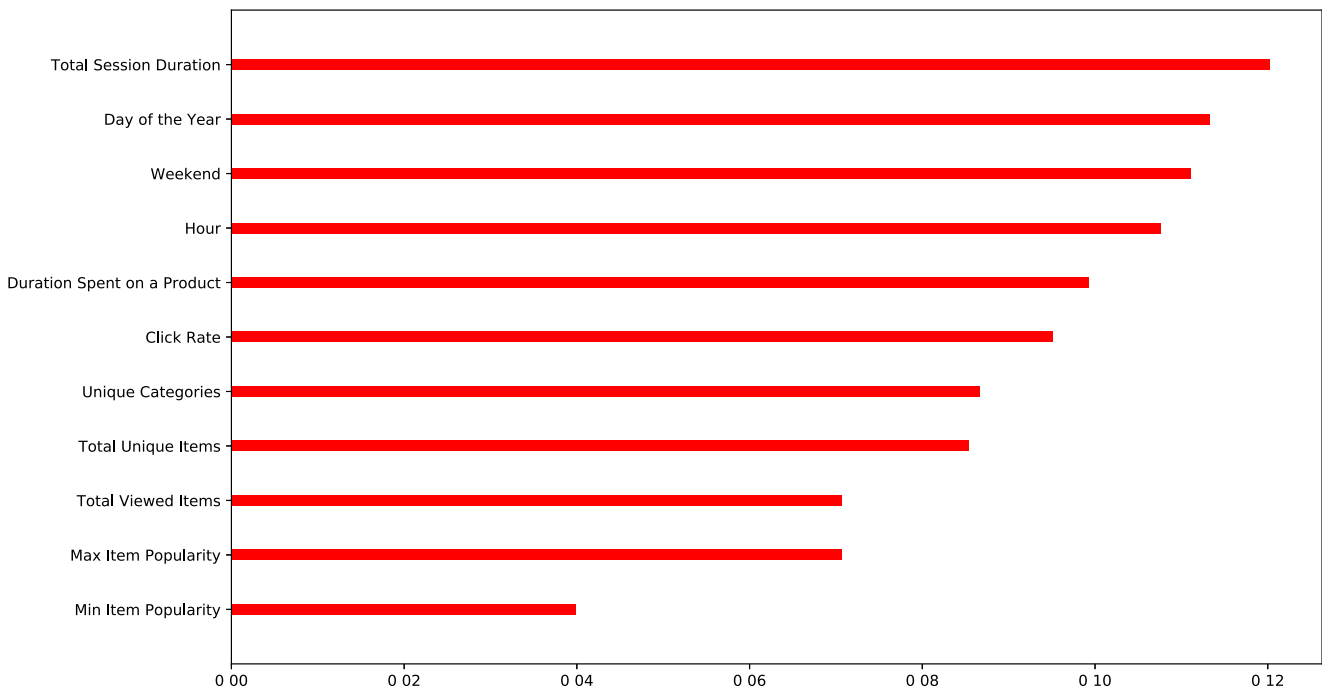
session logs and investigate how early purchase intention can be determined for a test session. In this section, we discuss our findings based on the questions asked in Section "Introduction". These are: (1) Given session data after a user's first interaction, how helpful can ML models be in predicting the likelihood of a purchase in an ongoing session? (2) What is the most critical session feature for early purchase prediction and how can it be identified? (3) How can ML models be evaluated to measure their performance on early purchase prediction? The findings based on the above questions are discussed in the next section (Section "Findings"). In Section "Contributions", we discuss the implications of these findings and our contribution to the body of knowledge.

## Findings

1. We found that ML models are useful in predicting the likelihood of purchase in an ongoing session after the first interaction, especially when the DT model is used. DT is able to predict early purchase intention with around % 97 AUC score. Bagging, KNN and RF achieved AUC scores of % 94, % 93, % 90 respectively. While NB showed % 75 AUC score (For more details refer to Table 5). We observed that based on our experiments, DT is the best performing ML model for predicting early purchase intention in an on-going session (Table 6). In addition, we analysed EPP utility scores for two best performing ML models (Figs. 10 and 11). Results show that the higher the number of sessions, the higher the EPP utility score that DT model can produce in comparison to Bagging model.

**Table 6** Confusion matrix analysing of DT, Bagging and KNN classifier models trained on category C undersampled dataset

			Predicted	
			No	Yes
DT	Actual	No	39,000	585
		Yes	236	5,178
Bagging	Actual	No	39,478	107
		Yes	903	4,511
KNN	Actual	No	39,285	300
		Yes	1,539	3,875



**Fig. 12** Scored feature importance using DT classifier

2. We found that “Session Duration” is the most important feature that gives critical signal for predicting users’ purchase intention. We believe that this could be attributed to the length of the duration that a user spends in a session as this could increase the probability of a purchase. On the other hand, “Min Popularity” feature that indicates the value of minimum popularity of the browsed items in the session is found as having the lowest importance to classify users’ purchase intention. We created the importance of features analyses (Fig. 12) during establishing the process of the best performed ML model, which is DT classifier.
3. To evaluate the performance of ML models on early purchase prediction, we develop an EPP utility scoring method that assigns efficiency scores to the ML models. The developed EPP scoring method applies defined rules for purchase-ended sessions and non-purchase sessions (Section “EPP scoring function”). A high utility score indicates that a ML model is efficient in early purchase prediction.

## Contributions

Although many works in literature investigated purchase intention prediction, they focused on using features and building ML models after sessions end (Park and Park 2016; Mokryn et al. 2019; Martínez et al. 2020; Wu et al. 2015; Kytö et al. 2019; Köcher et al. 2019). These works mainly investigate if users can purchase product

in their next sessions. Also, these works studied feature-based performance difference of the models on the purchase prediction after sessions end. For instance (Mokryn et al. 2019) examined performance improvement of the ML models when the product trendiness is added as a feature in the model training stage. In Kim et al. (2020), real-time purchase action was investigated by using cameras and object detection algorithms and analysed users’ physical movements in a product cabinet in order to detect purchase action. What differentiates our work from existing studies is that in our work, we designed an EPP framework that is able to predict purchase intention for an ongoing session on an e-commerce website. Also, we proposed a utility scoring method in order to investigate how early ML models can detect purchase in an on-going session before purchase action happens.

## Theoretical implications

Our work offers academic and industrial implications in line with how important predicting users’ purchasing intention is in digital marketing community (Qiu et al. 2015). Current studies about purchase prediction are primarily focused on predicting purchase intention from the next time a consumer visits the e-commerce website (Park and Park 2016; Mokryn et al. 2019; Martínez et al. 2020). Some of these studies applied ML models for purchase prediction after a session has ended (Park and Park 2016; Mokryn et al. 2019). On the other hand, research investigating

the purchase intention of a consumer when the session is active is still scarce albeit important to digital marketing research. In this regard, we extend the existing works of Park and Park (2016) and Mokryn et al. (2019) by applying the ML models to predict purchase intention in an on-going session. The purpose of analysing a consumer's purchase probability while the session is on-going is for targeting advertisements and recommending products in real-time since recommending products and advertisements or sending offers after a session has ended is difficult when a consumer is anonymous, i.e. when the consumer is not registered to the website (Jannach and Jugovac 2019).

To the best of our knowledge, this is the first work that applies and evaluates the performance of ML models in early purchase intention prediction. In this work, we develop an EPP utility scoring method to measure the performance of the ML models. We offer valuable contribution for information retrieval researchers by utilising EPP framework to identify shoppers who do not intend to purchase. This may provide the added benefit of recommending products and offering discounts to users in an on-going session. Since many studies show the positive value of offering discounts and product recommendations in aiding consumers' purchase decision (McColl et al. 2020; Liu et al. 2016; Leeftang et al. 2014; Ahmed et al. 2014), our EPP framework identifies users' intention in real-time before they abandon the website. This also provides high accuracy on early purchase intention prediction even after first product interaction. Another contribution of this work is to develop a framework that demonstrates the capability of ML models in determining early purchase intention regardless of a user's registration status (registered or unregistered). In the existing studies, it has been highlighted that returns from unregistered consumers is very low as providing personalised offers for unregistered user is a challenging task (Behera et al. 2020; Hallikainen et al. 2019). Consequently, this work shows that ML models are strong predictors for when the shoppers' behaviours are analysed appropriately and behavioural features are generated dynamically for active sessions. Conclusively, we showed that EPP utility score can be applied successfully in order to determine how early ML models can predict the true intention of a consumer for purchase prediction. Thus, researchers in other domains may find some benefit in adapting our proposed EPP utility scoring method to determine the probability of occurrence of an event. For instance, research in the airline operations shows that early determination of a failure of a system of an aircraft is important (Dangut et al. 2020) or in healthcare research, early diagnosis of a disease is very crucial in the successful treatment of the disease (Perveen et al. 2016).

## Practical implications

Notable practical implications can also be concluded from this work. Early online identification of shoppers with high purchase intention could allow marketers to implement different online strategies for consumer intention, personalised recommendations, sales-boosting and targeted offers. Also, providing better-personalised recommendations can make consumers feel special which can improve consumer experience and loyalty. The role of discounts and price history charts for shoppers' perception has been widely investigated by the digital marketing community (Drechsler and Natter 2011). It was found that giving smart discounts and providing price history chart can improve the persuasions of the consumer on purchasing products. Our framework can contribute to the digital marketing community by improving the quality of personalised recommendations and targeted advertisements, and discounts. It can also aid in identifying profitable consumers which can contribute to making marketing spending more efficient (Kumar et al. 2008).

## Conclusion and future work

In this work, we proposed an Early Purchase Prediction framework (EPP). The framework aims to predict the users' buying intention from their first few interactions during a given online session. The problem was modelled as a classification problem. We evaluated the proposed framework using five ML models. The framework was evaluated using a range of ML models and different features sets. The proposed EPP framework can be combined with existing methods for recommender systems. Our method uses only anonymous session logs as dataset. We built dynamic feature creation after each interaction in the session, and then we used that feature for prediction. In order to determine a given positive purchase prediction before or after purchase action, we designed an early purchase prediction utility score and selected the features that created the highest early purchase prediction utility score. We used various ML models, including ensemble model under different class imbalance methods to test algorithms' performance for the chosen features for each session. We found that DT model is the best ML model that can determine early purchase intention in an ongoing sessions. In order to apply proposed EPP framework, we assumed that shoppers interacted with at least two products when they visit e-commerce website since the scoring utility is activated after the second viewed product as described in Section "EPP scoring function". Therefore, the framework cannot determine whether the given purchase



prediction is early purchase intention prediction or not when a shopper browsed only one product. This can be viewed as a limitation of the EPP framework. Nevertheless, the EPP framework can still give a prediction for shoppers' purchase intention. However, after shoppers browsed two or more products, the EPP framework can determine successfully shoppers' early purchase intention. Several lines of future work are identified from this work. First, we propose our next step which aims to test the early purchase prediction framework in other session-based datasets to further substantiate the effectiveness and adaptability of our proposed framework. In addition, early purchase prediction framework can be integrated into a session-based recommendation system. For example, prediction results based on the created features from the session interactions can be utilised as a filtering option or a guide for choosing recommended items which can improve more personalised recommendation.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

Key terminologies used in this work are provided below.

1. Session Observation: Shows the data collection stage from an e-commerce website.
2. Utility Score: This is a function that measures how early a machine learning model can determine a purchase action in the ongoing session.
3. Attribute Selection: Shows a specification of the session that using this specification, other features can be created, such as timestamp that helps to find the day of the week and hour of the day features.
4. Class Imbalance: This shows in our work the imbalanced distribution of the purchase and not purchase ended sessions since most of the sessions in the datasets do not have purchase (Section "Dataset description").
5. Feature Generation: Features show the characteristics of a session such as session duration, the number of browsed products in the session and the day of the year when a session starts.
6. Machine Learning (Prediction) Algorithms or Machine learning (Prediction) models: These models can be interpreted as a function that can learn the parameters of the function based on the inputs and outputs (Russell and Norvig 2016). After parameters of the function are learnt in the training stage, the function can map a given input to output as a prediction.
7. Timestamp: Shows the time and date a session started.
8. Pre-processing: Pre-processing is a process to have a ready dataset in order to use in machine learning algorithms, for example, deleting sessions which have a missing timestamp (shows the time and date a session started).
9. Train Dataset: A separated dataset from the whole dataset for training the machine learning models.
10. Test Dataset: A separated dataset from whole dataset for testing the trained machine learning models.
11. Model Training: This shows the learning stage of the machine learning models using train dataset.

## References

- Ahmed, R.R., Parmar, V., Amin, M.A. (2014). Impact of product packaging on consumer's buying behavior. *European Journal of Scientific Research* 202X *European Journal of Scientific Research*, 120(2), 1450–216. <https://doi.org/10.13140/2.1.2343.4885>. [www.europeanjournalofscientificresearch.com](http://www.europeanjournalofscientificresearch.com).
- Batista, G.E.A.P.A., Prati, R.C., Monard, M.C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1), 20–29. <https://doi.org/10.1145/1007730.1007735>.
- Baumann, A., Haupt, J., Gebert, F., Lessmann, S. (2018). Changing perspectives: Using graph metrics to predict purchase probabilities. *Expert Systems with Applications*, 94, 137–148. <https://doi.org/10.1016/j.eswa.2017.10.046>.
- Behera, R.K., Gunasekaran, A., Gupta, S., Kamboj, S., Bala, P.K. (2020). Personalized digital marketing recommender engine. *Journal of Retailing and Consumer Services*, 53, 101799. <https://doi.org/10.1016/j.jretconser.2019.03.026>.
- Ben-Shimon, D., Shapira, B., Tsikinovsky, A., Rokach, L., Friedmann, M., Hoerle, J. (2015). RecSys challenge 2015 and the YOOCHOOSE dataset. In *Recsys 2015 - proceedings of the 9th ACM conference on recommender systems* (pp. 357–358). <https://doi.org/10.1145/2792838.2798723>.
- Berry, M., & Linoff, G. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Wiley. <http://portal.acm.org/citation.cfm?id=983642>.
- Bogina, V., & Kuflik, T. (2017). Incorporating dwell time in session-based recommendations with recurrent neural networks. In *CEUR Workshop proceedings*, (Vol. 1922 pp. 57–59).
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brodén, B., Hammar, M., Nilsson, B.J., Paraschakis, D. (2018). Ensemble recommendations via Thompson sampling: an experimental study within e-commerce. In *International conference on intelligent user interfaces, proceedings IUI* (pp. 19–29). <https://doi.org/10.1145/3172944.3172967>.

- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>.
- Corrigan, J.R., Alhabash, S., Rousu, M., Cash, S.B. (2018). How much is social media worth? Estimating the value of Facebook by paying users to stop using it. *PLoS ONE*, 13(12), e0207101. <https://doi.org/10.1371/journal.pone.0207101>.
- Dangut, M.D., Skaf, Z., Jennions, I.K. (2020). An integrated machine learning model for aircraft components rare failure prognostics with log-based dataset. *ISA Transactions*. <https://doi.org/10.1016/j.isatra.2020.05.001>.
- de Pechpeyrou, P. (2009). How consumers value online personalization: a longitudinal experiment. *Direct Marketing*, 3(1), 35–51. <https://doi.org/10.1108/17505930910945723>.
- den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research*, 166(2), 557–575. <https://doi.org/10.1016/j.ejor.2004.04.022>.
- Dou, X. (2020). Online purchase behavior prediction and analysis using ensemble learning. In *2020 IEEE 5th International conference on cloud computing and big data analytics, ICCCBDA 2020* (pp. 532–536). <https://doi.org/10.1109/icccbda49378.2020.9095554>.
- Drechsler, W., & Natter, M. (2011). Do price charts provided by online shophots influence price expectations and purchase timing decisions? *Journal of Interactive Marketing*, 25(2), 95–109. <https://doi.org/10.1016/j.intmar.2011.02.001>.
- Dutta, D., Paul, D., Ghosh, P. (2019). Analysing feature importances for diabetes prediction using machine learning. In *2018 IEEE 9th annual information technology, electronics and mobile communication conference IEMCON 2018* (pp. 924–928). <https://doi.org/10.1016/j.intmar.2011.02.001>.
- Esmeli, R., Bader-El-Den, M., Abdullahi, H. (2020). Using Word2Vec recommendation for improved purchase prediction. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). <https://doi.org/10.1109/ijcnn.2018.8614871>.
- Esmailian, P., & Jalili, M. (2015). Purchase prediction and item suggestion based on HTTP sessions in absence of user information. In *Proceedings of the international ACM recommender systems challenge 2015* (p. 6). <https://doi.org/10.1145/2813448.2813515>.
- Hallikainen, H., Alamäki, A., Laukkanen, T. (2019). Individual preferences of digital touchpoints: A latent class analysis. *Journal of Retailing and Consumer Services*, 50, 386–393. <https://doi.org/10.1016/j.jretconser.2018.07.014>.
- He, H., Bai, Y., Garcia, E.A., Li, S. (2008). ADASYN: adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of the international joint conference on neural networks* (pp. 1322–1328). <https://doi.org/10.1109/ijcnn.2008.4633969>.
- Huang, J., & Ling, C.X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3), 299–310. <https://doi.org/10.1109/tkde.2005.50>.
- Jannach, D., & Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems*, 10(4), 1–23. <https://doi.org/10.1145/3370082>.
- Kim, D.H., Lee, S., Jeon, J., Song, B.C. (2020). Real-time purchase behavior recognition system based on deep learning-based object detection and tracking for an unmanned product cabinet. *Expert Systems with Applications*, 143, 113063. <https://doi.org/10.1016/j.eswa.2019.113063>.
- Köcher, S., Jugovac, M., Jannach, D., Holz Müller, H.H. (2019). New hidden persuaders: an investigation of attribute-level anchoring effects of product recommendations. *Journal of Retailing*, 95(1), 24–41. <https://doi.org/10.1016/j.jretai.2018.10.004>.
- Kubat, M., & Matwin, S. (1997). Addressing the curse of imbalanced data sets: one-sided sampling. In *Proceedings of the fourteenth international conference on machine learning*, (Vol. 97 pp. 179–186). <http://sci2s.ugr.es/keel/pdf/algorithm/congreso/kubat97addressing.pdf>.
- Kumar, V., Venkatesan, R., Reinartz, W. (2008). Performance implications of adopting a customer-focused sales campaign. *Journal of Marketing*, 72(5), 50–68. <https://doi.org/10.1509/jmkg.72.5.50>.
- Kytö, E., Virtanen, M., Mustonen, S. (2019). From intention to action: predicting purchase behavior with consumers' product expectations and perceptions, and their individual properties. *Food Quality and Preference*, 75, 1–9. <https://doi.org/10.1016/j.foodqual.2019.02.002>.
- Leefflang, P.S., Verhoef, P.C., Dahlström, P., Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European Management Journal*, 32(1), 1–12. <https://doi.org/10.1016/j.emj.2013.12.001>.
- Lemaître, G., Nogueira, F., Aridas, C.K. (2017). Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(1), 1–5.
- Li, Q., Gu, M., Zhou, K., Sun, X. (2016). Multi-classes feature engineering with sliding window for purchase prediction in mobile commerce. In *Proceedings - 15th IEEE international conference on data mining workshop, ICDMW 2015* (pp. 1048–1054). <https://doi.org/10.1109/icdmw.2015.172>.
- Liu, X., Lee, D., Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, 56(6), 918–943. <https://doi.org/10.1177/0022243719866690>.
- Liu, Y., Liu, W., Yuan, P., Zhang, Z. (2016). Empirical research on how product advertising, time pressure and the discount rate effect on the sales of products in online group purchase. *Procedia Computer Science*, 91, 625–634. <https://doi.org/10.1016/j.procs.2016.07.156>.
- Lo, C., Frankowski, D., Leskovec, J. (2016). Understanding behaviors that lead to purchasing: a case study of pinterest. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 531–540). <https://doi.org/10.1145/2939672.2939729>.
- Louppe, G., & Geurts, P. (2012). Ensembles on random patches. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, (Vol. 7523 LNAI pp. 346–361). [https://doi.org/10.1007/978-3-642-33460-3\\_28](https://doi.org/10.1007/978-3-642-33460-3_28).
- Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, 281(3), 588–596. <https://doi.org/10.1016/j.ejor.2018.04.034>.
- McColl, R., Macgillchrist, R., Rafiq, S. (2020). Estimating cannibalizing effects of sales promotions: the impact of price cuts and store type. *Journal of Retailing and Consumer Services*, 53, 101982. <https://doi.org/10.1016/j.jretconser.2019.101982>.
- Mokryn, O., Bogina, V., Kuflik, T. (2019). Will this session end with a purchase? Inferring current purchase intent of anonymous visitors. *Electronic Commerce Research and Applications*, 34, 100836. <https://doi.org/10.1016/j.elerap.2019.100836>.
- Mu, W., Lennon, S.J., Liu, W. (2020). Top online luxury apparel and accessories retailers: what are they doing right? *Fashion and Textiles*, 7(1), 1–17. <https://doi.org/10.1186/s40691-019-0197-x>.
- Pálóvics, R., Szalai, P., Kocsis, L., Szabó, A., Frigó, E., Pap, J. (2015). Solving RecSys challenge 2015 by linear models, gradient boosted trees and metric optimization. In *Proceedings of the*

- international ACM recommender systems challenge 2015* (pp. 1–4). <https://doi.org/10.1145/2813448.2813513>.
- Park, C.H., & Park, Y.H. (2016). Investigating purchase conversion by uncovering online visit patterns. *Marketing Science*, 35(6), 894–914. <https://doi.org/10.1287/mksc.2016.0990>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. (2011). Scikit-learn: machine learning in {P}ython. *Journal of Machine Learning Research*, 12, 2825–2830.
- Perveen, S., Shahbaz, M., Guergachi, A., Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, 82, 115–121. <https://doi.org/10.1016/j.procs.2016.04.016>.
- Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, 26(1), 25–33. <https://doi.org/10.1007/s10489-006-0001-7>.
- Qiu, J., Lin, Z., Li, Y. (2015). Predicting customer purchase behavior in the e-commerce context. *Electronic Commerce Research*, 15(4), 427–452. <https://doi.org/10.1007/s10660-015-9191-6>.
- Romov, P., & Sokolov, E. (2015). RecSys challenge 2015: ensemble learning with categorical features. In *Proceedings of the international ACM recommender systems challenge 2015* (p. 1). <https://doi.org/10.1145/2813448.2813510>.
- Russell, S.J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia: Pearson Education Limited.
- Rust, R.T., Kumar, V., Venkatesan, R. (2011). Will the frog change into a prince? Predicting future customer profitability. *International Journal of Research in Marketing*, 28(4), 281–294. <https://doi.org/10.1016/j.ijresmar.2011.05.003>.
- Shu-Yue, M., Wen-Cai, L., Shuo, W. (2011). The study on the preprocessing in web log mining. In *Proceedings - 2011 4th international symposium on knowledge acquisition and modeling, KAM 2011* (pp. 315–317). <https://doi.org/10.1109/KAM.2011.90>.
- Stewart-Knox, B.J., Markovina, J., Rankin, A., Bunting, B.P., Kuznesof, S., Fischer, A.R. (2016). Making personalised nutrition the easy choice: creating policies to break down the barriers and reap the benefits. *Food Policy*, 63, 134–144. <https://doi.org/10.1016/j.foodpol.2016.08.001>.
- Suh, E., Lim, S., Hwang, H., Kim, S. (2004). A prediction model for the purchase probability of anonymous customers to support real time web marketing: a case study. *Expert Systems with Applications*, 27(2), 245–255. <https://doi.org/10.1016/j.eswa.2004.01.008>.
- Wang, Y., & Tseng, M.M. (2015). A Naïve Bayes approach to map customer requirements to product variants. *Journal of Intelligent Manufacturing*, 26(3), 501–509. <https://doi.org/10.1007/s10845-013-0806-2>.
- Wu, Z., Tan, B.H., Duan, R., Liu, Y., Mong Goh, R.S. (2015). Neural modeling of buying behaviour for E-commerce from clicking patterns. In *Proceedings of the international ACM recommender systems challenge 2015* (p. 12). <https://doi.org/10.1145/2813448.2813521>.
- Yagci, A.M., Aytekin, T., Gürgen, F.S. (2015). An ensemble approach for multi-label classification of item click sequences. In *Proceedings of the international ACM recommender systems challenge 2015* (p. 7). <https://doi.org/10.1145/2813448.2813516>.
- Yeo, J., Kim, S., Koh, E., Hwang, S.W., Lipka, N. (2017). Predicting online purchase conversion for retargeting. In *WSDM 2017 - Proceedings of the 10th ACM international conference on web search and data mining* (pp. 591–600). <https://doi.org/10.1145/3018661.3018715>.
- Zeng, M., Cao, H., Chen, M., Li, Y. (2019). User behaviour modeling, recommendations, and purchase prediction during shopping festivals. *Electronic Markets*, 29(2), 263–274. <https://doi.org/10.1007/s12525-018-0311-8>.
- Zhou, Y., Mishra, S., Gligorijevic, J., Bhatia, T., Bhamidipati, N. (2019). Understanding consumer journey using attention based recurrent neural networks. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 3102–3111). <https://doi.org/10.1145/3292500.3330753>.
- Zhuang, L., Kou, Z., Zhang, C. (2005). Session identification based on time interval in web log mining. In *IFIP Advances in information and communication technology*, (Vol. 163 pp. 389–396). [https://doi.org/10.1007/0-387-23152-8\\_50](https://doi.org/10.1007/0-387-23152-8_50).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.