

# Review of innovations in the South African collection industry

## AUTHORS:

Vinzent Blaszczyk<sup>1</sup>

Tanja Verster<sup>1</sup>

Alan Broderick<sup>2</sup>

## AFFILIATIONS:

<sup>1</sup>Centre for BMI, North West University, Potchefstroom, South Africa

<sup>2</sup>Shapiro Shaik Defries & Associates, Johannesburg, South Africa

## CORRESPONDENCE TO:

Tanja Verster

## EMAIL:

Tanja.Verster@nwu.ac.za

## DATES:

**Received:** 19 Oct. 2017

**Revised:** 11 Jan. 2018

**Accepted:** 06 Feb. 2018

**Published:** 30 July 2018

## KEYWORDS:

debt collecting; credit risk modelling; machine learning; collection strategies; social media

## HOW TO CITE:

Blaszczyk V, Verster T, Broderick A. Review of innovations in the South African collection industry. *S Afr J Sci.* 2018;114(7/8), Art. #2017-0360, 9 pages. <http://dx.doi.org/10.17159/sajs.2018/20170360>

## ARTICLE INCLUDES:

- × Supplementary material
- × Data set

## FUNDING:

None

The objective of this review was to provide an overview of new developments and innovations within the collections industry that could possibly enhance the performance of collection agencies, specifically in South Africa. A literature study was conducted to determine current practices in the collections industry, as well as possible future innovations. A significant trend identified throughout the literature study was the increasing prioritisation of automated digital communication in several aspects of debt collection. It is reasonable to assume that this trend will continue to become the industry standard. Four recommendations are made based on the findings of the literature study. Firstly, South African collection agencies should investigate the feasibility of developing an app-based solution to performing collections. Secondly, collection agencies should supplement traditional modelling techniques with other tools, such as those developed in the field of machine learning. Thirdly, collection agencies could consider using speech analytics to obtain insights into call centre agents' performance and adherence to business rules. Lastly, the usage of social media data in collections as well as credit risk modelling in general is recommended as a topic for future study.

### Significance:

- A review of the various techniques currently employed in the field of debt collections may serve as useful reference for both academics and those working in debt collections.
- Recommendations are provided to assist businesses in aligning the operational models of their debt collection units to industry best practice.
- Topics for future research in this crucial sector of the economy, which brings together such fields as risk governance, predictive modelling, human psychology, debt management, legal compliance and business analysis, are provided.

## Introduction

Collections agencies normally act as external debt collecting agencies for a diverse selection of institutions that may, during the normal course of their operations, encounter defaulters. These defaulters have breached their legal requirements and failed to fully repay the debt extended to them. This scenario may have several adverse effects on the company – the two most direct being a financial loss and the opportunity cost of the capital extended to the debtor. Debt collection is the process of trying to minimise these losses by recouping some of the debt. Companies may choose to make use of internal or external debt collection strategies, or a combination of the two. It is usual that external debt collectors receive those debtors who have proven unresponsive to internal collection efforts. Debtors referred to external collectors usually differ significantly in their repayment characteristics from those who respond to internal efforts. It has been shown that the difference can be as extreme as 10% non-payers for internal collections versus 83% for external collections, and 30% full repayments versus 0.7%, respectively.<sup>1</sup> Thus, with such a low probability of collecting on outstanding debt, it is critical that the collections process be as efficient as possible for external debt collectors to operate profitably.

As the purpose of this paper was to explore current best practices and possible future developments in the collections industry, as per the academic and business literature, the main research topics addressed are:

- traditional and innovative methods of creating debt collections queues;
- innovative strategies employed by leading debt collections agencies to optimise their operational profitability;
- methods for assessing and comparing the performance of different collections strategies;
- data sources not currently used by collections agencies in South Africa, which might have significant predictive power; and
- machine-learning techniques capable of modelling and improving the collections process.

We aimed to provide South African collection agencies with innovative thinking and strategies. From a brief review of the collection process, we identified four areas that became the focus of our research: (1) innovative uses of data sources in collections; (2) innovative uses of machine learning (and other statistical modelling techniques) in collections; (3) innovative ways to enhance collection strategies; and (4) innovative ways to evaluate collection strategies.

## The collection process

The collections process is described briefly. Debtors purchase products or services on credit. Those who default on their repayment and are not collected on by internal collections are amalgamated into debtor lists. These lists are then provided to collection agencies which combine the information contained therein with additional data

sources to create collections queues. Collections agents are assigned to these queues with the aim of recovering as much of the debt as possible. Collection agencies typically levy a commission on the recovered debt and pay over the remainder to their clients.

The collection process is demonstrated visually in Figure 1. Our research will focus on the four main operational phases in this collection process.

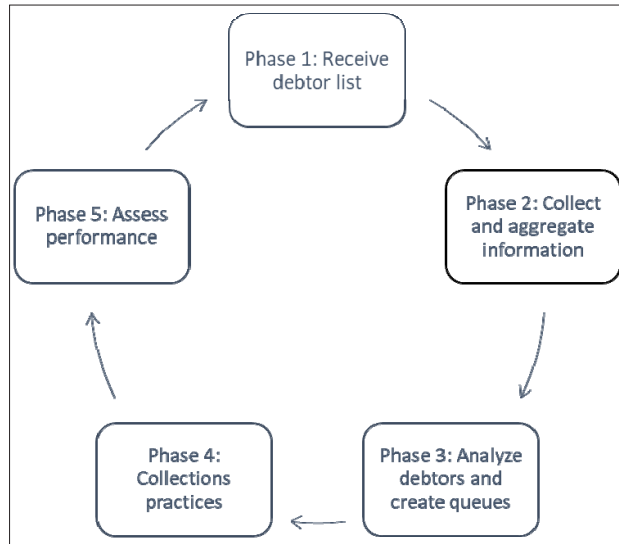


Figure 1: Operational phases of a debt collections agency.

### Phase 1: Receiving debtor list

The first of these, receiving the debtor lists, is not elaborated upon specifically, but research pertaining to the other phases was conducted.

### Phase 2: Collecting and aggregating information

This critical operational phase pertains to the collection and aggregation of information on those individuals on the debtor list. Several data sources can be used during this phase. Traditional data sources include data from the clients and credit bureaus. Additional data sources are summarised below.

### Phase 3: Analysing debtors and creating collection queues

This phase involves the analysis of debtors and then creation of collection queues. Based on a specific collection strategy, a call agent can be assigned to a specific group of debtors, that is, a queue. This queue may also include a prioritisation of debtors.

In this phase, the data collected during the previous stage are used as inputs to a chosen model. Typically this model should return scores of individual debtors. This score should, at a minimum, provide a collection agency with estimates of the likelihood of obtaining a promise-to-pay, the conversion thereof to recovery-rate and expected amount collected, the likelihood of zero payment as well as expected collections cost and the expected collections period. It is important to note that the above aspects may be modelled separately or jointly, and that several different models can be used, and each has its own respective strengths and weaknesses. Consequently, traditional methods can be complemented by techniques from other fields such as machine learning. The collection queues are then created using the scores as generated by the model. Logically, higher scores should be associated with better repayment behaviour and thus higher operational profit.

### Phase 4: Collections practices at the operational level

During this phase, the output from the previous phase is used to assign collection queues to agents, who attempt to recover as much of the debt as possible. The collections strategy becomes effective

during this phase, as it determines the communications method used and provides guidelines for agent interaction with debtors. A holistic collections strategy might also contain recommendations regarding write-off of debtors, as well as the method whereby agents are assigned to queues.

### Phase 5: Evaluating collections strategies and queue building models

This phase occurs concurrently with all the other phases and involves the monitoring of each phase whilst identifying possible irregularities and inefficient strategies. The performance of strategies and models should be assessed regularly and corrective measures taken as and when necessary. Collecting data on strategy performance will allow management to make data-driven business decisions and optimise operational efficiency on many fronts. It is critical that the evaluation methods be established up front and be applied consistently.

As our aim was to identify emerging practices, our research was not restricted specifically to the collections industry. Several possible new strategies and ideas emerged in the broader financial sector. It is reasonable to assume that companies might also be hesitant to publish extensive technical details pertaining to their internal systems, as considerable resources are spent developing these systems and companies will want to protect their investments. Business and academic publications were used, as they tend to have different focuses: academic publications contain more theoretical and technical details whilst business publications focus on operational implementations and business results. Case studies were preferred to highly theoretical publications as these address 'real-life' scenarios and report proven results. Note also that these case studies – specifically those from vendors – are perhaps less reliable, as they are not subject to peer review and exist primarily to serve as marketing tools; the results reported can reasonably be assumed to be biased and selectively chosen to reflect positively on the respective business.

In each of these four operational phases, specific research topics were identified as shown in Table 1. The rest of the paper will be presented around the research topics listed in Table 1.

Table 1: Overview of research topic per operational phase

Operational phase	Research topic
Phase 2: Collecting and aggregating information	Innovative uses of data sources in collections
Phase 3: Analysing debtors and creating collection queues	Innovative uses of machine learning (and other statistical modelling techniques) in collections
Phase 4: Collections practices at the operational level	Innovative ways to enhance collection strategies
Phase 5: Evaluating collections strategies and queue-building models	Innovative ways of evaluating collection strategies (in such a way that comparison between different strategies can be objectively made)

## Innovative uses of data sources in collections

During the research process, several different data sources were identified that could be used in the collections industry. The main data sources identified are listed in Table 2. It is important to consider that not all these data might have significant predictive power, and it is recommended that the predictive power be investigated before inclusion in the queue-building models to prevent unnecessary inflation of the data's dimensionality. The usability of these data in terms of current legislation should also be considered – there are rich sources of data which are precluded from use in collections strategies in terms of existing legislation.

**Table 2:** Innovative uses of data sources in collections

Data source	Possible application in collections
Social media data	Not used extensively in South Africa yet, but many case studies in the USA have shown possible extensions of standard models to include these data in collections <sup>2</sup>
Geo-location data	Can be used to identify areas with significant numbers of debtors; to identify possible relationships between location and repayment behaviour; and to identify which areas are connecting with a specific collection agency's social media
Device data	A collections strategy could be linked to device used, i.e. those who prefer mobile devices might be more responsive to telephonic or phone text communication, whilst desktop users might prefer email correspondence
Data brokers	Many case studies in the USA <sup>2</sup> have shown possible extensions of standard models to include the extensive types of data that data brokers can provide into predictive modelling (e.g. collections)

*Social media data* vendors, like Datasift and GNIP, provide services such as aggregating data across multiple social platforms (Facebook, LinkedIn, Twitter, etc.). Communication with Datasift suggests that privacy laws restrict the use of individuals' data. Only publicly available data may be used under these laws and the ethical implications thereof should be considered. Datasift<sup>3</sup> offers the provision of data starting at costs of USD15 000 per month. GNIP<sup>4</sup>, a daughter company of Twitter, provides packages of one million tweets, with prices starting at USD1250. A possible application is to link a collections strategy to debtor interaction with social media, i.e. the debtors might be more receptive after interaction with agent correspondence.

Step-by-step guides<sup>5</sup> have been published which describe methods to search and extract the social media data based on the criteria specified by individuals on popular social networks like LinkedIn, Facebook and Twitter. Ultimately, the goal is still to link customer behaviour with propensity to pay and selection of collections strategy.<sup>5</sup> Web structure mining, web content mining and web usage mining are three methods discussed in Pottharst<sup>6</sup> to extract data from any site in general.

Another innovative use of data in collections is the utilisation of *geo-location data* (inferred from cell or mobile phone data, or from Facebook etc.). Possible applications of geo-location analysis to collection agencies include: identifying areas with significant numbers of debtors; identifying possible relationships between location and repayment behaviour; and identifying which areas align with a specific collection agency's social media footprint.

The possible use of *device data* should also be investigated as inputs to the queue-building models. A collections strategy could be linked to preferred device, i.e. those who prefer mobile devices might be more responsive to app-based communication whilst desktop users might prefer more traditional methods such as email – although this is rapidly changing as emails are equally accessible on modern mobile devices. Another possible application of social media data and device data, is to investigate the relationship between the time of day when communication occurs and likelihood of obtaining a promise-to-pay.

*Data brokers*, also known as data aggregators, collect data on individuals from several domains. A comprehensive overview of data brokers was compiled by Clearinghouse<sup>7</sup>, and includes the following major data brokers based in the USA: Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Intellius, PeekYou, Rapleaf and Recorded Future. Similar companies operating in South Africa are Background Check International, Dracore and Blue Label Data Solution. Table 3 gives an overview of data provided by data brokers as well as the possible application to debt collection. Table 3 also includes some examples of the data in each data domain. The use of data brokers is already an existing practice in South African collection agencies (with compliance being an integral consideration), but the use of some data categories, specifically online behavioural data, is limited or undocumented. Many case studies in the

USA have shown possible extensions of standard models to include this type of data, e.g. an overview of data-mining techniques applied to analysis of social media data is provided by Injadat et al.<sup>2</sup>

In sum, collection agencies could investigate augmenting their existing data with additional data sources, which include social media data and other online behaviour, geo-location data, device data, and data provided by data brokers. These data sources can be used to augment the scoring of debtors and the determination of appropriate collections strategies.

### Innovative uses of machine learning in collections

Scorecards, such as those used for credit card applications, have traditionally employed the statistical method known as logistic regression. An introduction to the methodology commonly used during scorecard building is provided by Siddiqi<sup>8</sup>. Scorecards have some similarity to collection queues as both involve estimates of an individual's ability to pay. Logistic regression does, however, have very specific weaknesses (e.g. assumption of linear relationship between the logit of the target variable and other explanatory variables).

Machine-learning techniques are frequently used to overcome the limitations of logistic regression. Challenger models, based on machine-learning techniques, might outperform traditional techniques when the assumptions of those traditional techniques are not appropriate.

Currently, most collection agencies in South Africa use either expert knowledge or logistic regression or a combination to model collection aspects. For example, in an industry directed research project that was done for MBD in 2007, a collections scoring model was developed using logistic regression.<sup>9</sup> Another industry directed research project that was done for Absa Bank in 2012 also used logistic regression to develop a collection scorecard.<sup>10</sup> An additional example is a project done for Shapiro Shaik Defries & Associates in 2017 which also revealed (via the personal observation of A.B.) that this perception exists in South Africa. The same observation was made by DMC Debt Management (Myburgh S, 2018, email communication, February 1), that is, that most of the collection agencies in South Africa use either expert knowledge or logistic regression (or a combination).

The field of machine learning has seen a significant growth in industry interest over the past years.<sup>2</sup> Machine-learning techniques are applied to social media data in the literature and it is reported that the unstructured nature of the data, as well as some applications which require unsupervised learning, were significant contributors to the choice of using machine learning.<sup>2</sup> The most frequently employed techniques were support vector machines, classification trees and Bayesian networks. An important result to note here is that no single machine-learning technique will consistently outperform all others in all contexts. A summary of innovative uses of machine learning and other statistical modelling techniques is provided in Table 4.

It has been argued<sup>23</sup> that the use of machine learning has many possible advantages for debt collectors. Several considerations when building collection queues are raised by Davey<sup>24</sup> which could be easily solved by using machine-learning techniques. For example, a typical trend experienced by collectors is that debtors increasingly prefer to repay their debt over long periods of time rather than once-off or over short periods. Such a repayment structure requires the modelling of payment streams, in both amounts and timing. Consequently, the models used to score debtors need to take the collections period into account, as it is a significant factor affecting operational profitability (increased costs and time value of money). One of the abilities of machine-learning techniques is to find deep patterns in the data, which humans cannot reasonably be expected to identify. A critical point raised is the need to recalculate debtors' scores as regularly as possible<sup>24</sup>, as new data might become available which are not accounted for by older scores. Real-world results reported by Davey<sup>24</sup> for companies using First Data's PREDIGY<sup>®</sup> which is based on machine-learning techniques include a financial institution which reduced their losses on delinquent accounts by USD2.1 million and a management agency which increased their collection agents' performance by 18%. Note that these specific results do not convey any details on the specific machine-learning technique used, but generically refer to it as 'strategic predictive analytics'.

**Table 3:** Overview of data provided by data brokers

<b>Data domain</b>	<b>Application to debt collection</b>	<b>Examples</b>
Sensitive identity data	Possible, but ethical and legal implications	Social security number Driver's licence number Birth date Birth dates of children in household Birth dates of family members in household
Identifying data	Currently used in collections	Name Previously used names Address Address history Longitude and latitude Phone numbers Email address
Social media and technology data	Still a developing field and not used extensively in South Africa yet, but many case studies in the USA have shown possible extensions of standard models to include these data	Electronics purchases Friend connections Internet connection type Internet provider Level of usage Heavy Facebook user Heavy Twitter user (Twitter user with 250+ friends) Is a member of over five social networks Online influence estimation Operating system Software purchases Type of media posted (text, video, multimedia, etc.) Uploaded pictures Use of long-distance calling services Desktop computer access Use of mobile devices Social media and Internet accounts
Demographic data	Can be used as input to behavioural analytics models or be used to determine communications strategy	Gender Age Marital status
Court and public record data	Can be used as input to behavioural analytics models	Court orders Planned court appearances Criminal record
General interest data	Behavioural analytics and debtor profiling	Online activity Community participation
Financial data	Determine ability to pay	Employment status Length of current employment Income
Vehicle data	No case studies were found which utilised these data, but abstract underlying patterns might exist which could be used as part of the models	Vehicle ownership, including number and type of vehicles Purchase of new or second-hand vehicles
Travel data	Not used in collections currently, but applications should be investigated (e.g. perhaps people who travel extensively should rather be contacted via phone text/email and not via telephone calls)	Holiday club membership Airline loyalty member
Purchase behaviour data	Models to predict wallet prioritisation and models to profile debtors	Month-to-month variance in expenditure Fixed expenditure Blacklist status
Health data	Might determine ability to pay (fitness for work) and wallet prioritisation (e.g. medical bills)	Medical aid membership Age Parental status

**Table 4:** Innovative uses of machine learning and other statistical modelling techniques in collections

Machine learning / statistical modelling technique	Case study
Artificial neural network	Artificial neural networks were used to assess payment probability of customers <sup>11</sup> , and were used for classifying social media data <sup>12</sup> .
Bayesian network	Bayesian networks were used <sup>2</sup> in the field of social media data.
Markov chain	Markov chains were used for a 'payment sequence model' <sup>13</sup> which can be used to model repayment behaviour. A Markov decision process was also applied to improve the collections of the New York State Department of Tax and Finance. <sup>14</sup>
Support vector machine	Support vector machines were used to assess payment probability of customers <sup>11</sup> , and were used for classifying social media data <sup>12</sup> .
Text mining and speech analytics	The usage of text mining by collections agencies to assess future viability of collections is noted by De Almeida Filho et al. <sup>15</sup> The possibility of using speech analytics to assess the vast quantity of data generated by call centres which could not be feasibly done by human labour has been discussed by Ullrich et al. <sup>16</sup>
Classification tree	Classification trees were used to assess payment probability of customers <sup>11</sup> , and for classifying social media data <sup>12</sup> .
Random forest	Random forests were used to assess the payment probability of customers. <sup>11</sup>
Self-organising map	Self-organising maps combined with Cox-proportional hazard models were used to model repayment patterns of credit card debt. <sup>17</sup>
Multivariate kernel regression	Multivariate kernel regression was used to predict the probability of repayment. <sup>18</sup>
Memory-based reasoning	Memory-based reasoning was used in collections. <sup>19</sup>
Recency, frequency and monetary (RFM) analysis	Collection agencies can investigate applications of RFM analysis to their collections. <sup>20</sup> Recency might be 'How long ago was the last repayment?' Frequency might be 'How frequently does the debtor repay?' Monetary value might be 'Average repayment amount or proportion of debt repaid per repayment'. RFM analysis can then be used to segment debtors and obtain a better understanding of the factors influencing repayment behaviour, as well as as an input to a queue-building model. <sup>21</sup>
Multivariate regression	Collections were modelled with a multivariate regression model. <sup>22</sup>

Additional studies<sup>25</sup> used machine learning and other statistical techniques to increase debt collection. It might be of interest to collection companies to know which variables are typically used. Examples of the variables include financial ratios (e.g. age of account, size of account, size of repayment, liquidity of debtor); qualitative factors (e.g. account's business nature, market outlook); and behavioural factors (e.g. account transactions, balances, wallet priority).

In many cases it might be possible to quantify the monetary implication of a superior model. One such example is an interesting case study in which three machine-learning techniques (artificial neural networks, classification trees and memory-based reasoning) were compared with logistic regression.<sup>19</sup> These techniques were applied to a list of debtors whom the firm had declined to collect on. Their results showed that approximately 35% of these accounts were classified by the best model as 'good', indicating that the firm had probably misclassified a significant proportion of 'good debt' as 'bad debt' in their records. The unrecovered debt associated with these misclassified accounts totalled USD423 000. In this case, the logistic regression outperformed these three machine-learning techniques.

As mentioned in Table 4, multivariate kernel regression has also been used to predict the probability of repayment and infer the strengths of the relationships between the predictor variables and the response.<sup>18</sup> Some of the results included that homeowner status and self-employment status were not significant while bank account holding status was significant. The most interesting result of this study is that intuition might not necessarily be supported by the data. As an example, whilst it might be intuitive to assume that homeowners have better repayment behaviour than non-homeowners, this assumption was not supported in the analysed data. Therefore, it is strongly recommended that South African collection agencies evaluate the predictive performance of

the data used to create collection queues and ensure that expert-based systems are also supported by data.

The specific case study referred to in Table 4 that used a multivariate regression included a thorough analysis of the variables that can predict collections.<sup>22</sup> Data obtained from the German market, relating to more than 150 000 accounts from 78 creditors, were used to build a multivariate regression model. Again, it might be of interest for collection companies to know which variables were used, as a common challenge in modelling is identifying and acquiring appropriate input. In the German study,<sup>22</sup> the main variables that played a role in predicting collections were: exposure at default; age of account (fewer collections on older accounts); previous experience in collections; macroeconomic factors (unemployment rate for country and GDP growth); legal nature of debtor (corporations tend to have better repayment than individuals); intensity of customer complaint (an increase in intensity associated with lower collection rate); repayment method (instalments versus once-off) and bankruptcy (collections on accounts by debtors formally in bankruptcy were significantly lower).

To summarise, we found that different modelling techniques have been successfully used in collections. Some of these techniques include artificial neural networks, support vector machines, classification trees, random forests, Markov chains, text mining, self-organising maps, multivariate kernel regression, 'recency, frequency and monetary' analysis, and multivariate regression models.

### **Innovative ways to enhance collection strategies**

After analysing debtors and creating collection queues, the queues are assigned to agents who then communicate with debtors and attempt to collect as much as possible on the accounts in arrears. A summary of innovative ways to enhance collection strategies is provided in Table 5.

**Table 5:** Overview of data provided by data brokers

Innovative ways to enhance collection strategies	Example
Machine learning	Machine-learning models are used to create queues which are assigned to agents using a method dubbed the 'snake-draft method'. <sup>11</sup> A combination of optimisation theory and data modelling has been used <sup>26</sup> to develop two possible approaches to agent scheduling that incorporate random arrival and abandonment rates.
Optimised scheduling	Traditional methods of agent scheduling can be optimised using mixed-integer formulations, e.g. a spreadsheet-based solution <sup>27</sup> was implemented which takes into consideration: call volume during a time period; customer abandonment of the call; business rules determining acceptable waiting times; and staff availability, preference and seniority. This solution produced 15–20% lower aggregate personnel requirements when compared to previous scheduling methods used by the centre at no significant cost to the call centre's performance.
App-based solutions (Avatars)	The stigma surrounding bad debt and delinquent accounts affects the approachability of debtors by collections agents. Avatars have been proposed as a method for circumventing this stigma. Note that the industry leader for avatar-based collections is the AvaPay system developed by BeGuided <sup>28</sup> .
Interactive websites (online debt collection)	TrueAccord, a US software company, has developed a collections system that is reported to use behavioural analytics, machine learning and a humanistic approach to automate the collections process. An increase in collections by 20%, as well as a tenfold increase in agent productivity, has been reported. <sup>29</sup>
Real-life stories	It is recommended that real-life stories by debtors who have improved their lives by repaying bad debt should be included in collections strategies. This engagement might ease the communications process with debtors and promote the value of good repayment behaviour.
Ring-less voicemail drops	The DirectDrop Voicemail product <sup>30</sup> allows collections agents to send voicemails to debtors, without having to call. The two main benefits of such a system are that regulations restricting telephonic communication do not apply and debtors are able to access the messages in their own time, without intrusion into their personal or business lives. This approach might improve debtors' perception of collection agencies.

According to Melo<sup>31</sup>, a representative of the analytics software company FICO, the collection strategies of the future will make use of digital, automated communication. Melo's results showed that debtors prefer digital, automated interaction to agent-based interaction. Taking this result into consideration, and the fact that South African collection agencies typically perform the majority of their communication through call centres, for which personnel salaries account for up to 70% of collection costs<sup>27</sup>, innovative ways to enhance call centres are required (and possible ways are provided in Table 5).

Innovative ways to enhance collection strategies include the use of machine learning, optimised scheduling, app-based solutions, avatars, interactive websites (online debt collection), real-life stories and ringless voicemail drops. When considering these developing trends, it is reasonable to assume that the collection strategies of the future will mostly utilise digital, automated communication.

The impact of automated communication on the call centre industry is recommended as a crucial research topic, as estimates of the number of people employed in call centres in South Africa<sup>32</sup> range from 54 000<sup>33</sup> to 210 000<sup>34</sup>. Studies in the Philippines, home of the world's largest call centre industry, are already reporting reduced growth rates and increasing job complexity for call centre agents as artificial intelligence takes over simple jobs.<sup>35</sup>

### Innovative ways of evaluating collection strategies

A collection strategy encompasses the communications strategy, write-off strategy, queue assignment and more. The integration of the collection strategy with debtor scores and, consequently, the dependence on the data sources and models used to generate these scores greatly increase the challenge in fair and practical evaluations. The importance of accurately assessing collection strategies can be summarised by the phrase 'you can't fix what you don't know is broken'. Objective evaluation

and comparison of multiple collection strategies is a common challenge faced in the collections industry. Such evaluation requires an assessment of the model's performance and an unbiased comparison of the efficiency of multiple collection strategies. Causes of bias might include the application of different strategies to different debtor demographics or during periods of different macroeconomic circumstances.

There are two factors to consider when evaluating collection strategies. Firstly, before comparing the performance, it is important that macroeconomic factors be taken into account when the collection strategies under comparison are implemented at different times, as macroeconomic factors affect collections.<sup>22</sup> Secondly, it is also crucial that the performance evaluation criteria for a strategy be specified in advance, before the strategy is implemented.

Some innovative ways to evaluate strategies are given next. The first innovative way to evaluate such strategies is the use of speech analytics software<sup>36</sup> to analyse the large volume of telephonic conversations and associated data which are generated by call centres<sup>16,37</sup>. Because it is impractical for managers to listen to every phone call, a digital solution capable of identifying the quality of a phone call is a practical alternative. Two companies which offer commercial products capable of speech analytics are Nexidia's software analytics and CallMiner. We recommend that South African collection agencies investigate the usage of speech analytics to analyse debtor-agent interaction and identify key phrases associated with good and bad repayment behaviour. Different approaches to speech analytics are available.<sup>38</sup> As significant research has been done in text analytics, the usage of speech-to-text software is proposed as a method to obtain the necessary data for analysing automated collection strategies. Secondly, a specific best practice when evaluating a collection strategy is the use of a champion-challenger strategy.

Based on the literature, it is clear that separate evaluation of the critical components mentioned earlier is possible, as described in Table 6.

**Table 6:** Methods for separate evaluation

Methods for separate evaluation	Proposals and comments
Evaluation of the predictive power of data through the use of variable selection techniques	Automatic variable selection techniques are not recommended – variable selection should include considerations of both predictive power and the cost of obtaining the relevant data and consequently expert interpretation is required.  Note that this evaluation should form part of the ‘model development’ phase.
Evaluate model performance: ‘comparing apples with apples’	Apply ‘through-the-cycle’ versus ‘point-in-time’ model estimates to compensate for transitory macroeconomic conditions.  Establish reference periods for comparing models, i.e. model comparison should be based on the same data.
Challenger models can be used to evaluate the accuracy of the current score-generating model	It is critical that these challenger models be developed on the same data sets.
Evaluating model accuracy through back-testing	The scores generated by the model should be reflective of the profit a debt collector can expect to make by targeting a specific debtor for collections. Back-testing, which is the practice of comparing model predictions to actual values, is a suitable technique for evaluating the model’s accuracy. Note that this back-testing should occur at both the portfolio level and debtor level, with consideration of the individual debt amounts. Similar to account-level vs segment-level evaluation and exposure weighted vs number weighted evaluation.
Evaluating agent assignment and communication strategies	It is proposed that the existing collections strategy, specifically the method whereby agents are assigned to queues, and the communications strategy employed can be evaluated analogously to the champion–challenger methodology. The details are as follows: <ul style="list-style-type: none"> <li>• Identify different segments in an existing collections portfolio.</li> <li>• Randomly divide each segment into two uneven populations (i.e. use an 80–20 split) and apply the current strategy to the large population and the challenger strategy to the small population.</li> <li>• Statistical inference can be used to evaluate whether the challenger strategy results in significant improvements over the current strategy.</li> </ul>

## Recommendations for South African collection agencies

### *Innovative uses of data sources in collections*

The following is a list of our recommendations with respect to data sources:

- Collection agencies should consider broadening their communications strategies to include social media and digital platforms (such as applications). User interaction with such platforms generates data, which could also be used to profile debtors.
- When debtors are business owners, web-mining techniques can be used to extract data directly from individual websites. Such cases, however, might be sparse, and consequently a cost–benefit analysis of such an undertaking is advised.
- We emphasise that the predictive power and legality of any data need to be carefully assessed before data are included in any model, especially data sources from data brokers who collect data from several different domains, such as corporate (employment history, qualifications, etc.) and civil (dependants, marital status, etc.) records.
- Several social media data vendors are active in the developed world, most notably in the USA and UK. These firms provide data streams of social media data, aggregated across multiple platforms and filtered to an extent. Communication with one of these firms revealed that privacy legislation might impede the usage of such data sources to profile individual debtors. Because of these restrictions, we cannot advocate for the use of social media data, as provided by these vendors, at the current stage. Social media data, as they are currently available, may only provide a general overview of debtor demographics and high-level segmentation (such as analysing debtor repayment behaviour for rural versus urban debtors).

A notable exception is mobile phone data: despite the prevalence of mobile phones, no studies could be found which utilised mobile phone data for collections. Despite this fact, we are of the opinion that South African collection agencies should investigate the use of mobile phone data in this context.

### *Innovative uses of machine learning in collections*

The recommendations regarding debtor analysis and queue building are:

- As traditional techniques can be implemented relatively quickly and at low cost, it is recommended that these be used as a baseline.
- A critical examination of the validity of the assumptions underlying these models is advised. Should these assumptions be violated, remedial measures need to be taken. Should remedial measures not suffice, it is suggested that alternative techniques be investigated.
- When highly non-linear or abstract patterns exist in the data, machine-learning techniques such as artificial neural networks have been shown to outperform traditional techniques.
- No single modelling technique will outperform all others in all contexts: consequently, the usage of champion–challenger strategies is advised.
- Models can be built from the outputs of more than one technique; for example, a classification technique could be used to identify segments of similar debtors, thereafter a logistic regression model could be used to estimate individual debtors’ probability of payment. The collections strategy can then be aligned to the debtors’ segment, whilst an agent can be allocated according to the estimated probability of payment.
- Existing models should be used, not only for collections, but also for statistical inference so that data-driven understandings of operations can be obtained.
- Collection agencies should investigate offering longer pay-back periods to debtors, in conjunction with the agency’s clients.

### *Innovative ways to enhance collection strategies*

The recommendations for collections practices at operational level are as follows:

- Mobile applications have grown in popularity, and studies have shown that an increasing number of consumers use these apps for business. Collections-specific applications already exist and have been well received by debtors; it is reported that these applications also improve collection rates and agent productivity. It is recommended that collection agencies investigate the development of a mobile application for usage by both debtors and agents, as it can provide an easily accessible, innovative and standardised platform for collections and communication.
- Studies have shown that the performance of call centres can be accurately assessed using speech analytics. It is recommended that a cost–benefit analysis be performed to evaluate whether existing software should be licenced for this purpose.
- The usage of avatars for communication purposes has been shown to be effective in reducing the stigma associated with bad debt, thus easing the communications process. It is advised that the usage of avatars for guiding debtors through the collections be investigated experimentally.
- Several different assignment strategies of agents to collections queues are possible, and recommendations in this regard are provided in the section titled 'Innovative ways to enhance collection strategies'.
- Linear programming – a commonly used method in decision support – can be utilised to optimise agent scheduling for call centres.
- Collections strategies of the future will most probably pivot towards digital, automated communication.

### *Innovative ways to evaluate collection strategies*

Our recommendations regarding performance strategies are:

- It is critical that the collections strategy (queue-building model, business rules, etc.) used at any specific time be well documented to allow for fair comparisons of the performance of different collections strategies.
- The metric used to assess the collections strategy's performance – such as collections rate, collections period or collections cost – needs to be specified upfront.
- When assessing the performance of two different collections strategies, it is necessary to make sure that a valid comparison is made. For example, investigate the debtor profiles of the two different collection strategies. Also, the specific time period could be compared to determine a valid comparison (different macro-economic scenarios can make two strategies incompatible). Champion–challenger strategies are commonly used to evaluate and improve model performance.

The implementation of these recommendations will be critical to ensuring that South African collections agencies operate profitably, ethically and in line with international standards.

## **Conclusions**

Research on cutting-edge innovations for collection agencies was summarised. Key phases of the collections cycle were identified, and critical aspects of those phases were researched to determine the latest developments. A comprehensive overview of the research findings was provided and theoretical studies, as well as practical applications through case studies, were included. The understanding of the literature and the collections industry in general was used to identify possible applications of these findings to South African collection agencies and recommendations were provided.

It is clear that the field of collections is a complex one, but several slight improvements in the various aspects of a collection agency's operations can, in aggregate, produce significant improvements to the efficiency and profitability thereof. The collections strategies of the future will most likely mature into digital, automated interaction. The recommendations will most probably find their way into future best practices of the collections industry.

## **Recommendations for future research**

We have identified the following topics as viable future research prospects:

- The development of an operation evaluation strategy, comprising methods for the evaluation of all relevant aspects affecting a collections agency's performance, i.e. evaluation of data sources, score generating models, agent assignment strategies, communication methods, etc.
- The development of industry specific back-testing techniques.
- Investigation of the state of South African debt collections agencies with respect to the recommendations made in this paper.

## **Authors' contributions**

V.B.: conceptualisation; methodology; data collection; data analysis; validation; data curation; writing revisions; project leadership. T.V.: conceptualisation; sample analysis; data analysis; writing the initial draft; writing revisions; student supervision; project management. A.B.: conceptualisation; methodology; project leadership.

## **References**

1. Thomas L, Matuszyk A, Moore A. Comparing debt characteristics and LGD models for different collections policies. *Int J Forecast*. 2012;1(28):196–203. <https://doi.org/10.1016/j.ijforecast.2010.11.004>
2. Injadat M, Salo F, Nassif AB. Data mining techniques in social media: A survey. *Neurocomputing*. 2016;214:654–670. <http://dx.doi.org/10.1016/j.neucom.2016.06.045>
3. DataSift. Human data intelligence [webpage on the Internet]. c2017 [cited 2017 Sep 12]. Available from: [www.datasift.com](http://www.datasift.com)
4. Gnip. Release the power of social media data [webpage on the Internet]. c2017 [cited 2017 Sep 12]. Available from: [www.gnip.com](http://www.gnip.com)
5. Informatica Marketplace. Extract data from LinkedIn, Facebook, and Twitter [webpage on the Internet]. c2016 [cited 2017 Apr 12]. Available from: [https://marketplace.informatica.com/solutions/extract\\_data\\_from\\_linkedin\\_facebook\\_and\\_twitter](https://marketplace.informatica.com/solutions/extract_data_from_linkedin_facebook_and_twitter)
6. Pottharst A. Webmining: Examples and applications [webpage on the Internet]. c2017 [cited 2017 Apr 12]. Available from: [http://dewiar.staff.gunadarma.ac.id/Downloads/files/11103/Datamining-Webmining\\_Arne.pdf](http://dewiar.staff.gunadarma.ac.id/Downloads/files/11103/Datamining-Webmining_Arne.pdf)
7. Clearinghouse PR. Data brokers [webpage on the Internet]. c2013 [cited 2017 Apr 10]. Available from: <https://www.privacyrights.org/data-brokers>
8. Siddiqi N. Credit risk scorecards: Developing and implementing intelligent credit scoring. Hoboken, NJ: John Wiley & Sons; 2006.
9. De Vries L. Client specific collections scoring model (BMS9105/2008) [mini-dissertation]. Potchefstroom: North-West University; 2008. Available from: <http://library.nwu.ac.za/>
10. Mdalana S. Development of a model for predicting customers who accept and stay in arrangement collections (BMS9226/2012) [mini-dissertation]. Potchefstroom: North-West University; 2012. Available from: <http://library.nwu.ac.za/>
11. Jongmyoung K, Pilsung K. Late payment prediction models for fair allocation of customer contact lists to call center agents. *Decis Support Syst*. 2016;85:84–101. Available from: [http://nemo.yonsei.ac.kr/wp-content/uploads/2016/11/20161107\\_Late-payment-prediction-models-for-fair-allocation-of-customer-contact-lists-to-call-center-agents.pdf](http://nemo.yonsei.ac.kr/wp-content/uploads/2016/11/20161107_Late-payment-prediction-models-for-fair-allocation-of-customer-contact-lists-to-call-center-agents.pdf)
12. Al-Saggaf Y, Islam MZ. Data mining and privacy of social network sites' users: Implications of the data mining problem. *Sci Eng Ethics*. 2014;21:941–966. <https://doi.org/10.1007/s11948-014-9564-6>



13. Thomas LC, Matuszyk A, So M, Mues C, Moore A. Modelling repayment patterns in the collections process for unsecured consumer debt: A case study. *Eur J Oper Res.* 2016;249:476–486. <https://doi.org/10.1016/j.ejor.2015.09.013>
14. Abe N, Melville P, Pendus C, Jensen DL, Thomas VP, Bennett JJ, et al. Optimizing debt collections using constrained reinforcement learning. In: *Proceedings of the 16<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2010 July 25–28; Washington DC, USA. New York: ACM; 2010. p. 75–84. <https://doi.org/10.1145/1835804.1835817>
15. De Almeida Filho AT, Mues C, Thomas LC. Optimizing the collections process in consumer debt. *Prod Oper Manag.* 2010;6:698–708.
16. Ullrich M, McCracken D, Craft M. The delicate balancing act: Using speech analytics to optimize collections and mitigate risk. Waltham, MA: Callminer Inc.; 2010. Available from: <https://callminer.com/wp-content/whitepapers/Using-Speech-Analytics-to-Optimize-Collections-and-Mitigate-Risk.pdf>
17. Ho Ha S, Krishnan R. Predicting repayment of the credit card debt. *Comput Oper Res.* 2012;39:765–773. <https://doi.org/10.1016/j.cor.2010.10.032>
18. Grant C, Madula M. Using bounds to investigate household debt repayment behaviour. *Res Econ.* 2013;67:336–354. <https://doi.org/10.1016/j.rie.2013.09.007>
19. Zurada J, Lonial S. Comparison of the performance of several data mining methods for bad debt recovery in the healthcare industry. *J Appl Bus Res.* 2005;21(2):37–54. Available from: <https://pdfs.semanticscholar.org/c4d5/a5934769a6c9e851004514c01bc95a84becb.pdf>
20. Ivanovic S, Mikinac K, Perman L. CRM development in hospitality companies for the purpose of increasing the competitiveness in the tourist market. *UTMS J Econ.* 2011;2:59–68. Available from: <http://utmsjoe.mk/files/Vol.%202%20No.%201/0-2-1-6-IvanovicS-Mikinac.pdf>
21. Dursun A, Caber M. Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Manage Perspect.* 2016;18:153–160. <https://doi.org/10.1016/j.tmp.2016.03.001>
22. Beck T, Grunert J, Neus W, Walter A. What determines collection rates of debt collections agencies? *Finan Rev.* 2017;(52):259–279. <https://doi.org/10.1111/fire.12131>
23. De Jager P. How does data analytics help debt collection [webpage on the Internet]. c2017 [cited 2017 Apr 16]. Available from: <http://insights.principa.co.za/how-does-data-analytics-help-debt-collection>
24. Davey D. Collect more for less: Strategic predictive analytics is key to growing collections and reducing costs [document on the Internet]. c2009 [cited 2017 Apr 17] Available from: [https://www.firstdata.com/downloads/marketing-fs/fd\\_collectmoreforless\\_whitepaper.pdf](https://www.firstdata.com/downloads/marketing-fs/fd_collectmoreforless_whitepaper.pdf)
25. Nadolny M. The use of predictive modelling to boost debt collection [document on the Internet]. c2017 [cited 2017 Apr 17]. Available from: [https://www.business-school.ed.ac.uk/waf/crc\\_archive/2013/116.pdf](https://www.business-school.ed.ac.uk/waf/crc_archive/2013/116.pdf)
26. Bertsimas D, Doan XV. Robust and data-driven approaches to call centres. *Eur J Oper Res.* 2010;207:1072–1085. <https://doi.org/10.1016/j.ejor.2010.05.040>
27. Dietz DC. Practical scheduling for call centre operations. *Omega.* 2011;39:550–557. <https://doi.org/10.1016/j.omega.2010.12.001>
28. BeGuided Inc. AvaPay: Payment services with a smile [webpage on the Internet]. c2017 [cited 2017 May 31]. Available from: <https://www.beguided.com/#avapay>
29. TrueAccord. Debt collection, reimagined [webpage on the Internet]. c2017 [cited 2017 May 31]. Available from: <https://www.trueaccord.com>
30. VoApps Inc. VoApps DirectDrop voicemail home [webpage on the Internet]. c2017 [cited 2017 May 31]. Available from: <http://www.voapps.com/>
31. Melo D. How mobile data technologies are changing debt collection and transforming businesses [webpage on the Internet]. c2013 [cited 2017 Jun 01]. Available from: <http://www.itproportal.com/2013/12/17/how-mobile-technologies-are-changing-debt-collection-and-transforming-businesses/>
32. Callrica – Call centre innovation [homepage on the Internet]. c2016 [cited 2017 Sep 29]. Available from: <http://www.callrica.co.za/>
33. Keeble J. SA Commercial [webpage on the Internet]. c2014 [cited 2017 Sep 27]. Available from: <http://www.sa-commercial.co.za/sa-commercial-direct-the-growing-call-centre-sector-in-south-africa/>
34. Cohen M. South Africa ideal for call centres, accent 'sounds British'. *Mail & Guardian.* 2013 December 02; Business [cited 2017 Sep 27]. Available from: <https://mg.co.za/article/2013-12-02-south-africa-ripe-for-call-centre-expansion>
35. Cusi VS. Mid- to high-level BPO jobs to be in-demand in the Philippines [document on the Internet]. *Contact Center World.* 2017 September 08 [cited 2017 Sep 2017]. Available from: <https://www.contactcenterworld.com/view/contact-center-news/mid-to-high-level-bpo-jobs-to-be-in-demand-in-the-philippines.aspx>
36. Kendrick S. The beginner's guide to speech analytics technology [webpage on the Internet]. c2016 [cited 2017 Jun 01]. Available from: <https://callminer.com/blog/the-beginners-guide-to-speech-analytics-technology/>
37. Davey N, Hill-Wilson M. The speech analytics guide [document on the Internet]. c2017 [cited 2017 Jun 05]. Available from: <http://www.nexidia.com/media/1318/nexidia-the-speech-analytics-guide.pdf>
38. Mason D, Turner P. Beginner's guide to speech analytics [webpage on the Internet]. c2013 [cited 2017 Jun 01]. Available from: <https://www.callcentrehelper.com/beginners-guide-to-speech-analytics-57123.htm>

