

This item is the archived peer-reviewed author-version of:

Loyal to your city? A data mining analysis of a public service loyalty program

Reference:

De Cnudde Sofie, Martens David.- Loyal to your city? A data mining analysis of a public service loyalty program
Decision support systems - ISSN 0167-9236 - 73(2015), p. 74-84
DOI: <http://dx.doi.org/doi:10.1016/j.dss.2015.03.004>
Handle: <http://hdl.handle.net/10067/1243480151162165141>

Loyal to your city? A data mining analysis of a public service loyalty program

Sofie De Cnudde^{a,*}, David Martens^a

^a*Department of Engineering Management, Prinsstraat 13, B-2000 Antwerp, Belgium*

Abstract

Customer loyalty programs are largely present in the private sector and have been elaborately studied. Applications from the private sector have found resonance in a public setting, however, simply extrapolating research results is not acceptable, as their rationale inherently differs. This study focuses on data from a loyalty program issued by the city of Antwerp (Belgium). The aim of the loyalty card entails large citizen participation, however, an active user base of only 20% is reached. Predictive techniques are employed to increase this number. Using spatial behavioral user information, a Naive Bayes classifier and a Support Vector Machine are used which result in models capable of predicting whether a user will actively use its card, whether a user will defect in the near future and which locations a user will visit. Also, a projection of spatial behavioral data onto even more fine-grained spatio-temporal data is performed. The results are promising: the best model achieves an AUC value of 92.5%, 85.5% and 88.12% (averaged over five locations) for the predictions, respectively. Moreover, as behavior is modeled in more detail, better predictions are made. Two main contributions are made in this study. First, as a theoretical contribution, fine-grained behavioral data contributes to a more sound decision-making process. Second, as a practical contribution, the city of Antwerp can now make tailored strategic decisions to increase its active user base.

Keywords: knowledge discovery, data mining, CRM, behavioral data, loyalty card

*Corresponding author. Tel: +32 3 265 50 14

Email addresses: sofie.decnudde@uantwerp.be (Sofie De Cnudde), david.martens@uantwerp.be (David Martens)

1. Introduction

Since the evolution from a product-centric to a customer-centric approach, managing customer relationships has become a necessity for organizations [1]. Customers need to be understood on a per-customer basis in terms of their needs, preferences and potential [2, 3]. Maintaining close ties to customers may lead to a competitive advantage and to customer loyalty [4]. In order for organizations to initiate and maintain connections to their customers, Customer Relationship Management (CRM) systems are used. Data mining forms an integral part of these systems and provides insight into raw customer data [5]. Consequently, organizations can closely monitor their customers and anticipate accordingly through proactive marketing such as offering personalized incentives [5]. An important source for CRM data constitutes the output of loyalty programs and more specifically from loyalty cards [6]. Loyalty card data is typically characterized by a large sample of transactional data connected to the customers, for which also socio-demographic data is available [7].

The analysis of loyalty programs has been largely researched in literature related to the profit sector, especially in retail [8]. In contrast, the behavior of customers in a non-traditional customer-supplier relationship such as the relationship between a customer and the government has not received that kind of attention [2]. The results from analyses performed in the profit sector cannot simply be extended to the non-profit sector due to their differing rationale. To our knowledge, the analysis of loyalty card data in the context of a public institution has not yet been researched. This study performs an analysis of such data and thereby is a first attempt to contribute to this gap. We state the contributions as follows: (1) gaining insight into the behavior of users of a government loyalty card and (2) attempting to learn non-trivial user information from this behavior with the help of data mining techniques. Both should contribute to the design of a decision support system able to lead to high-quality information for decision makers.

Concretely, in this design science paper [9], we analyze the behavior of users of the *A-Card*, a loyalty

25 card issued by the city of Antwerp (Belgium) that can be used in public institutions and partnering
institutions such as libraries and musea. The purpose of the card consists of promoting participa-
tion in cultural services offered by the city. The data consists of 4 million transactions concerning
approximately 177,000 persons visiting one of 102 locations. Currently, there is no widespread use of
the benefits that come along with the loyalty card. Moreover, users visit a relatively unilateral set of
30 locations. This research presents an analysis of this government data that should empower the city to
take appropriate strategic decisions.

The rest of this paper is organized as follows. Section 2 discusses relevant background and previous
related work. Section 3 gives a detailed description of the data, the used data mining techniques and
the evaluation criteria. Next, in Section 4, the experimental set-up of the research is presented. The
35 results of the analysis are presented and discussed in Section 5. Lastly, Section 6 concludes our work
and presents avenues for further research.

2. Background and related work

2.1. Customer Relationship Management (CRM)

In the 1990s, organizations' customer focus shifted from a transactional view to a relational view
40 [10]. Customer Relationship Management (CRM) has since become a pivotal part of organizations. An
understanding of customers' preferences may lead to the possibility of tailoring products and services
to their wishes, which leads to higher customer satisfaction. This subsequently results in behavioral
customer loyalty [11] and ultimately in profit. In order for these CRM systems to be of any value for
the organization, data has to be collected and techniques have to be available to analyze this data
45 [12]. Loyalty programs are one way of tying customers to an organization through incentives [4]. They
also are a source for capturing behavioral patterns of the users which results in a large amount of
transactional data.

2.2. High-dimensional, behavioral data

Since data collection and storage have become cheap, data is being gathered in vast amounts
50 [13, 14]. This data often consists of fine-grained behavioral data, such as click behavior of website
visitors, payment transactions of a client of a bank or locations visited by a mobile phone user. These
fine-grained features are testimonies of an individual’s behavior. Using this data, predictions can be
made concerning individual users as to whether they would be interested in a certain news article [15],
a certain banking product [16] or a mobile ad [17]. An application in the public sector entails fraud
55 detection in companies using their payment data [18, 19]. These big high-dimensional feature sets lead
to more complexity in the analysis [20], for which traditional dimensionality reduction techniques are
not relevant given the low redundancy of the features. Traditional summarizing features such as socio-
demographic variables contain more information on a per feature basis in comparison to behavioral
features. The latter, however, each gradually add more information which leads to a more informative
60 whole [21].

2.3. Analysis techniques in CRM

Once the data is collected, techniques have to be identified to extract information and knowledge.
Data mining techniques are techniques that are able to extract patterns from the data which can
subsequently lead to vital knowledge for the organization [22]. Ngai, Xiu and Chau performed a
65 comprehensive literature review of data mining studies in the CRM field [8]. Here, the CRM field is
subdivided into four stages: identification, attraction, retention and development of customers. The
bulk of the research is positioned in the customer retention phase, and more specifically in personalizing
marketing campaigns dependent on user’s behavior (one-to-one marketing) and with the goal of tying
customers to the organization (loyalty programs). The techniques mostly used in these studies consist of
70 neural networks, decision trees and association rule techniques. Applications of data mining techniques
in the context of CRM can be found in different sectors such as retail, the banking sector and telco

companies [22]. In retailing, a typical application is market basket analysis in order to determine which products are frequently bought together. The banking sector is mainly interested in segmenting its customers in order to focus their marketing efforts towards the right customers. Another important
75 application is fraud detection where suspicious transactions are identified. In the telecommunications sector, churn prediction is of vital importance due to the highly competitive environment.

2.4. CRM in a public setting

The public sector deals with distinct challenges with respect to its private counterpart, among others providing basic services to its citizens. Lately, their modernization process has resulted in them
80 adopting a more privatized character [23] along with private-like principles such as a focus on efficiency, competitiveness and profitability [24]. As time evolves, this transformation will result in applications originating from the private sector, migrating towards the public sector [25]. CRM applications in the public field, referred to as e-government, have until now mostly focused on providing equally accessible, transparent and omnipresent public services [26], on citizen engagement [27] and on policing [28]. To
85 this end, diverse data can be used such as citizen feedback and complaints, social media data [29] and government data from legacy systems. The data traditionally used in big data government applications, however, is mostly textual [30]. Introducing behavioral data of some sort could open a new avenue of research and lead to surprising results, as it did in advertising research.

Despite some resemblances between the public and the private sector, two major differences have
90 to be kept in mind. First, the 'customer' base of public services is diverse and dynamic [31, 32]. Namely, the 'customer' base encompasses all citizens that belong to the governing institution. The customer base in the private sector only is a portion of this citizen base. Thus, the data collection opportunities are intrinsically larger and more diverse [33]. Secondly, the 'customers' cannot be valued based on profitability. In case of CRM in a government setting, customers must not only be served on
95 an individual basis, an extra equality constraint is introduced [2]. Although the principles of CRM can

be applied across sectors and industries [34, 35], care must thus be taken as to whether results from research in a private setting can be applied to its public counterpart [35]. A government-issued loyalty card cannot be evaluated solely on its financial benefits for the institution. The element of so-called public value has to be taken into account [36].

2.5. *Relevance*

Underpinning the relevance of our study are the following arguments. First, the public sector is an omnipresent institution in everyday life and offers its services to all aspects of society with a diverse set of stakeholders with whom unique relationships are held, i.e. no profitability element is involved [24]. This distinct character results in interesting research opportunities. In the past, research concerning strategy [31], management practices [37], information technology [38] and information systems [39] has been conducted in the public sector and has been compared with practices in the private sector. Extending this research effort to loyalty programs may result in interesting discoveries. Secondly, the use of Customer Relationship Management in the public sector is quite unilateral at present. Using fine-grained behavioral data has proven to lead to good predictive results [21] and has been employed in advertising research [17, 15, 16, 40]. Hence, investigating whether and to what extent these results are feasible with behavioral data originating from the public sector may give rise to interesting research conclusions. Finally, concerning the overall relevance of this work with respect to the government's cultural responsibility, cultural participation can help connect individuals and lead to civic engagement and social cohesion [41]. One of the goals of the *A-Card* is to encourage people to take part in the cultural life of the city of Antwerp and to this end analyze the behavior of users visiting public institutions in order to tailor the cultural offer. Getting insight into this behavior may thus lead to benefits for the city as a whole.

3. Research methodology

3.1. The city of Antwerp loyalty card: A-card

120 The *A-Card* was introduced on 18 December 2010 by the city of Antwerp (Belgium). Antwerp is the most populated city in Belgium with approximately 510,610 inhabitants [42]. With the *A-Card*, the city wishes to inform its citizens about its cultural offers, simplify access to cultural venues and convince people to visit cultural venues they have not yet visited. Overall, the goal is to promote participation of inhabitants and visitors in cultural activities organized by the city of Antwerp and for
125 the city to brand itself as attractive for people who live in, work in or visit Antwerp [43].

Inhabitants and visitors of Antwerp can register for an *A-Card* and use it when visiting certain public venues or venues from partnering institutions. The card can be used in 102 different places located over the city, that are grouped in 9 categories i.e. libraries, swimming pools, musea, cultural centra, the city shop, events, meeting centra, youth centra and digital registration events. The loyalty
130 card costs three euros and is currently only required when lending items in a library. At registration time, users provide some mandatory socio-demographic background and contact information, i.e. name, address, date of birth, gender, nationality, e-mail address, etc. Optionally, they may also provide some extra information regarding their personal interests or the number of members in their family. At registration time, each user immediately receives five starter advantages that can be used in each of
135 the following locations: library, swimming pool, museum, cultural centre and the city shop. Each subsequent visit yields a point, given that the card is scanned at a so-called *A-Card*-pillar. These pillars are located in the entrance hallway of the venues and visitors scan their card upon entrance. After collecting 10 points, users can cash these out in return for an advantage such as a discount on an entrance ticket. Points collected from 1 January to 31 August expire on the first of January the next
140 year. Points collected from September to December remain active during the next year and expire along with points collected from January to August of that next year. During the last week of October

and the first week of November, subscribers are reminded of this imminent expiration through two action weeks. The goal is to encourage users to cash out their collected points by offering them extra advantages during these weeks such as discounted or free entrance tickets.

3.2. Data characteristics

The data originating from the *A-card* consists of the visiting behavior of 177,761 users, which in total amount to 4,015,091 transactions that took place between 18 December 2010 and 5 February 2014. During 2011, the standard library card was replaced with the *A-Card*, which resulted in a high number of registrations that year, i.e. 50.20% of all users registered then. Table 1 shows these and some other general statistics regarding the loyalty card data. Most transactions take place on Wednesday, while Sunday and Monday are less busy. This can be explained by the fact that school ends at noon on Wednesdays and that most public institutions are closed or only open for half a day on Sunday and Monday. October, November, December and January are busy months. This can be explained by the action weeks that take place during October and November and potentially also the renewal of library subscriptions during January. During holiday periods, such as April, July and August, the transactions experience a low.

3.3. Problem introduction

The city of Antwerp would like its *A-Card* users to actively use their card and cash out the offered benefits on a regular basis. However, 56.55% of all users have never taken advantage of those loyalty benefits. Moreover, the active user base, i.e. users collecting and cashing out points on a regular basis, only consists of 18.62% of the total user population. Figure 1 shows the users that have enough points to cash out, the so-called potentials, and the users that actually cash out every month. The graph shows the number of potentials rising during 2011 and 2012 and, then, in January 2013 experiencing a significant drop. This drop is attributed to the expiration of points collected from the start until

Total number of users		177,761	Transactions per weekday	Monday	8.97%	
Total number of transactions		4,015,091		Tuesday	19.12%	
Number of users registered	in 2010	1.75%		Wednesday	21.04%	
	in 2011	59.25%		Thursday	15.08%	
	in 2012	32.19%		Friday	13.05%	
	in 2013	20.22%		Saturday	14.39%	
Usage by	inhabitants of Antwerp	85.43%		Sunday	8.35%	
	visitors from outside Antwerp	14.40%		Transactions per month	January	6.92%
	foreign tourists	0.14%			February	7.72%
Gender (female)	54.63%	March			8.66%	
Users with age between	1-10	1.05%	April		7.85%	
	11-20	21.71%	May		8.61%	
	21-30	14.88%	June		8.09%	
	31-40	13.57%	July		7.49%	
	41-50	11.31%	August		8.22%	
	51-60	10.08%	September	9.39%		
	61-70	8.94%	October	12.18%		
	71-80	5.57%	November	8.54%		
	81-90	2.22%	December	6.33%		

Table 1: General characteristics of the loyalty card data.

165 August 2012 (the city of Antwerp decided not to let points expire in January 2012 due to it being the first year). After the expiration of points, the potentials group starts rising again until a new imminent expiration. More importantly, Figure 1 makes clear that a minority of users having collected enough points cash out these points. Possible explanations for this might be that users are not informed about the cashing-out process, that they do not know what benefits the city offers them or that these benefits do not appeal to them. With respect to loyalty programs in the private sector, Vittal et al. [44] report an average of 35% of program users who make use of their rewards. We have not found a benchmark in literature which states that a ratio of 20% active users equals too low a program participation. Moreover, the loyalty card has only been in place for three years. However, we believe it is safe to state that this percentage, along with the findings in Figure 1, are not highly satisfying, considering 175 the context of this particular citizen loyalty program.

Figure 2 shows the visiting behavior of subscribers of the *A-Card* in terms of the number of locations and the number of *types* of locations one visits. Nearly 45% of all users only visit one location, while almost 50% also only visits one type of location. The visiting behavior of most users is thus quite unilateral. This finding is not satisfactory, as the city of Antwerp wishes to get people to visit a diverse

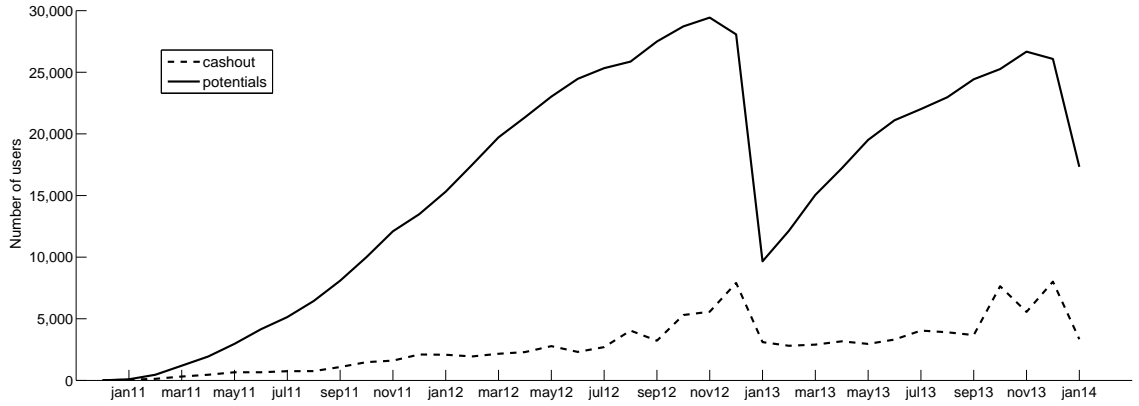


Figure 1: Number of users who have collected enough points to cash out (potentials) and number of users who actually cash out their loyalty points.

set of cultural venues. Figure 3 also confirms this finding via the degree distribution for the users and the locations. The degree distribution for the users is clearly skewed. Most users only visit few locations. Its counterpart degree distribution for the locations shows that the same holds, but since only 102 locations are used, the distribution does not demonstrate the same skewness with respect to that of the users.

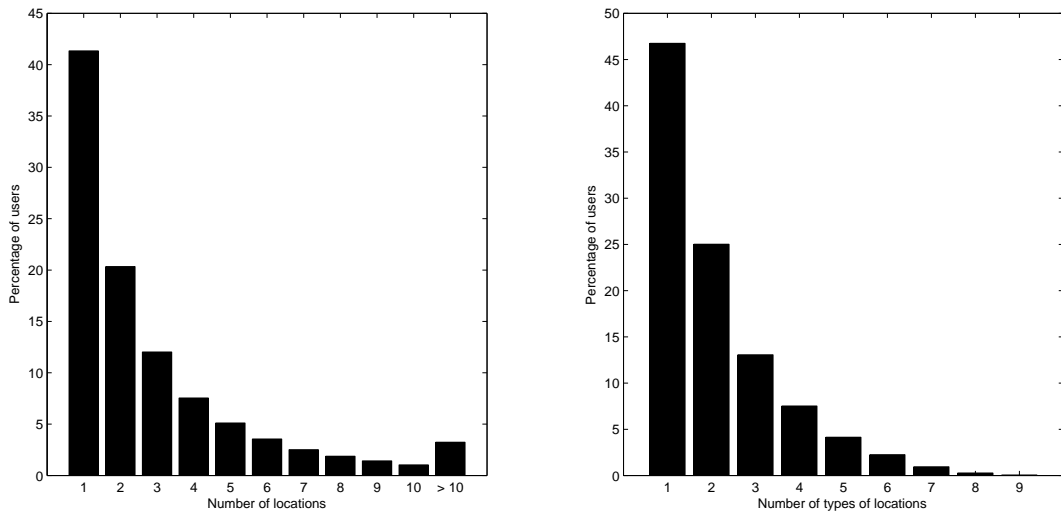


Figure 2: Number of locations (left) and number of types of locations (right) visited by users.

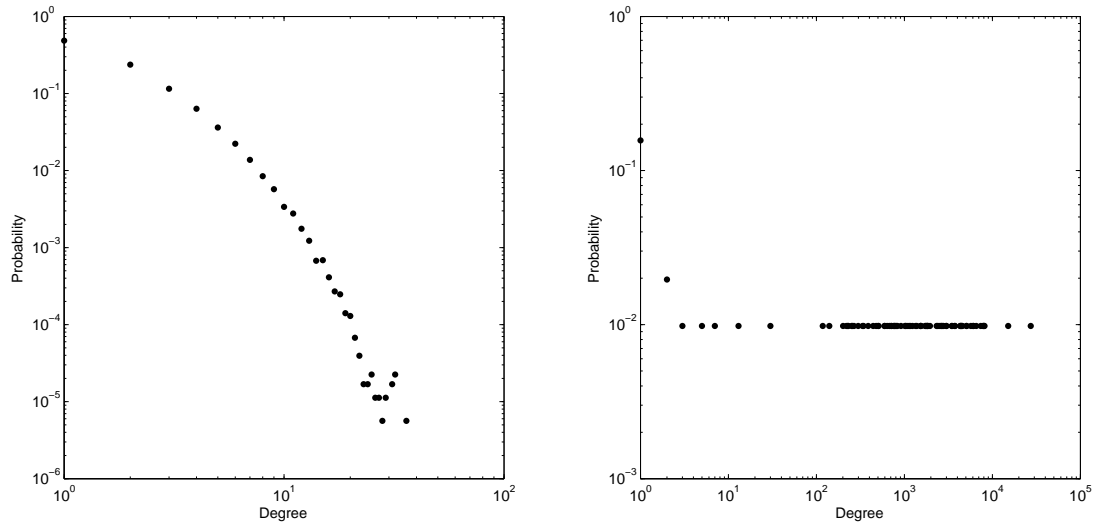


Figure 3: Degree distribution for users (left) and locations (right).

In conclusion, the city of Antwerp presently has a loyalty card that does not quite reach the objectives as stated in the mission statement. Figure 4 schematically demonstrates where strategic decisions should be made related to the two above mentioned problems. In order to achieve active use of the benefits of the loyalty card, in the first place, users must be encouraged to become active users (cashout their collected points). Secondly, users should be identified in time as likely to become inactive and consequently must be re-engaged (defect). At last, when a user has reached the active status, he should be retained by visiting (diverse) locations (visit locations, diverse locations). This decision-making process can be facilitated by a decision support system. The goal of this sytem is to provide insight into unstructured problems, represented by raw data [45]. These insights should then lead the way to high-quality decisions. In this work, we build a predictive and a descriptive model that will serve as the decision support system. Concretely, the objectives of this study are stated as follows: (1) build a predictive model able to predict which users make use of the benefits and which users do not (cashout behavior), (2) build a predictive model able to predict future visiting behavior (visit locations), (3) build a descriptive model able to identify locations frequently visited together

(visit diverse locations), (4) build a predictive model able to predict which users are likely to become inactive (defect behavior).

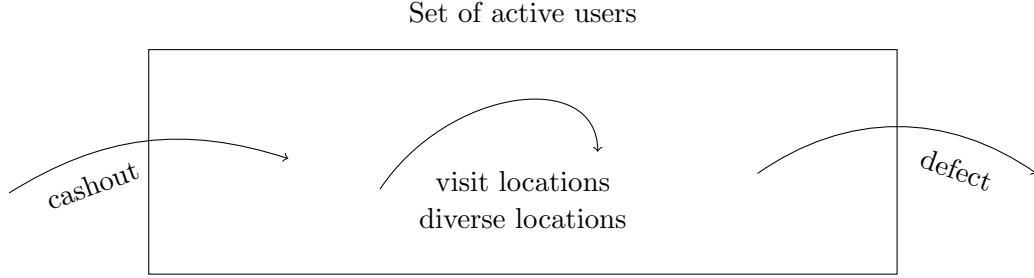


Figure 4: Schematic overview of where predictive models can help to get people in, retain them in and prevent them from leaving the active set of users.

200

3.4. Data mining techniques

Of the top ten algorithms in data mining [46], the relevant ones in this case are C4.5, Support Vector Machines, AdaBoost, k-nearest neighbor classification, Naive Bayes classification and CART. Regarding the fact that our problem deals with high-dimensional data, tree classifiers are not very
 205 suitable [47]. AdaBoost might overfit the training data in the presence of noise. In the light of big data in both dimensions, k-nearest neighbor classification does not scale with respect to execution time. Support Vector Machines (SVM) and Naive Bayes classifiers (NB) are two appropriate choices able to deal with the big dimensions of the data. We wish to state that we do not claim to have used the two best performing classification techniques. We believe, however, that we have selected two appropriate
 210 techniques taking into account the specifics of this study.

Both models take as input a data set of the following form: (\mathbf{x}_i, y_i) , with $(i = 1, \dots, n)$, $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \{-1, +1\}$, n the number of data points and m the number of features of each data point. This data set is called the training set. The classifier builds a predictive model, which is then used to predict y -values of data points in a test set.

The Naive Bayes (NB) classifier [48] uses Bayes' rule with naive assumptions to build a predictive

model. Bayes' theorem is defined as follows:

$$p(y_i|\mathbf{x}_i) = \frac{p(y_i)p(\mathbf{x}_i|y_i)}{p(\mathbf{x}_i)}. \quad (1)$$

Since the denominator is not dependent on the class variable y_i , it is not taken into account. Then, making use of the naive assumption that the features are mutually conditionally independent, the above equation can be rewritten as follows:

$$p(y_i|\mathbf{x}_i) \propto p(y_i) \prod_{j=1}^m p(x_{i,j}|y_i). \quad (2)$$

215 Data points from the test set are classified by selecting the class with the highest probability as defined in Equation 2. Despite the naive assumption, the NB classifier is able to build predictive models in a time-efficient manner providing competitive results with respect to more advanced classifiers.

The more complex Support Vector Machines (SVM) [49] project data points into a high-dimensional feature space with the help of a kernel function. In that space, the SVM looks for a maximal margin hyperplane that optimally separates the data points. This hyperplane takes the following form:

$$H = \mathbf{w}^T \mathbf{x} + b = 0, \quad (3)$$

with \mathbf{w} a weight vector for the features of the data points and b a bias. Concretely, the goal for the hyperplane is to maximally separate the data instances of the two distinct classes that are nearest to
220 the hyperplane. These data points are called support vectors. Maximizing this distance leads to higher generalizability and to lower variance. The Support Vector Machine finds the hyperplane by solving the following optimization problem:

$$\min_w \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i, \quad (4)$$

$$\text{s.t.} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad (5)$$

$$\xi_i \geq 0, \quad (6)$$

with C a penalty parameter that represents a trade-off between complexity and error rate and ξ_i ($i = 1, \dots, n$) slack variables that allow misclassifications. The Linear SVM (L-SVM) uses a linear kernel function to transform the data points, which results in a lower execution time compared to a non-linear-kernel SVM. The L-SVM is also considered more appropriate for large dimensional datasets [50]. We use L2-regularization and thus the loss function is defined as follows:

$$L(\xi_i) = \max(0, 1 - y_i \mathbf{w}^T \mathbf{x}_i)^2. \quad (7)$$

L2 regularization is chosen here due to the nature of the features i.e. fine-grained and low redundant.

Predictions are now made by the L-SVM by determining the side of the hyperplane the test points

are. Although faster than the non-linear variant, the Linear SVM runs slower than the Naive Bayes classifier.

In trying to identify locations frequently visited together, the well-known and widely used Apriori algorithm is used [51]. The algorithm looks for frequent item sets in a transactional database, where a transaction equals a collection of items such as items bought together or applied to our case, locations visited together. Then, the algorithm finds rules of the form $X \rightarrow Y$, with X and Y item sets. Two constraints are used while looking for these item sets and rules. The first is the support of an item set, $support(X)$, and is defined as the ratio of transactions containing the items in X . The confidence of a rule, $confidence(X \rightarrow Y)$, is defined as $support(X \cup Y)/support(X)$ and represents the strength with which one can state that if X is present also Y is present in a transaction.

3.5. Evaluation criteria

Accuracy is a fairly intuitive measure of performance: it expresses the percentage of correctly predicted instances, i.e. the ratio of true positives and true negatives. However, this measure is known to be influenced by class imbalance [52]. If 1 positive case occurs in a set of 100 instances, a model predicting all instances as negative has an accuracy of 99%, despite the fact that the model does

not make use of any information in the data. Accuracy can thus not be used as conclusive evidence regarding the model’s performance. Therefore, the AUC measure is used to determine the predictive performance of the models: the Area Under Receiving Operator Curve [53]. This measure is insensitive to class imbalance. The AUC expresses the ability of the model to rank the instances in a descending fashion in terms of their prediction score and thus represents the probability of a classifier to rank a randomly chosen positive instance higher than a randomly chosen negative instance. An AUC of 50% corresponds to a model performing no better than random.

The problem environment, being the public management area, implies that not only predictive power is important, an insight into the predictions made by the model also is relevant. Therefore, confusion matrices for the test sets are derived. A confusion matrix is a 2×2 -matrix of which Table 2 shows the form. The confusion matrix shows the possible classifications of instances in a test set. A positive instance is preferably classified as positive (TP), otherwise it is referred to as a false negative (FN). Analogous for the negative test instances, when correctly classified, the instance is referred to as a true negative (TN), otherwise it is called a false positive (FP). A good model preferably has a large amount of true positives and true negatives, but a trade-off is inherently present.

When comparing the AUC results achieved for each technique and for all datasets, a statistical significance test has to be used to ascertain whether the difference in performance results is significant. We employ ten-fold cross-validation, thus the AUCs are stabilized over different runs of the algorithms. The Wilcoxon signed rank test [54] is then used to compare the performance of two classifiers. As we

		True class	
		+	–
Predicted class	+	True Positive	False Positive
	–	False Negative	True Negative

Table 2: Confusion matrix.

project spatial behavior onto spatio-temporal behavior, more detailed analyses are needed for which

260 the Friedman test is used followed by the Nemenyi post-hoc test [54]. For further detail regarding these statistical significance tests, we refer to Demšar’s work [54].

4. Experimental set-up

4.1. Data set

Two choices are made when transforming the transactional dataset into a behavioral dataset. Modeling users’ visiting behavior can be done using the frequency of the visits or a weighted frequency measure. However, the choice was made to build binary matrices, omitting any frequency information. If this would not lead to good predictions, subsequently frequency information can be included. A second choice concerns the decision related to the time period along which user behavior is observed. It is assumed that the longer one goes back in time, the more irrelevant that action becomes in the present. Imagine a location visited as a child and how irrelevant it would be to use it to predict the next location visited by the same person as an adult. Concretely, we first transform the transactional data set into a sparse behavioral binary input matrix. The matrix is sparse due to the limited behavioral capital [21]. This denotes the fact that there are only so much actions a person can take from the entire set of possible actions. This matrix X is a $n \times m$ -matrix with n the number of users (177,761) and m the number of features, here locations (102). Each element $x_{i,j}$ is defined as follows:

$$x_{i,j} = \begin{cases} 0 : & \text{user } i \text{ has not visited location } j \text{ in a certain time period } t, \\ 1 : & \text{user } i \text{ has visited location } j \text{ in a certain time period } t. \end{cases} \quad (8)$$

Since a student will probably visit locations during off-school hours while an unemployed user may
265 visit venues during work hours, it is apparent that the time of day when people visit locations can be informative when trying to make predictions. Since more behavioral features lead to better predictive results [21], breaking down mere spatial behavioral information into spatio-temporal behavioral

information might result in more sound predictions. Hence, in a second set of finer grained behavioral data matrices, different time dimensions are taken into account. With increasing granularity, these

270 are:

- whether the location was visited during work hours (from 08.00-17.00 on Monday until Friday),
- whether the location was visited during school hours (from 08.30-16.00 on Monday, Tuesday, Thursday and Friday and from 08.30-12.00 on Wednesday),
- whether the location was visited in the week or the weekend,
- 275 • what day of the week the location was visited,
- during which hour of the day the location was visited.

Table 3 shows an example of a binary behavioral input matrix and its projection in terms of the week granularity level. One binary behavioral feature is projected onto seven day variables.

4.2. Predicting target variables

280 Regarding the second choice to be made with respect to transforming the transactional dataset into a behavioral one, we state that the underlying premise of this study encompasses the fact that past behavior is predictive for future behavior [55, 56]. However, since a person changes over time, the further one goes back in time, the less relevant this behavior is for the person he is now. Hence, we include the visiting behavior of users only from the past year. More concretely, we use the visiting
285 behavior of users from January to September and make predictions regarding their cashout, visit and defect behavior for the period of October to December. Concretely, a user is labeled as cashing out points or visiting a location, if he cashes out points or visits that location in the period from October to December (Figure 5).

Regarding location prediction, five locations are selected for which to predict whether a user will
290 frequent it i.e. the Permeke library, the Wezenberg swimming pool, the museum MAS, the Roma

	Permeke library	MAS Museum	...
Ann	1	0	...
Bill	0	1	...

↓

	Permeke (Monday)	Permeke (Tuesday)	...	Permeke (Sunday)	MAS (Monday)	...
Ann	1	1	...	0	0	...
Bill	0	0	...	0	1	...

Table 3: Example of a binary behavioral input matrix and its projection onto the temporal behavioral variant using the week granularity level. One binary behavioral feature is projected onto seven day variables. Ann’s visit to the Permeke library in the binary variant is projected onto more detail in the week variant, i.e. Ann visited the library on both Monday and Tuesday.

cultural centre and the Zoo of Antwerp. In this way, we believe we have a diverse set of types of locations that are also visited on different scales by different sets of users. The Permeke library and the Wezenberg swimming pool are the most visited locations: i.e. 13.97% and 7.35% respectively of all transactions involve these venues. The Permeke library and the MAS museum are the top two locations when it comes to the number of unique visitors, i.e. 32.70% and 19.14%. In order to check the predictive value for locations not that frequently visited, also the Zoo and the Roma cultural centre are included. For the case of predicting whether a person will become inactive, a user is labeled as such in an adaptive fashion. This seems an intuitive approach considering that some people might visit

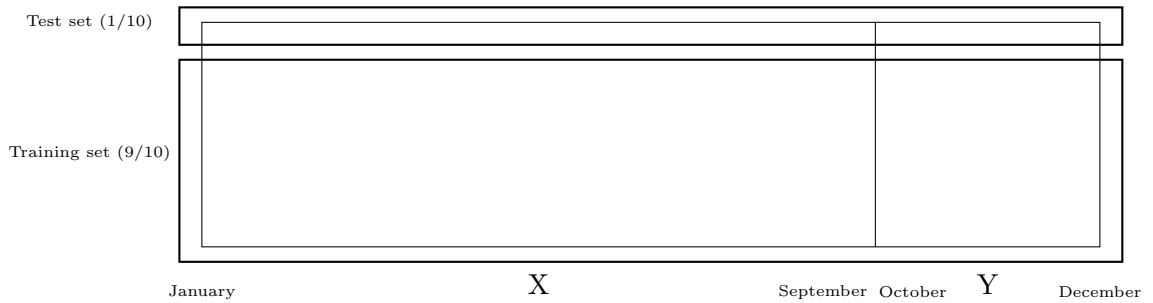


Figure 5: Visualization of the predictive modeling procedure.

the library every week and others might lend books only once every three months. While one month
 300 might be too long a period to wait for the first user, sending re-activation e-mails after one month of
 inactivity might be perceived annoying by the second user. Figure 6 shows the number of users that,
 after a certain period of inactivity (shown along the x-axis), become active again. As can be seen,
 there are users becoming active again after a period of 125 days (approximately three months) of being
 inactive. Therefore, for each user, his/her average period of inactivity is used to determine whether
 305 he/she is likely to become inactive. The prediction is thus tailored to each individual user and his
 visiting behavior. Considering Figure 6, a period of three months for determining whether a user is
 going to churn is too short. Therefore, the target variable must be determined over a longer period. We
 opt for a period of six months. A second factor to deal with consists of the action weeks. Considering
 users are then persuaded to use their card more frequently, the prediction might be biased. Thus we
 310 choose the six months preceding these two weeks: from May until October 2013. The prediction is
 then based on the visiting behavior in the year preceding this period: from April 2012 to April 2013.

4.3. Set-up of the data mining techniques

We use ten-fold cross-validation when evaluating the predictive techniques. The data set is ran-
 domly divided in 90% training set to build the predictive model, which is subsequently tested on the
 315 remaining 10% of the data set. This procedure is then performed ten times in order to obtain rigor
 in our evaluation and to not overfit the model to a particular training set [57]. We report the average
 performance results of these runs. From the training set, a 10% validation set is derived to tune the
 C parameter of the Linear SVM, based on the best AUC value.

The majority of the data sets demonstrate class imbalance as can be seen in Table 4. Some of these
 320 high class imbalances lead to poor classification performance when using the Linear Support Vector
 Machine. In that case, the decision boundary is biased towards the minority class, which results in a
 high number of false negatives and a lower classifier performance. Solutions to this problem include

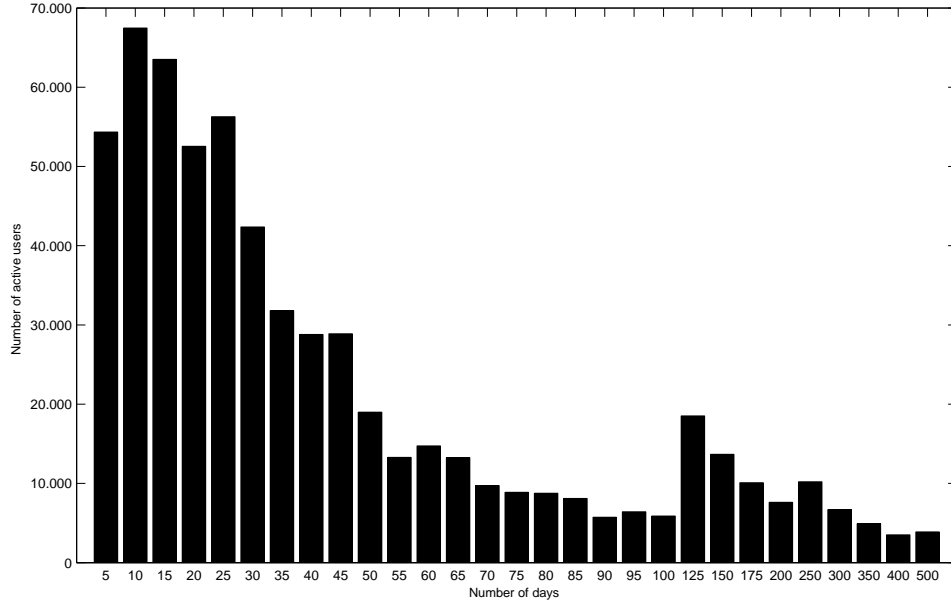


Figure 6: Number of active users after inactive period.

undersampling and oversampling [58]. When randomly undersampling the majority class of the data set, information loss might occur. Here, we randomly oversample the minority class. The multivariate

325 Naive Bayes implementation tailored to big data is used [21]. For the Linear Support Vector Machine, the LIBLINEAR-library is used [50].

Data set	% of positive cases
Museum MAS	1.82%
Roma cultural centre	0.96%
Permeke library	7.29%
Zoo	0.85%
Wezenberg swimming pool	2.18%
Cash out	6.71%
Defect	13.20%

Table 4: Imbalance of the datasets for the considered target variables.

5. Experimental results

The results obtained by the predictive models are presented in Table 5. In short, future visiting behavior, the visiting behavior of users that cash out collected points, and the visiting behavior of users that will defect, can be modeled in an accurate manner and hence these target variables are quite predictable. An evaluation of statistical significance of the results along three dimensions is now performed i.e. between the techniques, between the time granularity levels of the behavioral data and the nature of the prediction (whether future visiting behavior versus loyalty and defect behavior is predicted). We perform the Wilcoxon test when comparing the Naive Bayes classifier and the Linear Support Vector Machine over all datasets and over the two types of predictions. The Friedman test is used when defining whether a significant difference exists over the six time dimensions. We present the following statistically significant results:

- When comparing the performance results over all datasets and over all time granularities, the Naive Bayes classifier performs best with a z-value of -3.59 at the $\alpha = 0.05$ significance level (Wilcoxon). As the datasets have diverse time level characteristics and distinct predictive purposes (location vs. loyalty and defection), this result is investigated in more detail.
- When comparing the results for the non-temporal datasets, the Naive Bayes classifier performs better due to a T-value of 0 that is smaller than the critical Wilcoxon value at the $\alpha = 0.05$ confidence level.
- When using the Linear SVM technique, the use of hourly data performs better than the use of non-temporal data and better than the time levels of work, school and weekend with a critical difference value of 2.85 from the Nemenyi post-hoc test. There is no significant difference between the non-temporal variant on the one hand and the work, school and weekend variants on the other when employing the L-SVM technique. Also, no significant difference was found between

the hour level and the weekday level.

- When performing loyalty or defect prediction, the use of the Naive Bayes classifier performs better than the use of the L-SVM with a z-value of -2.90 at the $\alpha = 0.05$ confidence level (Wilcoxon).
- When performing location prediction, the NB classifier performs best with non-temporal data, while the L-SVM performs better with behavioral data on an hourly level, both at the $\alpha = 0.05$ confidence level with a zero T-value.

Overall, the NB classifier performs best except for location prediction on an hourly time level. Independent of the type of prediction, i.e. whether it concerns location prediction or loyalty or defect prediction, it is apparent that adding finer-grained behavioral data in terms of hourly behavior contributes to a higher performance when employing the Linear SVM. When performing loyalty or defect prediction, the behavior of the users is best modeled by the NB classifier for all time levels. This might originate from the fact that no dependence is relevant between the features when modeling behavior of users cashing out or defecting. For location prediction, the NB classifier performs better when using non-temporal data and the L-SVM performs better in case hourly data is used. The hourly level seems to add the most useful information with respect to the other time granularities.

It may seem obvious that high AUC values are obtained in the case of location prediction: predicting whether a user will visit a location X in the future is probably largely based on whether or not the user visited that location in the past. However, when looking at the top ten weights of the features determined by the L-SVM (Table 6) for the Permeke library and the MAS museum, it is apparent that the information not only comes from visiting the location under consideration.

In order to get a more intuitive insight into the quality of the predictions made by the models, confusion matrices are shown in Table 7. These result from predictions made by the best performing model on the highest behavioral granularity level (L-SVM for location prediction, NB for loyalty and

Target variables	Naive Bayes						Linear SVM					
	non-temporal	hour	weekday	weekend	work	school	non-temporal	hour	weekday	weekend	work	school
Permeke library	85.60	83.85	83.85	83.17	83.40	83.47	83.42	85.51	84.77	83.64	83.52	83.83
Wezenberg swimming pool	88.52	88.10	88.30	88.21	88.43	88.43	86.75	90.69	88.51	86.89	86.17	86.70
Museum MAS	78.97	78.50	78.75	79.10	78.97	79.02	75.05	81.56	78.58	73.91	73.76	74.23
Cultural centre Roma	89.19	88.87	88.98	89.19	89.20	89.18	83.61	91.33	88.51	81.77	81.70	81.83
Zoo	88.57	88.52	88.77	89.05	88.87	88.94	83.26	91.97	90.04	81.72	81.27	81.51
Cashout	89.08	92.50	91.9	90.28	90.50	90.68	87.58	92.36	91.55	87.91	89.11	89.26
Defect	82.66	85.50	85.14	81.35	81.60	81.53	82.82	83.73	83.71	79.53	79.67	79.51

Table 5: Predictive performance of the models in terms of AUC (highest achieved performance in boldface).

Permeke library	Museum MAS
Cultural Centre Ekeren	Youth - Teenage offer
Museum Middelheim	Museum Mayer van den Bergh
Antwerp Sport Event	CC Luchtbal theatre
Cultural Centre De Schelde	Digipoint
Meeting Centre Merksemdok	Museum Rubens
Library Luchtbal	Topvolley
Museum Vleeshuis	Ecohouse
Swimming pool De Molen	Swimming pool IJspiste
Permeke library	Museum Plantin-Moretus
Swimming pool De Schinde	Library Elsschot

Table 6: Top ten feature-weights for two location prediction variables.

defect prediction). Regarding location prediction, wrongly labeling a user as likely to visit a certain venue, might be disturbing for users if future communication is based on this fact. The same holds for persons that are labeled as likely to defect, while they are not planning on becoming inactive. Targeting these users with re-activation e-mails might be perceived annoying and in the extreme case might result in judicial actions. It is apparent that the number of false positives must be reduced in these cases. The false positives ratio in Table 7 for the location prediction models is located in a small interval around 11%. For the defect target variable this ratio amounts to 15.49%. The ratio of false negatives are 'lost' customers that are not targeted. This trade-off seems fair since a low ratio of false positives is considered a priority in this public management setting. Regarding cashout prediction, the false negative ratio is of importance, since these are the users that are wrongly labeled as not cashing out. Communicating with these users in function of convincing them to cashout might result in annoyance and potentially in lost active users. As can be seen in Table 7, this ratio only amounts

to 8.29%. The false positive ratio here is relatively higher i.e. almost 20% of non-active users might be lost since no targeting campaign includes them in persuading them to become active. At first sight, this does not comply with our previously stated goal of attempting to increase the active user base. However, a trade-off has to be made between the public character of the decision support process on the one hand (a government spamming its citizens is not desirable) and the increase of the active user base on the other. We believe this trade-off has been taken into account in the presented results.

Table 8 shows the results from applying the Apriori algorithm with a support of 0.02 and a confidence value of 0.6. This means that the item sets present in the rules only occur in 2% of the transactions and that for these transactions the rule holds in 60% of the cases. The reason for the low support originates from the majority of users only visiting one location, as is seen in Figure 2. This leads to a smaller pool of possible frequent item sets and thus to a lower support. The highest confidence that is gained for association rules in this dataset is 0.7. Taking into account the fact that close to 50% of all transactions consist of one item, the achieved confidence for these rules seems acceptable. As the active user base increases along with the number of venues that support the *A-Card*, and subsequently the number of users frequenting two or more locations increases, the association rules can become stronger and more diverse.

<u>Permeke library</u>			<u>Wezenberg swimming pool</u>			<u>Museum MAS</u>			<u>Cultural centre Roma</u>		
	+	-		+	-		+	-		+	-
+	74.5	11.85	+	75.6	11.12	+	70.2	11.09	+	71.3	11.25
-	25.5	88.15	-	24.4	88.88	-	29.8	88.91	-	28.7	88.75
			<u>Zoo</u>			<u>Cashout</u>			<u>Defect</u>		
	+	-		+	-		+	-		+	-
+	72.1	10.59	+	91.7	19.55	+	72.9	15.49	+	72.9	15.49
-	27.9	89.41	-	8.29	80.45	-	27.1	84.51	-	27.1	84.51

Table 7: Confusion matrices for the target variables as resulting from their best predictive model.

X →	Y	Confidence
Swimming pool Sportoase	Permeke library	0.6
Cultural centre Roma	Museum MAS	0.6
Museum Middelheim	Museum MAS	0.6
Permeke library and City shop	Museum MAS	0.6
Museum Plantin-Moretus	Museum MAS	0.6
Kaeck library	Bist library	0.6
Ecohouse	Permeke library	0.6
Meeting centre Atlas	Permeke library	0.7
Museum MAS and Swimming pool Sportoase	Permeke library	0.7

Table 8: Association rules resulting from the Apriori algorithm.

6. Conclusion and future work

6.1. Conclusion

CRM applications in the public area mostly consist of simplifying citizen access to government’s services. In the past, the public sector has shown a habit of adopting business-like practices. Notwithstanding obvious resemblances between both sectors, their rationale clearly is different. Private companies operate from a profit point of view, while governments need to take into account all citizens. We believe the research conducted in this study contributes to the gap existing in CRM-applications in the public sector.

This design science research paper presented a data mining analysis of spatio-temporal behavioral data of users of a government loyalty card of the city of Antwerp (Belgium). The mission of the *A-Card* mainly consists of promoting participation in cultural activities offered by the city. Whilst analyzing the data, it was found that the bulk of the card users did not actively use their card and that the set of locations visited was not very diverse.

The result of the study is a decision support system in the form of a predictive model which is capable of distinguishing between different visiting behaviors of the loyalty card users. Concretely, we are now able to predict whether a user will make use of his received rewards, whether a user will visit a

certain location, and whether a user is likely to become inactive in some time in the near future. The data consists of a transformation of transactional data into fine-grained spatio-temporal behavioral data. For location prediction ends, the best predictive model is the Linear Support Vector Machine.

420 For cashout and defect prediction, the Naive Bayes classifier performs best. These results are achieved in case the behavioral data is projected onto finer-grained hourly visiting behavior, which is the highest granular behavioral level.

6.2. Contributions

The novelty of this research does not lie in one particular element such as the data or the used
425 method. The innovation comes from a diverse interplay of different elements. First of all, the data is found in an atypical context: loyalty card data mostly originates from the private sector. More importantly, this transactional data has been transformed to fine-grained behavioral data in the form of spatio-temporal behavioral data. The results gained from this projection hint that the larger the number of features available to model a person’s behavior, the better the predictions that can be made.

430 It is highly probable that if the set of cultural venues is extended or the time granularity increased or extra behavioral information added, that even better predictions can be made. Also, as the user base grows, more data instances could also improve the predictive performance of the models. These findings correspond to the results described in [21].

Concretely, as a theoretical contribution, we state that an increase in levels when modeling a
435 person’s behavior, in this case temporal levels, results in better predictions. As a second theoretical contribution, we have demonstrated the need for more diverse CRM-related data research in a government context. Further privatization of public institutions, along with the presence of a vast amount of data available to these institutions, will lead to a growing need for research of private-like decision-support systems in a public setting.

440 We state our second contribution as a practical one. The city of Antwerp, along with other

organizations disposing of this kind of data, is now able to predict loyalty behavior, defect behavior and future visiting behavior based solely on past visiting behavior of their users. This should capacitate them to proactively change the behavior of their citizens. The interventions can be made at four points with the goal of increasing the active user base (visualized in Figure 4). First, since users can now
445 be identified as active or not, the non-active users can be targeted using their preferences or using information regarding their visiting behavior. Personalizing communication and offers might convince these users to actively use their card. Second, when users have joined the active customer base, the city wants them to stay there. This can be done by tailoring communications and/or benefits to locations that users will visit in the future, but also by surprising them with relevant suggestions regarding
450 other venues. Getting users to visit other locations then works self-reinforcing as more users visit more different places and more relevant suggestions can be made. Third, the city wishes to minimize the flow of users leaving the active user base. The designed decision support system is capable of identifying likely future defectors, and thus the city can intervene in time. This can be done by offering tailored benefits based on preferences or previous visiting behavior. Since the problem of the low active user
455 base might be rooted elsewhere, first a general survey might be conducted with the goal of finding out if and to what extent users are fully aware of the benefits offered by the loyalty card. Moreover, the nature of the loyalty benefits might be investigated. It was mentioned by the city of Antwerp that these benefits might not be convincing enough for some users. For users frequently going to the swimming pool, the collected points can be used as free entrance ticket, whereas the benefits for people
460 wanting to cash out their points in a library or a cultural centre are not that straightforward.

6.3. Limitations of our study

One of the main limitations of our study consists of the fact that we only perform this analysis in one city. We cannot claim that the results are generalizable to cultural loyalty programs of other cities. Further data collection in this context is necessary. Subsequently, this data must be thoroughly

analyzed and only then can conclusions be made as to whether these results hold in general. Second, the locations that are recognized by the *A-Card* are limited to cultural venues. Whether the conclusions will hold for cards supporting other types of locations such as docking stations for bike rental or other non-cultural public areas, can only be ascertained if more data is available and subsequently analyzed. Obviously, since bigger is better, having more data instances and more dimensions to our disposal, would also greatly benefit the study. A third limitation of our study is related to location prediction. We have evaluated the predictive performance of our models on five locations out of the total of 102 locations. As doing the analysis for all locations would be more valid when formulating conclusions, it would also be too elaborate to discuss in a research paper and moreover, in the event that more features are added, not feasible. We have therefore selected five locations which demonstrate distinct characteristics such that their results can be extrapolated to the remaining locations.

6.4. Avenues for further research

The opportunities for further research at the government-CRM intersection are vast. With respect to the *A-Card* or other cultural loyalty cards, we believe that in time, the vast behavioral data can be used to customize to a large extent the city's cultural offer to its citizens' preferences. In particular in this research, we have focused on loyalty cards in the context of cultural venues. More generally, there can be a diverse array of other behavioral CRM data in the public sector. Illustrations of these entail city cards for tourists, bike rental cards or public transportation cards. In time, these different research avenues with respect to user-defining behavioral data could lead to the consolidation of all this data, which could in turn lead to more sound and more diverse predictions. Furthermore, projecting this vast and diverse amount of information onto an even more fine-grained temporal level, might lead to stunning results. Investigating whether dimensions other than temporal levels lead to equally good or better results, might also lead to interesting conclusions. Making strong predictions about users and reacting in an adequate fashion may lead to higher citizen satisfaction, which in turn may benefit the

community as a whole. More broadly, these results may lead to the introduction of CRM in other
490 public institutions and to subsequent government-tailored analyses of these data.

7. Acknowledgments

We would like to thank the city of Antwerp for letting us use the data set of the *A-card* and for their valuable input.

Bibliography

- 495 [1] W. Reinartz, M. Krafft, W. D. Hoyer, The customer relationship management process: its measurement and impact on performance, *Journal of marketing research* 41 (3) (2004) 293–305.
- [2] S.-L. Pan, C.-W. Tan, E. T. Lim, Customer relationship management (CRM) in e-government: a relational perspective, *Decision Support Systems* 42 (1) (2006) 237–250.
- [3] D. L. Goodhue, B. H. Wixom, H. J. Watson, Realizing business benefits through CRM: hitting
500 the right target in the right way, *MIS Quarterly Executive* 1 (2) (2002) 79–94.
- [4] G. R. Dowling, M. Uncles, Do customer loyalty programs really work?, *Research Brief* 1.
- [5] S. Lessmann, S. Voß, A reference model for customer-centric data mining with support vector machines, *European Journal of Operational Research* 199 (2) (2009) 520–530.
- [6] C. Mauri, Card loyalty. A new emerging issue in grocery retailing, *Journal of Retailing and*
505 *Consumer Services* 10 (1) (2003) 13–25.
- [7] B. Berman, Developing an effective customer loyalty program, *California management review* 49 (1) (2006) 123.

[8] E. W. Ngai, L. Xiu, D. C. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, *Expert systems with applications* 36 (2) (2009) 2592–2602.

[9] R. H. von Alan, S. T. March, J. Park, S. Ram, Design science in information systems research, *MIS quarterly* 28 (1) (2004) 75–105.

[10] E. Garbarino, M. S. Johnson, The different roles of satisfaction, trust, and commitment in customer relationships, *the Journal of Marketing* (1999) 70–87.

[11] A. S. Dick, K. Basu, Customer loyalty: toward an integrated conceptual framework, *Journal of the academy of marketing science* 22 (2) (1994) 99–113.

[12] S. Lessmann, S. Voß, A reference model for customer-centric data mining with support vector machines, *European Journal of Operational Research* 199 (2) (2009) 520–530.

[13] A. Jacobs, The pathologies of big data, *Communications of the ACM* 52 (8) (2009) 36–44.

[14] D. Boyd, K. Crawford, Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, *Information, Communication & Society* 15 (5) (2012) 662–679.

[15] J. Liu, P. Dolan, E. R. Pedersen, Personalized news recommendation based on click behavior, in: *Proceedings of the 15th international conference on Intelligent user interfaces*, ACM, 2010, pp. 31–40.

[16] D. Martens, F. Provost, Pseudo-social network targeting from consumer transaction data.

[17] K. Li, T. C. Du, Building a targeted mobile advertising system for location-based services, *Decision Support Systems* 54 (1) (2012) 1–8.

[18] T. Fawcett, F. Provost, Adaptive fraud detection, *Data mining and knowledge discovery* 1 (3) (1997) 291–316.

- 530 [19] E. Junqué de Fortuny, M. Stankova, J. Moeyersoms, B. Minnaert, F. Provost, D. Martens, Corporate residence fraud detection, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 1650–1659.
- [20] L. Yu, H. Liu, Feature selection for high-dimensional data: A fast correlation-based filter solution, in: ICML, Vol. 3, 2003, pp. 856–863.
- 535 [21] E. Junqué de Fortuny, D. Martens, F. Provost, Predictive modeling with big data: Is bigger really better?, *Big Data* 1 (4) (2013) 215–226.
- [22] C. Rygielski, J.-C. Wang, D. C. Yen, Data mining techniques for customer relationship management, *Technology in society* 24 (4) (2002) 483–502.
- [23] P. R. Devadoss, S. L. Pan, J. C. Huang, Structural analysis of e-government initiatives: a case study of sco, *Decision support systems* 34 (3) (2003) 253–269.
- 540 [24] M. S. Haque, The diminishing publicness of public service under the current mode of governance, *Public Administration Review* 61 (1) (2001) 65–82.
- [25] D. Dufner, L. M. Holley, B. Reed, Can private sector strategic information systems planning techniques work for the public sector?, *Communications of the Association for Information Systems* 8 (1) (2002) 28.
- 545 [26] C. G. Reddick, The adoption of centralized customer service systems: A survey of local governments, *Government Information Quarterly* 26 (1) (2009) 219–226.
- [27] R. P. Lourenço, J. P. Costa, Incorporating citizens' views in local policy decision making processes, *Decision Support Systems* 43 (4) (2007) 1499–1511.
- 550 [28] H. Chen, J. Schroeder, R. V. Hauck, L. Ridgeway, H. Atabakhsh, H. Gupta, C. Boarman,

K. Rasmussen, A. W. Clements, Coplink connect: information and knowledge management for law enforcement, *Decision Support Systems* 34 (3) (2003) 271–285.

[29] J. C. Bertot, P. T. Jaeger, J. M. Grimes, Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies, *Government Information Quarterly* 27 (3) (2010) 264–271.

[30] H. Chen, R. H. Chiang, V. C. Storey, Business intelligence and analytics: From big data to big impact., *MIS quarterly* 36 (4) (2012) 1165–1188.

[31] G. A. Boyne, Public and private management: what’s the difference?, *Journal of management studies* 39 (1) (2002) 97–122.

[32] J. Rowley, E-government stakeholders - who are they and what do they want?, *International journal of Information management* 31 (1) (2011) 53–62.

[33] F. H. Cate, Government data mining: The need for a legal framework, *Harvard Civil Rights-Civil Liberties Law Review* (CR-CL) 43 (2).

[34] S. F. King, Citizens as customers: Exploring the future of crm in uk local government, *Government Information Quarterly* 24 (1) (2007) 47–63.

[35] A. Schellong, Citizen relationship management: A study of CRM in government, Vol. 560, Peter Lang, 2008.

[36] M. Raus, J. Liu, A. Kipp, Evaluating it innovations in a business-to-government context: A framework and its applications, *Government Information Quarterly* 27 (2) (2010) 122–133.

[37] S. Bretschneider, Management information systems in public and private organizations: An empirical test, *Public Administration Review* (1990) 536–545.

[38] W. Cats-Baril, R. Thompson, Managing information technology projects in the public sector, *Public Administration Review* (1995) 559–566.

[39] P. S. Ring, J. L. Perry, Strategic management in public and private organizations: Implications
575 of distinctive contexts and constraints, *Academy of Management Review* 10 (2) (1985) 276–286.

[40] F. Provost, Geo-social targeting for privacy-friendly mobile advertising: Position paper.

[41] M. S. Jeannotte, Singing alone? The contribution of cultural capital to social cohesion and
sustainable communities, *The International Journal of Cultural Policy* 9 (1) (2003) 35–49.

[42] Federal Public Service (FOD) Economics Belgium, The Belgian Provinces 2001, Statistics Belgium
580 (2014).

[43] City of Antwerp, Minutes of ‘College van Burgemeester en Schepen’, [http://notulus.
antwerpen.be/rs-bin/RightSite.exe/getcontent/C2010-007227.pdf?DMW_OBJECTID=
090f45ab80540dbb&DMW_FORMAT=pdf&fext=.pdf](http://notulus.antwerpen.be/rs-bin/RightSite.exe/getcontent/C2010-007227.pdf?DMW_OBJECTID=090f45ab80540dbb&DMW_FORMAT=pdf&fext=.pdf) (2011).

[44] S. Vittal, E. Murphy, Benchmarking customer loyalty program efforts, *Forrester* (2012).

585 [45] R. H. Sprague Jr, A framework for the development of decision support systems, *MIS quarterly*
(1980) 1–26.

[46] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu,
S. Y. Philip, et al., Top 10 algorithms in data mining, *Knowledge and Information Systems* 14 (1)
(2008) 1–37.

590 [47] T.-N. Do, P. Lenca, S. Lallich, N.-K. Pham, Classifying very-high-dimensional data with random
forests of oblique decision trees, in: *Advances in knowledge discovery and management*, Springer,
2010, pp. 39–55.

- [48] I. H. Witten, E. Frank, Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2005.
- 595 [49] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.
- [50] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, Liblinear: A library for large linear classification, The Journal of Machine Learning Research 9 (2008) 1871–1874.
- [51] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, in: ACM SIGMOD Record, Vol. 22, ACM, 1993, pp. 207–216.
- 600 [52] F. J. Provost, T. Fawcett, R. Kohavi, The case against accuracy estimation for comparing induction algorithms., in: ICML, Vol. 98, 1998, pp. 445–453.
- [53] T. Fawcett, An introduction to roc analysis, Pattern recognition letters 27 (8) (2006) 861–874.
- [54] J. Demšar, Statistical comparisons of classifiers over multiple data sets, The Journal of Machine Learning Research 7 (2006) 1–30.
- 605 [55] K. Coussement, D. V. d. Poel, Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers, Expert Systems with Applications 36 (3) (2009) 6127–6134.
- [56] S.-Y. Hung, D. C. Yen, H.-Y. Wang, Applying data mining to telecom churn management, Expert Systems with Applications 31 (3) (2006) 515–524.
- 610 [57] P. Domingos, A few useful things to know about machine learning, Communications of the ACM 55 (10) (2012) 78–87.
- [58] M. Farquad, I. Bose, Preprocessing unbalanced data using support vector machine, Decision Support Systems 53 (1) (2012) 226–233.