EAI Endorsed Transactions
on Security and Safety
Research Article EAI.EU

How data-sharing nudges influence people's privacy preferences: A machine learning-based analysis

Yang Lu[1,*], Shujun Li[2], Alex Freitas[2], Athina Ioannou[3]

[1]School of Science, Technology and Health, York St John University, UK
[2]School of Computing, University of Kent, UK
[3]School of Hospitality and Tourism Management, University of Surrey, UK

Abstract

INTRODUCTION: Many online services use data-sharing nudges to solicit personal data from their customers for personalized services.
OBJECTIVES: This study aims to study people's privacy preferences in sharing different types of personal data under different nudging conditions, how digital nudging can change their data sharing willingness, and if people's data sharing preferences can be predicted using their responses to a questionnaire.
METHODS: This paper reports a machine learning-based analysis on people's privacy preference patterns under four different data-sharing nudging conditions (without nudging, monetary incentives, non-monetary incentives, and privacy assurance). The analysis is based on data collected from 685 UK residents who participated in a panel survey. Their self-reported willingness levels towards sharing 23 different types of personal data were analyzed by using both unsupervised (clustering) and supervised (classification) machine learning algorithms.
RESULTS: The results led to a better understanding of people's privacy preference patterns across different data-sharing nudging conditions, e.g., our participants' preferences are distributed in a space of 48 possible profiles more sparsely than we expected, and the unexpected observation that all the three data-sharing nudging strategies led to an overall negative effect: they led to a reduced level of self-reported willingness for more participants, comparing with the case of no nudging at all. Our experiments with supervised machine learning models also showed that people's privacy (data-sharing) preference profiles can be automatically predicted with a good accuracy, even when a small questionnaire with just seven questions is used.
CONCLUSION: Our work revealed a more complicated structure of people's privacy preference profiles, which have some dependencies on the type of data nudging and the type of personal data shared. Such complicated privacy preference profiles can be effectively analyzed using machine learning methods, including automatic prediction based on a small questionnaire. The negative results on the overall effect of different data-sharing nudges imply that service providers should consider if and how to use such mechanisms to incentivise their consumers to share personal data. We believe that more consumer-centric and transparent methods and tools should be used to help improve trust between consumers and service providers.

Received on 11 September 2021; accepted on 8 December 2021; published on 21 December 2021

Keywords: Privacy, Nudging, Persuasive Technology, Data Sharing, User Segmentation, User Profiling, Machine Learning

doi:10.4108/eai.21-12-2021.172440

1. Introduction

Behavioral nudging refers to "any aspect of the choice architecture that alters people's behaviors in a predictable way without forbidding any options or significantly changing their economic incentives" [1, 2]. When being implemented in a digital environment, behavioral nudging is normally done via the use

*Corresponding author: Yang Lu (y.lu@yorksj.ac.uk). Most of Yang Lu's work was done when she was working at the University of Kent, UK.

of some specific user interface (UI) elements aiming to guide people's decision towards a very specific direction [3–6].

When applied to privacy-related applications, nudging can influence people's behaviors to disclose personal information to online services, e.g., to complete a transaction with an e-commerce website or to make a hotel booking via a travel agent [7, 8]. In this context, there are two different types of behavioral nudging depending on its purpose: 1) privacy protection nudges that help people to behave more securely in order to achieve better privacy protection (e.g., sharing less personal data to reduce privacy risks), and 2) data-sharing nudges used by online service providers to encourage their customers to share more personal data in exchange for more personalized services and/or benefits. Note that (over-)sharing personal data with service providers is the source of many privacy problems, so the second type of nudges have profound implications on privacy. Therefore, people's privacy behaviors have been extensively studied in the context of personal data-sharing decisions, e.g., research on privacy paradox (people's actual data-sharing behaviors deviate from their self-reported attitude) [9]. Data-sharing nudges are often implemented as a range of monetary or non-monetary incentivizing methods, e.g., cash returns, price discount, loyalty points, but can also be implemented in other ways, e.g., privacy assurance via a trust seal provided by an independent trusted party. To ensure that data-sharing nudging messages are effective at the individual level, some online service providers have also considered tailoring the user interfaces of their online services to present more targeted incentives mapping the user's personal preferences [10–12].

Although there is a growing number of studies focusing on both types of privacy-related nudging, it is still less understood how different types of data-sharing nudges can influence people' willingness to disclose personal information. Some researchers have argued that "one-size-fits-all" interventions should be tailored by leveraging individual differences in decision making and personality [13], which calls for personalized nudges – that in turn requires a better understanding of how people can be segmented into different profiles according to their privacy attitudes towards different types of data-sharing nudges.

In this work, we use both unsupervised (DBSCAN, *k*-means, and hierarchical agglomerative clustering) and supervised (decision tree, random forest, and naïve Bayes) machine learning algorithms to investigate the affect of three typical types of data-sharing nudges, monetary, non-monetary incentives and privacy assurances, on people's willingness to disclose personal data to service providers. The analysis is based on data collected from 685 UK residents who participated in a panel survey measuring their privacy attitudes towards personal data sharing. The data is analyzed following a three-step procedure, while the user segmentation and profiling is performed by examining how individual privacy preferences change across different data-sharing nudging conditions. The machine learning algorithms were chosen among widely used ones that tend to perform well in tasks with relatively small datasets. Multiple algorithms were considered to allow us to identify the best method for each step of the analysis procedure. The results led to a more complete understanding of people's privacy (data-sharing) preference patterns, e.g., their preferences concentrate more on a small number of profiles out of 48 possible ones, and the unexpected result that all the three data-sharing nudging strategies actually led more participants to report a reduced willingness level comparing with the no-nudging condition. Our work also showed that people's privacy (data-sharing) preference profiles could be automatically predicted with good accuracy, even if just seven features are used, which suggests that a small questionnaire with just seven questions can be used to profile users. This can help simplify the development of private data management tools while configuring the initial preference for each individual user.

The rest of the paper is organized as follows. In Section 2, we briefly review some closely related work. Section 3 explains the data used in detail and the proposed three-step procedure for automatic segmentation and profiling of individual users. The results of applying the three-step procedure to the collected data are discussed in Section 4. We cover further discussions and some limitations of our work in Section 5. Finally, Section 6 concludes the paper.

2. Related work

In this section we discuss some closely related work in two areas: how data-sharing nudges are used by service providers and how they influence people' behaviors, while also how people can be segmented based on their privacy attitudes (i.e., research on privacy typologies). We would like to highlight that, while many researchers have studied privacy attitudes and behaviors under different data-sharing nudging conditions and privacy typologies, we have not seen any work on applying machine learning to study privacy typologies in the context of different data-sharing nudging strategies, which is the focus of this paper.

2.1. Data-Sharing Nudges

Service providers have been offering different types of incentives such as monetary rewards, price discounts, loyal points, free products and services to encourage their consumers to share more personal data for more personalized services, with different

effects achieved [14]. It has been found that monetary incentives can exert positive influences in certain contexts. For instance, Hui et al. (2007) found that monetary incentives worked for a Singaporean company to boost personal data disclosure [15]. Mukherjee et al. (2013) reported that monetary incentives positively influenced not only privacy-disclosure preferences but also actual self-disclosure [16]. Similarly, Shibchurn and Yan (2014) showed that offering a monetary reward could increase the willingness levels of online social network (OSN) users to share personal data [17]. However, some other studies showed different results and below we give two examples. While requiring sensitive information, Lee et al. (2015) found that offering customers monetary benefits resulted in an increase in their privacy concerns, therefore it did not seem an effective way to encourage data sharing. Instead, they found that building trust with their customers can be a more effective mechanism for organizations to encourage their customers to share personal data [14]. In another study, Lu et al. (2018) found that monetary incentives were not better than simple email reminders for encouraging self-disclosure of personal data, in the context of an online car-sharing platform [18].

In addition to monetary incentives, engagement in online activities and self-disclosure of personal information are affected by non-monetary incentives and other factors. For instance, it was found that the use of Facebook is strongly associated with the benefits of social capital, such as self presence, accumulating friends, joining virtual groups etc. [19]. There are negative results in the literature as well, for instance, Ward et al. (2005) pointed out that neither price discounts nor personalized services had any effect in incentivizing customers to share personal information [20], and that, in addition, participants were found to be more sensitive when sharing financial information (e.g., online transaction records), while being relatively willing to share personal information.

Among all non-monetary factors studied, many researchers have studied if enhancing trust between customers and service providers via mechanisms such as self-stated privacy statements (e.g., privacy policies) and third-party trust seals can play an important role in influencing people' decisions on data disclosure. Gerlach et al. (2015) reported that the effect of a privacy policy's permissiveness on users' willingness to disclose personal information on OSNs is mediated by users' privacy risk perceptions [21]. In another study, Kobsa and Teltzrow (2004) found that people were more willing to share personal data when purchasing from online shops that attach certain descriptions of privacy practices [22]. Comparing with monetary rewards, Gabisch and Milne (2013) found that "safety cues", such as a statement reassuring users about the existence of a privacy policy and a privacy seal on a website, could be more effective in encouraging self-disclosure online [23]. Similarly, in the context of location-based social network services, Zhao et al. (2012) also reported that privacy policies (in addition to privacy controls) could help in reducing privacy concerns when sharing location-based information [24]. Hui et al. (2007) reported that by using proper privacy statements a local firm in Singapore could collect more personal data than using privacy seals [15]. In [25], it is demonstrated that neither advanced privacy conditions (i.e., *Confidential*, *Anonymized-Envelope* and *Anonymous-Postcard*) nor monetary incentives could result in higher disclosure rates of sensitive information. A similar result was reported in [26], which suggests that none of many types of privacy assurances had a direct or a moderating effect on personal information disclosure by online users in Saudi Arabia.

2.2. Privacy Typology

Research on privacy typology has shown that people could be segmented into different segments or profiles based on their privacy attitudes. For example, Harris and Westin's classic work on this topic lead to the so-called Westin's Privacy Segmentation Index, i.e., citizens can be grouped into three segments – *Fundamentalists*, *Unconcerned* and *Pragmatists*, due to different trust levels in existing laws and organizations' processes and collection of their personal data [27]. Examining Internet users' privacy concerns regarding the collection and usage of personal information [28], the original *Pragmatist* group could be further split into two subgroups, resulting in four user segments: *Unconcerned Internet*, *Circumspect Internet*, *Wary Internet* and *Alarmed Internet* users. To examine whether Westin's Privacy Segmentation Index could represent users' actual behaviors, Woodruff et al. (2014) conducted an online survey of 884 Amazon Mechanical Turk participants with detailed statements about privacy scenarios and privacy-sensitive consequences, leading to the finding that Westin's segments are not well correlated with behavioral intents or consequences [29]. Such conflicting results called for more research into this topic.

An immediate consequence of the existence of multiple user segments is that a "one-size-fits-all" solution will not be ideal for privacy protection. Therefore, prediction of the user's privacy attitude patterns (privacy profiling) is helpful in many applications. Due to the reported context-dependence attitude-consequence gap, it was also suggested to combine the contextual and cost-benefit analyses when aiming to predict privacy choices. For instance, in the context of mobile applications, in [30], user groups including *Conservatives*, *Unconcerned*, *Fence-Sitters* and *Advanced* users could be identified based on their comfort levels

towards requested app permissions with certain purposes. Towards web-based services such as the location-based service (LBS), Poikela et al. (2014) reported that users could be segmented based on the frequency and the level of accuracy of real-time shared location [31]. By inviting participants to rank privacy behaviors while using a technology service, Morton and Sasse (2014) suggested a five-group segmentation to describe users' information-seeking preferences and inform the construction of default privacy settings [32].

Privacy typology has also been studied by many researchers in the context of Internet of Things (IoT) and OSNs. For instance, for 14 IoT scenarios varying across eight factors, Naeini et al. (2017) found that over 86% of privacy (data-sharing) preferences of users can be modeled accurately [33]. Wisniewski et al. (2014) studied user privacy typology based on self-reported behaviors on privacy settings on user interfaces of Facebook, and reported six groups of user privacy behavioral profiles – *Privacy Maximizers*, *Selective Sharers*, *Privacy Balancers*, *Self-Censors*, *Time Savers* and *Privacy Minidists* [34]. In 2017 [35], they studied further profiling of Facebook users' privacy attitudes according to their awareness on privacy features, ranging from *Experts* to *Novices*. In [36], Lankton et al. (2017) examined privacy segmentation of Facebook users using two datasets on self-reported privacy strategies from undergraduate students in information systems and general users, respectively. They argued that their results could better interpret cluster differences through demographics, trust, privacy, and technology-usage perceptions. Considering that users tend to apply default privacy settings, some researchers investigated how well default privacy settings could meet users' expectations. For instance, in [37], Watson et al. (2015) concluded that using personal characterizations from relevant community data to create default privacy settings could better match users' expectations on OSNs. Based on a survey of 337 Internet users in Germany, Schomakers and Lidynia (2019) identified three user clusters with different privacy attitudes: *Privacy Guardians*, *Privacy Cynics* and *Privacy Pragmatists* [38].

In addition to the contextual dependency of user segmentation of privacy attitude and actual behaviors, some researchers also studied how the user segmentation differs across different types of data items. For instance, Knijnenburg et al. (2013) found that such data item dependency did exist in their analysis of three datasets of online information disclosure intentions and behaviors [39]. They also argued that more accurate user profiling algorithms should consider this effect to be able to help tailor the needs of different users.

Finally, we would like to mention that as a preliminary study of the present research, we conducted a clustering-based analysis of 685 UK travelers based on their self-reported willingness to share personal data, which led to two different user segments: *Privacy Pessimists* and *Privacy Rationalists* [40]. The present research reports our efforts of applying more advanced machine learning algorithms to the user segmentation problem, covering more complicated aspects such as how users' privacy attitudes change depending on the type of data-sharing nudging condition, and if and how we can automatically classify a given user into a specific user profile.

3. Data Used and Methodology

In this section, we first explain the data we used in detail, and then describe the proposed three-step procedure for segmenting and profiling individual participants based on the data.

3.1. Data Used

Our study aims to segment people according to their self-reported willingness to personal data disclosure to online travel service providers. To gather data needed for the study, we conducted an online survey with a panel of UK residents recruited through a professional survey company in May 2019, as part of the PriVELT project[1]. Participants were requested to state their level of willingness to share 23 types of personal information online. In this study, four data-sharing nudging strategies were tested: (1) no incentive; (2) monetary incentives (e.g., cash), (3) non-monetary incentives (e.g., discounts), and (4) privacy assurances (e.g., third-party trust seals). Each data item is scored on a five-point Likert scale: 1 = "strongly disagree", 2 = "disagree", 3 = "neither disagree nor agree", 4 = "agree", and 5 = "strongly agree". The questions we asked for the four different conditions are illustrated in the Appendix (Table C.1). For each of the four conditions, the participants reported how willing they would be to share the 23 different types of personal data shown in Table 1, so in total each participant reported $23 \times 4 = 92$ levels of data-sharing willingness. Note that we also include acronyms of the 23 data types in Table 1, which are used in Figures 2 and 3 to keep those figures more compact.

After cleaning the data, responses from 685 participants were considered valid. The 23-D willingness responses collected from the 685 participants were divided into four datasets, $\text{Dataset}_{\text{NI}}$, $\text{Dataset}_{\text{MI}}$, $\text{Dataset}_{\text{PA}}$ and $\text{Dataset}_{\text{NMI}}$, as the input of our data analysis process explained later, where the subscriptions in the datasets' names refer to the four data-sharing nudging conditions: NI = No Incentive; MI = Monetary Incentives; NMI = Non-Monetary Incentives;

[1] PRIvacy-aware personal data management and Value Enhancement for Leisure Travelers (https://privelt.ac.uk/)

Table 1. 23 types of personal data covered in our survey and the four datasets.

Personal Data	Meaning/Format
Name (N)	*First name(s) & last name*
Date of Birth (DoB)	*MM-DD-YYYY*
Home Address (HA)	*Street & number, city & postcode*
Email Address (EA)	*e.g., xxx.yyy@org*
Phone Number (PhN)	*11 digits XXXX-XXX-XXXX*
Profession (P)	*Organization & job title*
Education (E)	*Highest level of degree*
Credit Card Information (CCI)	*Card number, expiration date, card holder, etc.*
Bank Account Information (BAI)	*Account number, account holder, bank name, etc.*
Contacts in Address Book (CAB)	*Name, phone number, email address, etc.*
Passport Number (PaN)	*9 digits on UK passport*
Driver License Number (DLN)	*18 characters*
Fingerprint (F) Voice Sample (VS) Face Scan/Image (FS/I) Iris/Retina Pattern (I/RP)	*Biometric data*
Social Media Profile Data (SMPD)	*Username, communities, city of living etc.*
Hobbies/Personal Interests (H/PI)	
Personal Preferences (PP)	*E.g., hotel booking with requirements on room types*
Real-time Position (RP)	*GPS coordinates*
Smartphone Search History (SSH)	*Cookies*
Activity Sensor Data (ASD)	*Smartphone data such as time stamped movements*
Specific Expenses (SE)	*E.g. credit card, Paypal transaction records*

PA = Privacy Assurances. These acronyms will be used in the remaining of the paper for the sake of brevity. The demographic profiles of the 685 participants can be found in the Appendix (Table C.2).

3.2. Methodology

To analyze the above-described data for user segmentation and profiling purposes, we extended our preliminary work reported in [41] by applying a more advanced three-step procedure described below and detailed in the following three sub-subsections.

- **Step 1: Clustering** – Unsupervised clustering algorithms are applied to the four datasets to identify the best performing algorithm.
- **Step 2: Cluster Analysis** – Based on the clustering results, we look at three aspects of cluster analysis: participant distributions to clusters and user profiles, effectiveness of nudging strategies, and behavioral variances across different personal data types.
- **Step 3: Automatic Profiling** – Supervised machine learning algorithms are used to evaluate if cluster labels and user profiles across all four nudging conditions can be automatically predicted.

Step 1: Clustering. As stated earlier, each dataset includes participants' willingness responses towards 23 data types under one specific data-sharing condition, which can be represented as a 23-D vector, $\vec{w} = (w_i)_{i=1}^{23}$, where $\forall i, w_i \in \{1, 2, 3, 4, 5\}$. Since the dimensionality is relatively high, before conducting the cluster analysis, we firstly apply a Principal Component Analysis (PCA) for the purpose of dimensionality reduction. According to one of the commonly used criteria for selecting significant factors (principal components) – retaining factors with an eigenvalue greater than 1.0 [42], six factors are kept for actual clustering.

For the actual clustering part, a number of candidate clustering algorithms that may perform well should be selected and tested in order to identify the best method for further analysis. A number of clustering-evaluation metrics are needed here to compare the candidate clustering algorithms.

For non-deterministic clustering methods (such as k-means, which we actually used), the clustering result depends on the random initial condition. Therefore, it is necessary to examine the stability of the produced clusters under different random initial conditions, where the stability refers to the level of consistency of the clustering results, i.e., the same points stay in the same cluster. The procedure used to perform this stability evaluation is explained below.

For each dataset X and each number of clusters (k), the non-deterministic clustering method is run for n times varying the random initial conditions used to initialize the non-deterministic algorithm. To measure how stable the clustering results are across the n rounds, we need a quantitative metric to allow comparison and selection of the best parameters. Considering that the actual cluster labels of different runs for each (x, k) do not align and there are no ground truth labels, we define a pair-based stability metric $\mathrm{SM}_{X,k,n}$ as follows:

$$\mathrm{SM}_{X,k,n} = \frac{N_s(n)}{N_{x,k}}, \tag{1}$$

where $N_{X,k}$ is the total number of unique data-point pairs for the database X and a specific number of clusters k, and $N_s(n)$ is the number of pairs of data points that fall into the same cluster consistently for all n runs of k-means. This metric has a natural range between 0 and 1, and could be interpreted as the probability of a randomly sampled data-point pair in the dataset x never being split into two separate clusters across all n runs of clustering, when k clusters are pre-defined. A higher value of $\mathrm{SM}_{X,k,n}$ indicates a higher level of stability of the clustering results. With the above-defined stability metric, for each x we can find the best value of k giving the highest value of $\mathrm{SM}_{X,k,n}$.

By clustering a user's responses under four different data-sharing nudging conditions, a more complete picture of the user's privacy preferences to data-sharing nudging can be inferred by identifying user segments over four datasets, corresponding to their data-sharing willingness in four different conditions: "sharing for no additional benefit", "sharing for additional monetary returns", "sharing for additional non-monetary returns" and "sharing due to third-party privacy assurances". In this way, there are maximally $k_{\mathrm{NI}} \times k_{\mathrm{MI}} \times k_{\mathrm{NMI}} \times k_{\mathrm{PA}}$ possible user profiles.

Step 2: Cluster Analysis. Step 1 gives a set of clusters for each of the four data-sharing nudging conditions. In Step 2 we conduct a detailed analysis of the clustering results, focusing on the following three different but related aspects.

For the first aspect, we examine how all participants distribute to different clusters under each nudging condition and also to the $k_{\mathrm{NI}} \times k_{\mathrm{MI}} \times k_{\mathrm{NMI}} \times k_{\mathrm{PA}}$ different user profiles considering their overall behaviors across all four nudging conditions. The latter allows us to examine how some participants may "migrate" from one cluster with a lower level of data-sharing willingness to a higher one under a specific nudging condition, or vice versa, therefore providing some useful evidence about effectiveness of different nudging strategies. To capture the "cluster-migration rate", i.e., the portion of participants in one cluster X_i (the i-th cluster under the nudging condition X) who "come from" another cluster Y_j (the j-th cluster under the nudging condition $Y \neq X$), we define a similarity metric as follows:

$$S_{X_i,Y_j} = \frac{\#(X_i \cap Y_j)}{\#(Y_j)}, \tag{2}$$

where $\#(X)$ denotes the cardinality of a set X. This similarity has a natural range of $[0, 1]$, and a higher value indicates a higher level of membership overlap between these two clusters. Note that this similarity metric is asymmetric, i.e., $S_{X_i,Y_j} = S_{Y_j,X_i}$ does not hold in general, since the cluster-migration rates for both directions can be different. We focus on the cluster migration rates from clusters under the no-nudging condition NI to other three nudging conditions because this migration direction can give more useful information about how data-sharing nudges can influence participants' behaviors.

The second aspect will extend analysis on the effectiveness of the three nudging conditions by looking at to what extent they are able to successfully nudge participants towards a *higher* willingness level to share personal data. Different from the first aspect, which focuses more on collective behaviors inferred from the average data-sharing level per cluster, here we examine how data sharing nudging strategies influence individual participants' data-sharing willingness (increase, decrease or no change).

The third aspect is about how participants' data-sharing willingness levels vary across all 23 types of personal data. We expect participants will have different privacy (data-sharing) preferences on different data types, and it is interesting to know how such type-specific behaviors may affect our overall analysis results.

Step 3: Profile Prediction. In this step, we look at how to build four separate classifiers, each automatically predicting the cluster label of a given 23-D data point $\vec{w}$ (i.e., an individual participant's response) under a specific nudging condition. The four classifiers jointly can predict the participant's overall user profile. Having such classifiers also allows us to have a different set of evaluation criteria for the clustering performance in Step 1 because we would like to choose the clustering method and relevant parameters to optimize the prediction accuracy, too. In our experiments, we decided to choose decision tree (DT), random forest (RF) and naïve Bayes (NB) as the candidate classification algorithms because they can be trained based on a smaller dataset to achieve a reasonably good performance. The first two algorithms can also return indicators of importance of different features for selecting a smaller number of important features for classification, which will allow using a small questionnaire to profile users

and therefore improve usability of such user profiling / personalization systems. Once the most important features for classification are determined, we re-train the classification models using only those features, and measure their predictive accuracy.

4. Results

4.1. Step 1: Clustering

We decided to choose three well-established clustering algorithms of different types as candidates: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [43], k-means [44] and hierarchical agglomerative clustering (HAC) [45]. We used implementations of these algorithms in the widely used library scikit-learn[2] (0.21.3) running with Python 3.7. We set the parameters $k_{\max}$ (the maximum number of clusters) and $t_{\max}$ (the number of runs for each (x,k)) both to 10. For cluster-evaluation metrics, we decided to use four widely adopted indices: silhouette index (S), Calinski-Harabasz index (CH), Davies-Bouldin (DB) and S-Dbw (SD) index [46–51]. In order to determine the value of $k_{\max}$, we ran one of the three clustering algorithm DBSCAN with $k_{\max} = 10$ and it failed to produce any clustering results for $k = 9$ and $k = 10$ under two nudging conditions MI and NMI (see Tables C.4 and C.5). We considered that these results indicate that meaningful clusters cannot be identified for $k > 10$, so we decided to set $k_{\max} = 10$ for our clustering experiments. Tables C.3–C.6 show the comparison results of the 3 clustering algorithms for each of these indices and each value of k in $\{2, 3, \ldots, 10\}$. We identified the **best** clustering algorithm for each of the four datasets by choosing the algorithm with the most "wins" among all the performance metrics. For instance, Table C.5 shows that for Dataset_NMI, k-means is the best when $k = 4$ given that it achieved more best scores than others, and it is the best across all k values for the same reason.

Here we focus only on a summary of these results, as reported in Table 2. In this table, Columns 1 and 2 show the number of clusters (k) and the clustering algorithm, respectively. In Columns 3–6, each cell in this table shows, for each combination of k and algorithm, the number of clustering quality indices (out of 4) for which the algorithm was ranked the best, for a given dataset (as identified in the heading of Columns 2–6). The rightmost column shows the total number of times the algorithm was ranked the best across the 4 datasets, out of 16 times (4 datasets times 4 index values per dataset), for each value of k.

The results clearly show that the k-means algorithm achieved the best overall performance, which is stable across all values of k. More precisely, considering all k values shown in Table 2, k-means is ranked the best in 107 cases, whilst HAC and DBSCAN are ranked the best in only 20 and 17 cases, respectively. In addition, k-means outperformed the other two clustering methods for all 9 values of k. Hence, k-means was selected for further analysis.

The next step of the result analysis consists of selecting the best value of k for k-means. This is defined in Eq. (1) for computing the value of $\mathrm{SM}_{x,k}$ of each dataset x and each value of k in $\{2, 3, \ldots, 10\}$. The results are shown in Table 3, where the number of clusters corresponding to the largest possible value of $\mathrm{SM}_{x,k}$ (meaning the most stable cluster results – no change at all across all 10 runs) for each dataset is highlighted in light gray. In general the most stable clustering results were obtained with smaller k values, between 2 and 5.

In addition to training classifiers to do automatic profiling of participants in our survey, we are also interested in how good the clustering results are for supporting the automatic profiling in the Step 3. Therefore, to further validate the results in Table 3, we built some classifiers (based on decision tree, random forest and naïve Bayes, as mentioned in the previous section) and compared their predictive performance indicators. One classifier was built for each best clustering clustering result produced by the k-means algorithm (i.e., those gray cells shown in Table 3). The predictive accuracy of these classifiers was evaluated through a five-fold cross validation. Each classifier was trained using a training set, which includes 80% of all data points in one of the four datasets and the corresponding cluster IDs as the class labels. The remaining 20% data points in each dataset was used as the test set to calculate the performance indicators of the trained classifier.

In Table 3, there is a single best number of clusters (k) for the datasets NI and PA – $k = 2$. However, there is a tie between the two best values of k for datasets MI and NMI. Therefore, by using the classifiers to further validate the clustering results, we can also determine the best k value for the MI and NMI datasets between the two values with a tie.

Table 4 shows the results of three widely used predictive accuracy measures – the accuracy (i.e., the average percentage of correct predictions across all classes), the weighted-average F1-score and the Area Under the ROC Curve (AUC) after binarizing class labels – for all three classifiers we built. As observed in Table 4, for all three performance metrics and all three classifiers, the better value of k is 3 for the MI dataset and 4 for the NMI dataset. Hence, these k values are selected for further analysis for these two datasets. Besides, all the three classifiers trained with the selected k values obtained a high predictive performance, but random forest and naïve Bayes models performed better

[2] Also known as sklearn: https://scikit-learn.org/

Table 2. Comparison of clustering performances.

k	**Algorithm**	$\text{Dataset}_{\text{NI}}$	$\text{Dataset}_{\text{MI}}$	$\text{Dataset}_{\text{PA}}$	$\text{Dataset}_{\text{NMI}}$	**Sum**
2	DBSCAN	0	1	0	0	1
	k-means	1	3	4	2	10
	HAC	3	0	0	2	5
3	DBSCAN	1	0	0	0	1
	k-means	3	3	4	2	12
	HAC	0	1	0	2	3
4	DBSCAN	1	0	0	1	2
	k-means	3	2	3	2	10
	HAC	0	2	1	1	4
5	DBSCAN	2	0	1	1	4
	k-means	2	2	3	3	10
	HAC	0	2	0	0	2
6	DBSCAN	0	0	1	1	2
	k-means	4	4	3	3	14
	HAC	0	0	0	0	0
7	DBSCAN	1	0	1	0	2
	k-means	3	4	2	4	13
	HAC	0	0	1	0	1
8	DBSCAN	0	0	1	0	1
	k-means	3	4	3	3	13
	HAC	1	0	0	1	2
9	DBSCAN	1	0	1	0	2
	k-means	3	4	3	4	14
	HAC	0	0	0	0	0
10	DBSCAN	1	0	1	0	2
	k-means	2	3	2	4	11
	HAC	1	1	1	0	3

Table 3. Stability metrics of clustering results of k-means.

k	NI	MI	NMI	PA
2	1.00000	0.99766	0.97244	1.00000
3	0.98932	1.00000	1.00000	0.98298
4	0.99791	0.97932	1.00000	0.99456
5	0.99809	1.00000	0.97436	0.97653
6	0.97797	0.98920	0.96131	0.99602
7	0.96659	0.98788	0.97084	0.98713
8	0.92148	0.94397	0.96886	0.95987
9	0.94470	0.95525	0.95978	0.94978
10	0.93811	0.95281	0.94728	0.93280

than decision trees with a significant margin for all settings.

4.2. Step 2: Cluster Analysis

In this subsection, we report results of our cluster analysis on all the three aspects described in Section 3.2.

Participants vs. Clusters and User Profiles. How participants distribute to the different clusters under each nudging condition (produced by the k-means clustering algorithm in Step 1) is shown in Table 5, from which we can see that under each nudging condition different clusters have significantly different values of $\overline{w}$ (the average data-sharing willingness level across all participants belonging to a cluster). We also conducted a one-way (between-subject) ANOVA to check if the differences are statistically significant, and the results are positive for all with $p < 0.001$: NI – $F(1, 683) = 401$; MI – $F(2, 682) = 1295$; NMI – $F(3, 681) = 854$; PA – $F(1, 683) = 1233$. For MI and NMI, since there are more than two groups a Tukey's post-hoc test was run and the results showed significant differences between all group pairs. Based on the statistical results, we name the clusters according to the values of $\overline{w}$: L (low) or H (high) for the nudging conditions NI and PA ($k = 2$); L (low), M (medium) or H (high) for MI ($k = 3$); L (low), ML (medium low), MH (medium high) or H (high) for NMI ($k = 4$). Accordingly, we use these names to differentiate the clusters under the same conditions, e.g., NMI_{ML} refers to the ML cluster under the nudging condition NMI. We also conducted a separate one-way ANOVA to compare differences between all 11 clusters across different nudging conditions, which

Table 4. Performance indicators of all classifiers we built for predicting and further validating the user clustering results of k-means. For MI and NMI datasets, we highlight the higher accuracy in bold face for the two possible values of k.

Classifier	Metric	NI	MI		NMI		PA
		$k=2$	$k=3$	$k=5$	$k=3$	$k=4$	$k=2$
DT	Accuracy	0.892	**0.880**	0.799	0.876	**0.880**	0.879
	F1-score	0.890	**0.880**	0.799	0.876	**0.880**	0.879
	AUC	0.885	**0.925**	0.895	0.922	**0.940**	0.900
RF	Accuracy	0.955	**0.924**	0.877	0.899	**0.921**	0.921
	F1-score	0.954	**0.924**	0.876	0.899	**0.922**	0.921
	AUC	0.995	**0.990**	0.989	0.985	**0.993**	0.983
NB	Accuracy	0.942	**0.927**	0.893	0.909	**0.928**	0.949
	F1-score	0.942	**0.927**	0.895	0.910	**0.929**	0.949
	AUC	0.984	**0.987**	0.983	0.979	**0.987**	0.990

gave $F(10, 2729) = 540$ with $p < 0.001$. A Tukey's post-hoc test revealed that a significant difference exists between all cluster pairs except for the following ones: $(\mathrm{NI_H}, \mathrm{PA_H})$, $(\mathrm{MI_H}, \mathrm{NMI_H})$, $(\mathrm{NMI_H}, \mathrm{PA_H})$, $(\mathrm{MI_L}, \mathrm{NMI_L})$, and $(\mathrm{NMI_{ML}}, \mathrm{PA_L})$. Based on the results, in the following analysis we do not consider there are significant differences between these cluster pairs.

If we consider participants' overall behaviors across all four nudging conditions, we can observe some more interesting behavioral patterns as shown in Figure 1. Surprisingly, the distribution is much sparser than we expected: only 9 profiles stand out with a significant portion of participants, and each of the other 39 profiles has just a few (no more than 12) participants so can be considered background noises. Probably not surprisingly, two (the most and the third most) "popular" profiles turned out to be $(\mathrm{NI_L}, \mathrm{MI_L}, \mathrm{NMI_L}, \mathrm{PA_L})$ and $(\mathrm{NI_H}, \mathrm{MI_H}, \mathrm{NMI_H}, \mathrm{PA_H})$, representing *un-incentivizable privacy fundamentalists* and those *privacy-unconcerned*, respectively. The remaining 7 profiles can be classified into four sub-groups: 1) 4 profiles (those close to the left bottom corner) represent *moderately incentivizable privacy fundamentalists*; 2) the profile $(\mathrm{NI_H}, \mathrm{MI_H}, \mathrm{NMI_{MH}}, \mathrm{PA_H})$ representing those with *slight privacy concerns over non-monetary incentives*; 3) the profile $(\mathrm{NI_M}, \mathrm{MI_M}, \mathrm{NMI_{MH}}, \mathrm{PA_H})$ representing *privacy pragmatists happy with privacy assurance more than monetary and non-monetary incentives*; and 4) the profile $(\mathrm{NI_L}, \mathrm{MI_M}, \mathrm{NMI_{MH}}, \mathrm{PA_H})$ that represents *pragmatic privacy fundamentalists who are incentivizable, more so by non-monetary incentives and privacy assurance than monetary incentives*. To summarize, the above behavioral patterns lead to the following observation: 70.5% of participants fall into two ends of the spectrum – over half of all participants (51.2%) are more like privacy fundamentalists and 18.4% have no or only slight privacy concerns; only two profiles (which cover only 9.6% of all participants) can be considered privacy pragmatists; and all the other 39 profiles cover the remaining 19.9% of participants.

Cluster-migration rates from the two NI clusters to cluster in other three nudging conditions are shown in Table 6. As a general trend, all three data-sharing nudging strategies have no (L to L or H to H) or only a moderate (L to ML/M or H to MH/M) influence on most participants, although a small portion of participants changed their data-sharing willingness level drastically (L to H or H to L). This observation is aligned with the results shown in Figure 1.

Data-Sharing Nudges vs. Individual Behaviors. The results discussed in the previous sub-subsection already give some indication on the lack of effectiveness of all the three data nudging strategies, but the analysis is based on collective behaviors at the cluster level. In this sub-subsection we look at how the three data nudging strategies influenced individual participants' self-reported data-sharing willingness levels, to get more direct evidence on their actual effectiveness. Denoting the difference of the self-reported data-sharing willingness level reported by a participant under a nudging condition C and that under no-nudging condition by $\Delta_C(w)$, Table 7 shows the median and mean values of $\Delta_C(w)$ for all the three nudging conditions, from which we can see a surprising result – all the three nudging strategies actually led more participants to report a reduced willingness level comparing with the no-nudging condition. The observed failure of all nudging strategies even holds for participants in $\mathrm{NI_H}$.

One may argue that the direction of change in the value of w (i.e., no change, increased or decreased) matters more than how much it changed. Table 8 shows the percentages of participants who fall into these three categories. The results are indeed more revealing: for all nudging strategies, there are more participants with a decreased willingness level than those with an increased level. This implies that, while all the three nudging strategies can work for some participants, they failed for more participants so the overall effect is a failure (according to their purpose of

Table 5. The size # (i.e., the number of participants), the percentage and the average data-sharing willingness level $\overline{w} \in [1, 5]$ of all clusters (C = nudging condition).

C	NI		MI			NMI				PA	
i	L	H	L	M	H	L	ML	MH	H	L	H
#	482	203	204	306	175	202	187	169	127	404	281
%	70%	30%	30%	45%	26%	30%	27%	25%	19%	59%	41%
$\overline{w}$	2.19	3.45	1.31	2.38	3.57	1.27	2.04	2.82	3.54	1.97	3.38

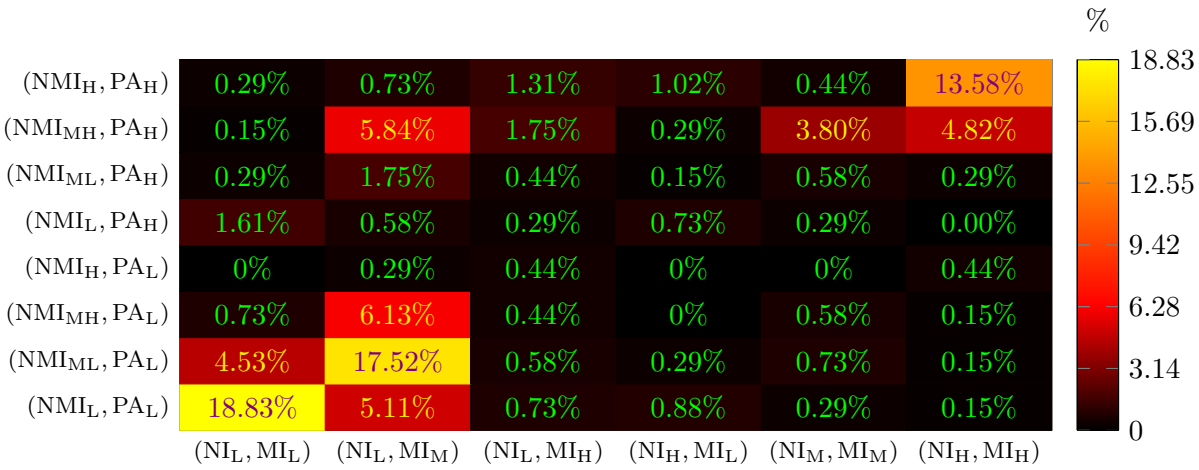

Figure 1. Distributions of participants to all 48 user profiles.

Table 6. Similarity scores between the two NI clusters and the clusters under the other three nudging conditions, calculated according to Eq. (2) (C = nudging condition).

C	MI			NMI				PA	
i	L	M	H	L	ML	MH	H	L	H
S_{C_i,NI_L}	0.38	0.54	0.09	0.39	0.36	0.21	0.04	0.79	0.21
S_{C_i,NI_H}	0.11	0.23	0.66	0.08	0.07	0.33	0.52	0.12	0.88

Table 7. Median and mean (± standard deviation) values of $\Delta_C(w)$ under different nudging conditions (C).

C	Participants in NI_L	Participants in NI_H	All participants
MI	-0.13 (−0.14 ± 0.69)	-0.09 (−0.34 ± 0.87)	-0.13 (−0.20 ± 0.75)
NMI	-0.22 (−0.26 ± 0.64)	-0.17 (−0.33 ± 0.73)	-0.22 (−0.28 ± 0.67)
PA	-0.04 (0.01 ± 0.53)	0 (−0.09 ± 0.62)	-0.04 (−0.02 ± 0.56)

increasing the overall data-sharing willingness level). This result came as a surprise to us, and has a profound implication on if and how service providers should use data-sharing nudges at all to solicit personal data from their customers.

While all the three nudging strategies failed to work as a whole, data in Tables 7 and 8 also show the order of overall effect of the three nudging strategies: PA > MI > NMI. The fact PA works better than MI implies that service providers should consider how to increase trust of their customers on their services rather than relying on monetary or non-monetary incentives.

Data Type vs. Data-Sharing Willingness. All our previous analysis is based on the data-sharing willingness level averaged across all 23 different types of personal data. Figure 2 shows how the willingness level varies across those data types. Despite the visible variations, for all 23 data types and all four nudging conditions, participants in a L cluster always reported a lower average level of willingness than those in the corresponding H cluster. Under the MI condition, participants in the H cluster reported a lower average level of willingness than those in the M cluster for all data types except for one (N). Under the NMI condition, results are more mixed: differences are less clear among

Table 8. Percentages of participants with an unchanged (=), increased (+) and decreased (-) value of w under different nudging conditions (C).

C	Participants in NI_L			Participants in NI_H			All participants		
	=	+	-	=	+	-	=	+	-
MI	7%	31%	**62%**	12%	28%	**60%**	8%	30%	**61%**
NMI	6%	25%	**69%**	7%	30%	**63%**	7%	26%	**67%**
PA	9%	39%	**52%**	15%	39%	**46%**	11%	39%	**51%**

ML, MH and H clusters for 10 data types (N, DoB, HA, EA, PhN, CCI, P, E, PaN, PP), showing the boundary between these clusters are not always a clear cut for these data types. As a whole, we felt the results based on averaging across 23 data types should still give largely reliable results since the mixed results have a limited effect. However, we believe that more future research is needed to further explore the data type-specific aspects.

Looking at all the results, another observation stands out: across all four nudging conditions, participants in H-clusters and those in non-H-clusters behaved particularly differently for the following types of personal data: BAI, CAB, PaN, DLN, F, VS, FS/I, I/RP, SMPD, SSH. Comparing with other data types, the above types seem generally more sensitive such as biometric features (F, VS, FS/I, I/RP), finance-related (CCI, BAN), and more private data (PaN, DLN, SSH). SMPD seems a notable exception – many people have public profiles on social media so it remains unclear why some participants had more concerns on this type of personal data. More future research will be required to understand more about this particular observation.

4.3. Step 3: Profile Prediction

In Section 4.1, we have shown how an individual's profile can be predicted by training four classifiers, one for each nudging condition, and then used the profile predictor to help further validate the clustering results and determine the the "best" k values for k-means clustering. We have also shown the performance of the classifiers are generally good (see Table 4). In this subsection we look at how the number of input features of the four classifiers can be reduced from $23 \times 4 = 92$ to a more usable number in real-world applications, i.e., a user needs to answer only fewer questions (ideally just up to a handful) to let the system set up his/her privacy (data-sharing) preference profile.

To help determine the most important features for all the four classifiers we built, we show all 23 features for each of the four trained classifiers in Figure 3, visualizing the significance level ([0,1]) of each feature (measured by running the method feature_importances_ in the scikit-learn library). From the results shown in Figure 3, we can see that, for decision trees (with parameter min_sample_leaf = 2) only a very small number of features have a high significance level, therefore we can easily choose the most important features. While for random forests (with parameter n_estimators = 1,000), it seems more difficult to select significant variables as their weights are more evenly distributed. The significance patterns of different nudging conditions are also significantly different, so we need to select the reduced feature subsets for the four classifiers separately.

When decision trees are used to build classifiers, we are able to select only 7 features out of 92 features across the four classifiers. For the no-nudging condition, the significance of willingness to share *Voice* data (significance level = 0.68) is much higher than the significance of any other features, and thus only this feature is selected for profile prediction. Similarly, a significantly reduced feature set can be determined for the other three nudging conditions: (*Email Address*, *Face Scan/Image*) (significance level = 0.33, 0.25) for monetary incentives, (*Name*, *Education*, *Fingerprint*) (significance level = 0.29, 0.2, 0.23) for non-monetary incentives, and *Activity Sensors* (significance level = 0.41) for privacy assurances. Hence, for these experiments with the most significant variables only, when using the decision tree algorithm, only 7 out of the 23 variables were used across the four datasets, in order to predict the class labels in four nudging conditions. In the case of random forests, we simply selected the five most significant features from each nudging condition to reduce the total number of features to 20. The prediction accuracy of the classifiers built with the reduced number of features can be seen in Table 9. While the performance drops in general, the accuracy is very good for the no-nudging condition (over 96% for both decision trees and random forests) and still reasonably high for the other three conditions: $\geq 79\%$ for decision trees and $\geq 88\%$ for random forests. Note that in real-world applications, a relatively accurate prediction is often sufficient and better than what is currently being used (a default profile for all or multiple profiles the user has to manually choose).

5. Further Discussions and Limitations

5.1. Further Discussions

The results reported in the previous section lead to a number of interesting observations about privacy

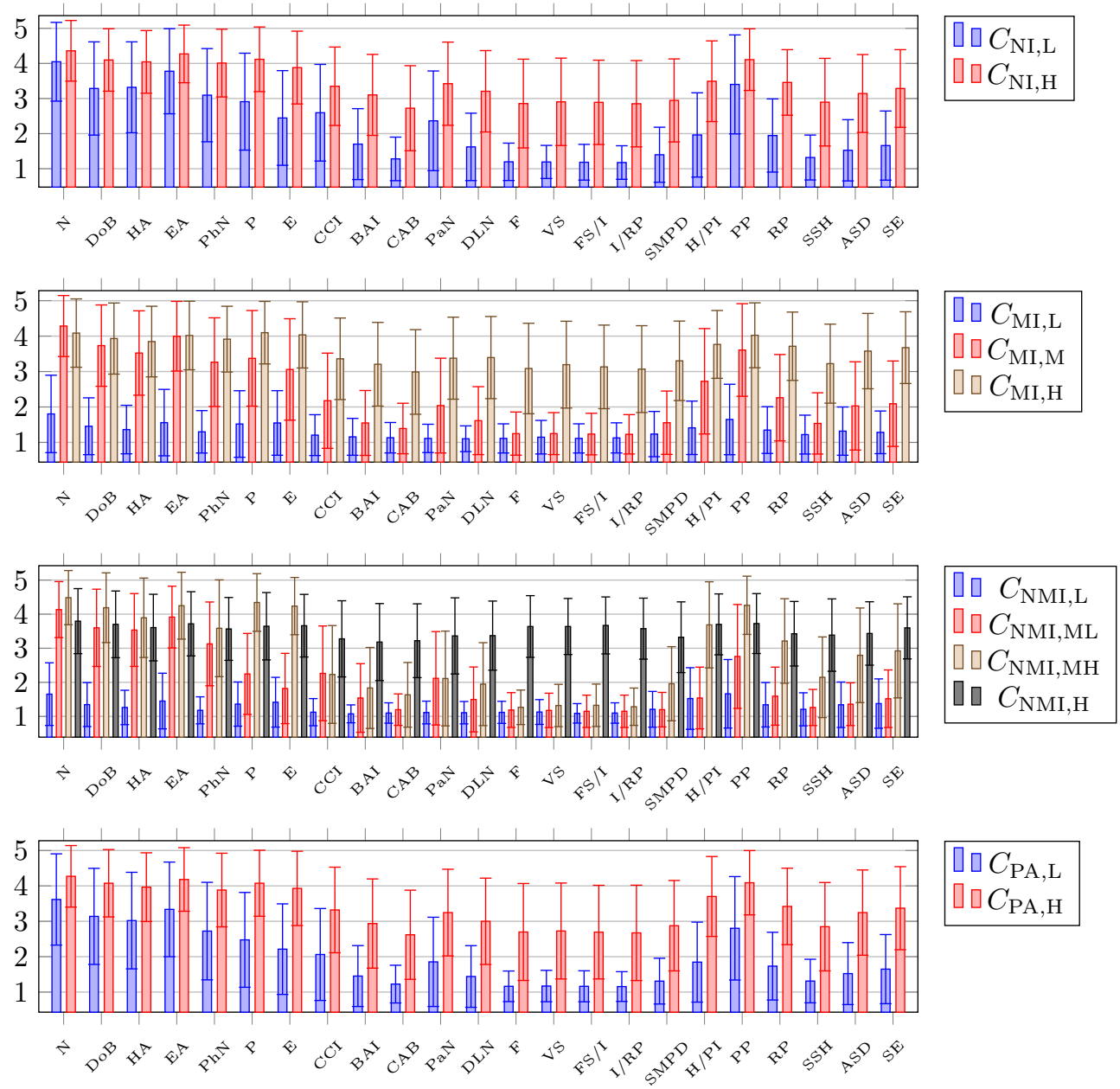

Figure 2. Means and standard deviations of the data type-specific data-sharing willingness levels w, under the four different nudging conditions, shown as error bar charts.

Table 9. Prediction accuracy of classifiers trained with a reduced number of important features only. The number within brackets indicates the reduced number of features used.

	NI, $k = 2$	MI, $k = 3$	NMI, $k = 4$	PA, $k = 2$
DT	0.964 (1)	0.854 (2)	0.861 (3)	0.796 (1)
RF	0.978 (5)	0.891 (5)	0.883 (5)	0.920 (5)

(data-sharing) preferences of individuals. First, our study draws a more comprehensive picture of 48 user profiles of individual privacy (data-sharing) preferences, showing that people's privacy preference patterns could be changed by data-sharing nudging strategies in different ways, which suggest that personalized solutions are needed. The surprising observation on the lack of overall effectiveness of all three data-nudging strategies indicates that people's privacy (data-sharing) preferences can be more complicated and service providers should re-consider if and how monetary and non-monetary incentives should be used. As we reviewed in Section 2.1, conflicting results on data-sharing nudges have been reported in the literature, so more research is needed to clarify if

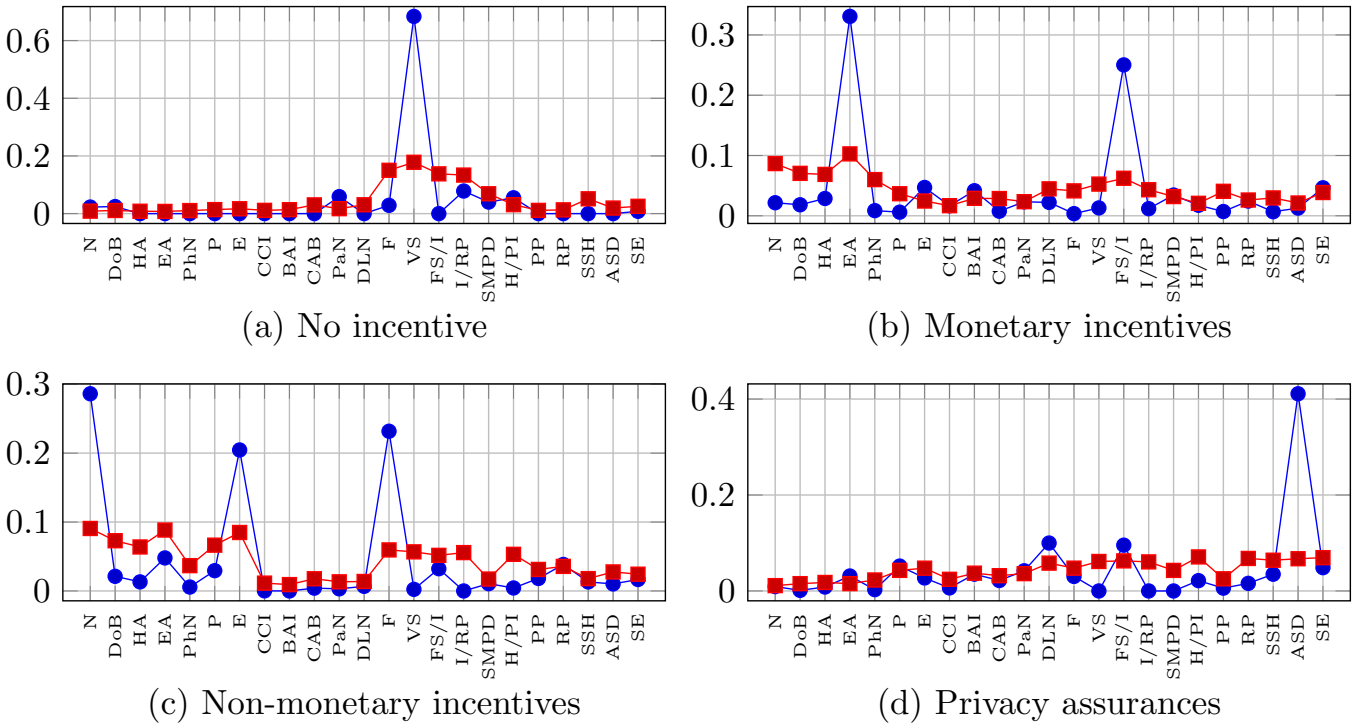

Figure 3. Feature significance levels of decision trees ( ) and random forests ( ).

what we observed in this research can be reproduced in other experiments and different contexts. Second, our results on automatic profile prediction show that it is possible to use four separate classifiers with a small number of features (as few as 7 in total) to capture the complicated user profiling problem, therefore potentially offering service providers a better tool to serve their customers with enhanced usability.

In addition to the direct implications of our results for service providers mentioned above, our work can provide useful insights to more parties of the larger data economy ecosystem. For instance, many independent personal data management platforms (PDMPs)[3] have been developed to give users more control of their own data and to empower users to trade data with service providers. Such PDMPs can use the user profiling method reported in this paper to engage users and better serve users, including recommending services to users according to their privacy (data-sharing) preferences and special requirements on utility. In addition, new businesses opportunities can be generated around user-facing tools to help individual users using the work reported in this paper, e.g., a data sharing awareness tool can help users better understand how they are sharing data with multiple entities online and in the physical world, and a service comparison tool can assist users to choose the best service based on their personal privacy (data-sharing) preferences, including switching to physical services to avoid sharing more sensitive data online. Furthermore, our work could help policy makers such as national data protection authorities, e.g., to define more granular regulations and guidelines for different businesses sectors and user groups.

Due to the sensitive nature of many personal data and privacy as a basic human right[4], our work has important links to issues around ethics, transparency, trust and legal obligations of different parties in the data economy ecosystem. For instance, ignoring consumers' privacy wishes and constantly nudging them for more personal data is likely unethical if not illegal. Our results reported in this paper re-confirmed previous findings that some people would become more alerted and less willing to share personal data if they see any data-sharing nudges, which could be explained by the lack of trust between some consumers and service providers. Such a lack of trust is often caused by the lack of transparency about service providers' data collection and processing practices.

[3]Some examples: Solid (https://solid.mit.edu/), HAT (https://www.hubofallthings.com/), Databox (https://www.databoxproject.uk/) and digi.me (https://digi.me/).

[4]As defined in the Universal Declaration of Human Rights (UDHR, https://www.un.org/en/universal-declaration-human-rights/) and human right laws in many jurisdictions.

Service providers should therefore increase the level of transparency of their processes, and provide more diverse solutions to people with different privacy (data-sharing) preferences, which can help build a more privacy-friendly ecosystem. Having a more user-centric ecosystem will eventually benefit service providers especially those who respect consumers more and engage them more actively. Our work can offer service providers tools towards such a direction.

5.2. Limitations

Since our analysis is based on data collected from an online survey, it can reflect only self-reported privacy attitudes rather than actual behaviors in real world. However, as widely reported in past studies around the privacy paradox theory [9], the self-reported willingness may not be consistent with actual data-sharing behaviors, which can be affected by various factors such as how information is presented and the context of the data sharing. We plan to investigate actual data-sharing behaviors in selected real world scenarios and see if how behavioral profiles may change. This will allow us to look at how people's privacy attitudes and behaviors evolve over time, e.g., via a longitudinal study.

As with all empirical studies, some biases may have been introduced in the data collection procedure and the experimental design. For instance, in the survey we conducted, for all questions the four nudging conditions were presented in one fixed order, i.e., "No Nudge" → "Monetary Incentive" → "Privacy Assurances" → "Non-Monetary Incentive". To identify and avoid such biases, we will consider re-validating our work under different experimental setups.

In addition, the panel survey we conducted may have attracted people in different base-line clusters (NI_L and NI_H) unevenly so the results on the overall population may be biased. Despite this uncertainty, aggregated results on both clusters in Section 4.2 showed that the affects of the three nudging strategies on the two NI clusters are largely aligned with just a small margin, so we believe that our main conclusions should hold.

While we believe the data we collected are sufficiently representative and the main insights are reliable, there are some factors that may affect the generalizability of the reported results especially at the lower level, e.g., which profiles out of the 48 ones are more dominating. For instance, our pool of participants was limited to a single country (UK) and the number of participants (685) may not be large enough to allow examination of some profiles, especially those that are less common. In addition, our survey was put into a more realistic context of data sharing with online travel service providers, which may not be able to capture different responses in other contexts. We will conduct more future work to further validate our reported results with participants from other countries.

When performing the first step, i.e., the clustering, a prior step was taken to reduce the dimensionality using PCA. Since our datasets contain 23 dimensions ($n = 23$), using PCA was helpful to reduce the dimensionality to 6 ($n = 6$) and thus avoid problems with the computation of distances between examples (samples) in datasets, where all examples would be essentially far away from each other due to the high dimensionality (if $n = 23$ was used). Such a pre-processing step has been commonly used by researchers for clustering high-dimensional data [52–57]. PCA has the drawback of being global so it cannot preserve pairwise distance between points in the original space. To address this problem, we can use a dimensional reduction algorithm that can preserve such distances better [58], e.g., t-SNE [59]. We plan to test such distance-preserving dimensional reduction algorithms in our future work and compare the results with what are reported in this paper with PCA.

For clustering we tested three algorithms, and for automatic profile prediction we tested three algorithms – decision trees, random forests and naïve Bayes. While those chosen algorithms gave good results, other algorithms may perform even better. We will explore these possibilities in future work.

6. Conclusions

This paper presents our work on utilizing both unsupervised and supervised machine learning algorithms to analyze 685 UK residents' self-reported willingness levels to share 23 types of personal data with online travel companies, under four different data-sharing nudging conditions (no nudge, monetary and non-monetary incentives, and privacy assurances). By applying a three-step data analysis process, we revealed more comprehensive user segmentation results when we consider how different types of data-sharing nudges influence people's data-sharing behaviors. We also showed that, using four classifiers and a small number of features (as few as seven), people's data-sharing behavioral profiles can be predicted with good accuracy. The results reported provide new insights on how people's privacy preferences interact with data-sharing nudges and how machine learning methods can be used to analyze and capture such interactions. They can find direct applications in many real-world scenarios, e.g., online booking for flights and hotels, OSN-based communication, and web forum discussions, to help better balance people's wishes for privacy protection and service providers' desires to provide more personalized services.

Acknowledgement. This work was funded by the EPSRC (Engineering and Physical Sciences Research Council), part of the UKRI (UK Research and Innovation), under the grant number EP/R033749/1.

References

[1] Thaler RH, Sunstein CR. Nudge: Improving decisions about health, wealth, and happiness. Penguin; 2009.

[2] Sunstein CR, Thaler RH. Libertarian paternalism is not an oxymoron. The University of Chicago Law Review. 2003:1159-202.

[3] Ridley-Siegert T. Data privacy: What the consumer really thinks. Journal of Direct, Data and Digital Marketing Practice. 2015;17(1):30-5.

[4] Weinmann M, Schneider C, vom Brocke J. Digital nudging. Business & Information Systems Engineering. 2016;58(6):433-6.

[5] Acquisti A, Adjerid I, Balebako R, Brandimarte L, Cranor LF, Komanduri S, et al. Nudges for privacy and security: Understanding and assisting users' choices online. ACM Computing Surveys. 2017;50(3):44.

[6] Schneider C, Weinmann M, vom Brocke J. Digital nudging: guiding online user choices through interface design. Communications of the ACM. 2018;61(7):67-73.

[7] Gómez-Barroso JL. Experiments on personal information disclosure: Past and future avenues. Telematics and Informatics. 2018;35(5):1473-90.

[8] Weinmann M, Schneider C, vom Brocke J. Digital nudging: guiding online user choices through interface design. Communications of the ACM. 2018;61(7):67-73.

[9] Gerber N, Gerber P, Volkamer M. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. Computers & Security. 2018;77:226-61.

[10] Schöning C, Matt C, Hess T. Personalised Nudging for more Data Disclosure? On the Adaption of Data Usage Policies Format to Cognitive Styles. In: Proceedings of 52nd Hawaii International Conference on System Sciences. University of Hawaii at Manoa, USA; 2019. p. 4395-404.

[11] Kankane S, DiRusso C, Buckley C. Can we nudge users toward better password management?: An initial study. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. ACM; 2018. p. LBW593:1-LBW593:6.

[12] Acquisti A, Brandimarte L, Loewenstein G. Privacy and human behavior in the age of information. Science. 2015;347(6221):509-14.

[13] Warberg L, Acquisti A, Sicker D. Can Privacy Nudges be Tailored to Individuals' Decision Making and Personality Traits? In: Proceedings of 18th ACM Workshop on Privacy in the Electronic Society. ACM; 2019. p. 175-97.

[14] Lee H, Lim D, Kim H, Zo H, Ciganek AP. Compensation paradox: the influence of monetary rewards on user behaviour. Behaviour & Information Technology. 2015;34(1):45-56.

[15] Hui KL, Teo HH, Lee SYT. The value of privacy assurance: an exploratory field experiment. MIS Quarterly. 2007:19-33.

[16] Mukherjee S, Manjaly JA, Nargundkar M. Money makes you reveal more: consequences of monetary cues on preferential disclosure of personal information. Frontiers in Psychology. 2013;4:839.

[17] Shibchurn J, Yan XB. Investigating effects of monetary reward on information disclosure by online social networks users. In: Proceedings of 2014 47th Hawaii International Conference on System Sciences. IEEE; 2014. p. 1725-34.

[18] Lu Y, Ou C, Angelopoulos S. Exploring the effect of monetary incentives on user behavior in online sharing platforms. In: Proceedings of 51st Hawaii International Conference on System Sciences. University of Hawaii at Manoa, USA; 2018. p. 3437-44.

[19] Ellison NB, Steinfield C, Lampe C. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication. 2007;12(4):1143-68.

[20] Ward S, Bridges K, Chitty B. Do incentives matter? An examination of on-line privacy concerns and willingness to provide personal and financial information. Journal of Marketing Communications. 2005;11(1):21-40.

[21] Gerlach J, Widjaja T, Buxmann P. Handle with care: How online social network providers' privacy policies impact users' information sharing behavior. The Journal of Strategic Information Systems. 2015;24(1):33-43.

[22] Kobsa A, Teltzrow M. Contextualized Communication of Privacy Practices and Personalization Benefits: Impacts on Users' Data Sharing and Purchase Behavior. In: Privacy Enhancing Technologies: 4th International Workshop, PET 2004, Toronto, Canada, May 26-28, 2004. Revised Selected Papers. vol. 3424 of Lecture Notes in Computer Science. Springer; 2004. p. 329-43.

[23] Gabisch JA, Milne GR. Self-disclosure on the web: Rewards, safety cues, and the moderating role of regulatory focus. Journal of Research in Interactive Marketing. 2013;7(2):140-58.

[24] Zhao L, Lu Y, Gupta S. Disclosure intention of location-related information in location-based social network services. International Journal of Electronic Commerce. 2012;16(4):53-90.

[25] Murdoch M, Simon AB, Polusny MA, Bangerter AK, Grill JP, Noorbaloochi S, et al. Impact of different privacy conditions and incentives on survey response rate, participant representativeness, and disclosure of sensitive information: a randomized controlled trial. BMC Medical Research Methodology. 2014;14(1):90.

[26] Al-Jabri IM, Eid MI, Abed A. The willingness to disclose personal information. Information & Computer Security. 2019.

[27] Louis H, Westin AF. Harris-Equifax consumer privacy survey 1991; 1991. A report by Equifax Inc. by Louis Harris & Associates.

[28] Sheehan KB. Toward a typology of Internet users and online privacy concerns. The Information Society. 2002;18(1):21-32.

[29] Woodruff A, Pihur V, Consolvo S, Brandimarte L, Acquisti A. Would a Privacy Fundamentalist Sell Their DNA for $1000... If Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences. In: Proceedings of 10th Symposium On Usable Privacy and Security. USENIX Association; 2014. p. 1-18.

[30] Lin J, Liu B, Sadeh N, Hong JI. Modeling users' mobile app privacy preferences: Restoring usability in a sea of

permission settings. In: Proceedings of 10th Symposium On Usable Privacy and Security. USENIX Association; 2014. p. 199-212.
[31] Poikela M, Schmidt R, Wechsung I, Möller S. Locate! – When do Users Disclose Location? In: Proceedings of Workshop on Privacy Personas and Segmentation (PPS) at 10th Symposium On Usable Privacy and Security. USENIX Association; 2014. p. 1-5. Available from: https://cups.cs.cmu.edu/soups/2014/workshops/privacy/s3p2.pdf.
[32] Morton AJ, Sasse MA. Desperately seeking assurances: segmenting users by their information-seeking preferences: AQ methodology study of users' ranking of privacy, security & trust cues. In: Proceedings of 2014 12th Annual International Conference on Privacy, Security and Trust. IEEE; 2014. p. 102-11.
[33] Naeini PE, Bhagavatula S, Habib H, Degeling M, Bauer L, Cranor LF, et al. Privacy expectations and preferences in an IoT world. In: Proceedings of 13th Symposium on Usable Privacy and Security. USENIX Association; 2017. p. 399-412.
[34] Wisniewski P, Knijnenburg BP, Lipford HR. Profiling Facebook users privacy behaviors. In: Proceedings of SOUPS 2014 Workshop on Privacy Personas and Segmentation. USENIX Association; 2014. p. 1-6.
[35] Wisniewski PJ, Knijnenburg BP, Lipford HR. Making privacy personal: Profiling social network users to inform privacy education and nudging. International Journal of Human-Computer Studies. 2017;98:95-108.
[36] Lankton NK, McKnight DH, Tripp JF. Facebook privacy management strategies: A cluster analysis of user privacy behaviors. Computers in Human Behavior. 2017;76:149-63.
[37] Watson J, Lipford HR, Besmer A. Mapping user preference to privacy default settings. ACM Transactions on Computer-Human Interaction. 2015;22(6):32.
[38] Schomakers EM, Lidynia C, Ziefle M. A Typology of Online Privacy Personalities. Journal of Grid Computing. 2019:1-21.
[39] Knijnenburg BP, Kobsa A, Jin H. Dimensionality of information disclosure behavior. International Journal of Human-Computer Studies. 2013;71(12):1144-62.
[40] Lu Y, Ioannou A, Tussyadiah I, Li S. Segmenting travelers based on responses to nudging for information disclosure. e-Review of Tourism Research. 2019;17(3):394-406.
[41] Yang S, Wang K. The influence of information sensitivity compensation on privacy concern and behavioral intention. ACM SIGMIS Database: the DATABASE for Advances in Information Systems. 2009;40(1):38-51.
[42] Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960;20(1):141-51.
[43] Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Transactions on Database Systems. 2017;42(3):1-21.
[44] Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010;31(8):651-66.
[45] Day WHE, Edelsbrunner H. Efficient algorithms for agglomerative hierarchical clustering methods. Journal of Classification. 1984;1(1):7-24.
[46] Thalamuthu A, Mukhopadhyay I, Zheng X, Tseng GC. Evaluation and comparison of gene clustering methods in microarray analysis. Bioinformatics. 2006;22(19):2405-12.
[47] Starczewski A, Krzyżak A. Performance evaluation of the silhouette index. In: Artificial Intelligence and Soft Computing: 14th International Conference, ICAISC 2015, Zakopane, Poland, June 14-18, 2015, Proceedings, Part II. vol. 9120 of Lecture Notes in Computer Science. Springer; 2015. p. 49-58.
[48] Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics-theory and Methods. 1974;3(1):1-27.
[49] Davies DL, Bouldin DW. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979;PAMI-1(2):224-7.
[50] Halkidi M, Vazirgiannis M. Clustering validity assessment: Finding the optimal partitioning of a data set. In: Proceedings 2001 IEEE International Conference on Data Mining. IEEE; 2001. p. 187-94.
[51] Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation measures. In: Proceedings of 2010 IEEE International Conference on Data Mining. IEEE; 2010. p. 911-6.
[52] Parsons L, Haque E, Liu H. Subspace clustering for high dimensional data: a review. ACM SIGKDD Explorations Newsletter. 2004;6(1):90-105.
[53] Kriegel HP, Kröger P, Zimek A. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Transactions on Knowledge Discovery from Data. 2009;3(1).
[54] Sebzalli YM, Wang XZ. Knowledge discovery from process operational data using PCA and fuzzy clustering. Engineering Applications of Artificial Intelligence. 2001;14(5):607-16.
[55] Wang XD, Chen RC, Yan F, Zeng ZQ, Hong CQ. Fast adaptive *K*-means subspace clustering for high-dimensional data. IEEE Access. 2019;7:42639-51.
[56] Rathore P, Kumar D, Bezdek JC, Rajasegarar S, Palaniswami M. A rapid hybrid clustering algorithm for large volumes of high dimensional data. IEEE Transactions on Knowledge and Data Engineering. 2018;31(4):641-54.
[57] Lee S, Kim T. Search Space Reduction for Determination of Earthquake Source Parameters Using PCA and-Means Clustering. Journal of Sensors. 2020;2020.
[58] van der Maaten L, Hinton G. Distance-preserving dimensionality reduction. WIREs Data Mining and Knowledge Discovery. 2011;1(5):369-80.
[59] van der Maaten L, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research. 2008;9:2579-605. Available from: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.

Appendix A. Survey questions for disclosure willingness

Table C.1 shows the questions we used to collect self-reported levels of willingness to share the 23 types of personal data from human participants in the panel survey.

Appendix B. Demographic information of human participants

Table C.2 shows some basic statistics on the demographic information of human participants of the user study.

Appendix C. Determining the best clustering method

Tables C.3–C.5 show the detailed performance metrics of all clustering algorithms we applied to all the four databases corresponding to the nudging conditions, for determining the best clustering method.

Table C.1. Questions we used to capture the level of willingness to disclose personal information with online travel companies. Each question is followed by a list of 23 data types and a 5-point Likert scale.

Nudging Condition	Items
No Nudge	"How willing are you to share the following information with online travel companies?"
Monetary Incentives	"Should you receive monetary incentives (i.e., cash), how willing are you to share the following information with online travel companies?"
Non-Monetary Incentives	"Should you receive non-cash incentives (such as coupons and discounts), how willing are you to share the following information with online travel companies?"
Privacy Assurances	"If the online company is providing privacy assurances (such as an easy to read privacy policy) about the protection of your personal data, how willing are you to share the following information with online travel companies?"

Table C.2. Demographic profiles of all human participants (N = 685).

Demographic characteristic	Features	Percentage (%)
Gender	Male	47.15
	Female	52.41
	Other	0.44
Age (years)	25 or younger	4.82
	26–35	23.94
	36–45	12.26
	46–55	17.23
	56–65	22.19
	over 65	19.56
Education	Less than high school	2.92
	High school	38.83
	Bachelor's degree	34.45
	Master's degree	14.31
	PhD/Doctoral degree	3.94
	Other	5.55
Frequency of travel	1-2 times per year	33.87
	3-4 times per year	36.93
	More than 4 times per year	29.20
Frequency of online shopping	Daily	9.78
	Several times a week	21.02
	Several times a month	41.32
	Roughly once a month	23.50
	Almost never	4.38

Table C.3. Performance of user clustering ($Dataset_{NI}$).

Metric		DBSCAN (D)	k-means (K)	HAC (H)	Best
$k = 2$	S	0.008	0.298	0.349	H
	CH	44.559	310.220	204.861	K
	DB	1.593	1.292	0.960	H
	SD	1.076	1.196	1.053	H
$k = 3$	S	-0.076	0.260	0.198	K
	CH	57.372	305.391	225.800	K
	DB	1.200	1.342	1.301	D
	SD	1.072	1.048	1.224	K
$k = 4$	S	-0.034	0.241	0.205	K
	CH	55.837	274.006	237.771	K
	DB	1.392	1.371	1.401	K
	SD	1.112	1.178	1.272	D
$k = 5$	S	-0.125	0.227	0.200	K
	CH	46.030	247.373	212.119	K
	DB	1.334	1.380	1.521	D
	SD	1.243	1.464	1.293	D
$k = 6$	S	-0.120	0.230	0.182	K
	CH	53.244	234.26	193.284	K
	DB	1.689	1.337	1.439	K
	SD	1.316	1.224	1.279	K
$k = 7$	S	-0.160	0.222	0.165	K
	CH	38.300	222.745	180.731	K
	DB	1.144	1.325	1.469	D
	SD	1.256	1.122	1.273	K
$k = 8$	S	-0.159	0.221	0.150 6	K
	CH	44.481	209.744	169.485	K
	DB	1.479	1.345	1.497	K
	SD	1.463	1.233	1.226	H
$k = 9$	S	-0.17	0.216	0.146	K
	CH	36.055	199.121	160.126	K
	DB	1.106	1.378	1.49 1	D
	SD	1.445	1.222	1.250	K
$k = 10$	S	-0.161	0.203	0.145	K
	CH	37.571	185.255	152.538	K
	DB	1.315	1.399	1.566	D
	SD	1.650	1.615	1.306	H
End of Table					

Table C.4. Performance of user clustering ($Dataset_{MI}$).

Metric		DBSCAN (D)	k-means (K)	HAC (H)	Best
$k = 2$	S	0.012	0.346	0.332	K
	CH	81.010	463.115	425.800	K
	DB	1.582	1.158	1.205	K
	SD	0.989	1.046	1.066	D
$k = 3$	S	0.018	0.324	0.271	K
	CH	70.737	427.812	331.866	K
	DB	1.549	1.185	1.205	K
	SD	1.069	.992	0.971	H
$k = 4$	S	0.033	0.313	0.276	K
	CH	105.662	377.193	315.944	K
	DB	1.278	1.309	1.268	H
	SD	1.179	0.942	0.914	H
$k = 5$	S	0.005	0.320	0.278	K
	CH	88.851	350.365	297.708	K
	DB	1.265	1.291	1.240	H

Continuation of Table C.4					
Metric		**DBSCAN (D)**	***k*-means (K)**	**HAC (H)**	**Best**
	SD	1.209	1.079	0.950	H
$k = 6$	S	-0.004	0.326	0.274	K
	CH	79.416	334.567	284.012	K
	DB	1.293	1.190	1.391	K
	SD	1.342	1.054	1.140	K
$k = 7$	S	-0.028	0.329	0.252	K
	CH	72.697	319.228	261.962	K
	DB	1.268	1.161	1.358	K
	SD	1.255	1.034	1.134	K
$k = 8$	S	0.043	0.329	0.255	K
	CH	62.776	319.228	248.332	K
	DB	1.258	1.161	1.389	K
	SD	1.411	1.034	1.115	K
$k = 9$	S	NA	0.322	0.265	K
	CH	NA	285.074	239.369	K
	DB	NA	1.270	1.380	K
	SD	NA	1.039	1.097	K
$k = 10$	S	0.008	0.321	0.268	K
	CH	66.721	270.381	229.357	K
	DB	1.331	1.225	1.334	K
	SD	1.334	1.084	1.072	H
End of Table					

Table C.5. Performance of user clustering ($\text{Dataset}_{\text{NMI}}$).

Metric		**DBSCAN (D)**	***k*-means (K)**	**HAC (H)**	**Best**
$k = 2$	S	0.028	0.356	0.297	K
	CH	91.533	472.667	340.444	K
	DB	1.539	1.130	0.972	H
	SD	1.001	1.037	0.922	H
$k = 3$	S	0.038	0.334	0.271	K
	CH	106.643	423.579	321.449	K
	DB	1.390	1.195	0.984	H
	SD	1.117	0.956	0.926	H
$k = 4$	S	0.005	0.331	0.273	K
	CH	99.461	391.390	320.465	K
	DB	1.227	1.294	1.321	D
	SD	1.196	1.031	0.949	H
$k = 5$	S	-0.004	0.334	0.307	K
	CH	81.566	361.857	318.903	K
	DB	1.125	1.253	1.237	D
	SD	1.134	0.933	0.965	K
$k = 6$	S	-0.001	0.331	0.303	K
	CH	73.022	338.671	296.393	K
	DB	1.133	1.184	1.355	D
	SD	1.254	1.001	1.211	K
$k = 7$	S	0.006	0.335	0.295	K
	CH	94.013	330.169	280.170	K
	DB	1.426	1.205	1.353	K
	SD	1.238	0.979	1.173	K
$k = 8$	S	-0.045	0.335	0.295	K
	CH	84.165	315.980	264.054	K
	DB	1.490	1.201	1.300	K
	SD	1.370	0.995	0.984	H
$k = 9$	S	NA	0.334	0.271	K
	CH	NA	301.091	253.002	K
	DB	NA	1.194	1.283	K

Continuation of Table C.5					
Metric		**DBSCAN (D)**	***k*-means (K)**	**HAC (H)**	**Best**
	SD	NA	1.219	1.138	K
$k = 10$	S	NA	0.332	0.270	K
	CH	NA	287.920	242.973	K
	DB	NA	1.214	1.298	K
	SD	NA	1.113	1.146	K
End of Table					

Table C.6. Performance of user clustering ($\text{Dataset}_{\text{PA}}$).

Mertic		**DBSCAN (D)**	***k*-means (K)**	**HAC (H)**	**Best**
$k = 2$	S	.0180	0.318	0.293	K
	CH	73.695	389.328	350.373	K
	DB	2.050	1.245	1.320	K
	SD	1.125	1.079	1.094	K
$k = 3$	S	-0.015	0.281	0.257	K
	CH	75.688	349.108	279.382	K
	DB	1.437	1.306	1.353	K
	SD	1.027	0.991	1.118	K
$k = 4$	S	-0.062	0.277	0.204	K
	CH	71.926	314.660	257.368	K
	DB	1.496	1.356	1.405	K
	SD	1.214	1.037	1.032	H
$k = 5$	S	-0.089	0.277	0.227	K
	CH	51.031	293.509	253.661	K
	DB	0.999	1.278	1.356	D
	SD	1.251	1.046	1.106	K
$k = 6$	S	-0.077	0.258	0.224	K
	CH	60.531	279.631	242.466	K
	DB	1.044	1.249	1.313	D
	SD	1.327	1.082	1.109	K
$k = 7$	S	-0.076	0.262	0.228	K
	CH	63.501	268.403	240.777	K
	DB	1.125	1.212	1.281	D
	SD	1.423	1.043	1.026	H
$k = 8$	S	-0.088	0.265	0.226	K
	CH	64.001	254.786	224.765	K
	DB	1.145	1.268	1.379	D
	SD	1.291	1.076	1.209	K
$k = 9$	S	-0.090	0.260	0.219	K
	CH	64.865	238.436	210.855	K
	DB	1.218	1.248	1.459	D
	SD	1.279	1.004	1.207	K
$k = 10$	S	-0.084	0.260	0.217	K
	CH	63.402	225.423	199.442	K
	DB	1.259	1.318	1.455	D
	SD	1.386	1.235	1.219	H
End of Table					