Leveraging fine-grained transaction data for customer life event predictions

Arno de Caigny, Kristof Coussement, Koen de Bock

Leveraging Fine-Grained Transaction Data for Customer Life Event Predictions

Arno De Caigny[1,2], Kristof Coussement[1,2], and Koen W. De Bock[3]

[1]IESEG School of Management, 3 Rue de la Digue, F-59000, Lille, France

[2] UMR 9221-LEM-Lille Economie Management, 3 Rue de la Digue, F-59000 Lille, France

[3]Audencia Business School, 8 Route de la Jonelière, F-44312 Nantes, France

E-mail addresses: a.de-caigny@ieseg.fr (Arno De Caigny); k.coussement@ieseg.fr (Kristof Coussement), kdebock@audencia.com (Koen W. De Bock)

Corresponding author: Kristof Coussement, 3 rue de la Digue, F-59000 Lille, France. Tel. (+33) 3 20 54 58 92; e-mail k.coussement@ieseg.fr.

Leveraging Fine-Grained Transaction Data for Customer Life Event Predictions

Abstract

This real-world study with a large European financial services provider combines aggregated customer data including customer demographics, behavior and contact with the firm, with fine-grained transaction data to predict four different customer life events: moving, birth of a child, new relationship, and end of a relationship. The fine-grained transaction data—approximately 60 million debit transactions involving around 132,000 customers to more than 1.5 million different counterparties over a one-year period—reveal a pseudo-social network that supports the derivation of behavioral similarity measures. To advance decision support systems literature, this study validates the proposed customer life event prediction model in a real-world setting in the financial services industry; compares models that rely on aggregated data, fine-grained transaction data, and their combination; and extends existing methods to incorporate fine-grained data that preserve recency, frequency, and monetary value information of the transactions. The results show that the proposed model predicts life events significantly better than random guessing, especially with the combination of fine-grained transaction and aggregated data. Incorporating recency, frequency, and monetary value information of fine-grained transaction data also significantly improves performance compared with models based on binary logs. Fine-grained transaction data accounts for the largest part of the total variable importance, for all but one of the life events.

Keywords

Life event prediction, predictive modeling, pseudo-social networks, customer relationship management (CRM), big data, data science

1 Introduction

Successful companies build long-term relations with their customers [1], often through formal customer relationship management (CRM) strategies [2]. As customers' needs change, managing these relations becomes challenging though, so firms seek to analyze customer lifecycles to gain predictive indicators of customers' behavior and needs [3]. According to family lifecycle theory, customers behave similarly when they are in the same life stage [4,5], and their movement from one stage to another over time usually is triggered by detectable events (e.g., new relationship, birth of a child). Such trigger events are typically important moments in life and we will refer to them as *life events*. Life events impose acute stressors on most people [6,7], so to adjust to the related changes, they reevaluate their consumption priorities and develop coping behaviors [8]. Thus, sudden shifts in customer needs and behaviors often occur after life events [8], such as when new parents start spending more on healthy food and less on eating out [9]. Managing customer relationships during such life events has crucial strategic importance and can align the company's actions better with future customer behavior [10].

Accordingly, companies invest considerably in methods to detect life events; for example, approximately 40 life events, including moving and the birth of a child, can be detected from textual analyses of Twitter data [11]. Customer life event information is stored in databases that can be used for many CRM applications [12]. Correct detection of life events enables companies to engage customers with appropriate cross- or up-selling offers [13] and reactively target them with relevant marketing campaigns [14]. With Google ads for example, insurance companies can target customers who have recently moved with appropriate renters' or homeowners' insurance products [15]. Furthermore, recent innovations in big data analytics suggest options for moving beyond such detection efforts and toward life event predictions, which could support proactive rather than reactive targeting of customers [15].

Predictive models are instrumental to score customers based on the likelihood of experiencing a future life event. They use historical data to anticipate the future state of a customer [16]. Predictive modeling accordingly supports customer scoring applications in many CRM domains [17], such as customer churn prediction [e.g. 18], credit scoring [e.g. 19], and response modeling [e.g. 20]. These models rely on predictive variables that are engineered from customer data stored in large relational databases [21,22]. Typically, these variables reflect information about customer demographics, variables summarizing the customer relationship, such as length of relationship or purchase history, and customer–company interactions; they generally are referred to as aggregated or structured variables [23]. Although big data analytics promise the use of other sources of (unstructured) data [15], including in predictive models [23], previous predictive modeling studies in CRM mainly focus on structured data [18,24].

To complement such insights, we propose and test, for the first time, a method for customer life event prediction (CLEP), which represents an especially challenging problem for most companies for three main reasons. First, most customer life events take place outside the scope of the relationship with the company [25]. For example, to predict whether a customer will move in coming months, the company often lacks any direct insights obtained from prior interactions with the customer, which differs from the situation that arises for other applications of predictive modeling. Customer churn predictions, for example, relate directly to the customer's prior behavior. Second, CLEP requires information about prior life events to train the predictive machine learning model, and gathering such information requires additional effort and investment by companies, because they rarely gather such details in the course of normal business practices [26]. Third, most life events are rare, which affects data preprocessing and the modeling process [25].

In general, predictive CRM applications depend on three types of variables aggregated at the customer level: those that refer to socio-demographical characteristics; those that describe the contact between the client and the company; and historical customer behaviors, including RFM

variables (i.e., recency, frequency and monetary value) derived from the customer relationship. We refer to these types as *aggregated variables*. Although they offer good performance in other CRM applications, they might be insufficient for accurate CLEP, for the reasons we just listed. Instead, recent attempts have demonstrated the added value of alternative aggregation strategies based on acknowledging the underlying network structures of *fine-grained data* that describe all interactions of customers with counterparties, on the most detailed level. Typical examples of fine-grained data include Facebook likes [27], payment history [23], and online display ad networks. Such data have great power to predict personal attributes [27] and improve targeted marketing [23]. They provide logs of every event (e.g., likes, payments) involving each counterparty (e.g., Facebook page, merchant), which can reveal novel insights into customers' lifestyles [27]. In section 2, a detailed overview of fine-grained transaction data is presented, which is the type of fine-grained data used in this study. By leveraging these data for CLEP, we seek to contribute to research into the use of big data in CRM.

However, because these fine-grained data are characterized by high dimensionality and sparsity, a different method than for aggregated data is required to summarize information on the client level while still maintaining the richness of the data. A scalable way to leverage fine-grained data is to calculate behavioral similarity scores among customers in a pseudo-social network (PSN) [23], as we detail in Section 2. Calculations of behavioral similarity measures solely rely on the existence of transactions with the same counterparties [23], but we attempt to extend this methodology to capture more information from fine-grained data. Noting that RFM variables, as contained within a customer's purchase history, effectively summarize transaction history and can predict future customer behavior [28,29], we propose an extended approach to behavioral similarity measures that reflect not just whether customers engage in transactions with shared counterparties but also whether they transact in similar fashion with them by incorporating RFM information.

In this study, four life events (moving, birth of a child, new relationship, end of relationship), are predicted using both aggregated and fine-grained transaction data, gathered from a real-life data

set provided by a financial services provider. A new methodology for the fine-grained transaction data is proposed that extends previous work of [23] by incorporating RFM information of the transaction logs. We benchmark the results against those obtained from existing methods. Companies can use such predictions for different applications; these predictions can be used to set up a targeted marketing campaign which can be very successful if products are correlated with life events (e.g. fire insurance and moving). A more subtle way to use these predictions is via personalization of the service. A client who enters the website of the company, for example, might be shown different products or promotions based on his predicted probability to encounter a life event.

The objective of this study is to explore whether life events can be effectively predicted using customer profile and transaction data, and how fine-grained transaction data is best leveraged for this purpose. Previous research demonstrated that the outcomes of predictive models can help managers in making better decisions which improves the CRM [e.g. 30,31]. In this regard, predictive modeling is crucial for pro-active, data-driven decision making [32,33]. Companies that deploy data-driven decision systems perform better in objective measures of financial and operational results [34]. Customer life events are used in many CRM applications such as customer targeting, segmentation [35] and customer life time value estimations [5]. Therefore, CLEP will enable pro-active decision making in these areas and help companies in achieving a better CRM.

This study contributes to the business analytics research stream in decision support systems literature in three ways. This study contributes to the literature in three ways. First, previous DSS literature have explored the potential of new analytical application domains [e.g. 31]. Therefore, this paper contributes to this literature stream that find innovative DSS by demonstrating the value of CLEP in a real-world setting. Second, extant research [e.g. 23] has focused on methodological development that proof the added value of fine-grained data in a predictive modeling setting. We extent the existing methodology for fine-grained data, which only allowed for binary input data,

for continuous input data. This allows us to explore and demonstrate the added value of the incorporation of RFM information of transactions in behavioral similarity measures derived from the PSN. Thirdly, this study investigates which data to include in CLEP; by constructing models with aggregated data and/or fine-grained data. Our results show that aggregated data and fine-grained transaction data should be combined to obtain the highest predictive performance. Insights are provided in the importance of the different variable categories.

~~First, this is the first study to demonstrate CLEP in a real-world setting. Second, existing methodology for fine-grained data, which only allowed for binary input data, is extended for continuous input data. This allows us to explore and demonstrate the added value of the incorporation of RFM information of transactions in behavioral similarity measures derived from the PSN. Thirdly, this study investigates which data to include in CLEP; by constructing models with aggregated data and/or fine-grained data. Our results show that aggregated data and fine-grained transaction should be combined to obtain the highest predictive performance. Insights are provided in the importance of the different variable categories.~~

After we elaborate on the proposed methodology and the use of fine-grained data in the next section, we provide an overview of our experimental design in Section 3. Section 4 presents the results. Section 5 offers conclusions, and Section 6 ends with limitations and directions for further research.

2 Methodology

This section provides an overview of the methodology. The first part presents a generic two-stage framework and discusses how fine-grained data and aggregated data can be combined and integrated in a predictive model. The second part zooms in on fine-grained data, which is the key aspect of this study. Note, that we focus mainly on fine-grained data in this section, as our methodological contribution specifically relates to such data. The treatment of aggregated data

follows well-known practices previously described in literature [E.g. 36,37] and is presented in detail in section 3, the experimental design.

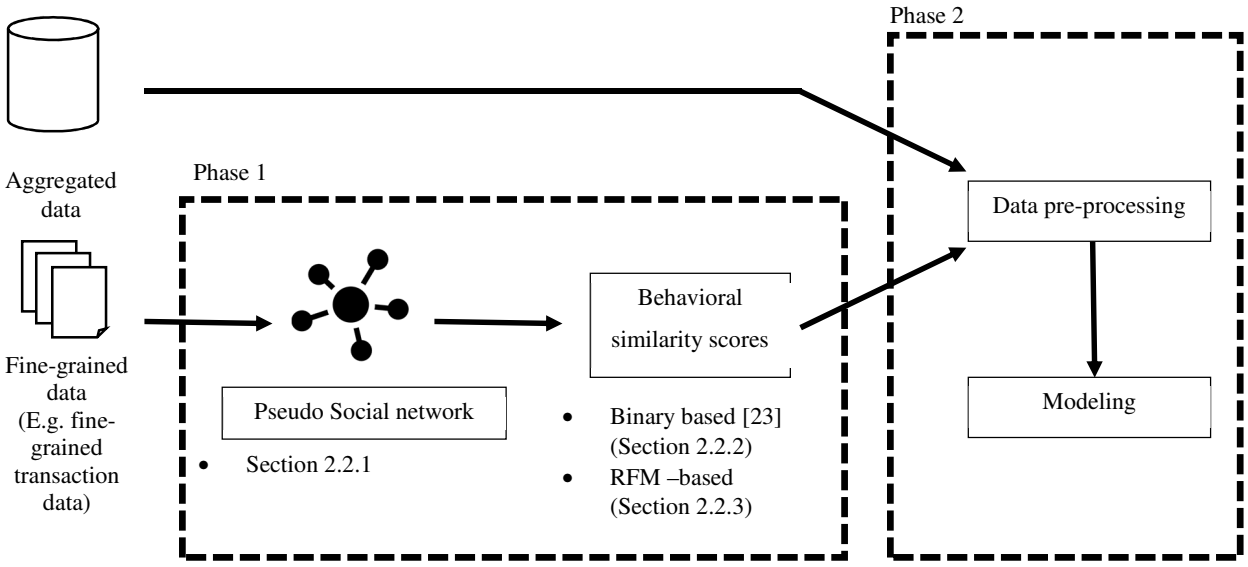
2.1 Framework for integrating fine-grained data in a predictive model

This section provides an overview of how aggregated data and fine-grained data are combined in a predictive model. Figure 1 presents a high-level overview of this framework which summarizes the two main phases. Aggregated data and fine-grained data are handled separately because there is a difference in data complexity of these two data sources, and which allows aggregated data to be passed directly passed to the second phase.

In the first phase, relevant variables are constructed for the fine-grained data. The fine-grained data requires, compared to the aggregated data, some additional steps to construct relevant variables. This entire process is detailed in section 2.2. First, the data are represented in a pseudo-social network, which is explained in section 2.2.1 in full detail using a small example. Second, from this network, relevant variables are calculated which is referred as *featurization* [38]. Section 2.2.2 explains how such variables are calculated based on existing methodology using a pseudo-social network constructed on binary information. In section 2.2.3, a new methodology is proposed that extends the existing one such that it allows for pseudo-social networks constructed on continuous data.

In a second phase, data pre-processing and modeling steps are included. The aggregated data requires data pre-processing steps to deal with categorical variables, outliers and missing values. As these steps are common practice in any predictive model and not the core of this study, they are only discussed in section 3 about the experimental design. Variables constructed on both data types can be combined in a predictive model. An appropriate modeling technique is chosen based on the problem at hand. Previous studies have already explored many algorithms in many different settings such as customer churn prediction [E.g. 18,39,40], customer acquisition [E.g. 41] and customer response modeling [E.g. 30]. We do base our choice for a modeling techniques on existing literature, which is detailed in section 3.3.

Figure 1: Framework for integrating fine-grained data in a predictive model



2.2 Methodology for fine-grained data

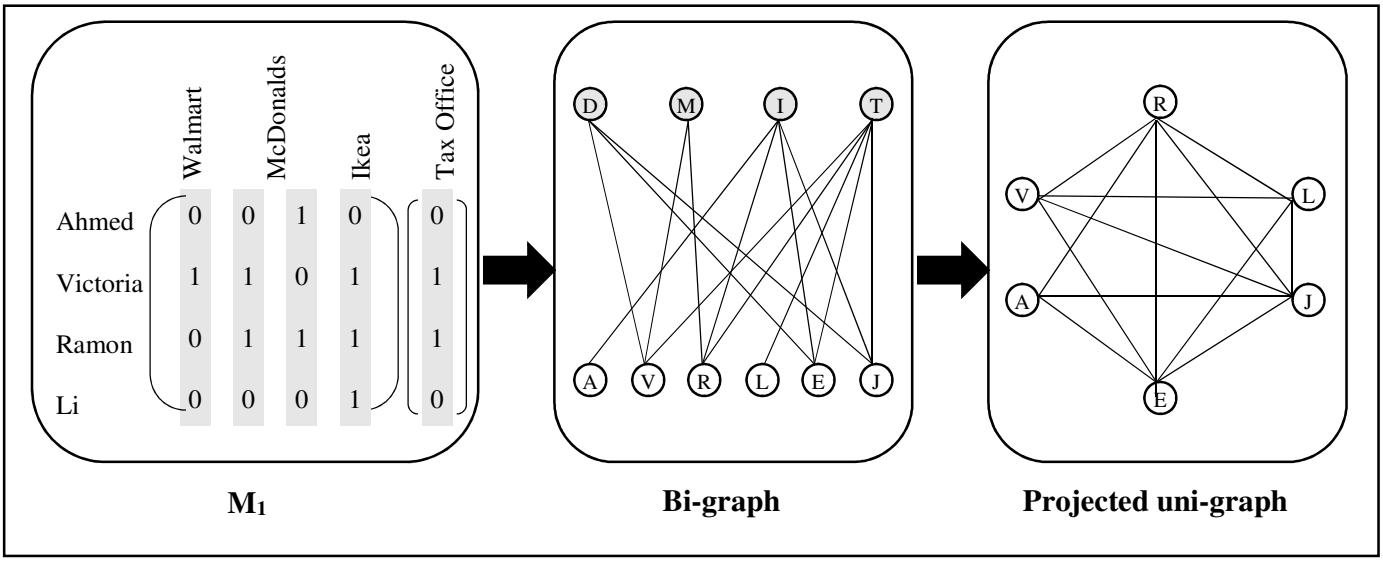
In this section, the methodology for fine-grained data is discussed. In elaborating the methodology for extracting variables from fine-grained data, we first present different representations of transaction data, which is the type of fine-grained data that are used in the focal study. Next the PSN is presented and we explain how it is derived from fine-grained transaction data. Then we outline a state-of-the-art method proposed by Martens et al. [23] for calculating behavioral similarity scores between customers, according to customers' transactions with counterparties. Finally, to extend this existing methodology, we propose incorporating RFM values resulting from the customers' payment behavior.

2.2.1 Fine-grained transaction data representation

Fine-grained transaction data contain the logs of transactions or payments from clients to several counterparties, such as companies, institutions, and other clients. This information can be represented in three ways, as illustrated through an example in Figure 2.Note that Figure 2 only serves illustrative purposes to help explain the main principles of the proposed methodology.

Real-world fine-grained data would typically contain thousands or even millions of records and thousands of counterparties. First, it can be represented in a large matrix $\mathbf{M_1}$, where element x_{ij} represents the occurrence of a transaction between client *i* and counterparty *j*. In Figure 2, x_{ij} is a binary indicator that signals whether a client made a payment to the counterparty, which we refer to as *binary logs*. Life event information is represented in a separate binary vector. Thus for every life event, a separate model is constructed. Second, the information in matrix $\mathbf{M_1}$ could indicate an adjacency matrix for a bi-graph; in our case, the clients and counterparties would be the two nodes in a bi-graph. Edges, defined by the transaction data, exist only between clients and counterparties with non-zero values. In Figure 2, client *Ahmed* appears only in the bi-graph connected with the counterparty *Ikea* because he only interacted with this counterparty. Third, a uni-graph that represents a PSN among clients, based on their similar transaction behavior toward counterparties, can be derived from the bi-graph. In Figure 2, *Ahmed* is connected to the clients *Ramon, Emma,* and *Jacob*, because they share *Ikea* as a counterparty. The PSN is a social network because customers are linked with one another, but it is "pseudo" in the sense that, unlike in real social networks, even the strongly linked customers might not know one another [23]. In Figure 2, clients *Ahmed* and *Ramon* are linked in the PSN because they both made a payment to counterparty *Ikea*, but they probably have never met.

Figure 2: Representations of transaction data adapted from Martens et al. [23].



2.2.2 Behavioral similarity

The graphical representation of fine-grained transaction data as a PSN provides links among customers, on the basis of the similarity of their counterparties. The similarity weights of these links incorporate two principles in the original method [23]. First, customers that share more counterparties earn higher weights. Second, popular counterparties are subject to down-weighting because large companies such as energy providers often establish little similarity information. The *Tax Office* counterparty in Figure 2, for example, needs to be down-weighted, because all but one of the clients made payments to it. This method, developed specifically to handle fine-grained transaction data, effectively trades off computational efficiency, scalability, and predictive performance [42].

Behavioral similarity scores, based on similarity weights between customers, then can be derived for customers within the network. In a predictive modeling context, customers who have recently experienced a life event serve as "seeds," and the scores indicate the similarity between a focal customer and the average seed customers. The underlying assumption is that customers who have strong links in the PSN, and thus transact with similar merchants, behave similarly in other ways too, such as encountering important life events or buying the same products [43]. In Figure 2, clients *Victoria* and *Ramon* have experienced a life event and serve as seed customers. The behavioral similarity score aims to represent how similar the payment behavior of any other customer, such as *Emma* or *Jacob*, is to that of these seed customers. As demonstrated by Martens et al. [23] and presented in Equation (1), the behavioral similarity scores can be regrouped algebraically by counterparty. For every counterparty j, the empirical probability E_j (ratio of seed customers that made a payment to the counterparty [NS_j] and the total number of customers [NC_j]) represents the behavioral similarity term. Therefore, the behavioral similarity score of client X_i is defined as the sum of empirical probabilities E_j of counterparties to which X_i has made at least one payment, as in Equation (2).

$$E_j = \left(\frac{NS_j}{NC_j}\right). \tag{1}$$

$$S_{besim}(X_i) = \sum_{j \mid x_{ij}=1} E_j. \tag{2}$$

Using the example in Figure 2, we can calculate the empirical probabilities E_j for every counterparty j, with the process we summarize in Table 1. In our simplified example, clients *Emma* and *Jacob* made payments to the same counterparties, namely *Walmart*, *Ikea*, and the *Tax Office*, so their behavioral similarity scores are identical:

$$S_{besim}(Emna) = S_{besim}(Jacob) = E_{Wallmart} + E_{Ikea} + E_{Tax\,Office} \tag{3}$$
$$= 0.33 + 0.25 + 0.40 = 0.98$$

Table 1: Calculation of empirical probabilities for counterparties in the example

Merchant *j*	**Clients**	NC_j	NS_j	E_j
Walmart	**Victoria**, Emma, Jacob	3	1	0.33
McDonalds	**Victoria**, **Ramon**	2	2	1
Ikea	Ahmed, **Ramon**, Emma, Jacob	4	1	0.25
Tax Office	**Victoria**, **Ramon**, Li, Emma, Jacob	5	2	0.40

Note: Clients in bold are the seed customers.

Equation (2) represents one option for calculating behavioral similarities, but other variants are available, as detailed in Equations (4)–(9) [23]. They differ in two main ways. First, the components pertaining to the seed customers, which we called behavioral similarity terms, are calculated differently. Second, two variants down-weight popular merchants based on inverse consumer frequency (ICF) or a cross-validated beta distribution. All calculations account only for the binary logs of the payment transactions as presented in matrix $\mathbf{M_1}$. Martens et al. [23] and De Cnudde et al. [42] provide more elaborate discussions of behavioral similarity scores based on PSN. These behavioral similarity scores can be used to score customers as presented in section 4.

$$ICF_j = log_{10}\left(\frac{n}{NC_j}\right). \tag{4}$$

$$S_{BIN\_NSNC}(\boldsymbol{x}) = \sum_{j|x_j=1} \frac{NS_j}{NC_J}. \tag{5}$$

$$S_{BIN\_ICF}(\boldsymbol{x}) = \sum_{j|x_j=1} \frac{NS_j}{NC_J}\, ICF(j). \tag{6}$$

$$S_{BIN\_B}(\boldsymbol{x}) = \sum_{j|x_j=1} \frac{NS_j}{NC_J} B(\alpha^*, \beta^*)(j). \quad (7)$$

$$S_{BIN\_NS}(\boldsymbol{x}) = \sum_{j|x_j=1} NS_j. \quad (8)$$

$$S_{BIN\_1}(\boldsymbol{x}) = \sum_{j|x_j=1} I(S_{NS}(\boldsymbol{x}) \neq 0). \quad (9)$$

where

n Total number of customers in the data set,

m All unique counterparts,

J A specific counterparty, such that $j \in m$,

NC_j The number of customers who made payments to the counterpart j,

NS_j The number of seed customers who made payments to the counterpart j, and

x_{ij} Binary variable that indicates whether customer i made a payment to counterpart j.

2.2.3. Extension of behavioral similarity with RFM

The derivation of behavioral similarity scores from a PSN built on fine-grained transaction data is a powerful way to incorporate this information in a predictive model, but it only accounts for whether transactions with counterparties occurred, such that the matrix $\mathbf{M_1}$ contains binary payment information. The existing method thus can only handle binary input data, which can cause an important loss of information. Hence, to contribute to this research domain, we seek to enrich the model with RFM dimensions gathered from fine-grained transaction data, because these variables reflect transactional information with great richness and have substantial importance for predictive marketing applications [e.g. 30,44]. Therefore, we propose a new methodology to incorporate RFM information of fine-grained transaction data to calculate behavioral similarity measures, as an extension of Martens et al. [23].

To predict life events, binary transactions might not tell the full story. In the example from Figure 2, clients *Emma* and *Jacob* earn the same behavioral similarity score because they are connected

to the same counterparties, but richer information might improve the accuracy of these behavioral similarity scores. Consider matrix $\mathbf{M_2}$ in Figure 3, in which every element reflects the total monetary value that a client spends with different merchants instead of the binary logs as presented in matrix $\mathbf{M_1}$. At counterparty *Ikea*, client *Ramon*, who experienced a life event, spent substantially more than client *Ahmed*, who did not. This important nuance is not captured in the binary logs. We propose that clients should earn higher similarity scores when they resemble seed customers, in that they have spent more at *Ikea*. In this example, client *Jacob* spent a lot more at *Ikea* than *Emma*, so *Jacob*'s behavior is more similar to the seed customer Ramon's, and he takes a higher score than *Emma* who spent less at *Ikea*. For this example, we use monetary value, but the same logic applies to recency and frequency variables.

The extension of the behavioral similarity scores to RFM-based behavioral similarity scores requires several adjustments to Equations (1)–(9). For our illustration, we focus on monetary value, but all presented formulas are similar for the recency and frequency dimensions. Note that these Equations are built step by step, such that elements of a previous Equation are sometimes used in later Equations. First, the behavioral similarity scores using RFM should capture whether the observed value is more similar to seed customers or non-seed customers, such as by calculating the deviations between a client's value and the average values for seed and non-seed clients. These deviations can be defined for a client *i* and counterparty *j* as:

$$DS_j(\boldsymbol{x}) = \left|AS_j - x_{ij}\right|, \text{ and} \tag{10}$$

$$DC_j(\boldsymbol{x}) = \left|AC_j - x_{ij}\right|, \tag{11}$$

where AS_j represents the average monetary value for all seed customers of a counterparty *j*, and AC_j is the average monetary value for all non-seed customers. In this case, x_{ij} does not represent the binary variable but the specific monetary value for customer *i* at counterparty *j*. Table 2 contains these values, calculated for monetary value using our ongoing example.

Figure 31: Fine-grained transaction data with monetary value

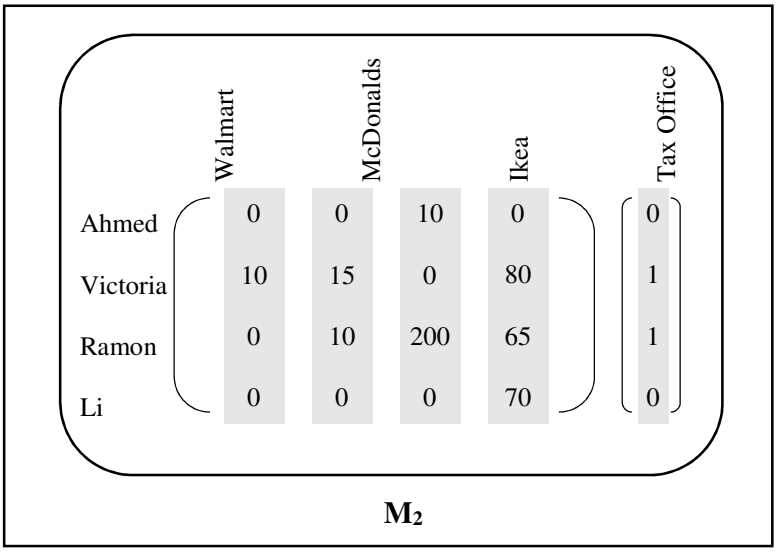


Table 2: Average monetary value for seed and non-seed clients

Merchant ***j***	**Clients**	***Total amount spent, non-seed clients***	***Total amount spent, seed clients***	AC_j	AS_j
Walmart	**Victoria**, Emma, Jacob	0	10	0	10
McDonalds	**Victoria**, **Ramon**	0	25	0	12.50
Ikea	Ahmed, **Ramon**, Emma, Jacob	10	200	10	200
Tax Office	**Victoria**, **Ramon**, Li, Emma, Jacob	70	145	70	72.50

Notes: Clients in bold are the seed customers.

Second, the calculation of the ratio R_{ij} , as presented in Equation (12), is similar to that of the empirical probability E_j [23] which was discussed before in Equation (1). Next, Equation (13) describes how the maximum value is calculated at a counterparty j and similarly Equation (14) provides the calculation for the minimum value. These minimum and maximum values are necessary in Equation (15) to describe how R_{ij} is normalized to R'_{ij} , which ensures the output always lies within the interval [0,1]. Clients with values of deviation from the average for seed

customers smaller than those from non-seed customers receive higher scores. In analogy with section 2.2, the final behavioral similarity score S_{besim_rfm} of a client *i* that includes one of the RFM indicators is therefore the sum of ratios R'_{ij} of the counterparties to which the client has made payments as shown in Equation (16). The respective equations are as follows:

$$R_{ij} = \left(\frac{\ln(DC_{ij}+1)}{\ln(DS_{ij}+1)}\right). \tag{12}$$

$$Max_j = \ln(|AC_j - AS_j| + 1). \tag{13}$$

$$Min_j = \left(\frac{1}{\ln(|AC_j - AS_j|+1))}\right). \tag{14}$$

$$R'_{ij} = \left(\frac{(R_{ij}-Min_j)}{Max_j-Min_j}\right). \tag{15}$$

$$S_{besim\_rfm}(X_i) = \sum_{j \mid x_{ij}>0} R'_{ij}. \tag{16}$$

Table 3 contains the behavioral similarity scores of clients from our example. Clients *Emma* and *Jacob* do not exhibit the same behavioral similarity score anymore; they differ in their spending at the different merchants. The difference between their scores mainly reflects the different monetary values in relation to *Ikea* for which *Jacob* is clearly more similar to the seed customers than *Emma*.

Third, analogous to the behavioral similarity scores derived from the binary PSN, we construct three variants to calculate behavioral similarity for each RFM dimension. These behavioral similarity scores account for deviance, so it is possible to detect whether a customer is more similar to the seed customers or the overall population. The following equations provide an overview of these measures. Equation (17) calculates the behavioral similarity as in the example and depends on the normalized ratio R'_{ij} defined in Equation (15). The two other variants also use the normalized ratio R'_{ij}, but adjust the weights of the counterparties, similar to Equations (6) and (7) respectively. The weighing of counterparties in Equation (18) are based on the ICF. In Equation (19), these weights are determined by a cross-validated Beta distribution.

$$S_{rfm\_ncns}(X_i) = \sum_{j \mid x_{ij}>0} R'_{ij}. \quad (17)$$

$$S_{rfm\_ICF}(X_i) = \sum_{j \mid x_{ij}>0} R'_{ij}\, ICF(j). \quad (18)$$

$$S_{rfm\_B}(X_i) = \sum_{j \mid x_{ij}>0} R'_{ij}\, B(\alpha^*, \beta^*)(j). \quad (19)$$

Table 3: Behavioral similarity scores based a PSN constructed with monetary value links

Clients i	$R'_{i,\text{Walmart}}$	$R'_{i,\text{Ikea}}$	$R'_{i,\text{tax office}}$	S_{besim_rfm}
Emma	0.7291	0.1681	0.5678	1.4650
Jacob	0.8358	0.7200	0.5506	2.0760

3 Experimental set-up

3.1 Data and experimental design

The data in this study came from a large European financial services provider. For all 132,703 customers included, this financial services provider is their primary bank; all customers thus exhibit bank transactions. Moreover, in line with the General Data Protection Regulations, all customers gave formal consent to the use of their data. Figure presents the timeline for the variable construction. The experimental design mimics previous research, and we apply a ten-fold cross-validation [23]. That is, we split the data into ten different folds, each used once as a test set, with the remaining nine folds as the train set. If required, we also can split the training data into a train and validation set (one third of the train data), to determine hyper-parameter settings. To derive the behavioral similarity scores, all clients in the train set that faced the particular life event in the independent period are considered seed customers for that particular life event. Based on the information that is included (aggregated, fine-grained or a combination), different models are defined for which further details about the specifics of each model are provided in section 3.3.

The independent variables come from two sources. First, 207 aggregated variables, as frequently used in other predictive scoring models [22,45,46], describe socio-demographic information (e.g., age, gender), customer purchase history (e.g., relationship length, monetary value, product

possession), and customer–company contacts (e.g., number of consults, number of complaints). Second, the raw transaction records, similar to the fine-grained transaction data described in Section 2, contain around 60 million debit transactions by all customers, involving more than 1.5 million different merchants, over a one-year period. Both sources provide information only for the independent period as indicated on the timeline in Figure 4.

The dependent variables are four life events, rigorously collected by the financial services provider, because it regards them as the most important determinants of its business. These variables are binary indicators that receive a value of 1 if the life event was registered or 0 otherwise. The models predict such life events for the next 6 months (Figure 4). In our setting we create separate models for each life event. All customers that experienced the considered life event in the train set are considered as seed customers for the PSN. In Table 4, which summarizes the life events, *considered customers* refers to the number of customers who theoretically might encounter the life event; for example, only customers currently in a relationship could experience an *end of relationship* event.

Figure 4: Timeline for variable creation

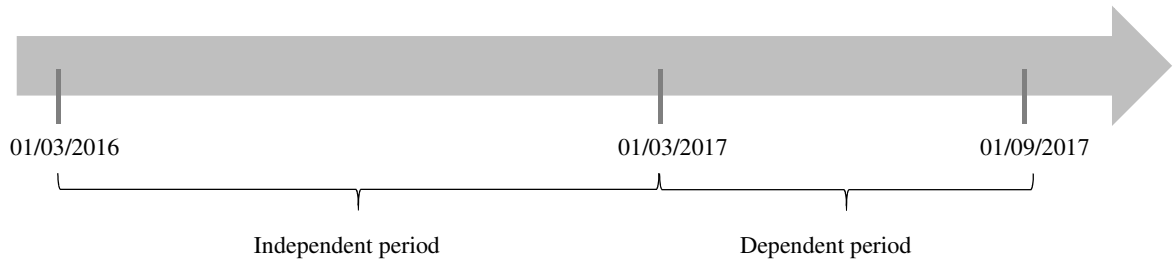


Table 4: Overview of dependent variables

Life moment	**Definition**	**# Considered Customers**	**Incidence**
Moving	Change in permanent address	132,703	9.44%
Birth	Birth of a child	132,703	1.64%
New relationship	Relationship change from single to married or officially living together	101,819	2.02%
End relationship	Relationship change from married or officially living together to	30,884	1.61%

	single		

3.2 Data preprocessing

The aggregated data preprocessing follows standards established by prior customer scoring literature [18,45]. The fine-grained transactional data contain anonymized transaction logs from payments by customers to merchants, which can be entered directly into the PSN and do not require further preprocessing.

First, we impute missing values, according to the specific characteristics of each variable. That is, the missing values might be imputed using zero, the median, or the modus [16]. The very fact that the variable is missing is potentially important information [47], so we use new dummy variables to flag these imputations. Second, outliers are extreme values that can distort predictive performance [48], so we use winsorization to transform them into less extreme values, within a three standard deviation range of the variable mean [49,50]. Third, we also create dummies for categorical variables, to encode the presence or absence of a particular category. The v unique values of the categorical variable are represented with $v - 1$ dummy variables [51]; every categorical variable thus transformed into dummy variables increases the total number of variables. Fourth, the incidences of the different life events range from 1.61% to 9.44%, so they are clearly imbalanced. Random undersampling [52] is frequently applied to remedy such imbalances for predictive modeling [53,54]. Practically, the number of customers of the majority class—that is, those without any change in the life event—gets reduced to the same level as the minority class—that is, customers who experienced the life event. Only the training data are altered; the test data must present the actual situation as realistically as possible.

3.3 Modeling

First, the five constructed models differ in the data used, as summarized in Table 5. The aggregated and fine-grained transaction data are modeled separately. As in previous research.

[23], the scores of the models that use both types of data are combined by a linear model, with the coefficients estimated using a validation set of the training data.

Second, all models are trained using logistic regression with forward variable selection. As one of the most popular algorithms for binary classification tasks, logistic regression supports many marketing applications [e.g. 16,45,55]; it combines comprehensibility with robust results and good predictive performance, often even better than more complex models [16,56]. A logistic regression estimates the interpretable posterior probabilities directly. For a binary dependent variable, the logistic regression estimates the probability $P(y=1|\mathbf{x})$ by:

$$P(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-(w_0 + \boldsymbol{w}\,\boldsymbol{x})}}, \tag{20}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an n-dimensional input vector (independent variables), $\mathbf{w}$ is the parameter vector (weights), and w_0 defines the intercept [57]. The calculation of the estimates for w_0 and $\mathbf{w}$ uses a maximum likelihood procedure.

The model that solely includes aggregated data is denoted mod_s. The models with fine-grained transaction data feature behavioral similarity scores based on the PSN (see Section 2). We select the best performing variant in terms of calculating behavioral similarity, using the performance of a validation set. The model *mod_psn* uses original PSN behavioral similarities, whereas *mod_rfm* uses Equations (17)–(19) to extend the PSN with RFM measures, combining these variables linearly according to the validation set of the training data.

Table 5: Overview of models

Models	**Data**		
	Aggregated Data	**Fine-Grained Transaction Data**	
		Binary	**RFM**
Mod_s	X		
Mod_psn		X	
Mod_rfm			X
Mod_s_psn	X	X	
Mod_s_rfm	X		X

3.4 Evaluation criteria

To measure the predictive performance of the different models, we use the area under the receiver operating characteristics curve (AUC) and top decile lift (TDL), as they are frequently applied to evaluate customer scoring models [e.g. 16,18,58]. The AUC is an independent performance measure of the discriminatory power of the predicted probabilities for the considered life event. For a binary classification, it provides a simple, one-value score between 0.5 and 1 [59]. The intuitive ranking offers a measure of posterior life event probabilities; in a CLEP context, it indicates the probability that a randomly chosen person with a life event is correctly ranked higher than a person without that life event. The TDL is a measure to assess the predictive performance of a model by expressing the predictive performance in the top decile as a number that indicates how much better (>1) or worse (<1) it performs than random guessing. It is commonly used to evaluate predictive performance in a binary classification setting [e.g. 45], ~~instead measures predictive performance for the top decile, relative to random guessing~~. Companies often focus only on some of their customer base, because it is too expensive to target all customers. From a managerial point of view, the TDL is highly relevant, because it pertains specifically to those customers who are most likely to experience a life event and thus the optimal ones to target [56]. We consider the top 10%, similar to prior literature [23].

To test whether observed differences in the performance measures across the models are statistically significant, the comparison of the classifiers uses an established testing framework [60]. The non-parametric Wilcoxon signed-ranks tests [61] rank performance differences between two classifiers and compare the ranks for positive and negative differences. We use this method to assess the pairwise differences among the models in this study.

4 Results and discussion

In presenting the results, we start with overall performance, then compare the predictive performance of the RFM-based behavioral similarity scores with that of the occurrence-based behavioral similarity scores. Third, we compare models that combine aggregated data and fine-

grained transaction data against models with single data sources, before detailing a meta-analysis to specify variations of behavioral similarity in the different models in the fourth part of this section.

4.1 Overall performance

The overall performance of the models can be compared with random guessing, which have, by definition, a value of AUC of 0.50 and a value of TDL of 1.Tables 6 and 7 list the average AUC and TDL values for the ten test folds, showing that all models perform better than random guessing for all considered life events; *birth*, *new relationship,* and *end relationship* each achieve AUC values of more than 71%. The combination of aggregated and fine-grained data produces a TDL of around 3. Predicting life event *moving* appears more difficult, according to our results, such that the predictive performance of *mod_s* for *moving* is considerably lower than that for the other three life events. From a managerial perspective, the fact that life events can be better predicted from data than random guessing can help to improve the customer experience through better service. In the financial services industry, companies can for example use these predictions to show the most relevant products first when a customer is browsing in its bank application. A more traditional way to use these predictions is for the bank advisor to propose a meeting when there is a high probability that the client will have a life event.

Table 6: Average AUC values (10-fold cross-validation) and standard errors

	Mod_s	**Mod_psn**	**Mod_rfm**	**Mod_s_psn**	**Mod_s_rfm**
Moving	0.633***	0.510***	0.656***	0.635***	<u>0.664</u>***
	(0.013)	(0.009)	(0.045)	(0.012)	(0.025)
Birth	0.725***	0.553***	0.667***	0.739***	<u>0.748</u>***
	(0.029)	(0.012)	(0.034)	(0.020)	(0.037)
New relationship	0.681***	0.542***	0.656***	0.712***	<u>0.730</u>***
	(0.035)	(0.019)	(0.036)	(0.016)	(0.039)
End relationship	0.690***	0.567***	0.640***	0.671***	<u>0.719</u>***
	(0.044)	(0.025)	(0.029)	(0.039)	(0.041)

Notes: The model with the highest average AUC over 10 folds is <u>underlined</u>.
The AUC of a random model is by definition equal to 0.50. Based on the Wilcoxon signed rank test, significant differences from the random model at the 90%, 95%, and 99% levels are indicated by *, **, and ***, respectively.

Table 7: Average TDL values (10-fold cross-validation) and standard errors

	Mod_s	**Mod_psn**	**Mod_rfm**	**Mod_s_psn**	**Mod_s_rfm**

Moving	1.866***	1.052*	1.391***	1.889***	<u>1.985</u>***
	(0.118)	(0.065)	(0.372)	(0.127)	(0.325)
Birth	2.880***	1.233**	1.544**	<u>3.097</u>***	3.000***
	(0.600)	(0.195)	(0.284)	(0.290)	(0.226)
New relationship	3.000***	1.133**	1.482***	3.167***	<u>3.242</u>***
	(0.516)	(0.142)	(0.252)	(0.290)	(0.267)
End relationship	2.509***	1.453**	1.737**	<u>2.851</u>**	2.799***
	(0.435)	(0.458)	(0.559)	(0.652)	(0.844)

Notes: The model with the highest average TDL over 10 folds is <u>underlined.</u>
The TDL of a random model is by definition equal to 1. Based on the Wilcoxon signed rank test, significant differences from the random model at the 90%, 95%, and 99% levels are indicated by *, **, and ***, respectively.

4.2 Fine-grained transaction data: RFM versus binary

The results show that *mod_rfm* consistently outperforms *mod_psn*, and the differences are significant for all assessments with the exception of the TDL for *end relationship*. When taking into account aggregated data, the positive effect of RFM logs over binary logs is mixed with the performance increase of the aggregated data. Nevertheless the AUC of *mod_s_rfm* is always significantly higher than the *mod_s_psn* for all life events. The performance measured by TDL for *mod_s_rfm* is significantly better for the *new relationship* life event, but the other TDL differences are not significant. The detailed Wilcoxon test statistics for these two analyses are in Table 8, and the results clearly confirm the value of incorporating RFM information into behavioral similarity scores derived from the fine-grained transaction data.

Table 8: Wilcoxon signed-rank test statistics over differences in performance

	Moving		Birth		New relationship		End relationship	
	AUC	TDL	AUC	TDL	AUC	TDL	AUC	TDL
Mod_rfm vs. mod_psn	0***	0***	0***	2**	0***	0***	0***	13
Mod_s_rfm vs. mod_s_psn	2***	14	16*	26	16*	15*	0***	9

*Significant at 90%. **Significant at 95%. ***Significant at 99%.

4.3 Combination of data sources versus single source

In comparing the performance of a model that combines both data sources (*mod_s_rfm*) with two single-source models (*mod_rfm* and *mod_s*), we find that the former performs better than a model

that features only fine-grained transaction data. These results are significant for all life events for both AUC and TDL, with the exception of the AUC for event *moving*. Furthermore, *mod_s_rfm* always performs better than *mod_s* too, and the differences in performance are significant for all life events in terms of AUC. However, the observed TDL difference between *mod_s_rfm* and *mod_s* is significant only for *new relationship*. Table 9 presents the detailed Wilcoxon test statistics, which clearly indicate that both data sources add some value for predicting life events and clarify that the fine-grained transaction data and aggregated data do not measure the same thing. Therefore, both data sources should be combined to achieve the best predictive performance.

Table 9: Wilcoxon signed-rank test statistics over the differences in performance

	Moving		Birth		New relationship		End relationship	
	AUC	TDL	AUC	TDL	AUC	TDL	AUC	TDL
Mod_s_rfm vs. mod_rfm	14	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}	0^{***}
Mod_s_rfm vs. mod_s	0^{***}	12	2^{**}	25	0^{***}	17^{*}	8*	15

*Significant at 90%. **Significant at 95%. ***Significant at 99%.

4.4 Meta-analyses

We conducted two meta-analyses, to determine the importance of the different data categories, as well as investigate the implications of the selected behavioral similarity variations.

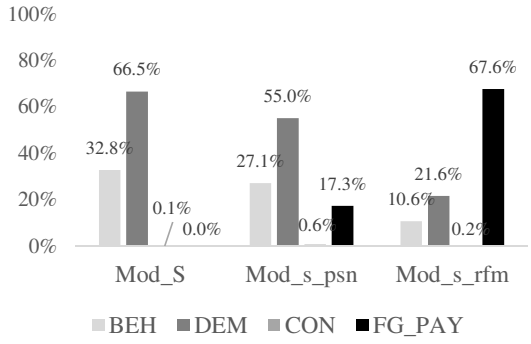
4.4.1 Variable importance

To determine the relative importance of the different variable categories, we use the absolute value of the Wald statistic for each individual predictor. The Wald statistic relates directly to the normalized β coefficients of the variables in the regression. Instead of a traditional logistic table, which can represent only a single model, we present the average importance of the aggregated variable categories in Figure 5, Panels a–d, across the multiple models in our ten-fold cross-validation experimental design.

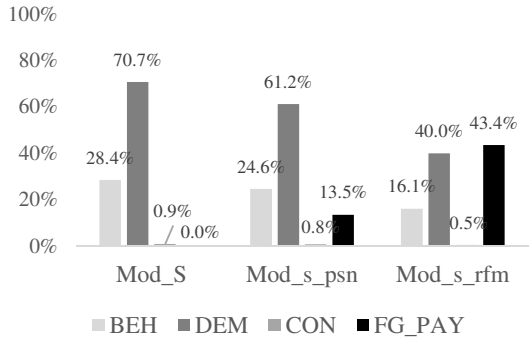
Among the three categories of aggregated variables (customer demographics, customer behavior, client–company contact), customer demographics are always the most important for predicting life events—a logical finding, in that customer demographics often feature variables that relate conceptually to life events. Customer behavior information also can help predict life events, with less importance. Yet in typical CRM applications, such as customer churn prediction, customer behavior is often the most important variable category [e.g. 18]. Finally, the variables that describe the contact between the client and company have the least importance. With regard to the fine-grained transaction data, we note substantial impacts in terms of predicting life events. They represent the most important source of information for moving, birth, and new relationships. The model that uses the RFM-extended behavioral similarity scores extracts far more predictive information out of the data, which then leads to heightened importance scores.

Figure 5. Relative importance of variable categories by model

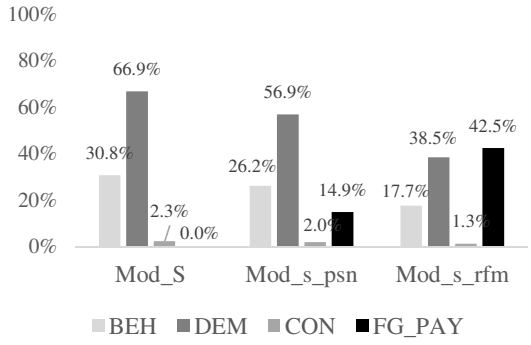
a. Moving life event



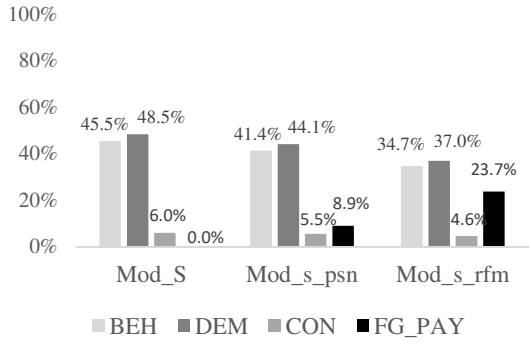
b. Birth life event



c. New relationship life event



d. End relationship life event



Notes: BEH = customer behavior, DEM = Customer demographics, CON = Client/company contact, FG_PAY = fine-grained transaction.

4.4.2 Behavioral similarity variants

By considering different approaches for calculating behavioral similarity, we derive counts of the selected behavioral similarity calculation variations per model, as depicted in Figure 6. The variations that adjust the weight of the counterparties are more frequently selected during the cross-validation. Then Figure 7, Panels a–d, presents the selected behavioral similarity variations per life event. Each life event indicates a dominant variation, which differs across events. For example, in *mod_s_rfm*, *moving* and *end relationship* are dominated by variation *b*, but for *birth* and *new relation*, the variation *icf* is preferable.

Figure 6: Count of selected behavioral similarity variations per model

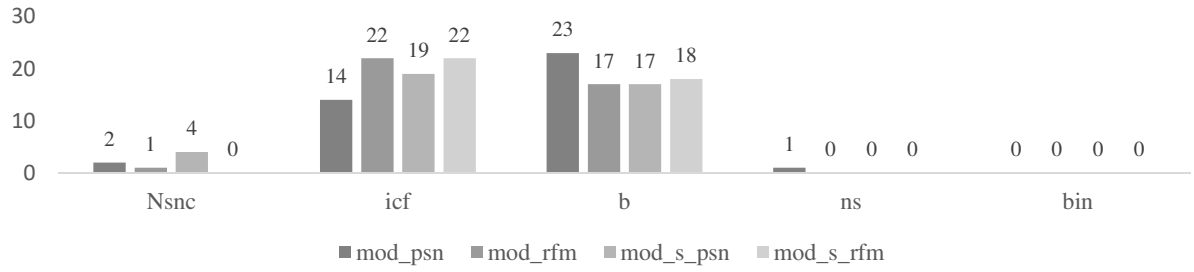
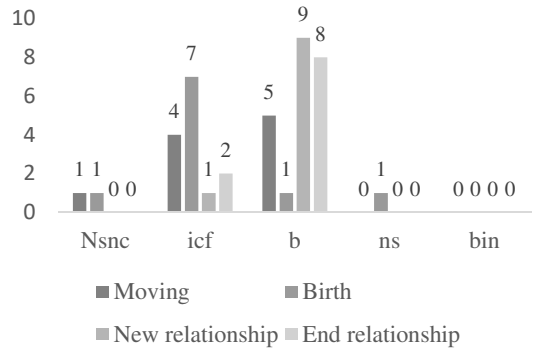
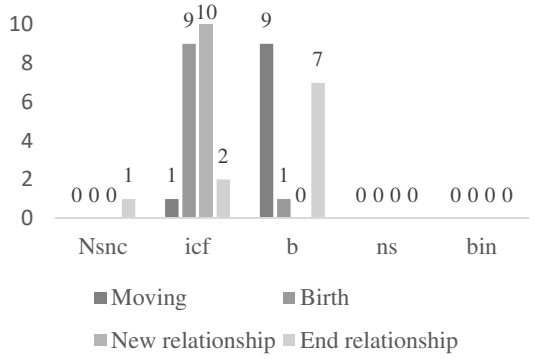


Figure 7: Count of selected behavioral similarity measures per life event

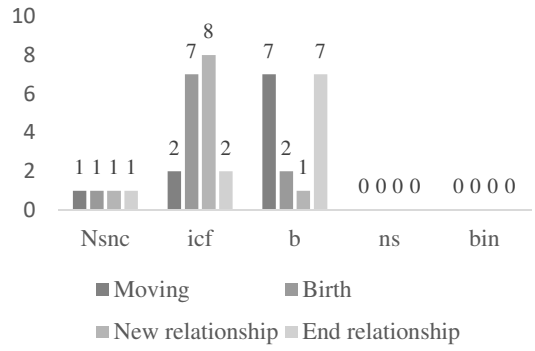
a. mod_psn



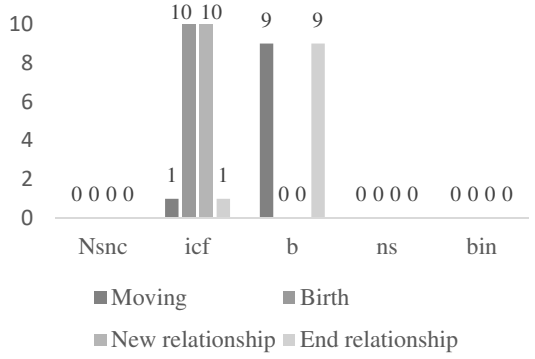
b. mod_rfm



c. mod_s_psn



d. mod_s_rfm



5 Conclusions

With this study, we sought to contribute to the business analytics research stream in decision support system literature in three main ways. First, we investigate CLEP in a real-world setting. In so doing, we demonstrate that life event prediction is feasible; all the life events we test can be predicted more accurately than random guessing. Therefore, these results should guide decision makers in their investments in life event prediction tools, as part of their broader CRM strategies.

Second, we propose and test an extension to a state-of-the-art method to incorporate fine-grained transaction data. Although PSN methods offer excellent performance in other applications [23], they also can be extended with behavioral similarity measures based on RFM variables obtained from payment transactions. The incorporation of RFM-based behavioral similarity measures significantly improves predictive performance. Third, by benchmarking models that incorporate different types of information, we derive insights about the importance of different variable categories. In particular, we show that optimal predictive performance requires a combination of aggregated and fine-grained transactional data. Practitioners thus can apply the guidelines that we establish herein to implement CLEP models, by leveraging fine-grained transaction data, customer demographics, and customer behavior as the most important data sources. Client–company contact information has less importance when it comes to predicting life events.

6 Limitations and future research

This study is the first to predict several life events using real-world data from the financial services industry and shows that the PSN can be successfully extended with behavioral similarity measures based on RFM, but it also features some limitations and directions for further research.

First, extending the behavioral similarity measures to account for RFM of payment transactions can improve predictive performance; further research also might consider alternative, innovative ways to calculate behavioral similarity or different ways to combine the RFM of the transactions to achieve even better improvements. Similarly, the results confirm the added value of fine-grained transactional data for predicting life events; other available sources of data might improve predictive performance even further. Because life events are personally significant, by definition, they likely are top of mind among customers, so communications with the company might be insightful, as they are for other applications [62].

Second, in demonstrating options for accurately predicting life events, this study is limited to only a portion of the potential CRM applications. For example, to understand how to target customers

about to experience specific life events, it would be interesting to investigate conversion rates for different offers. As demonstrated in uplift modeling studies [63], some customers might prefer not to be targeted by life event–specific campaigns. An interesting question involves whether their preferred targeting tactics differ across various life events.

Third, using real-world data has many advantages, but they often are difficult to obtain. We study the financial services industry, but other industries also might benefit from CLEP. In the telecommunications industry for example, customers' telecom needs might change after a life event, so the company might propose specific subscription formulas. Telecom operators also possess substantial, relevant data for such predictions and may have developed customer scoring models. Therefore, this sector would be an interesting context in which to test our proposed methodology. Also in the retailing industry, many applications could benefit from accurate life event predictions. Many retailers are already aware of the impact of life events on customer behavior, such as for example *Wallmart*, who dedicates a webpage to such events. Predicting life events might help to target the right customers with the right promotions, which could give a competitive advantage. Moreover, retailers possess store transaction data, which are a form fine-grained data and for which the methodology of this study could be used. In this regard, the legal context in the different industries should be considered, which can be different than in the financial services industry. Clients in the financial services industry have an obligation to provide life event for certain products for example. Especially in Europe, the storage and the use of personal data is strictly regulated. . Companies have to specify why certain data are collected and for which purpose; while customers have to formally accept this.

Fourth, our results demonstrate that predicting life events can be predicted using a combination of aggregated and fine-grained transaction data. However, as in any predictive model, the success of our models relies on good data quality. The real-world case study in the financial services industry that was discussed in this paper benefitted from rigorous data collection on life events, but in other industries accurate life event data might be difficult to obtain by companies. This could lead

to problems related to missing values and data accuracy, which in turn could compromise predictive performance. Therefore, further research could investigate the effect of missing values on the predictive performance of customer life event prediction and explore strategies to deal with such issues.

7 References

[1] W. Reinartz, V. Kumar, The mismanagement of customer loyalty, Harv. Bus. Rev. 80 (2002) 86–87.

[2] W. Reinartz, M. Krafft, W.D. Hoyer, The CRM process: Its measurement and Impact on performance, J. Mark. Res. 1 (2004) 293–305.

[3] W. Wells, G. Gubar, Life cycle concept in marketing research, J. Mark. Res. III (1966) 355–363. http://www.jstor.org/stable/3149851.

[4] M.C. Gilly, B.M. Enis, Recycling the family life cycle: A proposal for redefinition, Adv. Consum. Res. 9 (1982) 271–276.

[5] P.C. Verhoef, B. Donkers, Predicting customer potential value an application in the insurance industry, Decis. Support Syst. (2001). doi:10.1016/S0167-9236(01)00110-5.

[6] B. Wheaton, Life Transitions, Role Histories, and Mental Health, Am. Sociol. Rev. 55 (1990) 209.

[7] E. Lee, G.P. Moschis, A. Mathur, A study of life events and changes in patronage preferences, J. Bus. Res. 54 (2001) 25–38.

[8] A. Mathur, G.P. Moschis, E. Lee, A longitudinal study of the effects of life status changes on changes in consumer preferences, J. Acad. Mark. Sci. 36 (2008) 234–246.

[9] M. Solomon, G. Bamossy, S. Askegaard, M. Hogg, Consumer Behaviour : A European perspective Fourth Edition, 2009.

[10] E.C. Malthouse, Mining for trigger events with survival analysis, Data Min. Knowl. Discov. 15 (2007) 383–402.

[11] J. Li, A. Ritter, C. Cardie, E. Hovy, Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts, Proc. 2014 Conf. Empir. Methods Nat. Lang. Process. (EMNLP 2014). (2014) 1997–2007.

[12] A. Sen, A.P. Sinha, IT alignment strategies for customer relationship management, Decis. Support Syst. (2011). doi:10.1016/j.dss.2010.12.014.

[13] P.C. Verhoef, Understanding the effect of customer relationship management efforts on customer retention and customer share development, J. Mark. 67 (2003) 30–45.

[14] J. Peppard, Customer Relationship Management (CRM) in Financial Services, Eur. Manag. J. 18 (2000) 312–327.

[15] C. Lehrer, A. Wieneke, J. vom Brocke, R. Jung, S. Seidel, How Big Data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service, J. Manag. Inf. Syst. 35 (2018) 424–460.

[16] K. Coussement, S. Lessmann, G. Verstraeten, A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry, Decis. Support Syst. 95 (2017) 27–36.

[17] E.W.T. Ngai, L. Xiu, D.C.K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, Expert Syst. Appl. 36 (2009) 2592–2602.

[18] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, B. Baesens, New insights into churn prediction in the telecommunication sector: A profit driven data mining approach, Eur. J. Oper. Res. 218 (2012) 211–229.

[19] S. Lessmann, B. Baesens, H.-V. Seow, L.C. Thomas, Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, Eur. J. Oper. Res. 247 (2015) 124–136.

[20] K. Coussement, P. Harrigan, D.F. Benoit, Improving direct mail targeting through customer response modeling, Expert Syst. Appl. 42 (2015) 8403–8412.

[21] J. Ganesh, M.J. Arnold, Kristy E. Reynolds, Understanding the Customer Base of Service Providers: An Examination of the Differences Between Switchers and Stayers., J. Mark. 64 (2000) 65–87.

[22] D. Van den Poel, B. Larivière, Customer attrition analysis for financial services using proportional hazard models, Eur. J. Oper. Res. 157 (2004) 196–217.

[23] D. Martens, F. Provost, J. Clark, E. Junqué De Fortuny, Mining Massive Fine-grained Behavior Data to Improve Predictive Analytics, MIS Q. 40 (2016) 869–888.

[24] A. Knott, A. Hayes, S.A. Neslin, Next-product-to-buy models for cross-selling applications, J. Interact. Mark. 16 (2002) 59–75.

[25] M.J.A. Berry, G.S. Linoff, Data mining techniques For Marketing, Sales, and Customer Relationship Management. Second Edition, 2004.

[26] M. Khodabakhsh, M. Kahani, E. Bagheri, Z. Noorian, Detecting life events from twitter based on temporal semantic features, Knowledge-Based Syst. 148 (2018) 1–16.

[27] M. Kosinski, D. Stillwell, T. Graepel, Private traits and attributes are predictable from digital records of human behavior, Proc. Natl. Acad. Sci. 110 (2013) 5802–5805. doi:10.

[28] P.S. Fader, B.G.S. Hardie, K.L. Lee, RFM and CLV: Using iso-value curves for customer base analysis, J. Mark. Res. 42 (2005) 415–430.

[29] P.C. Verhoef, P.N. Spring, J.C. Hoekstra, P.S.H. Leeflang, The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands, Decis. Support Syst. (2002). doi:10.1016/S0167-9236(02)00069-6.

[30] D.L. Olson, B. Chae, Direct marketing decision support through predictive customer response modeling, Decis. Support Syst. (2012). doi:10.1016/j.dss.2012.06.005.

[31] T. Verbraken, F. Goethals, W. Verbeke, B. Baesens, Predicting online channel acceptance with social network data, Decis. Support Syst. 63 (2014) 104–114. doi:10.1016/j.dss.2013.08.011.

[32] A. Barfar, B. Padmanabhan, A. Hevner, Applying behavioral economics in predictive analytics for B2B churn: Findings from service quality data, Decis. Support Syst. (2017). doi:10.1016/j.dss.2017.06.006.

[33] S. Moro, P. Cortez, P. Rita, A data-driven approach to predict the success of bank telemarketing, Decis. Support Syst. (2014). doi:10.1016/j.dss.2014.03.001.

[34] A. McAfee, E. Brynjolfsson, Big data. The management revolution, Harvard Buiness Rev. 90 (2012) 61–68.

[35] S. Lian, Y. Xu, C. Zhang, Family profile mining in retailing, Decis. Support Syst. (2019). doi:10.1016/j.dss.2019.01.007.

[36] K. Coussement, D. Van den Poel, Improving customer complaint management by automatic email classification using linguistic style features as predictors, Decis. Support Syst. 44 (2008)

870–882. doi:10.1016/j.dss.2007.10.010.

[37] A. De Caigny, K. Coussement, K.W. De Bock, S. Lessmann, Incorporating textual information in customer churn prediction models based on a convolutional neural network, Int. J. Forecast. (2019). doi:https://doi.org/10.1016/j.ijforecast.2019.03.029.

[38] B. Baesens, V. Van Vlasselaer, W. Verbeke, Social Network Analysis for Fraud Detection, in: Fraud Anal. Using Descr. Predict. Soc. Netw. Tech., 2015. doi:10.1002/9781119146841.ch5.

[39] J. Moeyersoms, D. Martens, Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector, Decis. Support Syst. 72 (2015) 72–81. doi:10.1016/j.dss.2015.02.007.

[40] K. Coussement, S. Lessmann, G. Verstraeten, A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry, Decis. Support Syst. (2016).

[41] M. Meire, M. Ballings, D. Van den Poel, The added value of social media data in B2B customer acquisition systems: A real-life experiment, Decis. Support Syst. (2017). doi:10.1016/j.dss.2017.09.010.

[42] S. De Cnudde, D. Martens, T. Evgeniou, F. Provost, A benchmarking study of classification techniques for behavioral data, 2017. https://ideas.repec.org/p/ant/wpaper/2017005.html.

[43] D. Martens, F. Provost, Pseudo-social network targeting from consumer transaction data *, New York Univ. - Stern Sch. Bus. Work. Pap. CeDER-11-05. (2011) 1–31.

[44] K. Coussement, F.A.M.M. Van den Bossche, K.W. De Bock, Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees, J. Bus. Res. 67 (2014) 2751–2758.

[45] A. De Caigny, K. Coussement, K.W. De Bock, A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees, Eur. J. Oper. Res. 269 (2018) 760–772. http://www.sciencedirect.com/science/article/pii/S0377221718301243.

[46] E. Ascarza, P.S. Fader, B.G.S. Hardie, Marketing models for the customer-centric firm, in: Int. Ser. Oper. Res. Manag. Sci., 2017.

[47] W. Buckinx, D. Van Den Poel, Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting, Eur. J. Oper. Res. 164 (2005) 252–268.

[48] D.E. Jennings, Outliers and Residual Distributions in Logistic Regression, J. Am. Stat. Assoc. 81 (1986) 987–990.

[49] J. Freeman, D. Anderson, D. Sweeney, T. Williams, E. Shoesmith, Statistics For Business and Economics, 2nd ed., Cengage, 2010.

[50] D. Ghosh, A. Vogt, Outliers: An evaluation of methodologies, in: Jt. Stat. Meet., 2012: pp. 3455–3460.

[51] D. Pyle, S. Editor, D.D. Cerra, Data preparation for data mining, Morgan Kaufmann Publishers, San Francisco, 1999.

[52] G.M. Weiss, Mining with Rarity: A Unifying Framework, SIGKDD Explor. 6 (2004) 7–19.

[53] K.W. De Bock, D. Van den Poel, Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models, Expert Syst. Appl. 39 (2012) 6816–6826.

[54] C.X. Ling, C. Li, Data Mining for Direct Marketing: Problems and Solutions., Proc. 4th Int. Conf. Knowl. Discov. Data Min. (1998) 73–79.

[55] R.E. Bucklin, S. Gupta, Brand choice, purchase incidence, and segmentation: An integrated modeling approach, J. Mark. Res. 29 (1992) 201–215.

[56] S.A. Neslin, S. Gupta, W. Kamakura, J. Lu, C.H. Mason, Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models, J. Mark. Res. XLIII (2006) 204–211.

[57] D.W.J. Hosmer, S. Lemeshow, Applied logistic regression, 2nd ed., John Wiley, New York, 2000.

[58] A. Lemmens, C. Croux, Bagging and boosting classification trees to predict churn, J. Mark. Res. 43 (2006) 276–286.

[59] A.J. Hanley, J.B. McNeil, The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve, Radiology. 143 (1982) 29–36.

[60] J. Demsar, Statistical Comparisons of Classi ers over Multiple Data Sets, J. Mach. Learn. Res. 7 (2006) 1–30.

[61] F. Wilcoxon, Individual Comparisons by Ranking Methods, Biometrics Bull. 1 (1945) 80.

[62] K. Coussement, D. Van den Poel, Integrating the voice of customers through call center emails into a decision support system for churn prediction, Inf. {&} Manag. 45 (2008) 164–174.

[63] F. Devriendt, D. Moldovan, W. Verbeke, A Literature Survey and Experimental Evaluation of the State-of-the-Art in Uplift Modeling: A Stepping Stone Toward the Development of Prescriptive Analytics, Big Data. 6 (2018) 13–41. doi:10.