Forecasting Customer Invoice Settlement with Behavioural Analytics

B Meenal

January 22, 2025

Forecasting Customer Invoice Settlement with behavioural analytics

Meenal B
UT Dallas

Abstract

Empirical evidence across diverse sectors illustrates that organizations frequently encounter challenges with the collection of payments from customers. Research findings indicate that a significant proportion of invoices issued to small and medium-sized enterprises, as well as business-to-business entities, in the United States and United Kingdom are settled beyond their due dates. The primary objective of this study is to investigate customer payment behavior in relation to invoice payments and introduce an analytical framework for studying and projecting payment patterns. Our reasoning can subsequently be integrated into a decision support framework, enabling decision makers to formulate forecasts relating to future disbursements and to initiate appropriate measures aimed at recovering any outstanding liabilities, or to modify their financial strategies in accordance with projected cashflow figures. Our examination involves the utilization of a comprehensive dataset comprising over 1.6 million customers, encompassing their invoice details, payment history, and interactions, such as email communication, SMS, and phone calls, initiated by the company issuing the invoices in order to prompt payment. We employ both supervised and unsupervised learning methodologies to anticipate the likelihood of a customer settling their invoice or outstanding balance by the upcoming due date, drawing upon the interactions initiated by the company and the corresponding customer responses. We introduce an innovative behavioral scoring framework to be utilized as a predictive model input. The outcomes of logistic regression examined in the study demonstrate a maximum accuracy of 97%, irrespective of whether preclustering of customers was conducted, compared to the other two machine learning methodologies evaluated. This model exhibits considerable potential to assist decision-makers in formulating strategies that enhance the financial robustness of the organization through effective cash flow management and reduction of extraneous corporate credit lines.

Keywords: behavioral analytics; invoice collection; invoice to cash; logistic regression; machine learning; predictive analytics

Introduction

It is a frequent occurrence for companies operating within sectors such as telecommunications, energy, utilities, and comparable industries to encounter challenges in the collection of customer invoice payments, leading to delays in settling their obligations to suppliers. Research has indicated that a considerable proportion of invoices issued by small- and medium-sized enterprises (SMEs) in both the United Kingdom and the United States are subject to delayed payments, with approximately 46% and 42.5% being paid late, respectively. This has led to an estimated total of £212 billion in outstanding bills in the United Kingdom alone, with similar challenges being faced by businesses in Asia. The significance of this matter is particularly pronounced for SMEs, given the notable impact that delayed payments can have on their ability to effectively manage cash flows.

In addressing challenges within the invoice-to-cash cycle, companies employ proactive measures by engaging in communication with customers to prompt and expedite timely payments. For example, in the event of a missed payment deadline, customers are notified through reminder notifications sent via electronic means such as email or text message, or through direct telephone contact initiated by a representative of the company. If the customer remains unresponsive to payment requests, this may exacerbate the situation, prompting the company to implement escalated measures such as issuing more assertive communications, suspending services, or initiating legal proceedings. This process, identified as the Dunning process in literature, involves a methodical and expedited approach to engaging with customers in an effort to shorten the time taken for debt collection. Figure 1 depicts a theoretical illustration of the Dunning methodology.

Research indicates that these measures effectively decrease the time spent on collecting debts, thereby enhancing the financial sustainability of the organization. Furthermore, there exists a diverse array of commercial software solutions tailored for debt collection purposes. Examples include PowerCurve by Experian, Collect!, CollectMax, and WebAR Ace, which can be classified as specialized business process management tools for debt collection management. The majority of this software incorporates internal modules designed to conduct a comprehensive evaluation of customers with regards to their probability of settling overdue debts. These modules subsequently offer guidance to collection experts and managerial personnel in determining the most suitable customers to target and the appropriate actions to undertake. Nevertheless, these systems often exhibit a deficiency in their capability to assess and distinguish various customer behaviors, limiting their capacity to recommend tailored strategies for debt recovery.

The primary objective of this study is to explore customer behavior with respect to invoice payments, and to introduce an analytical methodology for studying and forecasting payment conduct. We employ three distinct machine learning methodologies, specifically logistic regression (LR), one rule (OneR), and support vector machines (SVMs), to forecast whether a client will settle the complete outstanding balance prior to the subsequent payment deadline. We present an innovative behavioural credit scoring framework that leverages the historical invoicing and payment records of clients, in conjunction with the interactions initiated by the company and how customers have engaged with them. The scores derived from this behavioural model are utilised as inputs for our machine learning models.The most effective performance technique identified in our study, the LR model, demonstrates a predictive accuracy of up to 97%. This development bodes well for the company, as it enables the implementation of tailored customer interactions to streamline payment processes promptly and establish strategic, enduring customer engagement strategies for those who exhibit persistent payment delays, leading to notable efficiencies in both time and financial resources.

The primary innovation of our study consists of two key elements: a unique behavioural customer scoring framework that generates discerning features for customer differentiation and subsequent action sequence customisation, alongside a machine learning deployment exhibiting exceptional accuracy in forecasting invoice payments by leveraging the behavioural scores derived for individual customers. The primary observation derived from our research is the discovery that historical payment records, in conjunction with the remedial measures enacted by the organisation to ensure prompt customer payments, serves as a reliable indicator of future payment conduct, without reliance on demographic data. Hence, enterprises encountering challenges in retrieving accounts receivables should consider their own conduct in conjunction with their interactions with various types of clientele. It is evident that current decision support systems, including the debt collection management software mentioned, stand to gain significant advantages from the integration of our predictive modelling methodology. By incorporating our logic into their modules, these systems can enhance their accuracy in estimating payment probabilities, resulting in improved customer scoring and prioritisation, ultimately leading to higher rates of success in debt collection endeavors.

Background

Numerous prior investigations have examined the challenge of collecting payments for overdue invoices. Although this issue intersects with customer credit assessment and predictive models for customer attrition, our research emphasizes studies that advance predictive and behavioral strategies in addressing the late invoice payment collection issue. The scholarly research that examines the topic of concern encompasses predictive models for insolvency and delayed payment. Moreover, various prediction techniques exist, including those employing Markov chains or hazard rate methodologies. Nevertheless, our focus is on supervised and unsupervised techniques for forecasting insolvency and delayed payment events, which serve as the primary source of inspiration for our research.

Daskalaki et al. (25) employ decision trees, neural networks, and discriminant analysis with the data from a telecommunications company to forecast customer insolvency. While their ability to forecast solvent customers demonstrated a high level of accuracy, the overall predictive performance is found to be suboptimal when accounting for both solvent and insolvent customers. Hsiao and Whang (26) demonstrate a high level of accuracy in forecasting customer insolvency by employing multiple discriminant analysis, artificial neural networks (ANNs), and logistic regression (LR). They utilize a dataset from the insurance industry and demonstrate that Artificial Neural Networks (ANNs) exhibit superior performance compared to their alternative methodologies. Zabkowski and Szczesny utilized a dataset from a Polish telecommunications company, comprised of approximately 100,000 customers and 205 explanatory variables. Decision trees and neural networks are utilized for forecasting customer insolvency, with the findings indicating that artificial neural networks (ANNs) demonstrate strong performance in insolvency prediction. Conversely, it is observed that the efficacy of decision

trees diminishes over time, suggesting that they may become less effective in predicting insolvency accurately as time progresses.
Zeng and colleagues utilize a variety of machine learning algorithms to forecast delays in invoice payments. Their objective is to anticipate the likelihood of receiving payments for unpaid invoices at varying time points following the specified due date. To achieve this, they employ cost-sensitive learning techniques utilizing partial decision trees (PART), C4.5, boosted decision stumps, logistic regression (LR), and naive Bayes classifiers. Hu13 puts forth a method that mirrors the approach taken by Zeng et al.,6 with the added nuance of incorporating customer clustering prior to classification. Chen et al. (2011) examined the clientele of a telecommunications company in Taiwan utilizing a methodology that encompasses three sequential stages: rule-based learning, clustering analysis, and predictive modeling. Kim and Kang (2014) utilized data provided by a Korean cable TV service provider company and applied four machine learning techniques to forecast the likelihood of payment. They employ a single classification tree, random forests, artificial neural networks (ANN), and support vector machines (SVM). Furthermore, they propose a hybrid approach wherein the likelihood of payment is determined as the mean of the probabilities computed by the remaining four machine learning techniques. Their findings indicate that the hybrid method outperforms the other four machine learning methods in accurately predicting late payments.
The review of relevant literature indicates that machine learning and other analytical methodologies exhibit significant promise in forecasting invoice-to-cash collection outcomes. We draw motivation from these studies and enhance our prediction model by incorporating behavioural factors related to the interactions between companies and customers, along with customer responses. The synergy between our proposed method and the customer behaviour scoring framework yields significantly enhanced outcomes.

Data

The dataset utilised in our investigation comprises the one-year (2014) invoice records of 1,659,083 de-identified clientele utilising the internet services of a global telecommunications corporation. The dataset comprises over 45 million individual invoice payment records, as well as data pertaining to the quantity of outstanding invoices and specifics regarding the company's efforts to prompt payment.The dataset is limited to customers with outstanding invoices in 2014, a period marked by monthly invoicing.
Dunning activities refer to the series of measures undertaken by a service provider organisation in response to customers failing to make timely payments on their invoices. The objective of these measures is to reduce the number of days that invoices remain outstanding and unpaid. In the present investigation, the terms action and Dunning action are utilised synonymously. A variety of actions with diverse contents are formulated for implementation depending on the quantity of outstanding invoices and the duration of payment delay.
The sequence of actions commences with the issuance of a brief message service (SMS) containing a subtle reminder to the customer regarding an outstanding invoice, culminating in a progression of legal measures. This timetable is standard and is uniformly utilised for all patrons. Several instances of actions encompass SMS, interactive voice response (IVR) messages, phone interactions facilitated by agents, and warning correspondences presenting varying tones such as gentle, stern, or severe.
Each customer is assigned a distinct and anonymised hashed customer identification code, with no accompanying demographic data or supplementary customer information disclosed.
The following is the list of attributes available in each record of the data set:

- CustomerID: a hashed unique number indicating each customer.
- Action type: such as SMS, IVR, phone calls, service cut related and legal actions.
- Action date: the date of the action's execution
- Stage changer flag: for system-generated or people-initiated events, such as payment through the banking system, unpaid invoice occurrence, or action timer flag.

Figure 2 shows the frequency distribution of all Dunning actions by category.

Methodology

Theoretical framework

Methodical examination of understanding and forecasting payment conduct. In order to accomplish this objective, it is necessary to delineate variables and attributes that can adequately characterise customer behaviour, which in turn will serve as the inputs for the predictive model. Hence, it is essential to create an environment in which the calculation of specified behavioural metrics and factors can be performed. In this context, all of the input factors are determined by a scoring mechanism that has been formulated on the basis of customer/company behaviours, responses, and associated expenditures.

Action cost

It is commonly agreed that implementing any Dunning action will incur costs to the company. Let Cikt = C(o) þ C(r) denote the combined cost ikt ikt of action of type k taken by the company for customer i in month t, where C(o) and C(r) denote the two types of costs incurred, namely, the operational costs of the action and the cost of the risk of customer churn. We refer to *Cikt* as the "Action Cost" and C(o)ikt and C(r)ikt typically are to be provided by the company experts.

Payment scores. Conversely, paying debts or outstanding balances will have a positive effect on the standing balances will have a positive effect on the customer/company relationship; thus, payments results in a positive score. Let *Pipt*(*sit*) be the positive score earned by customer *i* in month *t* as a result of payment *p*. Here, *sit* denotes the state customer *i* is in month *t* with respect to the number of outstanding unpaid invoices. Table 1 below provides an example of payment scores *Pipt* that can be earned by customer i in various states. Monthly score. Invoices are issued on a monthly basis and each customer's debit situation is likely to change at the end of each month, so we define a ''Monthly Score,'' which is the accumulation of all action and payment scores customer *i* earns in month *t*. The monthly score *MSit* indicates each customer's behaviour during a single month allowing us to interpret that customer's behaviour in terms of payments and the company's behaviour in terms of actions.

$$MS_{it} = \sum_k S_{ikt} + \sum_p P_{ipt}(s_{it})$$

Cumulative score. The cumulative score is conceptualised as the aggregation of individual customers' monthly scores across a specified time frame, enabling the analysis of the evolution of a customer's payment patterns and historical trends over time. This is denoted by

$$CS_{it} = \sum_{l \leq t} MS_{il}.$$

A positive cumulative score indicates than late payments
in his or her history filled with late payments and Dunning actions.

Table 1. Payment scores with respect to customer states

s_{it}	*0 invoice*	*1 invoice*	*2 invoices*	*3 and more invoices*
P_{ipt}	10	7	5	3

The cumulative score CSit encompasses details regarding a customer's entire payment history, yet it does not fully elucidate the customer's most recent payment conduct. Let us examine a scenario involving a client who has been a patron of a business for a duration of 2 years, consistently fulfilling all payments punctually in the initial 20 months, yet exhibiting a cessation in settling invoices in the concluding 4-month period. This customer's overall payment history will be reflected positively through a cumulative score. However, the representation of the customer's recent payment behaviour may not be adequately presented. To depict the recent behaviour of a customer, we introduce an additional variable that signifies whether a customer's pattern of on-time payments has been positive.

This categorisation enables users to recognise four primary customer segments, allowing for tailored approaches involving varied actions, content, communication tones, and other strategies to efficiently manage outstanding debt and enhance collection outcomes.
Figure 3 visually represents a sample of 700 customers selected at random from our dataset, along with the distribution of their CSit and bit values as of April 2014.
The ultimate variable in our study is denoted as the "Debt Settlement Indicator," serving as a binary variable determined through the implementation of an indicator function. Should the customer fully pay off all outstanding debts by the upcoming due date, the indicator will be assigned a value of 1; if not, it will be designated a value of 0.

$$\mathbf{1}_i^t = \begin{cases} 1 & \text{if customer } i \text{ settles all debt by the Due Date of month } t \\ 0 & \text{if customer } i\text{'s debt remains totally or partially unpaid by the Due Date of month } t \end{cases} \tag{1}$$

Trend. In this analysis, we utilise a basic linear regression model to evaluate the cumulative scores (CSit) of customers over a rolling window of the most recent h months, and establish a regression line. The coefficient of the regression line indicates the general trajectory of the fluctuation in the customer's aggregated score over the most recent h months. This slope value is denoted as the "Reliability Trend" or commonly referred to as the "Trend."

A downward trend suggests a reduction in the overall customer score, which implies that the customer is facing consequences due to non-payment of outstanding invoice balances. A positive correlation exists between the magnitude of the negative trend and the likelihood of accruing additional debts in subsequent periods. Conversely, a favourable pattern indicates an enhanced payment conduct exhibited by a customer throughout the past h months.
To elucidate how CSit and bit represent customers' behaviours, we cluster customers into four groups considering positive and negative values of CSit and bit:
1. Positive CSit and bit : reliable customers with good payment history and recent on-time payments.
2. Positive bit and negative CSit: recovering customers with a history of late payments exceeding on-time payments, but recently paying their debts.
3. Negative bit and positive CSit: deteriorating customers with a good payment history, but recently getting into trouble by not paying their invoices on time.
4. Negative CSit and bit : trouble customers.
It is important to acknowledge that, although our method of scoring may bear similarities to a conventional credit scoring model, we intentionally refrain from labelling it as such due to notable distinctions in two key aspects. Initially, the analysis examines the expenses incurred by companies which can be measured in numerical terms, such as the cost associated with sending an SMS or having an agent call a customer to prompt payment. Additionally, it also takes into account costs that are not readily measurable, such as the potential negative impact on customer satisfaction and retention resulting from excessive actions and reminders. Additionally, it fails to incorporate external transactions with third-party entities or additional features such as customer demographics.

Proposed prediction approach
Our modelling methodology involves a structured series of procedures encompassing data preprocessing, feature manipulation, model learning, and evaluation activities, culminating in the assessment and documentation of the predictive efficacy of our models. This procedure is depicted in Figure 4.
Our primary objective is to forecast clients' payment tendencies, encompassing the timely resolution of all accrued debt and pending invoices prior to the stipulated deadline. This aligns with the task of predicting the 1ti variable for every individual customer as elucidated in the Theoretical Framework segment. If the payment is made by the stipulated deadline, the customer is assigned an indicator value of 1; conversely, failing to meet the payment deadline results in the customer being assigned an indicator value of 0. This implies that we leverage the payment and behaviour history of customers, characterised by various variables and features, in order to forecast a binary output indicator. Due to the incremental nature of payment forecasting and the configuration of our dataset, we opt for

employing a time series methodology involving a rolling analysis of customers' monthly behaviour.
In pursuit of this objective, LR, a frequently employed classification method in the fields of statistics and machine learning, is utilised. Logistic regression calculates the likelihood of an event occurring by considering various predictors or independent variables in relation to a binary outcome variable. In this study, we utilise CSit and bit as the exogenous variables or regressors derived from the preceding period's activity up to and including month t in order to forecast the customer's debt settlement status for month t + 1.
We proceed on the data set as follows. First, we train the model using customers' historical data. To elucidate, consider the end of month t, where the aim is to predict the probabilities of debt settlement by the due date of month t þ 1. For each customer, we calculate bi,t-1 from month *t-theta*-1 to t-1 and then utilise the calculated CSi,t-1 and bi,t-1 as independent variables, with the debt settlement indicator of month t as the dependent variable to train the model.

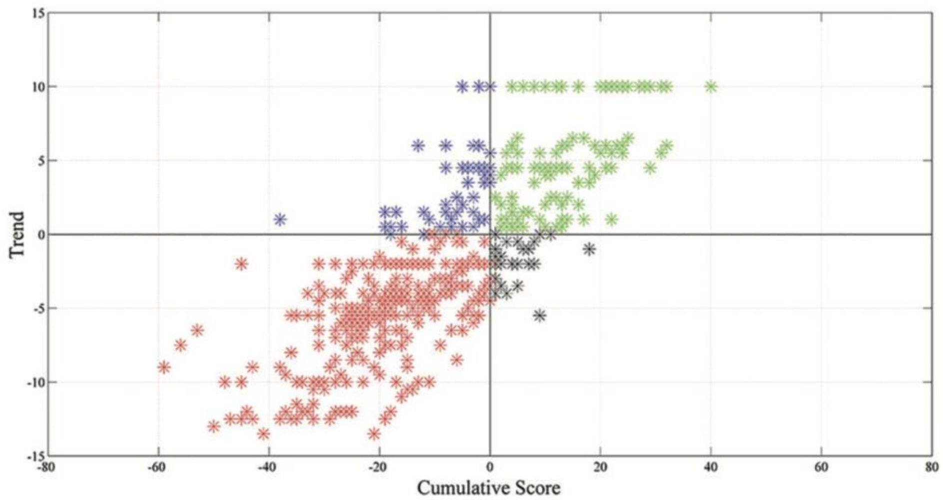

FIG. 3. Trend (β_{it}) and cumulative score (CS_{it}) distribution of 700 randomly chosen customers (shown by asterisks) as of April 2014. Color images are available online.

After training the model, we again calculate bi,t , this time using the cumulative scores of months t h to t, and use CSit and bit associated with month t as predictors of the model. The result of the final calculation will be the probability of debt settlement in month t þ 1. Figure 5 shows an example of how the predictive model uses a rolling window of h=5 months for the calculation of input features and 1-month payment information for labelling, to predict the payments in July.

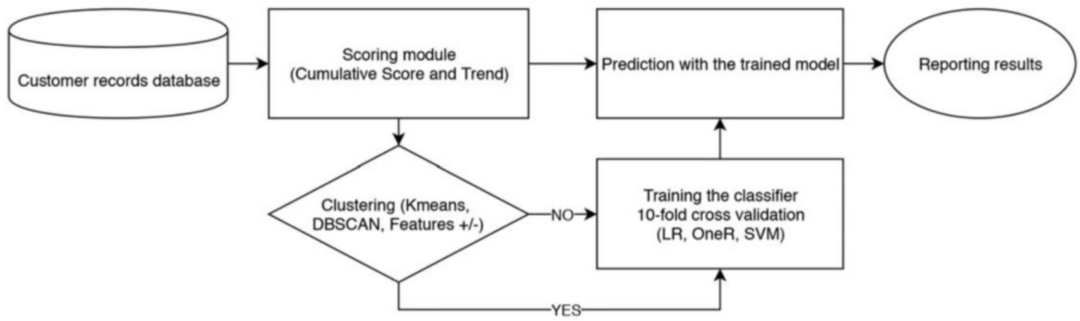

FIG. 4. The overall proposed prediction modeling framework. DBSCAN, density-based clustering; LR, logistic regression; OneR, one rule; SVM, support vector machine.

As an optional step, we cluster the customers using different methods based on their CSit and bit values and train the model separately for each cluster. We then calculate the input variables with respect to the new data, cluster the customers again, and use the predictive model. We suggest different variations of our LR model to be tested as follows:
1. Aggregated model (no customer clustering)
- Single predictor (trend)

- Bi-predictor (trend and cumulative score) 2. Clustered model

2. Clustered model
- Two clusters by reliability trend (+ vs.)
- Four clusters by reliability trend and cumulative score (+ vs.)
- Density-based clustering (DBSCAN) with several combinations of e and MinPts parameters28
- k-Means clustering (using different cluster numbers, k)

Finally, since the debt settlement indicator is a bi- nary variable and yet the logit model's output is a prob- ability value between 0 and 1, one must translate the probabilities into binary classes indicating customers' predicted payment behaviour in month t + 1. This enables one to compare the model's prediction with the actual data. To address this issue, we use a thresh- old, P*threshold*, for the probabilities calculated by the model and an indicator function, shown in Equation (2), that transforms the probabilities to binaries such that if the model's output for a customer is greater than or equal to the defined threshold, it will be trans- formed into 1, otherwise into 0.

$$\widehat{1}_i^{t+1} = \begin{cases} 1 & \text{If } P_i^{t+1} \geq P_{threshold} \\ 0 & \text{If } P_i^{t+1} < P_{threshold} \end{cases} \qquad (2)$$

One common approach for determining the optimal probability threshold involves the application of the OneR method following the calculation of probabilities through logistic regression during the training phase. The optimal probability threshold and duration of the training window are determined through an analysis of the training dataset with the goal of maximizing overall accuracy. Our approach emphasizes the significance of establishing the threshold prior to determining the optimal training window length. Specifically, we initially identify the most suitable probability threshold and subsequently proceed to selecting the best training window length based on the established threshold. To identify the optimal probability threshold, a stepwise OneR approach with a granularity of 0.01 is employed. This method involves adjusting the classification threshold incrementally from 0.1 to 0.9 following the calculation of payment probabilities using logistic regression, and evaluating the corresponding accuracy at each threshold value. The primary guideline is to select the threshold that maximizes accuracy when applied to the training dataset, and to subsequently utilize this threshold for forecasting invoice payments in the following month. It is important to note that in adherence to our modeling methodology, a minimum of three sequential periods (i.e., months) of past data should be utilized preceding the forecasted period. Specifically, the initial two months of data are employed for trend analysis, while the third month's debt settlement indicator serves as the dependent variable for model training. It should be noted that a consistent value of h (specifically, the training window length) is employe d for all custome rs during each training iteration.

ıg origin approach. Color images are available

A detailed algorithm is provided as follows:

Algorithm

1- Define t ($t+1$: prediction month)

2- Apply linear regression to cumulative scores of customer i for months $t-\theta-1$ to $t-1$

3- $\beta_{i,t-1} \leftarrow$ slope of the fitted line in step 2

o (optional) Cluster the customers

4- Train the logit model with $\beta_{i,t-1}$ and $cs_{i,t-1}$ as the independent variables and the binary variable 1_i^t as the dependent variable

5- Define best $P_{threshold}$ and θ which maximize total accuracy

6- Apply linear regression to cumulative scores of customer i for months $t-n$ to t

7- $\beta_{it} \leftarrow$ slope of the fitted line in step 6

8- Calculate probability (P_{t+1}^i) for customer i using the trained logit model with β_{it} and cs_{it} as predictors

9- If $P_{t+1}^i \geq P_{threshold}$, set the predicted debt settlement indicator of month $t+1$ to 1 ($\widehat{1}_i^{t+1}=1$) otherwise to 0 ($\widehat{1}_i^{t+1}=0$)

10- Compare $\widehat{1}_i^{t+1}$ and 1_i^{t+1}

At step 9 of the algorithm, an outcome is generated, comprising a binary score assigned to each customer, reflecting the anticipated payment conduct by the designated deadline in month t + 1. To assess the model's effectiveness, it is essential to conduct a comparison between the predicted and actual debt settlement indicators, a task that is carried out during the tenth step of the process.

Baseline method

In order to evaluate the effectiveness of our proposed machine learning method relative to a baseline approach, we have developed a basic linear classifier for comparison. We seek to identify an optimal line passing through the origin on a two-dimensional Euclidean plane that effectively divides customers into two categories - payers and non-payers based on the distribution of their cumulative scores and reliability trends with the highest accuracy possible on the training dataset.

In accordance with the data presented in Figure 3, there are discernible customer segments characterised by the polarity of their cumulative score and reliability trend. One could anticipate that customers positioned in the first quadrant - indicative of positive cumulative score and trend, are likely to settle their debts whereas those located in the fourth quadrant - characterised by negative cumulative score and trend, are unlikely to do so. The task at hand involves forecasting the payment behaviour of customers situated within the second and fourth quadrants. Therefore, we examine the equation representing the line defined by beta = -CS and consider perturbations of that line passing through the origin within the range belonging [90.180], where alpha = 90 represents the CS = 0 vertical y-axis line and alpha = 180 represents the beta = 0 horizontal x-axis line. In other words,

$$\beta = \tan\left(\alpha \frac{\pi}{180}\right) * CS, \qquad \alpha \in (90, 180] \qquad (3)$$

$$CS = 0, \alpha = 90 \qquad (4)$$

These lines extend across both the second and fourth quadrants.

And partition the Cartesian space into two segments, with clients located on the upper-right side of the line anticipated to be payers and clients situated on the lower-left side of the line projected as non-payers. Figure 6 illustrates a demonstration of the manner in which the baseline technique categorises customers into two distinct groups: payers (depicted in green) and non-payers (depicted in red). We conduct fine-tuning to optimise our model for achieving the highest classification accuracy on our designated training dataset. The resulting predictions serve as a reference point for evaluating the performance in comparison to LR, OneR, and SVM algorithms.

Analysis Results

To evaluate the effectiveness of our proposed scoring system and prediction modelling approach on the dataset, we initially employed a sampling method by randomly selecting subsets of 1000, 10,000, and 100,000 customers from the complete pool of 1.66 million customers in our dataset. The objective of this study was to assess the preliminary efficacy of our prediction methodology prior to its application to the entire dataset comprising 1.66 million customers.

Following initial examination, it was determined that the sample of 1000 customers proved insufficient, while the 10,000-customer subset was deemed adequate for generating dependable and significant outcomes. The sample size of 100,000 customers was deemed excessive for analysis, necessitating progression to the 1.66 million-customer dataset. The findings from the sample set of

10,000 customers exhibit considerable promise. To maintain conciseness, we have documented our discoveries with this sample set in the Supplementary Data section of our publication. "Subsequently, we present the findings derived from the 1.66 million customer dataset."
Figure 7 presents the trend and cumulative score distribution of the 10,000-customer sample utilised in our preliminary analysis as of September 2014, serving as an illustrative example. It can be noted that the preponderance of customers within this sample fall within quadrants 1 and 3, indicating they are either maintaining a positive trajectory or experiencing a deteriorating status. We have substantiated the presence of a comparable trend within the entirety of the 1.66 million-customer dataset.

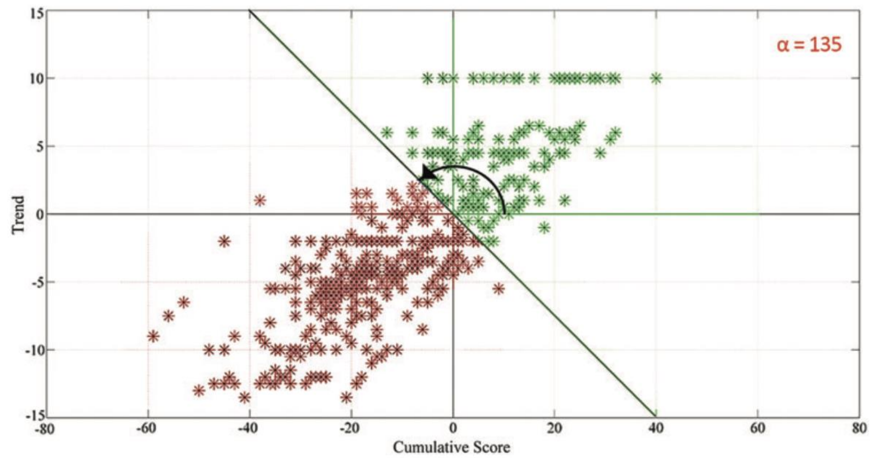

FIG. 6. An example of classification by the baseline method. The line $\beta = -CS$ separates a random sample of 700 customers (shown by asterisks) into two classes of payers (upper-right) and nonpayers (lower-left). Color images are available online.

Subsequently, clustering will be employed on customers utilising the two dimensions (trend and cumulative score) in order to enhance predictive accuracy following the training of the models individually for each cluster. Figure 8a–d illustrates the clustering of the 10,000-customer sample based on the clustering methodologies outlined in the preceding section.
Our study utilised a 10-fold cross-validation technique on the training dataset to train individual models on each fold, generate forecasts for period t + 1, and evaluate model performance based on metrics including accuracy, true positive rate (TPR), true negative rate (TNR), and area under the receiver operating characteristic (ROC) curve (AUC). We conducted multiple calculations for each segment, utilising various historical data periods for training across different probability thresholds. During the examination of each fold, we evaluate six prediction models as outlined in the Proposed

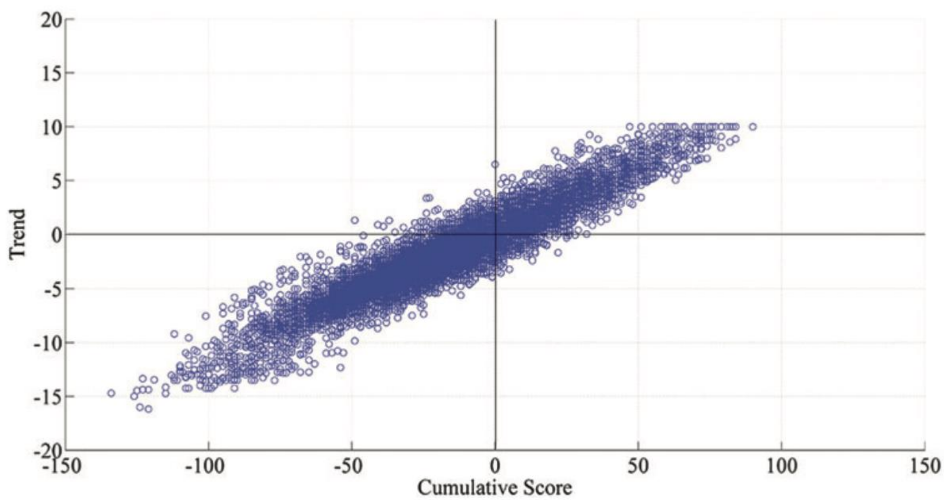

FIG. 7. Cumulative score (CS_{it}) and trend (β_{it}) distribution of a sample of 10,000 customers (shown by circles). Color images are available online.

Prediction Approach section, along with 45 variations of training window duration and prediction month combinations.

Table 2 presents the superior performance outcomes achieved by the 1.66 M-customer dataset across various permutations of training window duration, forecasted month, and clustering methodology. The optimal results are computed and articulated in the table.
The predictive performance and area under the receiver operating characteristic curve (AUC) obtained by each approach.Based on the findings presented in Table 2, it is evident that LR outperforms the other prediction methods analyzed. Consequently, we have opted to focus our subsequent analysis on LR exclusively. In the Baseline Method section, the prediction accuracy is the sole metric reported due to the absence of a defined AUC for this approach. Overall, the baseline algorithm demonstrates satisfactory performance, however, it exhibits comparatively lower accuracy values when compared to the nonlinear classifiers employed.

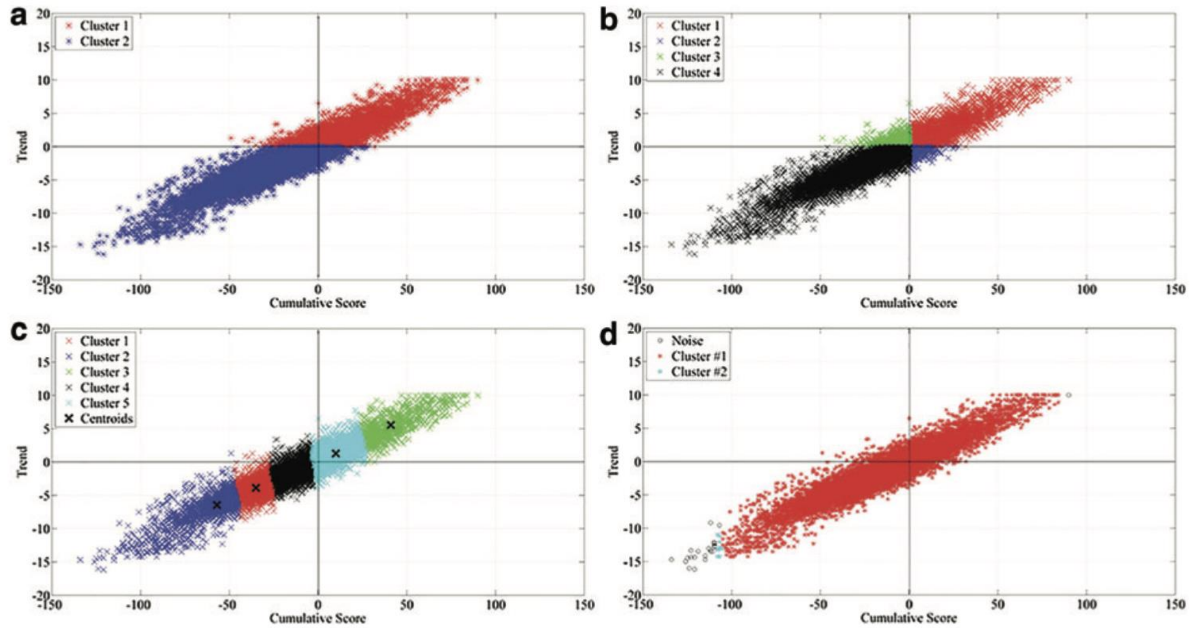

FIG. 8. Customers clustered based on different methods **(a)** Customers clustered by "Trend" into two clusters, **(b)** customers clustered by "Trend" and "Cumulative Score" into four clusters, **(c)** customers clustered by "k-Means" clustering method into five clusters, **(d)** customers clustered by "DBSCAN" method and $\varepsilon = 3$, MinPts = 20 into two clusters + noise. Color images are available online.

Table 2. Best performance comparison of prediction methods based on accuracy and area under the ROC curve

	Logistic regression		*Support vector machines*		*OneR*		*Baseline method*
Predicted month	*Accuracy*	*AUC*	*Accuracy*	*AUC*	*Accuracy*	*AUC*	*Accuracy*
April	89.80	0.83	86.63	0.59	81.67	0.56	75.60
May	90.35	0.86	86.20	0.65	80.21	0.54	76.48
June	91.33	0.55	85.61	0.60	79.93	0.53	75.16
July	92.95	0.54	87.57	0.59	82.97	0.52	76.30
August	91.37	0.86	88.62	0.62	82.77	0.52	78.03
September	95.98	0.60	88.67	0.61	83.17	0.52	77.76
October	93.08	0.59	88.89	0.67	82.85	0.51	76.79
November	97.39	0.58	91.96	0.63	89.08	0.51	75.19
December	96.84	0.94	89.38	0.61	87.62	0.51	82.72

AUC, area under the ROC curve; OneR, one rule.

Table 3. Logistic regression maximum prediction accuracy (in percentage)

	Single predictor (trend only)	*Two predictors*	*Two predictors*	*Two predictors*	*Two predictors*	*Two predictors*
Predicted month	*No clustering*	*No clustering*	*2 Clusters by trend*	*4 Clusters by trend and cumulative score*	*k-Means (k = 5)*	*DBSCAN (various ε and MinPts)*
April	82.76	89.80	84.94	88.77	87.71	88.23
May	81.62	90.35	84.68	88.89	91.02	86.48
June	81.63	88.28	84.03	91.20	91.33	84.68
July	84.27	89.74	86.31	92.09	92.95	86.82
August	84.52	91.37	87.10	90.41	91.20	87.39
September	84.91	90.96	86.87	90.45	95.98	87.42
October	84.81	91.18	86.49	89.72	93.08	87.62
November	89.34	92.86	89.83	93.70	97.39	92.52
December	88.91	96.84	88.14	91.43	90.18	88.30

DBSCAN, density-based clustering.

Table 3 presents a comprehensive overview of the outcomes for LR across different permutations of prediction month and clustering methodology. In this study, we present the highest level of accuracy achieved by Logistic Regression (LR) for each prediction month and clustering method examined. Our analysis reveals that the logistic regression model with two predictors and no clustering exhibits superior performance. Additional comprehensive findings for this specific configuration can be found in Table 4.

Table 4 presents a comprehensive overview of prediction outcomes utilising two predictors for logistic regression without any clustering. The columns consist of the count of outstanding debts (representing the negative class) and the count of settled debts (representing the positive class), True Positive Rate (TPR), True Negative Rate (TNR), precision, F1-score, accuracy, and Area Under the Curve (AUC) metrics for the dataset comprising 1.66 million customers.

It is important to emphasise that while achieving success in predicting the true negative rate (TNR) holds significance, it is equally crucial to attain a respectable true positive rate (TPR) within this specific framework. This is due to the potential of causing dissatisfaction among valuable customers through unwarranted dunning actions, leading to their potential departure and a significant financial impact on the organisation. Conversely, attaining a high True Negative Rate (TNR) is essential, as it is imperative for the organisation to promptly undertake appropriate measures for non-paying customers to optimise revenue generation.

The findings presented in both tables demonstrate accuracy ranges of 80% to 97% and AUC values ranging from 0.80 to 0.95, which are regarded as acceptable according to our evaluation. These findings validate that our scoring methodology and feature design effectively capture and reflect customers' behaviours with a high level of information and representation.

Subsequently, we examine the impact of historical data periods utilised in forecasting. The data presented in Table 4 indicates that incorporating the complete payment history for a specific month may not invariably lead to heightened predictive accuracy. Our analysis reveals that, for the majority of months, the utilisation of customers' most recent transactional patterns results in improved rates of predictive success. It is important to acknowledge that the financial circumstances of customers may evolve over time, potentially impacting their payment behaviour. Therefore, leveraging recent payment history is more apt to enhance the accuracy of predictions.

Table 4. Detailed prediction results using logistic regression with two predictors and no clustering

Prediction month	*Rolling window length*	*Positive class (paid)*	*Negative class (unpaid)*	*TPR*	*TNR*	*Precision*	*F1-score*	*Accuracy*	*AUC*
April	2	206,107	1,452,976	38.43	97.08	65.14	56.60	89.79	0.83
May	2	224,347	1,434,736	58.65	95.31	66.16	62.18	90.35	0.86
June	2	254,318	1,404,765	55.51	94.20	63.44	59.23	88.28	0.82
July	3	218,863	1,440,220	52.26	95.43	63.50	57.33	89.74	0.87
August	3	209,376	1,449,707	51.79	97.09	72.00	60.25	91.37	0.86
September	3	209,026	1,450,057	50.94	96.73	69.17	58.67	90.96	0.83
October	3	205,710	1,453,373	48.50	97.22	71.21	57.70	91.18	0.85
November	6	124,503	1,534,580	24.54	98.40	55.45	34.03	92.86	0.80
December	2	176,898	1,482,185	70.38	99.98	99.97	82.61	96.84	0.94

TNR, true negative rate; TPR, true positive rate.

Discussion and Conclusion

The prompt collection of invoices is a crucial concern for numerous organisations, as it has a direct impact on their cash flow and, subsequently, their ability to sustain operations. The analytical framework presented in this research seeks to resolve this concern through an examination of individual customer data, including their invoice and payment records, the strategies employed by the organisation to encourage timely payments, and the corresponding reactions from customers. We introduce a machine learning predictive modelling technique designed to forecast invoice payment outcomes with high accuracy, and implement it in conjunction with the inputs generated by our innovative behavioural customer scoring framework. We argue that the predictive model utilised in this study, particularly due to its behavioural focus, demonstrates high efficacy in forecasting whether a customer will make payment on their next invoice or settle the entire outstanding balance by the following due date, achieving an accuracy rate of up to 97%. Our evaluation system allows for differentiation among customers, enabling decision makers to tailor action sequences accordingly. This outcome holds promising implications for organisations, as it enables them to devise more precise strategies tailored towards various customer segments in order to enhance collection rates and bolster

the financial position of the company.
This study is anticipated to give rise to various extensions and novel research avenues. Enhanced precision in forecasting outcomes could be achieved by devising Markov chain-based methodologies that leverage individual-level payment and action histories, particularly through access to expanded data duration from the data source. One potential avenue for research could involve developing and executing an action strategy tailored to individual customers, taking into account varying payment patterns and responses to actions by customers. A field experiment could be organised in collaboration with the enterprise involving a specific group of clientele for the evaluation of the operational approach alongside the initial predictive framework. The insights derived from the field investigation could subsequently be leveraged to enhance and optimise the behavioural scoring mechanisms and prediction models. Ultimately, incorporating financial measures and/or incentives into the framework can help in developing strategic sequences of actions that are not only financially viable but also aligned with the target behavior.

References
1. Howorth C, WilsonN. Late payment and the small firm:An examination of case studies. J Small Bus Enterprise Dev. 1998;5:307–315.
9. Experian. Available online at www.experian.com/business-services/ debt-recovery.html/ (last accessed December 1, 2017).
3. Funding Gates. Available online at http://blog.fundinggates.com/2015/ 07/7-industries-most-prone-to-late-payments/ (last accessed December 1, 2017).
2. Shin HH, Soenen L. Efficiency of working capital management and corporate profitability. Financ Pract Educ. 1998;8:37–45.
5. Paul SY, Devi SS, Teh CG. Impact of late payment on Firms' profitability: Empirical evidence from Malaysia. Pac-Basin Financ J. 2012;20:777–792.
4. The Telegraph. Late payments put a quarter of small firms at risk of insolvency. Available online at www.telegraph.co.uk/finance/ businessclub/money/11787765/Late-payments-put-a-quarter-of-small- firms-at-risk-of-insolvency.html (last accessed December 1, 2017).
6. Zeng S, Melville P, Lang CA, et al. Using predictive analysis to improve invoice-to-cash collection. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY: ACM, August 24–27, 2008. pp. 1043–1050.
8. Capterra. Available online at https://www.capterra.com/debt-collection- software/ (last accessed December 1, 2017).
17. Baesens B, Van Gestel T, Viaene S, et al. Benchmarking state-of-the-art classification algorithms for credit scoring. J Oper Res Soc. 2003;54: 627–635.
10. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer Science & Business Media, 2009.
11. Chen CH, Chiang RD, Wu TF, Chu HC. A combined mining-based framework for predicting telecommunications customer payment behaviors. Expert Syst Appl. 2013;40:6561–6569.
12. Ha SH, Krishnan R. Predicting repayment of the credit card debt. Comput Oper Res. 2012;39:765–773.
15. Abdou HA, Pointon J. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. Intell Syst Account, Finance Manag. 2011;18:59–88.
7. Wikipedia. Dunning (process). Available online at http://en.wikipedia.org/ wiki/Dunning_(process) (last accessed December 1, 2017).
13. Hu P. Predicting and improving invoice-to-cash collection through machine learning. Master dissertation, Massachusetts Institute of Technology, 2015.
14. Kim J, Kang P. Late payment prediction models for fair allocation of customer contact lists to call center agents. Decis Support Syst. 2016;85: 84–101.
32. Thomas LC, Matuszyk A, So MC, et al. Modelling repayment patterns in the collections process for unsecured consumer debt: A case study. Eur J Oper Res. 2016;249:476–486.
16. Chen X, Wang X, Wu DD. Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China. Decis Support Syst. 2010;49:301–310.
18. Lessmann S, Baesens B, Seow HV, Thomas LC. Benchmarking state-of- the-art classification algorithms for credit scoring: An update of research. Eur J Oper Res. 2015;247:124–136.

19. Khandani AE, Kim AJ, Lo AW. Consumer credit-risk models via machine-learning algorithms. J Bank Finance. 2010;34:2767–2787.
29. Grimshaw SD, Alexander WP. Markov chain models for delinquency: Transition matrix estimation and forecasting. Appl Stoch Model Bus Industry. 2011;27:267–279.
22. Hadden J, Tiwari A, Roy R, Ruta D. Churn prediction: Does technology matter. Int J Intell Technol. 2006;1:104–110.
20. Coussement K, Lessmann S, Verstraeten G. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. Decis Support Syst. 2017;95: 27–36.
21. Farquad MAH, Ravi V, Raju SB. Churn prediction using comprehensible support vector machine: An analytical CRM application. Appl Soft Comput. 2014;19:31–40.
23. Idris A, Rizwan M, Khan A. Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. Comput Electr Eng. 2012;38:1808–1819.
25. Daskalaki S, Kopanas I, Goudara M, Avouris N. Data mining for decision support on customer insolvency in telecommunications business. Eur J Oper Res 2003;145:239–255.
27. Zabkowski TS, Szczesny W. Insolvency modeling in the cellular telecommunication industry. Expert Syst Appl. 2012;39:6879–6886.
28. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd. 1996; 96:226–231.
30. Deloffre A, Nait Abdallah R. Modelling of the debts collection process for service companies. IFAC Proc Vol. 2006;39:349–354.
24. Xie Y, Li X, Ngai EWT, Ying W. Customer churn prediction using improved balanced random forests. Expert Syst Appl 2009;36:5445–5449.
31. Lin CT, Yang SY. A forecasting model for the likelihood of delinquency, default or prepayment: The case of Taiwan. Int J Bus. 2003;8:203–212.
26. Hsiao SH, Whang TJ. A study of financial insolvency prediction model for life insurers. Expert Syst Appl. 2009;36:6100–6107.

Abbreviations Used

ANNs = artificial neural networks
AUC = area under the ROC curve
DBSCAN = density-based clustering
IVR = interactive voice response
LR = logistic regression
ML = machine learning
OneR = one rule
PART = partial decision trees
SMEs = small- and medium-sized enterprises
SMS = short message service SVMs 1⁄4 support vector machines
TNR = true negative rate TPR 1⁄4 true positive rate