

# Personalization in Context: Does Context Matter When Building Personalized Customer Models?

M. Gorgoglione  
Politecnico di Bari  
m.gorgoglione@poliba.it

C. Palmisano  
Politecnico di Bari  
c.palmisano@poliba.it

A. Tuzhilin  
New York University  
atuzhili@stern.nyu.edu

## Abstract

*The idea that context is important when predicting customer behavior has been maintained by scholars in marketing and data mining. However, no systematic study measuring how much the contextual information really matters in building customer models in personalization applications have been done before. In this paper, we address this problem. To this aim, we collected data containing rich contextual information by developing a special-purpose browser to help users to navigate a well-known e-commerce retail portal and purchase products on its site. The experimental results show that context does matter for the case of modeling behavior of individual customers. The granularity of contextual information also matters, and the effect of contextual information gets diluted during the process of aggregating customers' data.*

## 1. Introduction

The director of personalization of one of the major on-line retailing companies once received a nasty email from the CEO telling him that he should either fix his personalization system or lose his job. The CEO's email was prompted by a customer's complaint that the company's personalization system was making offensive assumptions about the lifestyle of this customer and was recommending inappropriate products to that person. Upon a closer examination, it was discovered that the customer once bought an item as a gift for his friend, and the personalization system started recommending related products to that customer making implicit assumptions about his lifestyles, which infuriated that customer.

This true story is very symptomatic of problems pertaining to many personalization systems that often infer customer behavior from the registration and the purchasing information of online customers without studying the *contexts* in which these purchases are made. In the previous example, if the system knew that the purchase was made in the context of a gift, this transaction should have been discarded from inferring that customer's behavior, and the whole problem would have been avoided.

Getting such contextual data characterizing the circumstances in which purchasing or other transactions took place, such as the intent of a purchase, special payment conditions, economic climate and the customer's geographic location, is not easy in many marketing applications. For instance, it may not be practical to ask the customer in the previous example about the purpose of his/her purchase. Therefore, before insisting on requesting such contextual information, it is necessary to provide hard scientific evidence that this contextual information indeed makes a significant difference in building better customer models in some marketing and e-commerce applications to justify extra costs of obtaining this contextual information.

In this paper, we address this problem and try to answer the question *whether this additional contextual information matters*, i.e., does it lead to building better personalized predictive models of customer behavior, where by "better" we assume superior predictive performance. This problem is not trivial because it entails a tradeoff between transaction homogeneity and data sparsity: by providing contextual information, customer transactions pertaining to this particular context are reduced, making fewer data points to fit the model, while homogeneity of these transactions increases, making it easier to predict more accurately customer behavior in similar contexts. In data mining terms, this problem is related to the well-known bias-variance tradeoff [13]. Therefore, the main problem studied in this paper is which effect dominates the other, given additional contextual information: increased variance due to insufficient data or decreased bias due to the homogeneity of transactions associated with the specified context.

The idea that the contextual information is important when predicting customer behavior is not new and much anecdotal evidence, similar to our prior example, has been collected in the popular press. In the marketing and data mining literature, related issues have also been studied.

Scholars in marketing have maintained that the purchasing process is contingent upon the context in which the transaction takes place. The same customer can adopt different decision strategies and prefer different products or brands depending on the context [7]-[19]. According to [18], "consumers vary in their decision-

making rules because of the usage situation, the use of the good or service (for family, for gift, for self) and purchase situation (catalog sale, in-store shelf selection, sales person aided purchase).” Therefore accurate prediction of consumer preference undoubtedly depends upon the degree to which we have incorporated the relevant contextual information. The importance of including contextual information in recommendation systems has been demonstrated in [3]. If the concept of customer and that of transaction are broadened to embrace any user interacting with a company or an application to get a service, then the importance of knowing the context is recognized in other fields and applications. For instance, context-aware systems are designed to exploit the contextual information available (e.g., where the user is, who is with him/her, what devices are accessible) to better serve the user [11], and to adapt to changes in the context.

Context has several alternative definitions in different fields and applications. The Webster’s dictionary defines context as “conditions or circumstances which affect some thing.” In the data mining community, context is defined in [6] as those events which characterize the life of a customer and can determine a change in his/her preferences, status (e.g., prospect to actual), and affect the customer’s value for a company. Examples of context include a new job, the birth of a son, marriage, divorce, retirement, a disease. In the context-aware systems literature, context was initially defined as the location of the user, the identity of people near the user, the objects around, and the changes in these elements [22]. Other factors have been added to the previous definition. For instance, Brown, Bovey, and Chen [8] consider the day date, the season, the temperature as part of context. Ryan, Pascoe, and Morse [21] add the physical and conceptual statuses of interest for a user. Dey [11] includes the user’s emotional status in the definition of context. Dey, Abowd and Salber broaden the definition of context to any information which can characterize and is relevant to the interaction between a user and an application. Some scholars associate the context with the user [8]-[12], others emphasize how context relates to the application [20]-[23]. The concept of context is usually referred to the present situation, but sometimes the history of past contexts should be considered as well [11]. A critical issue when modeling the behavior of a customer is to determine how to acquire the contextual information [10]-[11]. In some situations this contextual information could be readily available. For example, an individual has to communicate to the company that he/she is moving to another city or report the change in the surname after a marriage. Changes in the context can also be inferred from other data, such as a change in the location of a user detected by a mobile telephone company, and a change in the way a bank account is used. In general, the ability of

exploiting the knowledge of context is expected to increase the potential of many applications aimed at delivering services to users [1]. Much work on contextual information has been done in various applications. However, no systematic attempts to conduct scientific studies measuring how much the contextual information really matters in building customer models in personalization applications have been done before. In this paper, we address this problem and conduct an empirical study of how much the contextual information matters. Since one needs the right type of data to answer this question, and this data is not readily available in most of the industrial applications, we collected our own data containing rich contextual information and studied the main research question, stated above, using this data.

As a result of this study, we conclude that:

1. Context does matter for the case of modeling behavior of *individual* customers (i.e., the 1-to-1 case): knowing the context in which a customer does the purchase increases the ability to predict the customer’s behavior.
2. Granularity of contextual information also matters. The more we know about the context of a transaction, the better we can predict the customer’s behavior.
3. The effect of contextual information gets diluted during the process of aggregating customers’ data. Context does matter for individual customers, as explained above; but does not significantly matter when predicting behavior of the whole customer base. The contextual effects get stronger when we build progressively smaller segments of customers.

The findings reported in this paper have significant implications for marketers, data miners and industrial researchers working on designing recommender systems. They show that contextual information does matter in personalization applications and that the finer it is – the better. Therefore, data miners and marketers should insist on collecting this important contextual information, whenever possible, since this information significantly helps to build better personalized models of customer behavior. In the next section we formulate the problem of comparing context-based vs. regular models of customer behavior.

## 2. Problem formulation

Let  $C$  be the customer base represented by  $N$  customers. Each customer  $C_i$  is defined by the set of  $m$  demographic attributes  $A = \{A_1, A_2, \dots, A_m\}$ , and a set of  $r$  transactions  $Trans(C_i) = \{TR_{i1}, TR_{i2}, \dots, TR_{ir}\}$ , where each transaction  $TR_{ij}$  performed by customer  $C_i$  is defined by a set of transactional attributes  $T = \{T_1, T_2, \dots, T_p\}$ . In addition, we also have contextual information  $K$  associated with each transaction  $TR_{ir}$ . The table

specifying all this demographic, transactional and contextual information about customers is presented in Figure 1.

		Demographic attributes $A$			Transactional Attributes $T$			Context $K$			
Trans ( $C_i$ )	TR <sub>i1</sub>	A <sub>1</sub>	...	A <sub>m</sub>	T <sub>1</sub>	...	T <sub>p</sub>	K <sub>g1</sub> , K <sub>g2</sub>	K <sub>g11</sub> , K <sub>g12</sub> , K <sub>g13</sub> , K <sub>g14</sub>		
	TR <sub>i2</sub>	A <sub>1</sub>	...	A <sub>m</sub>	T <sub>1</sub>	...	T <sub>p</sub>	K <sub>g1</sub> , K <sub>g2</sub>	K <sub>g11</sub> , K <sub>g12</sub> , K <sub>g13</sub> , K <sub>g14</sub>		
	TR <sub>ir</sub>	...									
Trans ( $C_j$ )	TR <sub>j1</sub>	...									
	TR <sub>j2</sub>	...									
	TR <sub>jr</sub>	...									
Trans ( $C_N$ )	TR <sub>N,1</sub>	...									
	TR <sub>N,2</sub>	...									
	TR <sub>Nr</sub>	A <sub>1</sub>	...	A <sub>m</sub>	T <sub>1</sub>	...	T <sub>p</sub>	K <sub>g1</sub> , K <sub>g2</sub>	K <sub>g11</sub> , K <sub>g12</sub> , K <sub>g13</sub> , K <sub>g14</sub>		

Figure 1. Data structure

In general the *domain*  $K$  of this contextual information can have a complicated structure reflecting the complex nature of this information. However, in this paper, we assume that domain  $K$  is defined by a *single discrete* variable having a *hierarchical structure* associated with it, the lower levels of the hierarchy defining *finer* (more *granular*) and the higher levels *rougher* (*coarser*) contexts of variable  $K$ .

For example, a customer  $C_i$  can be defined by the demographic attributes  $A = \{IDuser, Name, Age, Income\}$ , by the set of five transactions made by  $C_i$ ,  $Trans(C_i) = \{TR_1, TR_2, TR_3, TR_4, TR_5\}$ , each transaction defined by the transactional attributes  $T = \{ProductID, StoreID, Price, TransactionTime\}$ . The contextual variable  $K$  defines for whom the product was purchased. Its values are defined as a hierarchy shown in Figure 2. The highest (coarsest) level of the hierarchy for  $K$  consists of values  $\{Personal, Gift\}$ , meaning that the purchase was made for the customer him/herself or as a gift. At the next level of the contextual hierarchy  $K$ , value Personal consists of  $\{PersonalWork, PersonalOther\}$ , meaning that the personal purchase is either work-related or bought for other purposes. Similarly, Gift can be for a Partner/Friend or a Parent/Other. Finally, the customer base  $C$  can be partitioned into several *segments* [16] by computing  $h$  summary statistics  $S_i = \{S_{i1}, S_{i2}, ..., S_{ih}\}$  for customer  $C_i$  over the transactions made by that customer, each  $S_{ij}$  being defined as a statistics on some of the attribute in  $T$  across the transactions  $Trans(C_i)$ . For instance, for the transactions made by the customers in the previous example, the statistics can be  $S_i = \{Average time spent\}$  or  $S_{i+1} = \{Average price\}$ . Then customers can be clustered into segments in the space defined by these

statistics. A model of customer behavior can be built in the following general form:

$$Y = f(X_1, X_2, ..., X_p) \quad (1)$$

where  $X_1, X_2, ..., X_p$  are some of the demographic attributes from  $A$  and some of the transactional attributes from  $T$ , and  $Y$  is the dependent variable to be predicted. Function  $f$  is a predictive function learned via different types of machine learning methods. For instance, one may try to predict in which store the customer  $C_i$  will make a purchase, or which product will be bought, or the product's price.

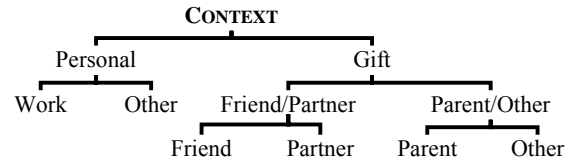


Figure 2. Contextual information for a purchasing transaction

Moreover, the model of type (1) can be built for a *segment* of customers, meaning that only the transactional data for customers  $C_i$  belonging to a segment is used for building that model. In two extreme cases, there can be only one segment for the whole customer base  $C$ , defining the *aggregated* case and one segment per each customer, defining the *individual* (or *1-to-1*) case. There can also be cases of building intermediate segments with finer or coarser granularities. The predictive models of type (1) do not assume any contextual information since the contextual variable  $K$  is not a part of these models. Therefore, we call the models of this type *un-contextual*. In addition, we define *contextual* counterparts of predictive models (1) of the form

$$Y = f(X_1, X_2, ..., X_p, K=a) \quad (2a)$$

$$Y = f_{K=a}(X_1, X_2, ..., X_p) \quad (2b)$$

where the two models (2a) and (2b) constitute two separate notations for the same concept: only transactions associated with the context  $K = a$  are used for building the model. For example, if the model is built for the computer science faculty from University X, where  $K = \text{"gift"}$ , this means that only the *gift-related* transactions made by the CS faculty are used for building the model. Because of the hierarchical structure of the contextual variable  $K$ , the number of contextual models can vary depending on how fine the contextual knowledge is, i.e., at which level of the hierarchy the value  $a$  is defined in  $K = a$ . Then the meaning of the expression "*context matters*" is that the contextual predictive models of type

(2) *significantly* outperform the un-contextual models of type (1) across different experimental conditions. The degree to which the former outperforms the latter determines how significantly the context matters.

### 3. Experimental setup

Since contextually rich datasets suitable for building personalized customer models are not readily available, as was explained in Section 1, we had to collect such data ourselves in order to conduct our study. To collect this data, we developed a special-purpose browser to help users to navigate a well-known e-commerce retail portal and purchase products on its site. This browser was made available to a group of students for browsing and purchasing purposes. While navigation was real, purchasing was simulated (no real money was spent). The user could visit any page of the portal and use all the browsing and navigational activities on the portal except the actual purchase function. Once a product was selected by a student to be purchased, the browser recorded the selected item, the purchasing price and other useful characteristics of the transaction. In addition, the student specified the *context* in which the purchase was made. The browser was directly linked to a database that we created, where all the customer information about browsing and purchasing activities of the customer on the portal was automatically recorded. We had 749 students from an Italian university who participated in this study during the period of two months. The overall number of purchasing transactions recorded is 34,957.

No restrictions were imposed on the participants either in terms of the products they can purchase or the amount of money they can spend. They were only recommended to buy on the order of 50 items. The data was pre-processed by excluding the students who made less than 40 transactions. The resulting number of students having at least 40 transactions was 556, and the total number of purchasing transactions for these students was 31,925. For each customer (student) we collected the following *demographic* data: age, previous studies, marital status, composition of the family, place of living, hobbies and whether the student owned a car. The car ownership was used as a proxy for the income. The *transactional* data included item purchased, price, day, time, session duration, number of clicks per connection, and the time elapsed for the web page. Table 1 reports the selection of the attributes  $X_1, X_2, \dots, X_p$  used in the predictive models.

The intent of purchase was gathered as contextual information. At the beginning of each browsing session, the user was asked to specify whether the purchase would be intended for *personal* purposes or as a *gift*, for which specific personal purpose, and for whom the gift is intended. The overall contextual purchasing options were defined by the following hierarchical structure  $K = \{K_\omega,$

$K_\beta\}$  presented in Figure 2, where  $K_\alpha = \{K_{\alpha1}, K_{\alpha2}\}$ ,  $K_\beta = \{K_{\beta1}, K_{\beta2}, K_{\beta3}, K_{\beta4}\}$ . The user had to specify the contextual option each time for a new browsing session.

After collecting all the purchasing data for all the students, we built predictive models of their purchasing behavior for the contextual and un-contextual cases (using predictive models of type (1) and (2)) under different experimental settings obtained by varying the following parameters:

1. *Degree of contextual information.* The contextual models can be built by considering few values for  $K$  (rough knowledge of context) or as many as available (finer knowledge.)
2. *Granularity of customer segments.* The unit of analysis can range from the single aggregated customer base to segments to individuals, i.e., a predictive model is built for the specified groups of customers.
3. *Types of predictive models.* We considered different types of data mining classifiers modeling function  $f$ , including decision trees and decision rules.
4. *Dependent variables.* Several variables can be used for predicting customers' behavior, such as the day of purchase or the store where the customer will purchase.
5. *Performance measures.* We used the predictive accuracy and the area under the ROC curve [15] as performance measures of predictive models  $f$ .

Each of these settings is discussed in detail now.

**Table 1. Demographic and transactional data**

Demographic data	Type	Values/range
1. Gender	Boolean	Male/Female
2. Age	Numerical	18-31
3. High School description	Nominal	Grammar, Professional, Private
4. Student description	Nominal	Outside, Traveling, Resident
5. Personal Car	Boolean	Yes/No
6. Hobby	Nominal	Reading, Dancing, Music, Electronics, Sports, Movies, Traveling, Informatics, Cooking, Cars, Arts, Photography, Collections, Fashion
Transactional data	Type	Values/range
1. Weekday	Boolean	Weekday/weekend
2. Visit Duration	Numerical	0-919 sec.
3. N. of clicks	Numerical	1-35
4. Price	Numerical	1-2000 \$
5. Store	Nominal	Electronics, home/garden, featured, Kid/baby, book/music, new
6. Purchase description	Boolean	Yes/No

### 3.1. Degree of contextual information

The contextual information  $K$  that we gathered is structured in a three-level hierarchy, as shown in Figure 2, from a rough to a finer degree of knowledge. In the first level, the contextual variable  $K$  takes two different values:  $K_\alpha = \text{"personal"}$  and  $K_\beta = \text{"gift"}$ . In the third and finer level, the “personal” context is split in  $K_{\alpha1} = \text{"personal for work"}$  and  $K_{\alpha2} = \text{"personal for other purposes"}$  and the “gift” in  $K_{\beta1} = \text{"gift for partner"}$ ,  $K_{\beta2} = \text{"gift for friends"}$ ,  $K_{\beta3} = \text{"gift for parents"}$  and  $K_{\beta4} = \text{"gift for others"}$ . In the second level, the partner and friend values are aggregated in one category, as well as parent and others values are aggregate resulting in  $K_{\beta12}$  and  $K_{\beta34}$ , respectively. The degree of knowledge of contextual information grows from the 1<sup>st</sup> level to the 3<sup>rd</sup> level. We used all the three levels of granularity of contextual information in our experiments.

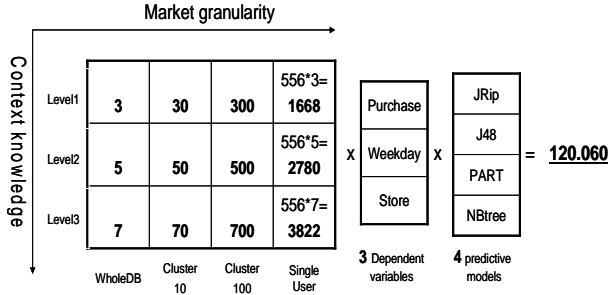


Figure 3. Overall number of models

### 3.2. Granularity of customer segments

The behavioral models can be built for different groups of customers (unit of analysis) depending on how finely the customer base is partitioned. As one progresses from the aggregate to the individual level of granularity, the groups of customers become increasingly more homogenous, making predictions potentially more accurate [14]. However, individual models run into the data sparsity problem. Thus, experimental analysis is crucial for studying under which conditions the knowledge of context counterbalances the sparsity effect. We considered all the following four level of analysis in our experiments:

- *The whole customer base*: the unit of analysis is the whole customer base. One overall model is built to predict the behavior of the whole customer base.
- *Cluster-10*: The units of analysis are ten macro-segments. One predictive model is built for each segment. These 10 macro segments have been generated by applying the Farthest First clustering method to the customer base, where each customer is

defined by a vector of three summary statistics (Average Price, Average Number of Clicks per Session, Average Time Elapsed per Session).

- *Cluster-100*: the units of analysis are 100 micro-segments. Both the summary statistics and the clustering algorithms are the same as in the Cluster-10 case.
- *Single customer*: the unit of analysis is a single user. A predictive model of customer behavior is then applied only to the transactions made by one particular customer.

### 3.3. Types of predictive models

Four types of classifiers for building predictive models were used: Naïve Bayesian Tree, J48, PART, and JRIP. We selected these four classifiers because they constitute popular classification methods and they are also computationally fast to generate. The latter is crucial because we ended up generating 120,060 models in total as part of our experiments, as shown in Figure 3.

The performance of these algorithms was determined using the 10-fold cross-validation process, as described in Section 3.5.

### 3.4. Dependent variables

The following dependent variables were chosen in our experiments from the set of transactional variables  $T$ :

- *Purchase*: the model predicts whether a customer or a group of customers will make a purchase or not.
- *Weekday*: the model predicts when a customer will perform a transaction, regardless of whether the session ends with a purchase or not.
- *Store*: the model predicts in which store a transaction will be made. The store is that one where the session is concluded, either positively or negatively.

### 3.5. Performance measure

We used two measures for determining performance of our models: the predictive accuracy and the area under the ROC curve (AUC) [15]. We calculated these two performance measures for each customer segment used in our studies. For example, if we used the *Cluster-100* segmentation, this means that we would end up with 100 predictive accuracy and 100 AUC measures, one for each segment and for each experimental setting. Since we are interested in comparing the performances of the un-contextual and contextual models, we really need to compare two distributions and see if there are statistically significant differences between them. The same situation is with the AUC measures. Since this entails to comparing the averages of two distributions, a test of

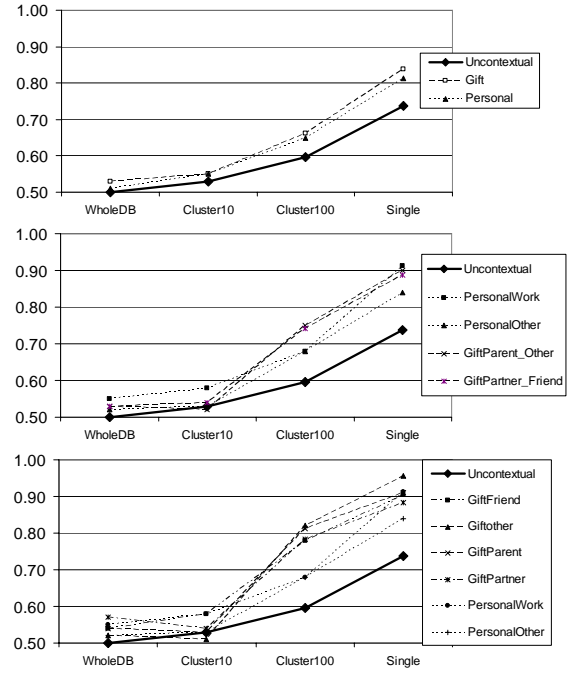
statistical significance needs to be performed with the null hypothesis stating that there is no difference between the two averages, i.e., the performance of the un-contextual model is equal to that of the contextual models. Because the expectation is that context matters, the test is directional: the expectation is that each contextual model outperforms the un-contextual. Since the two models are built for the same unit of analysis (single customer, segment, whole set of customers), the two samples can be considered related. Moreover, the distribution cannot be considered normal so a Wilcoxon test [5] was used for testing the null hypothesis. We compared predictive performance of our customer models for contextual and un-contextual settings across all the experimental conditions described in this section.

We performed these comparisons only on a single data set. We understand that it is important to do these comparisons across multiple data sets to make our results be generalized across different types of data. Unfortunately, it is very difficult to obtain data sets with annotated contextual information since this contextual information is not readily available. In fact, we had to generate such data set on our own for these experiments, which is an expensive and time-consuming endeavor. We hope that this study will generate interest in the industry and that it will eventually produce more data sets with contextual information that can be used in other studies.

#### 4. Results

In this section we present the results of comparing performance of contextual and un-contextual models of customer behavior across all the experimental conditions described in Section 3. Given the number of experimental settings (three degrees of contextual information, four customer granularities, four classifiers, three dependent variables, two performance measures), the total number of generated models was 120,060 and the number of tables reporting the performance comparison results is 288 (144 for each performance measure), which constitutes a challenging problem to present within the limited space of this paper. To give a “flavor” of the obtained results, Figure 4 reports three graphs generated by plotting the values of AUC for a specific classifier (JRIP) and dependent variable (purchase) for different degrees of contextual information (the three graphs in Figure 4 are presented in the order of progressively more refined contextual information) and different customer granularities (moving left to right on each graph). The graphs in Figure 4 show that for these particular experimental settings the predictive performance improves from the aggregate to the single case, except in one case (when  $K$  = “gift to the partner” in the third graph for the second level of granularity). Moreover, all the contextual models show a better predictive performance

compared to the un-contextual model, except two cases (also in the third graph and second level of customer granularity) where the difference is very small. The difference between the performance of un-contextual model and contextual models improves from the aggregate to the single case, and improves when using a finer degree of contextual information. All these differences are statistically significant ( $p < 0.01$ ), except one. Similar charts can be plotted for the remaining classifiers (J48, PART, Naïve Bayesian Tree), dependent variables (weekday, store), and performance measure (predictive accuracy). We observe the same type of behavior in these graphs as in Figure 4. Although the curves are not always monotonic, the predictive performance of the contextual models is usually higher than that of the un-contextual model for the finest level of customer granularity and degree of contextual information.



**Figure 4. Example of comparison of performance (AUC)**

Instead of plotting individual graphs, as in Figure 4, a more concise representation may be obtained by computing the average values of performance over some experimental settings. However, the two performance measures vary in different ranges, namely  $[0, 1]$  and  $[0.5, 1]$  for predictive accuracy and AUC, respectively. Therefore, a reasonable way to compare each contextual model to the un-contextual model is computing the relative difference between the performance values as

$$(Performance_{con} - Performance_{unc}) / Performance_{unc} \quad (3)$$

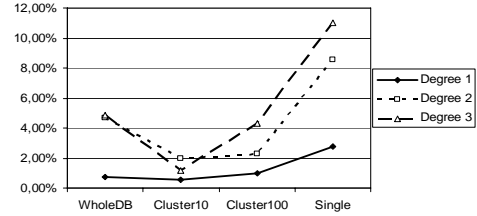
where *con* refers to the contextual models and *unc* to the un-contextual model. A positive value means that the contextual model outperforms the un-contextual. The formula in (3) can be computed as the average over some experimental settings. Table 2 reports the values of (3) computed as the average over the four classifiers and the two performance measures, for each dependent variable. Each row in the table reports the values for each customer granularity (increasing from left to right). Each column reports the values for each degree of contextual information (increasing from top to bottom): the first two comparisons refer to the roughest degree of contextual information, where *K* takes only two values (Gift and Personal), whereas the last six comparisons refer to the finest degree, where *K* takes six values. Figure 5 show the values in Table 2 after computing the average over the three dependent variables.

**Table 2. Relative difference of performance**

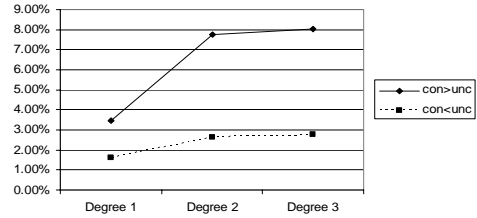
		WholeDB	Cluster10	Cluster100	Single
Dependent variable: Weekday	Gift vs. Uncontextual	-0.02%	-0.67%	-1.23%	1.56%
	Personal vs. Uncontextual	-0.02%	-0.52%	1.03%	2.96%
	GiftParent_Other vs. Uncontextual	1.99%	0.66%	-2.88%	5.37%
	GiftPartner_Friend vs. Uncontextual	3.00%	4.55%	-1.35%	2.64%
	PersonalWork vs. Uncontextual	4.37%	4.44%	-3.02%	8.24%
	PersonalOther vs. Uncontextual	2.79%	2.56%	1.37%	5.24%
	GiftFriend vs. Uncontextual	-0.26%	0.71%	-1.71%	7.41%
	GiftOther vs. Uncontextual	5.80%	2.41%	1.37%	18.07%
Dependent variable: Purchase	GiftParent vs. Uncontextual	1.24%	-0.37%	-1.64%	6.18%
	GiftPartner vs. Uncontextual	5.99%	5.41%	2.00%	9.68%
	PersonalWork vs. Uncontextual	4.37%	4.44%	-3.02%	8.24%
	PersonalOther vs. Uncontextual	2.79%	2.56%	1.37%	5.24%
	Gift vs. Uncontextual	0.76%	1.77%	3.99%	6.27%
	Personal vs. Uncontextual	0.25%	0.38%	3.44%	5.02%
	GiftParent_Other vs. Uncontextual	1.15%	0.08%	11.29%	10.95%
	GiftPartner_Friend vs. Uncontextual	1.60%	1.83%	10.20%	9.79%
Dependent variable: Store	PersonalWork vs. Uncontextual	3.30%	3.40%	5.54%	11.39%
	PersonalOther vs. Uncontextual	0.05%	0.04%	5.74%	6.60%
	GiftFriend vs. Uncontextual	2.24%	2.65%	12.83%	10.71%
	GiftOther vs. Uncontextual	2.35%	-0.92%	16.59%	15.15%
	GiftParent vs. Uncontextual	1.56%	2.16%	15.99%	11.65%
	GiftPartner vs. Uncontextual	4.03%	2.30%	13.87%	9.77%
	PersonalWork vs. Uncontextual	3.30%	3.40%	5.54%	11.39%
	PersonalOther vs. Uncontextual	0.05%	0.04%	5.74%	6.60%
Dependent variable: Store	Gift vs. Uncontextual	-2.98%	0.92%	-4.76%	-3.42%
	Personal vs. Uncontextual	6.30%	1.35%	3.61%	4.25%
	GiftParent_Other vs. Uncontextual	-6.94%	-7.40%	-8.30%	3.75%
	GiftPartner_Friend vs. Uncontextual	3.44%	1.94%	0.04%	6.43%
	PersonalWork vs. Uncontextual	31.11%	11.05%	3.96%	25.91%
	PersonalOther vs. Uncontextual	10.26%	0.82%	4.57%	6.35%
	GiftFriend vs. Uncontextual	4.11%	-4.66%	2.62%	9.66%
	GiftOther vs. Uncontextual	-0.83%	-4.46%	0.39%	15.57%
Dependent variable: Store	GiftParent vs. Uncontextual	-5.39%	-10.06%	-6.56%	5.33%
	GiftPartner vs. Uncontextual	14.41%	3.91%	3.96%	15.53%
	PersonalWork vs. Uncontextual	31.11%	11.05%	3.96%	25.91%
	PersonalOther vs. Uncontextual	10.26%	0.82%	4.57%	6.35%

Figures 6 and 7 represent the results in an even more concise computing the average value of (3) in all experimental settings except one, granularity of customer base and degree of contextual information, respectively. In order to have a clearer representation, the positive values of (3) are computed separately from the negative

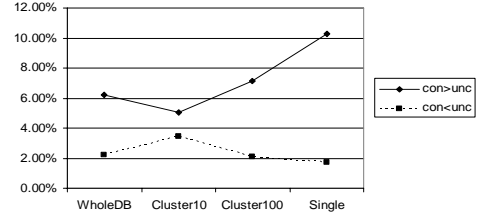
values, and the absolute values are plotted. This modification allows us to discuss the cases in which the contextual models outperform the un-contextual model separately from the cases where the contrary happens.



**Figure 5. Relative difference of performance (average over dependent variables)**

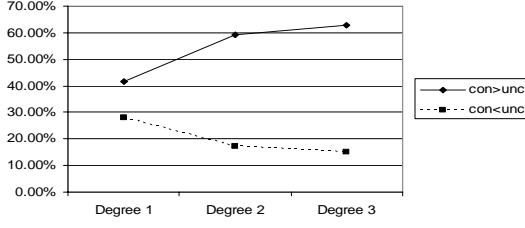


**Figure 6. Relative difference of performance per degree of context**

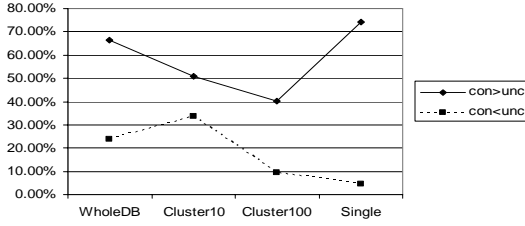


**Figure 7. Relative difference of performance per customer granularity**

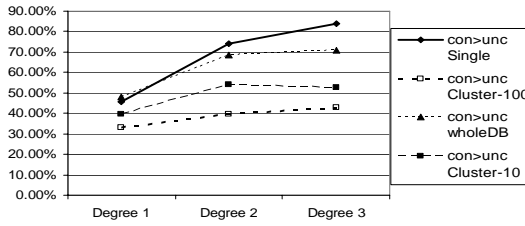
The last issue about the results representation concerns their statistical significance. Since reporting the statistical significance of each comparison would be impossible because of the large number of them (1152 in total), Figures 8 and 9 present a summary of these comparisons by reporting the percentage of comparisons with a statistical significance higher than 95%. The cases in which the contextual models significantly outperform the un-contextual are plotted separately (with the solid line) from the cases where the un-contextual models dominate the contextual one (dashed lines). The values in Figure 8 are computed as follows: for each customer granularity, the number of significant comparisons is divided by the overall number of comparisons. The same is done for each degree of contextual information in Figure 9.



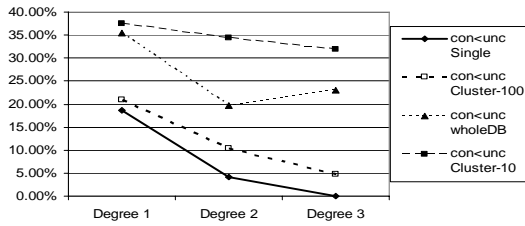
**Figure 8. Statistical significance per degree of context**



**Figure 9. Statistical significance per customer granularity**



**Figure 10. Statistical significance (contextual outperforms un-contextual)**



**Figure 11. Statistical significance (un-contextual outperforms contextual)**

Figure 10 reports the percentage of statistically significant comparisons for each degree of contextual information (values on the x-axis) and for each customer granularity (on different curves), for the cases in which the contextual models dominate the un-contextual. Finally, Figure 11 reports the same results for the cases in which the un-contextual model dominates the contextual models.

## 5. Discussions

The results described in Section 4 present empirical evidence that the models built by taking into account the contextual information usually provide better predictive performance. More specifically, the main conclusions of our study can be summarized as follows:

1. Context does matter for the case of modeling behavior of *individual* customers (i.e., the 1-to-1 case): knowing the context in which a customer does the purchase increases the ability to predict the customer's behavior.
2. The degree of contextual information also matters. The more we know about the context of a transaction, the better we can predict the customer's behavior.
3. The effect of contextual information gets diluted during the process of aggregating customers' data. Context does matter for individual customers, as explained above; but does not significantly matter when predicting behavior of the whole customer base.

Each of these points is discussed in detail below.

*1. Knowing the context matters for the personalization (1-to-1) models.* In the case of models built for single customers, the predictive performance of contextual models is almost always higher than that of the un-contextual model, as Table 2 and Figures 7 and 9 demonstrate. Looking at the results shown in Table 2, the last column presents positive values except one. Actually, in some cases the un-contextual model dominates the contextual. However, as reported in Figure 9, the number of these cases is quite low. The percentage of events in which the difference in performance is negative and statistically significant is around 5%, whereas the percentage of cases in which the difference is positive and statistically significant is around 75%. Moreover, Figure 7 shows that the average gain in performance obtainable by a personalized and contextual model is higher than 10%. This gain reaches a peak of 25.91% in specific experimental settings, as in Table 2 (average over the four classifiers and two performance measures). The highest value of gain in performance was 46.08%, obtained for the case where purchasing was done for personal work, predicting the store type using a Naïve Bayesian Tree model, and measuring the predictive accuracy. On the other hand, building a personalized and contextual model can lead to a decrease in the performance in some conditions, but the loss is lower than 2% on average. The negative peak value is -3.42% in Table 2 (average over the four classifiers and two performance measures). The lowest level, in very specific conditions, was -8.33%.

*2. The degree of contextual information matters:* the finer the knowledge about the context of a transaction,



the better the predictive performance of a customer's behavior. Looking at the right column of Table 2 (single customer), the only negative value is referred to the roughest degree of contextual information ( $K$  takes only two values, Gift and Personal). As we move to finer degrees of contextual information, we observe higher values of performance gain. This is true in general, when the unit of analysis becomes more aggregate, moving left in Table 2. On average, as shown in Figure 6, knowing the finer context leads to higher gains in performance, ranging from 3.5% (when  $K$  takes two values) to 8% (when  $K$  takes six values), in the cases in which the contextual models dominate the un-contextual. On the other hand, gathering finer degrees of contextual information can lead to decrease in performance in those settings in which the un-contextual model dominates the contextual. However, the loss is moderate, from 1.63% to 2.75% in absolute terms in Figure 6. The same evidence is provided by Figure 8, which shows the percentage of events in which the difference in performance is statistically significant. Figures 10 and 11 provide further insights. Whatever the unit of analysis (from the single customer to the aggregate customer base), the number of statistically significant events grows when the contextual models dominate the un-contextual, and decreases in the opposite case. The highest variations occur for single customer models. In the 1-to-1 approach, there is no case in which the un-contextual model outperforms the contextual models and the difference is statistically significant, as shown in Figure 11.

3. *The effect of contextual information gets diluted during the process of aggregating the customer base.* As shown in Figure 7, the gain in performance obtainable by a contextual model decreases when the unit of analysis moves from the single customer to the aggregate customer base (moving left in the graph). The results shown in Table 2 support this observation, as the number of negative values is higher in the left columns than in the right column or, more precisely, the number of cases in which the difference in performance is always more statistically significant for the 1-to-1 models than for other aggregation levels. This result can be interpreted in terms of the tradeoff between data sparsity and homogeneity. As mentioned above, by providing contextual information, customer transactions pertaining to this particular context are reduced, making fewer data points to fit the model, while homogeneity of these transactions increases, making it easier to predict more accurately customer behavior in similar contexts.

As a result of this research, we can conclude that the homogeneity induced by providing additional contextual information tends to dominate the effect of data sparsity in the 1-to-1 case. If the unit of analysis is a more aggregate customer base, the two effects tend to get mixed. The best way to visualize this effect is to look at

Figure 4. Although this graph refers to a specific experimental setting, it demonstrates that the gain in predictive performance obtained by using a contextual model instead of the un-contextual is high for the single case, and decreases when customers are aggregated. This observation can be generalized to similar settings. In fact, Figure 5 shows the average values supporting this effect: it demonstrates that the relative difference in performance is moderate for the most aggregate cases (whole customer base and customers aggregated in 10 clusters). Moreover, in the case of the whole customer base, the curves related to the two contextual models overlap. The difference increases when customers are grouped in 100 cluster and reaches the peak in the case of individual customers.

The findings reported in this paper have significant implications for data miners and marketers. They show that contextual information does matter in personalization applications. In particular, in the case of personalized applications, the finer the degree of contextual information the better the ability of predicting the customer's behavior. Therefore, data miners and marketers should insist on collecting contextual information, whenever possible, especially when their goal is to build personalized models of customer behavior.

The research presented in this paper can be enhanced in several directions. First, the reported findings are constrained by the type of data used, which was collected for the online purchasing environment. Although it is possible to construct artificial counter-examples, we believe that our conclusions are generalizable to various other applications, given prior scientific findings described in Section 1 and in the popular press that context actually affects the behavior of customers. Finally, further experiments should be carried out for different types of applications and the collected data. Unfortunately, this effort is currently limited by difficulty of obtaining such contextually-rich data in today's industrial applications. However, we hope that our and other studies will generate interest in the industry and, subsequently, more contextually-rich data will be generated by some industrial applications, which should advance the research on this topic further.

## 6. References

- [1] Aaltonen, A., P. Huuskonen, and J. Lehtikainen, Context Awareness Perspectives for Mobile Personal Media, In Aaltonen, A. and J. Lehtikainen (ed.) Personal and Ubiquitous Computing, Springer-Verlag, London, 2005.
- [2] G. Adomavicius, and A. Tuzhilin, "Expert-driven validation of rule-based user models in personalization

applications”, *Data Mining and Knowledge Discovery*, Vol. 5, No.1, 2001, pp. 33-58.

[3] G. Adomavicius, A. Tuzhilin, R. Sankaranarayanan and S. Sen, “Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach”, *ACM Transactions on Information Systems*. Vol.23, No. 1, 2005, pp. 103-145.

[4] Banavar, G., J. Black, R. Cáceres, M. Ebling, E. Stern, J. Kannry, *Deriving long-term value from context aware computing* In Aaltonen, A. and J. Lehtikainen (ed.) *Personal and Ubiquitous Computing*, Springer-Verlag London Ltd., 2005.

[5] Barnes, J.W., *Statistical Analysis for Engineers and Scientists*, McGraw Hill, Singapore, 1994.

[6] Berry, A., G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, Wiley Computer Publishing, New York USA, 1997.

[7] Bettman, J.R., M.F. Luce, and J.W. Payne, *Consumer Decision Making: A Constructive Perspective*, in Tedeschi M. (ed.), *Consumer Behavior and Decision Making*, 1991, pp.1-42.

[8] P.J. Brown, J.D. Bovey, Chen, X., “Context-aware applications: From the laboratory to the marketplace”, *IEEE Personal Communications*, Vol.4 No. 5, 1997, pp. 58–64.

[9] Byun, H.H., K. Cheverst, *Utilizing context history to provide dynamic adaptations*, Distributed Multimedia Research Group, Department of Computing, Lancaster University, Lancaster, U.K. *Applied Artificial Intelligence*, Vol. 18, 2004, pp. 533-548.

[10] Chen, G. and D. Kotz, *A Survey of Context-Aware Mobile Computing Research*, Department of Computer Science, Dartmouth College, Technical Report TR2000-381, 2000.

[11] A.K. Dey, G.D. Abowd, D. Salber, “A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications”, *Human Computer Interaction*, volume 16, No. 2, 2001, pp. 97-166.

[12] Franklin, D., J. Flischart, *All gadget and no representation makes jack a dull environment*, Proceedings of the AAAI 98 Spring Symposium on Intelligent Environments. Menlo Park, CA: AAAI Press, 1998.

[13] Hastie, T., Tibshirani, R. and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001

[14] T. Jiang and A. Tuzhilin, “Segmenting Customers from Population to Individuals: Does 1-to-1 Keep your Customers Forever?”, *IEEE TKDE*, To appear.

[15] N.M. Klein, and M. S. Yadav, “Context Effects on Effort and Accuracy in Choice: an Enquiry into Adaptive Decision Making”, *Journal of Consumer Research*. Vol. 15 No. 4, 1989, pp. 411-21

[16] Kotler, P., *Marketing Management*, Prentice Hall, London, 2003.

[17] Lehtikainen, J., *Interacting with Wearable Computers: Techniques and Their Application in Way finding Using Digital Maps*, Ph.D. thesis, University of Tampere, Department of Computer and Information Sciences, Report A-2002-2, 2002.

[18] Lilien, G.L., P. Kotler, and S.K. Moorthy, *Marketing Models*, pp. 22-23, Prentice Hall, USA, 1992

[19] J.G. Lussier, and R.W. Olshavsky, “Task Complexity and Contingent Processing in Brand Choice”, *Journal of Consumer Research*, 1979, pp. 154-165.

[20] T.Rodden, K. Cheverst, K. Davies, and A. Dix, “Exploiting context in HCI design for mobile systems”, *Proceedings of the Workshop on Human Computer Interaction with Mobile Devices*. Glasgow, Scotland, 1998.

[21] Ryan, N., J. Pascoe, and D. Morse, *Enhanced reality fieldwork: The context-aware archaeological assistant*, In V. Gaffney, M. V. Leusen, & S. Exxon(Eds.), *Computer applications in archaeology*, Oxford, United Kingdom, 1997.

[22] B. Schilit, M. Theimer, “Disseminating active map information to mobile Hosts”, *IEEE Network*, Vol. 8 No. 5, 1994, pp. 22–32.

[23] Ward, A. Jones, Hopper A. (1997), “A new location technique for the active office”, *IEEE Personal Communications*, Vol.4 No. 5, 1997, pp. 42–47.

[24] *Webster’s new twentieth century dictionary of the English language*. (1980). Springfield, MA: Merriam-Webster, Inc., 1980.