Title

Who Should Participate in DR Program? Modeling with Machine Learning and Credit Scoring

Authors
Lee, Kyungeun
Lee, Eunjung
Lee, Hyesu
et al.

Publication Date
2017-11-20

Who Should Participate in DR Program?
Modeling with Machine Learning and Credit Scoring

Kyungeun Lee, Eunjung Lee*, Hyesu Lee**, Hyoseop Lee**, Yoonjin Yoon[o], Wonjong Rhee**

* Department of Transdisciplinary Studies, Seoul National University, Seoul, Republic of Korea; ruddms0415@snu.ac.kr, ej-lee@snu.ac.kr, wrhee@snu.ac.kr

** Encored Technologies, Seoul, Republic of Korea; hyesu@encoredtech.com, hslee@encoredtech.com

[o] Department of Civil and Environmental Engineering, KAIST, Daejeon, Republic of Korea; yoonjin@kaist.ac.kr

Abstract: In this study, we consider a residential DR program of incentive-based peak power reduction where invitation for participation can be sent selectively. The selective process can be crucial for improving efficiency of the program for two reasons. First, there are customers who do not change their behavior at all but take rewards due to the natural variations in their life patterns. Second, too many notifications can cause adversarial effects where participants turn off the notification channels or make complaint calls. For the selective process, obviously the process needs to be made as efficient as possible, but it is also essential to maximize the explainability of the selection process such that the operation of the program can be made smooth. To address this problem, we propose a customer participation behavior prediction model considering both accuracy and explainability, where the accuracy advantage of Machine Learning (ML) and the explainability advantage of Credit Scoring (CS) are combined. For the study, data was collected from 15,091 households in Korea for one year in 2016. ML algorithms, with up to 56 features, were studied and showed a fairly high prediction performance (AUROC 0.9576), but they were too complicated to satisfy explainability. A CS method of classing with a scorecard was adopted, where its explainability has been heavily tested and proven in the financial sector already. Direct adoption of general CS, however, does not guarantee an acceptable accuracy performance because energy data is quite different from financial data. To this end, we define a modified CS method using general CS as the base but with additional rules for high prediction performance. While this modified CS method maintains its explainability via a well-defined scorecard, it also shows comparable prediction performance as ML's (AUROC 0.9509). The modified CS method is expected to affect residential DR in a positive way. Its high accuracy for predicting customer participation behavior means a large potential for improving efficiency. Its explainability means not only an easier interaction with customers but also less effort for educating call-center agents who need to deal with the customers.

Keywords: Incentive-based Peak Reduction, Residential DR, Participation Prediction, Explainability, Machine Learning, Credit Scoring

---

1. Introduction

Household sector's power consumption has been steadily increasing, and it is expected to become 41% of overall consumption within the next 20 years [1]. Besides the large share, household sector can be influential because the educational activities therein can play a significant role in improving social awareness of energy use. In these contexts, residential Demand Response (DR) is considered as an important tool for solving the imbalance problem of power supply and demand. The residential DR was first launched in the United States in 2010. Between the two types of residential DR programs, Price-Based DR (PBDR) and Incentive-Based DR (IBDR), the latter is known to provide both efficiency and reliability in the initial adoption period because it allows a flexibility in consumer choice and does not require an extensive investment on dedicated infrastructure. Note that IBDR utilizes incentives instead of fines, and therefore it can reduce conflicts during the early introduction of DR [2]. The IBDR usually operates in the form of an invitation-based event that is scheduled when a peak reduction is deemed required.

Because the IBDR cannot force a customer to participate or not to participate, a "selective process" can be helpful in two ways. First, a selective process can exclude customers who opportunistically take rewards. Because of natural variations in people's life patterns, some can occasionally take rewards without making any power-reduction efforts simply by participating in all events. This issue is pronounced for residential DR because more fluctuations in life patterns can occur for a residential customer than a business customer. Therefore, a DR provider can improve its operational efficiency by having an option to detect and exclude such customers. Second, a selective process can improve customer satisfaction by adaptively controlling how often invitation and reminder notifications should be sent. For IBDR to work, sending invitations is necessary. Furthermore, reminder notifications have been shown to be helpful for improving effectiveness of residential DR [3]. More notifications, however, can have a downside of irritating some customers, and eventually cause complaint calls or opt-outs. For this matter, a selective process can be used to determine and control if a particular customer should receive few or many notifications.

So far, most of the existing studies have focused on rebate amount optimization [4, 5], and not so much on selective process. Despite of possible improvement in operational efficiency and customer satisfaction, a selective invitation or action also means a discriminatory participation opportunity that can lead to fairness issues. As shown in [6, 7], energy customers must be protected whenever fairness can be, and providers have a legal obligation to make their processes transparent and take actions that have causal relationships. Furthermore, residential DR requires easy and concise explanations because it deals with a large number of customers who do not have vocational knowledge on energy industry. Therefore, explainability is an important requirement for residential DR.

In this study, we develop a prediction model that can provide both accuracy performance and explainability. Accuracy is an important factor for making selective process effective, and explainability is a basic requirement for customer protection and regulation compliance. We study three approaches to understand the tradeoffs. First, a simple rule-based approach is considered, where the algorithm is allowed to use only a single feature that can be obtained directly from the raw data. This approach provides a high level of explainability. Second, a Machine Learning (ML) approach is considered, where the algorithm fully utilizes extensive feature engineering and feature selection. This approach is known to provide a high level of accuracy performance. While ML has been a popular choice for solving a variety of prediction problems, its black-box nature means very low level of explainability [8]. Third, Credit Scoring(CS) approach is considered, where it can provide a superior balance between accuracy performance and explainability. CS is a widely adopted method in the finance discipline, and it can produce fairly simple models that have been shown to work well in financial sector by incorporating nonlinear relationships in an explainable way [8, 9]. For studying the three approaches, a residential DR dataset was used, where the dataset is from a pilot study conducted in Korea for one year.

2. Explainability & Credit Scoring

2.1 Explainability

2.1.1 Importance of Explainability in Energy Market

With the increase in automated decision makings using big data and machine learning, many customers are becoming concerned of personal information protection. To address the concerns, legal regulations are being implemented starting from EU's 'General Data Protection Regulation (GDPR)' [7] and 'Communication on Delivering a New Deal for energy consumers (New deal)' [6]. GDPR was passed in May 2016, and it will take effect in 2018. It is intended for providing transparency and fairness when personal data is used for an automated decision making. According to GDPR, the data subject has the right to monitor all the processes and to receive understandable explanations [7]. In other words, the model used in the automated decision making must be designed in an explainable way such that an answer can be provided when a customer raises a question or files a complaint. According to EU's New deal (2015) [6], an explainable modeling is required in the energy market for the same reason. In the energy market, all the decision processes related to energy services must be transparently disclosed to the customers, and the customers have the right to monitor and change their program options freely at any time. DR and a selective process also need to follow the same regulations. For residential DR programs, general public are the targets, and therefore even simpler explanations need to be made available.

2.1.2 Definition of Explainability

Prediction performance is easy to quantify, and a variety of metrics such as accuracy, AUROC (Area Under Receiver Operation Characteristics), and AUPRC (Area Under Precision-Recall Curve) are available. On the other hand, explainability is not easy to define because its requirements are dependent on the person's background and knowledge base. Because of this subjective nature of explainability, most of the existing studies define and evaluate explainability indirectly. For instance, explainability is defined as shown in Table 1 in the studies of [10, 11].

Table 1. Definition of explainability according to [10, 11]

Condition	Description	How to make it more explainable
Less Complex	Use less parameters in the model	Use less features/classes
More Reliable	Ensure prediction performance above a certain level	Should be predictable at least as much as the simplest model
More Usable	Provide information that assists users to accomplish a task in a brief form	Use a form that everyone already knows
More Applicable	Anyone can understand and implement in a few seconds/minutes	Do not need any additional knowledge or program to implement
More Transparent	Do not use sensitive features Use a white-box approach	Provide all information to customers in every step
More Trusted	Use a convincing method for everyone	Must have a reasonable explanation in every step

According to the previous studies, explainability can be defined using heuristics, but it cannot be evaluated using a quantitative metric. In particular, [11] points out that the only way to guarantee explainability is to use a model that is already known as explainable. Therefore, we adopt two known explainable approaches in this study – simple rule-based approach and Credit Scoring approach. Credit Scoring is known as an explainable classifier and it is also compatible with the definitions of explainability in [9, 12, 13].

2.2 Credit Scoring: Explainable Model

Credit Scoring(CS), originally devised by Durand in 1941, sets the score for each range through logistic regression or linear programming of features that were generated from the customer's historical data. CS uses a scorecard, the simple and familiar form [9, 12, 13], such that customers with no expertise in statistics can understand. Primarily, CS is used by banks to minimize the financial risks while providing explainability. For example, if a person visits a bank to obtain a mortgage loan, the banker predicts the person's loan repayment possibility through a CS model and decides how much the person can borrow, at what interest rate, and for how long. For the score evaluation, an example is to give -10 points for delinquency in the last 3 months, +50 points for average salary of $5,000 or more, +20 points for owning a car, etc. In this case, if a customer is denied of a loan, the customer might file a complaint on the bank's fairness. For the purpose of handling this issue, financial sector implemented legal regulations such as 'The National Credit Act of 2005 states in article 62' where banks have a legal obligation to disclose the decision-making process and to clearly explain the causal relationship in each process. In other words, an explainable model must be used in situations where fairness can be questioned by the customers.

Besides meeting the explainability requirement, CS has been actively used in the industry for over 60 years because of its high prediction performance. There are two reasons for the high prediction performance. First, overfitting is controlled by extracting only the features that are necessary for the prediction. Secondly, nonlinearity is handled by adaptively choosing the ranges for different scores. Since CS performs classing based on probabilistic basis rather than scale or class of raw data, it improves prediction performance compared to the simple case of using raw data as it is. To improve the prediction performance of CS, Machine Learning(ML) techniques can be applied to the feature selection or feature generation steps. When ML techniques are adopted, however, the limitation of simple scorecard should be maintained such that explainability is preserved. In our study, we also applied ML techniques in a few steps like feature selection and scoring, but the final model remains as an understandable scorecard.

3. Data & Feature Description

We used incentive-based residential DR pilot study data of 15,091 households in Korea. The dataset contains energy measurement data and DR participation log data for one year (2016.01.01~2016.12.31). For all of 65 DR events, a notification was sent out a day before through smartphone app, and 1~2 possible reminders were sent on the event day. In the pilot, each event lasted for an hour. Due to the data availability issue, the first 4 events were excluded from the study. Datasets for training, validation, and test were fixed and used in the same way for evaluating all three approaches. In the pilot, typical participation rate was around 10% because the events were not very actively advertised. Because of the skewness in the participation and non-participation ratio, AUROC (Area Under Receiver-Operator Curve) was used as the performance metric [14, 15].

Raw data consisted of three parts. First dataset included event day information that is common to all customers (date, start hour, weather, temperature, and amount of base incentive). Second dataset included each customer's personal event records (participation, peak-reduction goal successful, reduction goal, and usage). Third dataset contained the 15-minute electricity consumption data of each customer. The data was collected through smart-meters. Joining of the data was possible using unique customer ID. From the raw datasets, 56 features were selected or generated by processing the raw data. The list of features is shown in Appendix A. While the raw data of general CS in financial sector has more than 100 independent and user-specific features, DR data is less complex and we ended up with only 56 features in the final feature set. As will be shown later, 56 turns out to be sufficient to achieve a high accuracy.

4. Modeling

To compare performance and explainability tradeoffs, we investigate three different approaches with different objectives in modeling step. First, a simple rule-based approach uses only the features from the raw data without any processing at all. Therefore, a simple rule-based approach serves as the simplest baseline while meeting explainability requirement. Secondly, ML approaches freely use many features that were heuristically or systematically generated. While ML does not meet the explainability requirement, it serves as a benchmark and provides an upper limit of performance. Finally, the CS approach uses only a small number of features with scorecard such that the results are still well explainable. In terms of accuracy performance, CS is between simple rule-based and ML.

4.1 Simple rule-based Approach: Focusing on Explainability Only

As the baseline, the simple rule-based approach is allowed to use only one of the most important features from the raw dataset. We used Kolmogorov-Smirnov statistics (KS) in order to define important features for participation prediction. KS is a distance measure of the cumulative distribution functions of two groups, and it quantifies how well a feature discriminates the two groups [16]. After evaluating KS of all raw data features in the training dataset, we selected the top 10 as the important features. Then, the best prediction rule was determined according to AUROC of validation dataset as shown in Table 2. It turns out that the best single feature is the 'participation of the last event', and therefore simple rule-based was decided to make a prediction purely based on the last event's participation record.

Table 2. Model Selection in Simple Rule-based approach

Feature	KS	Rule (Predicted to participate in)	AUROC
Success of the last event	0.5132	Succeeded	0.6758

Participation of the last event	0.4937	Participated	**0.7674**
Success of the 2nd last event	0.4599	Succeeded	0.6653
Participation of the 2nd last event	0.4275	Participated	0.7450
Success of the 4th last event	0.3931	Succeeded	0.6261
Success of the 3th last event	0.3920	Succeeded	0.6361
Participation of the 4th last event	0.3837	Participated	0.7017
Participation of the 3th last event	0.3715	Participated	0.6985
Success of the 5th last event	0.3455	Succeeded	0.6110
Rebate of the last event	0.3340	More than 0 won	0.6731

4.2 ML Approach: Focusing on Prediction Performance Only

ML approach utilizes all the features listed in Appendix A, and it only focuses on performance without considering explainability at all. We investigated four popular ML algorithms: Logistic Regression (LR), Gradient Boosting Method (GBM), Neural Network (single-hidden-layer, NN), and Naïve Bayes (NB). In the training step, 5-fold cross validation was employed. Furthermore, we thoroughly investigated six feature selection methods to maximize the AUROC performance. Five of them were adopted from general ML feature selection methods – Stepwise Forward Selection, Stepwise Backward Elimination, Correlation-based Feature Selection, Chi-square Feature Selection, and Best First Search [8, 17, 18, 19]. The sixth one was from Credit Scoring modeling using Information Value (IV), and the details are explained in 4.3.

To understand the important features chosen by the feature selection methods, we define the top 10 features chosen by each method as the important features. Table 3 shows which features were frequently chosen as the important feature by the six feature selection methods. Not surprisingly, participation rate is chosen as an important feature by all six methods, and it is the only feature to be selected by six. 'Participation of the last event', that was chosen for simple rule-based approach, turns out to be selected five times as important features. 'Participation of the 2nd last event' was also chosen by five.

Table 3. Important features – 'important feature count' is the number of feature selection methods that have chosen the feature to be important (top 10).

Important feature count	Features
6	Participation Rate
5	Participation of the last event, Participation of the 2nd last event
4	Cumulative Rebate, Success of the last event, Past Maximum Rebate
3	Participation of the 3rd last event, The number of maximum consecutive success, Consecutive participation rate, Success of the 4th last event
2	The number of participation, The number of maximum consecutive participation, The number of success
1	Event date, The number of events, Consecutive non-participation rate, Consecutive success rate, Success rate, Recent rebate, Success of the 2nd last event, Success of the 3rd last event,

Participation of the 4th last event, Success of the 5th last event,

Success of the 6th last event, Success of the 7th last event

4.3 CS Approach: Focusing on Both Explainability and Prediction Performance

CS approach's modeling process is transparent and reasonable because the result is scorecard and because simple scorecards have been proven to be quite understandable by general public [9, 12]. CS approaches that are typically used in industry have a weakness where they apply a number of heuristics or assumptions. In this study, in order to strengthen the explainability for the energy data, we modify the original CS modeling steps in [8, 20] and add new rules based on mathematical and statistical perspectives.

4.3.1 Feature Selection

Feature selection in CS approach refers to the step of selecting a few important features necessary for achieving performance while maintaining explainability. General CS approaches select important features based on Information Value (IV) that is defined below as equation (1). IV measures how well two groups (Y=0 or 1) are separated by a feature (X). Features with IV value of 0.1 or greater are defined as important features [8, 20].

$$\mathrm{IV} = \sum\nolimits_i (\Pr(Y = 1|x_i) - \Pr(Y = 0|x_i)) * \log\left(\frac{\Pr(x|Y = 1)}{\Pr(x|Y = 0)}\right)_{x=x_i} \tag{1}$$

4.3.2 Fine Classing

Fine classing is the step of binning the continuous feature into 20 classes or less. Missing value is regarded as a separate class. For example, if 95% data of the "Annual Income" is between \$10,000 and \$200,000, it is broken down into 20 intervals with step-size of \$9,500 (=(200,000-10,000)/20). If a feature has less than 20 distinct values, all values are treated as separate classes. We adopted the typical linear binning as the fine classing method, and we binned each selected feature into 20 classes with equal intervals for the range of 95% of the data.

4.3.3 Coarse Classing

While fine classing provides a simple and linear way to define classes, having a good population count balance among the classes is important for explainability. In this way, customers can understand why the person was selected or not selected. Coarse classing consolidates classes into more stable and statistically significant classes. General CS approach approximates the probability of the two groups by logistic regression, and logit is in linear relationship with SCORE variable (2-4). The SCORE variable is the score of each feature.

$$P(Y = Participation \mid X = range1) = \frac{exp -(\beta_0+\beta_1*SCORE)}{1+exp -(\beta_0+\beta_1*SCORE)} \tag{2}$$

$$P(Y = Non - Participation \mid X = range1) = \frac{1}{1 + \exp -(\beta_0 + \beta_1 * SCORE)} \tag{3}$$

$$ln \frac{P(Y = Non - Participation \mid X = range1)}{P(Y = Participation \mid X = range1)} = ln\left(\frac{n(Non - Participation\ AND\ range1)}{n(Participation\ AND\ range1)}\right) = (\beta_0 + \beta_1 * SCORE) \tag{4}$$

In order to increase the explainability while using fewer classes, the logit values calculated in the fine classing step were rounded to integers and the classes with the same values were consolidated. In case of a continuous variable, the calculated logit value might not be monotonously increasing or decreasing due to the randomness and limited size of the data. To maintain explainability, such classes were consolidated with a nearby classes and monotonicity was forced.

Finally, we estimate the linear regression for the SCORE variable and logit calculated for the integrated class (i.e. intercept and slope in (4)). However, not only intercept and slope, but also SCORE variable is also unknown.

To determine SCORE value from this underdetermined equation, most CS methods apply many heuristic assumptions without any mathematical evidence, use empirical methods based on expert intuition, or just use the class value as it is without considering difference in features' importance. These directions are easy to implement, but they are not explainable method. Therefore, we set the SCORE value for each feature to an integer between 0 and 100, and the difference of features' importance is considered in next step (4.3.4) through the scaling factor.

4.3.4 Model Development

Model development can become arbitrary depending on the characteristics of the data and purpose when heuristics are used. In order to make model more rational and explainable, we eliminated unreasonable assumptions and established four well-founded rules. First, the most important feature from feature selection is regarded as the baseline and we use SCORE values derived from coarse classing. Secondly, starting from the second important feature, a scaling factor between 1 and 0 with 0.05 step-size is introduced. Accuracy performance by adding a new feature is evaluated by the product of SCORE value from coarse classing and scaling factor. We reflect the importance of features through the scaling factor. Thirdly, features are tried in ranking order, and a feature is added to the model only if it improves AUROC with respect to the previous model. Finally, for the purpose of making less complex model, features are admitted only when AUROC is increased by 0.001 or more.

5. Results

The simple rule-based approach using 'participation of the last event' only, ML approaches using 56 features, and the CS approach developed according to 4.3 were evaluated using test data. For ML, GBM showed the best performance and therefore only GBM's performance is included. For CS, five features were chosen in the final scorecard, and the features are shown in Table 4. Because of the enhanced rules in 4.3.4, only features that improved AUROC by more than 0.001 were chosen, and only five features were finally reflected in the scorecard. Note that each feature uses 2~6 ranges, and the total score can be up to 235. A decimal point means that the scaling factor is applied.

Table 4. Scorecard used for CS approach

Feature	Range	Score
Participation rate	0~0.05	0
	0.05~0.1	19
	0.1~0.25	41
	0.25~0.45	58
	0.45~0.7	75
	0.7~	100
Maximum number of consecutive participation	0	0
	1	19.25
	2~4	28.35
	5~	35
Cumulative rebate	0	0
	1~328	13.25
	329~3930	17.75
	3931~	25
Maximum number of consecutive non-participation	0	35
	1	20.3
	2~4	15.75
	5~6	9.1
	7~11	4.2
	12~	0
Success in the last event	Success	40
	Non-participation or fail	0

The AUROC for test dataset is shown in Table 5. Even the simple rule-based performs reasonably well, and achieves AUROC of 0.7643. This indicates that the participation of the last event is a reasonable indicator for predicting if a customer will participate in the immediately next event. An additional investigation showed that AUROC can be boosted to around 0.9 by using all the raw data of last 3~10 participation records and forming engineered features, but such a use raw data fields through feature engineering should be considered as ML or CS depending on the number of features and the level of explainability. ML with full use of 56 features achieved the best AUROC, and the value was 0.9576. The value is significantly better than simple rule-based, and it is the best performance we have achieved while studying the dataset of pilot DR program. CS approach achieved 0.9509, which is only slightly worse than ML approach. Even though CS used only 5 features in an easy-to-explain way as shown in Table 4, the DR data's structure allowed the CS to achieve almost the same performance as the ML's. Furthermore, CS actually performed better than ML when ML was forced to use only 5 features as CS did. Despite of fully utilizing feature selection methods and using the 5 features that generate the best AUROC value, still ML's performance was worse than CS. This can be explained as follow. From ML's perspective, the scorecard shown in Table 4 can be understood as the result of an extra feature engineering that followed the particular procedures discussed in 4.3. While ML used the 56 features as they are shown in Appendix A, CS had a chance to modify the features and then choose only 5. The feature engineering combined with scoring method turned out to be slightly better than ML when it was allowed to choose only 5 unmodified features from Appendix A.

Table 5. Result

Approach	Number of used features	AUROC
Rule-based	1	0.7643
ML	5	0.9498
	56	0.9576
CS	5	0.9509

6. Conclusions

Energy service providers are increasingly relying on energy data for improving visibility and efficiency. To improve efficiency, it can be helpful to use data and make selective dissemination of information or customer-targeted actions rather than applying the same to all the customers. Such a use of data for a selective process, however, needs to be sufficiently explainable to the customers for the purpose of protecting customer rights and complying with regulations. In this study, we have considered an incentive-based peak-reduction DR program, and investigated the tradeoff between accuracy performance and explainability. Simple rule-base, machine learning, and credit scoring approaches were inspected, and credit scoring was shown to achieve almost as good performance as machine learning while providing sufficient explainability. Our particular implementation of credit scoring used a scorecard with five easy-to-explain features. The result indicates that credit scoring might be a viable solution for many other selective actions that are based on energy data.

Acknowledgements

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20161210200410).

Appendix A: Feature Description

Group1 (Which data used)	Group2 (User-specific or not)	Group3 (Continuous or not)	Feature
Raw data	Not User-specific	Continuous	Average temperature of event date (Celcius degree)
Raw data	Not User-specific	Continuous	System Marginal Price (SMP, won)
Raw data	Not User-specific	Continuous	Base rebate (won)
Raw data	Not User-specific	Discrete	Event date (0000-00-00)
Raw data	Not User-specific	Discrete	Event start hour (0~23)
Raw data	Not User-specific	Discrete	DR Provider (A, B)
Raw data	Not User-specific	Discrete	Customer Baseline(CBL) calculation method (A, B, C, D)
Raw data	Not User-specific	Discrete	Weather of event date (Sunny, Cloudy, Rainy, Snowy)
Raw data	Not User-specific	Discrete	Day of the week of event date (Monday~Sunday)
Raw data	User-specific	Continuous	Customer Baseline (CBL, Wh)
Raw data	User-specific	Continuous	Goal reduction amount (Wh)
Raw data	User-specific	Continuous	Planned rebate (won)
Raw data	User-specific	Continuous	The number of events (times)
Raw data	User-specific	Continuous	Rebate in the last event (won)
Raw data	User-specific	Discrete	Participation/Success of the last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 2nd last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 3rd last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 4th last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 5th last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 6th last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 7th last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 8th last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 9th last event (0, 1, unknown)
Raw data	User-specific	Discrete	Participation/Success of the 10th last event (0, 1, unknown)
Processed data	User specific	Continuous	The number of solicitation (times)
Processed data	User specific	Continuous	The number of participation (times)
Processed data	User specific	Continuous	The number of success (times)
Processed data	User specific	Continuous	Cumulative rebate (won)
Processed data	User specific	Continuous	Past maximum rebate (won)
Processed data	User specific	Continuous	The number of maximum consecutive participation (times)
Processed data	User specific	Continuous	The number of maximum consecutive non-participation (times)
Processed data	User specific	Continuous	The number of maximum consecutive success (times)
Processed data	User specific	Continuous	The number of maximum consecutive failure (times)
Processed data	User specific	Continuous	Average hourly usage for last 4 weeks (Wh)
Processed data	User specific	Continuous	Average hourly usage for the same day of the last 4 weeks (Wh)
Processed data	User specific	Continuous	Average hourly usage for the same hour of the last 4 weeks (Wh)
Processed data	User specific	Continuous	Average hourly usage for the same day&hour of the last 4 weeks (Wh)
Processed data	User specific	Continuous	Participation rate (= Participation / Solicitation)
Processed data	User specific	Continuous	Success rate (= Success / Participation)
Processed data	User specific	Continuous	Consecutive participation rate (= Maximum consecutive participation / Participation)
Processed data	User specific	Continuous	Consecutive non-participation rate (= Maximum consecutive non-participation / Non-participation)
Processed data	User specific	Continuous	Consecutive success rate (= Maximum consecutive success / Success)
Processed data	User specific	Continuous	Consecutive failure rate (= Maximum consecutive failure / Failure)
Processed data	User specific	Discrete	First solicited (True, False)
Processed data	User specific	Discrete	Approximated tier using usage data for last 4 weeks in old system(A~F, unknown)
Processed data	User specific	Discrete	Approximated tier using usage data for last 4 weeks in revised system(A~C, unknown)

References

[1] Electric power annual 2010, U.S. Energy Information Administration, EIA Press, 2015.

[2] Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them, U. S. Department of Energy, A Report to the United States Congress Pursuant to Section 1252 of the Energy Policy Act of 2005, 2006.

[3] Ex Post Load Impact Evaluation for Pacific Gas and Electric Company's SmartRate Tariff Final Report, Freeman, Sullivan & Co., 2008.

[4] Coupon Incentive-Based Demand Response: Theory and Case Study, Haiwang Zhong et al., IEEE Transactions on Power Systems, 2013.

[5] Demand Response Modeling Considering Interruptible/Curtailable Loads and Capacity Market Programs, H.A. Aalami et al., Applied Energy, 2010.

[6] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, European Commission, 2015.

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union, 2016.

[8] Statistical Classification Methods in Consumer Credit Scoring: a Review, D.J.Hand&W.E.Henley, Journal of the Royal Statistical Society, Series A 160, 523-541, 1997.

[9] What's the Point of Credit Scoring?, Loretta Mester, Business Review, 1997.

[10] Learning Interpretable Models, S. Ruping, Dortmund Univ. (PhD Thesis), 2006.

[11] Towards a Rigorous Science of Interpretable Machine Learning, Finale Doshi-Velez, arXiv, 2017.

[12] FICO blog, http://www.fico.com/en/blogs/analytics-optimization/gdpr-and-other-regulations-demand-explainable-ai/, 2017.

[13] Credit Scoring and Its Applications, LC Thomas, SIAM Monographs on Mathematical Modeling and Computation, SIAM: Philadelphia, 2002.

[14] Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring, B.Baesens, Journal of the American Statistical Association, 2003.

[15] Measuring classifier performance: a coherent alternative to the area under the ROC curve, David J. Hand, Machine Learning Vol. 77, 2009.

[16] The Kolmogorov-Smirnov Test for Goodness of fit, FJ Massey Jr, Journal of the American Statistical Association, 1951.

[17] An Introduction to Variable and Feature Selection, Isabelle Guyon, Journal of Machine Learning Research, 2003.

[18] A review of feature selection techniques in bioinformatics, Yvan Saeys, Bioinformatics, 2007.

[19] Correlation-based Feature Selection for Discrete and Numerical Class Machine Learning, Mark A. Hall, International Conference on Machine Learning, 2000.

[20] Logistic regression and its application in credit scoring, Christine Bolton, 2009.