OPTIMIZING DEBT COLLECTION USING ARTIFICIAL INTELLIGENCE

R·D·S
RECOVERY DECISION SCIENCE

recoverydecisionscience.com
(888) 384-0734
contact@rdscientist.com

An Artificial Intelligence (AI) Based Suit-Decisioning Methodology to Debt Collection Optimization

Artificial Intelligence (AI) has provided a much-needed makeover to the debt collection process. Due to the availability of massive amounts of consumers' historical data, AI technologies can bring real-world business significance for banks and other financial institutions in their debt collection strategies. AI applications allow lenders to optimize their debt management services by prioritizing accounts in their litigation channel. Consumers' possible patterns of behavior can be revealed by designing AI-based predictive models, which ultimately provide optimal strategies for litigation decisions.

What's inside? We'll show you how one top10 bank could have recovered 96% of its collections by targeting only 30% of accounts.

Recovery Decision Science was founded to support the collections industry in two ways:

1. To create the best treatments possible
2. To generate completely new revenues out of accounts that hadn't yet paid a dime

Powered by decades of Unifund's industry leadership, RDS helps any business sitting on volumes of non-paying accounts by offering two levels of support:

- Products: Developed an ever-evolving suite of analytical products to enhance profitable returns through improvements in areas such as suit-decisioning and asset identification.
- Portfolio Servicing: Provide global creditors and investment companies with strategic recovery workflow and optimization solutions for portfolios of distressed assets.

In this whitepaper, we focus on RDS's extensive AI-driven analytics in consumer segmentation, clustering, likelihood prediction of consumer debt repayment, cost estimation of collection process, and litigation profit optimization. RDS' analytics are developed based on a large volume of multimodal data and we curate various types of data (e.g., image, text, categorical, ordinal and numerical data) that are effective in maximizing recovery. Our data are gathered through different resources such as vendors, third-party APIs, credit bureau data, public records, etc. We also gather data, such as litigation costs, from our own proprietary database.

SECTION I

SUIT-DECISIONING WITH PAYMETRIX PROFITABILITY INDEX (PI)

RDS has built AI-based suit-decisioning tools using our rich data to identify the most profitable accounts and prioritize the litigation process by focusing on those accounts. Suit-decisioning is mathematically formulated as a prediction task aiming at estimating expected profitability of accounts and maximizing litigation by scoring accounts based on the estimated profitability. RDS analyzes numerous input variables for this purpose; among these variables, Asset has been shown as one of the most significant variables. In fact, Asset significantly helps gauge the ability of debt repayment by looking into the consumer's home and job status. Different Asset scenarios are available with respect to the consumer's home and job status. If the consumer is a home owner and already employed, the Asset variable takes the value of Both (B). In case the consumer is only employed but renting his/her house, the asset variable is identified as Job (J). For the scenario that the consumer is not employed but is a home owner, the Asset is defined as Home (H). Finally, in case that consumer is not employed and is a renter, the Asset variable indicates None (N). Intuitively, consumers with a "Both" asset are more likely able to pay back their debts than consumers with a "None" asset. Statistical analysis on historical suit-decisioning events has also verified the predictive importance of Asset.

Intuitively, consumers with **BOTH** assets are more likely able to pay back their debts than consumers with **NONE** assets

Although Asset is a significant variable in our prediction task, verifying information on the home and job status of consumers is a timely and very expensive process (it can take up to 90 to 120 days). Some creditors may want to avoid this costly expense and make a decision quickly on their accounts, so waiting for verification of consumer asset information might not be the best option. Hence, RDS has also created an asset prediction model to accelerate suit-decisioning by predicting a consumer's Asset. The predicted Asset will then be used instead of an actual Asset in the original model. The overview of our suit-decisioning methodology with predicted Asset is shown in Figure 1.

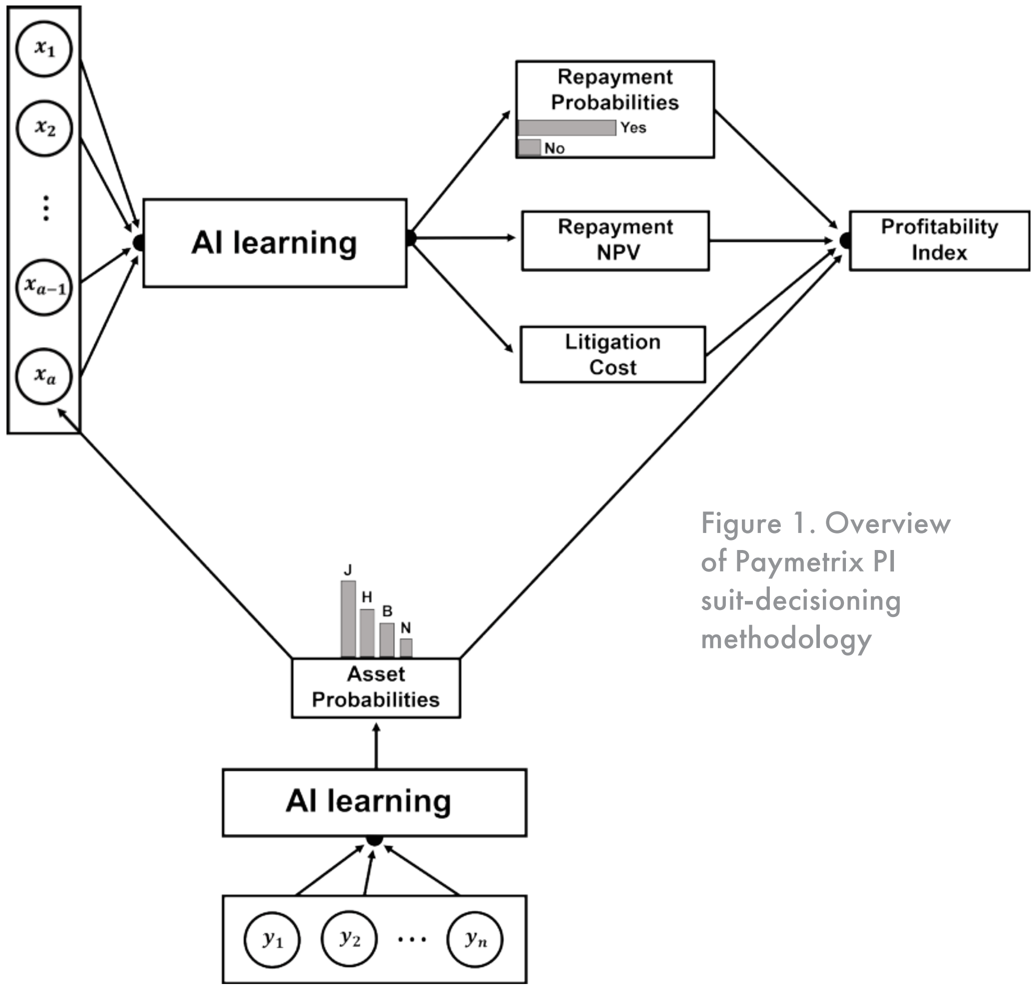


Figure 1. Overview of Paymetrix PI suit-decisioning methodology

Our methodology uses multiple input x_i's $(\forall i = 1, 2 \ldots, a)$ variables to predict repayment probabilities, repayment net present value (NPV), and litigation cost. There are three output variables. Repayment probability (P) measures how likely the consumers will repay the debt; thus, it is a probability value between 0 (0% chance of repayment) and 1 (100% chance of repayment). Repayment NPV (NPV) measures the repayment amount in the present value. Finally, litigation cost (LC) represents the expenses associated with the litigation processes such as court cost, regional costs, etc.

Our methodology adopts a multi-task learning regime, which includes the classification task of repayment probabilities and regression tasks related to repayment NPV and litigation costs.

The upper part of Figure 1 represents our suit-decisioning method assuming consumer Asset information (denoted by x_a) is known. Given this assumption, the predicted outputs will then be used to estimate the **profitability Index (PI)**:

$$PI = f(P, NPV, LC)$$

$$= \frac{P \cdot NPV}{LC} \qquad (1)$$

If x_a is not known and there is a motivation to run suit-decisioning without actual information on Asset (i.e. quick and cost-effective analysis) the bottom part of Figure1 applies. Asset prediction also uses multiple inputs y_i's $(\forall i = 1, 2 \ldots, a)$ to determine the probability of a consumer having Asset, which is designed as a multi-class classification problem with 4 classes (H, J, B, and N). The output of Asset prediction is shown as below:

$$p(x_a = \mathrm{J}|y_1, y_2, .., y_n) = p_J$$

$$p(x_a = \mathrm{H}|y_1, y_2, .., y_n) = p_H$$

$$p(x_a = \mathrm{B}|y_1, y_2, .., y_n) = p_B$$

$$p(x_a = \mathrm{N}|y_1, y_2, .., y_n) = p_N$$

where

$$p_J + p_H + p_B + p_N = 1 \qquad (2)$$

Due to the above probabilistic outputs, the Asset probabilities will be incorporated into the final PI analysis. In other words, we run the suit-decisioning methodology for all 4 Asset scenarios and then weight the resulted PIs with the corresponding Asset probabilities. This can be represented as an expectation of PI over x_a as shown in the following:

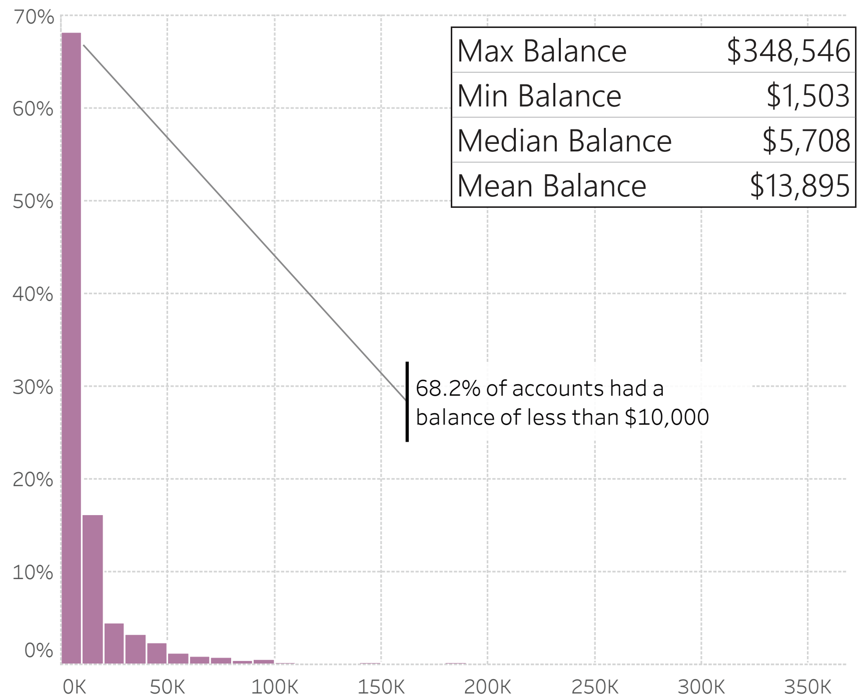
$$PI = E\left[f\ (P, NPV, LC)_{x_a}\right]$$

$$= \sum_{k \in \{J,H,B,N\}} p_{\mathrm{m}} \cdot f\ (P, NPV, LC|x_a = \mathrm{k}) \qquad (3)$$

The proposed methodology is designed to be flexible based on the client's goal. If the goal is more accurate suit-decisioning, which requires incurring cost and time to gain Asset information, Eq. (1) will be used, but if the goal is a quick (less than 24 hours) and cost-effective suit-decisioning then Eq. (3) will be applied (which will be less accurate).

CASE STUDY: APPLICATION OF SUIT-DECISIONING IN THE BANKING INDUSTRY

To verify the performance of our PI suit-decisioning methodology, a top 10 bank in the US selected a test pool of accounts that had already been collected upon. The requested data points included account information with charge-off date, balance, age, consumer name, and address. The bank did not provide any information regarding collection figures for the accounts. The test pool included 1,495 accounts of mortgage, unsecured, and auto deficiency account types. The charged-off dates of accounts ranged from June 2018 to February 2019. The balance distribution of accounts with some statistics are presented in Figure 2.



Max Balance	$348,546
Min Balance	$1,503
Median Balance	$5,708
Mean Balance	$13,895

Figure 2. Balance statistics and distribution of the test accounts

Given the requested information from the bank and our proprietary data pipelines, we gathered input variables to run our PI analysis. The client's goal was to have a quick and cost-effective analysis on their accounts, hence, Eq. (3), predicted asset, was applied as the basis for PI suit-decisioning. The PI score for each test account was calculated and then summarized into a range of 10 deciles. These deciles were 10 groups, each consisting of the same number of accounts, ranked by their PI values.

As an example, the first decile represented the 10% of accounts with the highest PI values, the second decile represented the next 10% with the next highest values, etc. The PI scores and deciles were sent to the bank so that they could evaluate the performance of our suit-decisioning methodology. The bank compared the account-level PI rankings to their actual account collection data. The bank then provided the collection data back to RDS to compile these results.

A summary of the collection on the accounts, using their own collection strategy, is provided in Table 1. The total value of the pool was $20,773,770. Overall, 22.7% of accounts made payments totaling $2,564,068, which represented 12.3% of the total value.

Total Pool Balance	$20,773,770
Number of Accounts	1495
Collected Amount	$2,564,068
Percentage of Collection	12.3%
Percentage of payers	22.7%

Table 1. Bank's collection strategy resuts for test pool

The next step was to determine how PI would have helped the bank to prioritize collections efforts on this pool of accounts. **Table 2 shows the accounts sorted by their PI deciles and compares that to actual collection amounts provided by the bank**.

PI Decile	Collections	% Recovered
10%	$ 1,785,000	69.6%
20%	$ 550,932	21.5%
30%	$ 119,156	4.6%
40%	$ 28,151	1.1%
50%	$ 21,819	0.9%
60%	$ 14,674	0.6%
70%	$ 15,431	0.6%
80%	$ 9,590	0.4%
90%	$ 10,509	0.4%
100%	$ 8,807	0.3%

Best 30% of Accounts Yielded 95.7% of Total Collections

Table 2. PI performance with respect to bank recovery

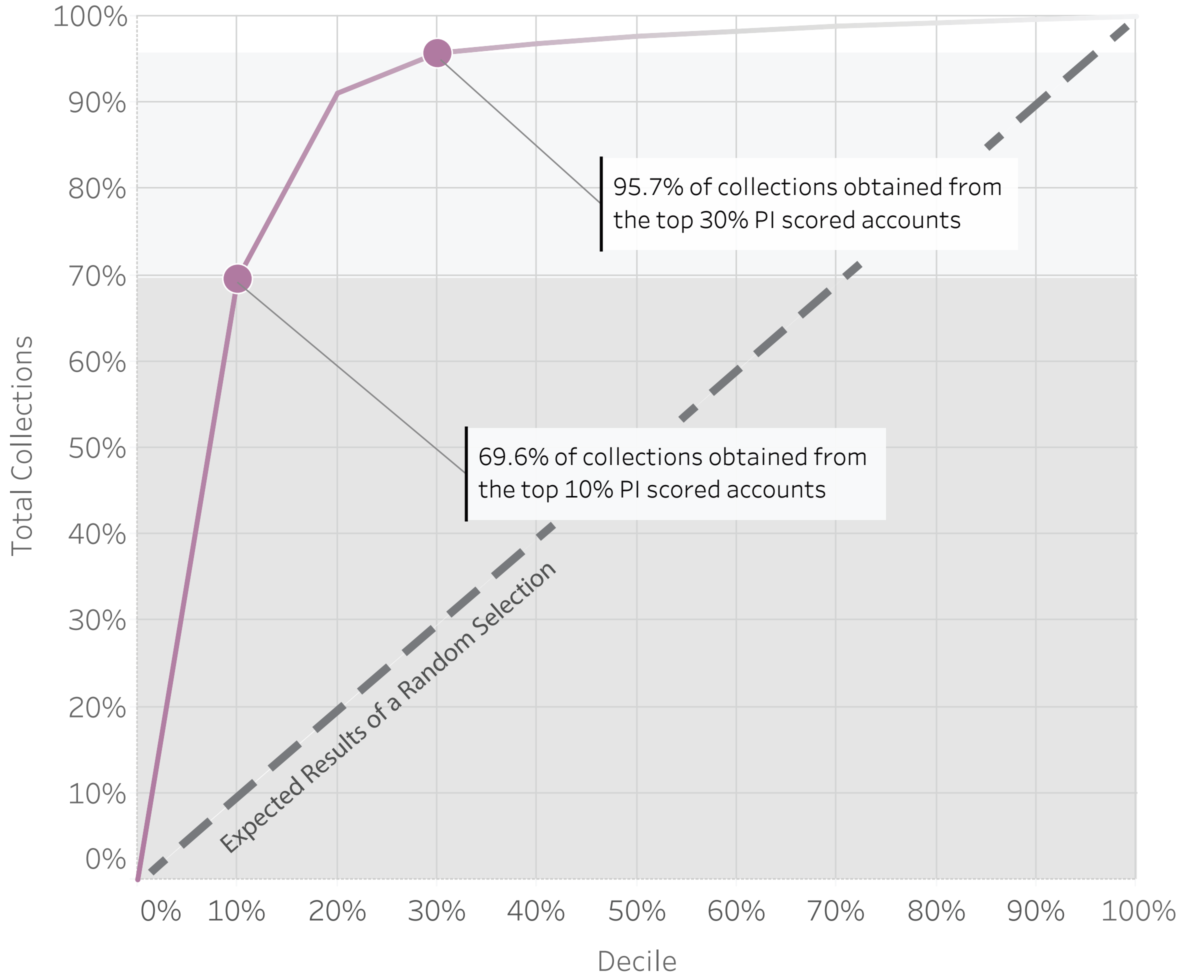


Figure 3. PI performance compared to random selection

Table 2 shows the results, that higher PI deciles (i.e. first 3 deciles - top 30%) have resulted in significantly higher collections vs. the accounts in lower deciles (i.e. bottom 70%). To be specific, the top decile (top 10%) alone accounts for nearly 70% of the bank's total collections. Figure 3 shows PI performance on this pool as compared to a random selection (where it would be expected that 30% of the accounts would yield 30% of the total collections). This means that if the bank had used PI, they could have targeted the top 30% of the accounts (i.e. top 449 accounts with highest PI values out of 1495 accounts) and yielded 96% of their total collections.

The top 3 deciles (top 30%) combine to cover nearly 96% of the bank's total collections on this pool.

The 70% of accounts with lower PI scores yielded a mere 4% of their total collections.

Clearly this performance is exceptional, and while we expected PI to outperform a random selection, this far exceeded the anticipated results. The performance of PI on in-house accounts is presented in Figure 4. From this figure, PI is expected to outperform random sampling that eventually verifies the significance of PI in cost effective suit decisioning optimization. However, the outperformance of PI over the bank's accounts is still exceptionally significance.

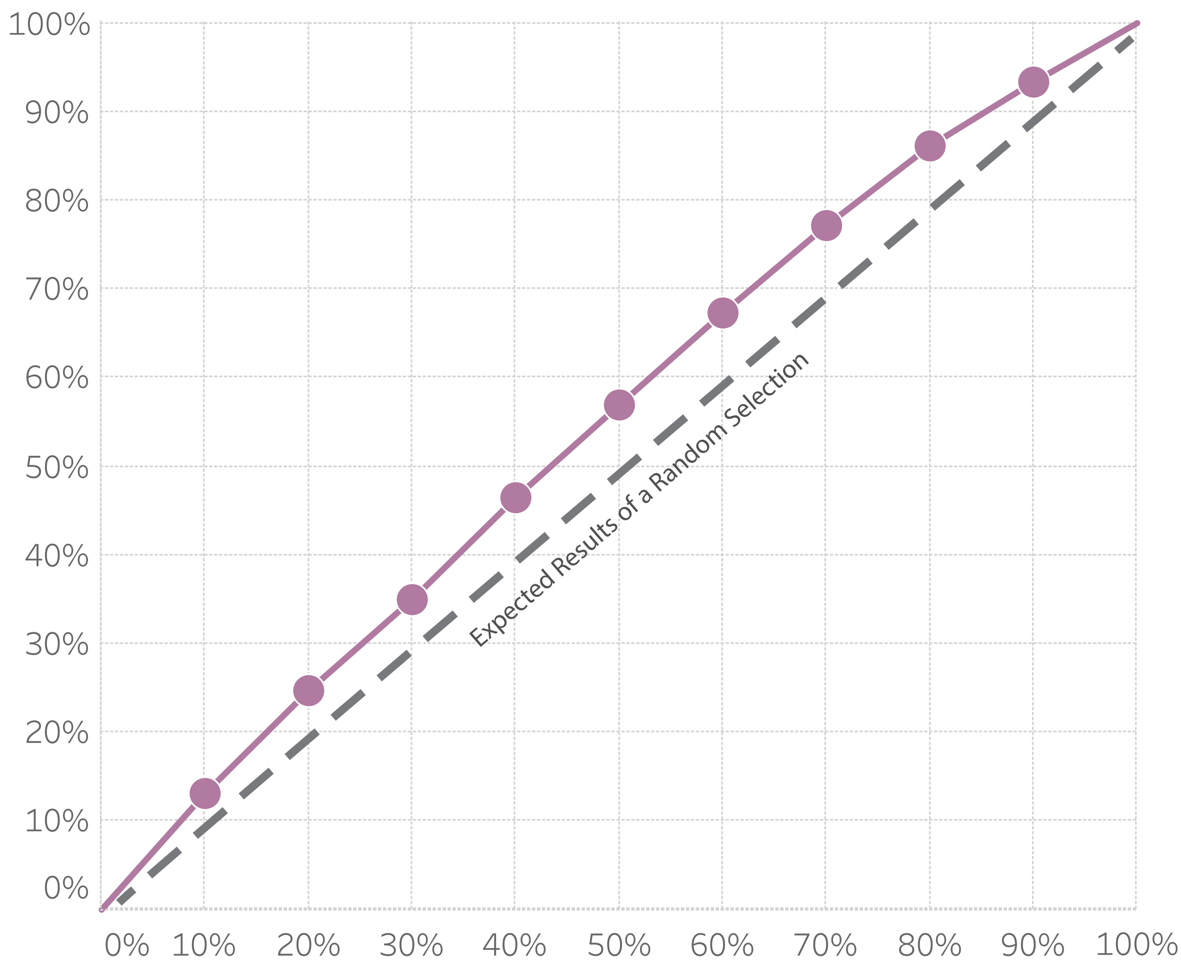


Figure 4. PI performance compared to random selection on in-house accounts

CONCLUSION

This paper introduced an AI based suit-decisioning methodology named PI. The proposed PI allows banks and other financial institutions to optimize their debt collection strategies by prioritizing their litigation decisions. PI was developed based on multi-task learning regime using multimodal dataset. The proposed methodology is equipped with an asset predictive modeling that allows for quick and cost-effective analysis. The effectiveness of the proposed methodology was verified based on analysis of a test pool adopted from one of the top 10 banks in the US. The results indicated that PI would significantly improve the bank litigation outcomes. **Targeting only 30% of the accounts with the highest PI values, they would be able to recover 96% of their original collections**. Although we do not believe that results such as these are typical, the results do confirm that PI is a powerful tool that financial institutions can use to prioritize their accounts in their collection strategies, maximizing collections and lowering costs.

R·D·S
RECOVERY DECISION SCIENCE

TO LEARN MORE ABOUT RECOVERY DECISION SCIENCE:

Call Jonathon Wall at **(888) 384-0734,**
visit **recoverydecisionscience.com**
or email **contact@recoverydecisionscience.com**