

Collection scoring via técnicas de Machine Learning

Amanda Xavier

11 de julho de 2021

Conteúdo

1	Introdução	3
2	Contextualização e Revisão teórica	4
2.1	Collection Score	4
2.2	Regressão Logística	4
2.3	Floresta aleatória	4
2.4	SVM - Suport Verctor Machine	4
2.5	Análise discriminate	4
2.6	Avaliação de modelos de classificação	4
2.7	Avaliação de modelos de tempo **** Melhorar esse nome	4
3	Simulação dos dados	4
A	Anexo	5

Lista de Figuras

Lista de Tabelas

1 Introdução

No dia a dia dos consumidores é comum ouvir falar em crédito. O crédito trás aos consumidores uma ampliação de recursos financeiros, possibilitando tanto a aquisição de novos bens quanto o pagamento de dívidas e financiamentos. Esta ampliação de recursos em diversos setores é extremamente importante para a economia de um país, influenciando diretamente no PIB.

As instituições financeiras tem grande interesse no ramo de concessão de crédito, devido ao alto retorno associado ao capital investido. No entanto a concessão de crédito também está associada a diversos riscos.

Quando falamos em riscos, há diversos aspectos a serem analisados. Tanto os relacionados á instituição que irá ceder o crédito quanto aos clientes que receberão. Do ponto de vista das instituições um dos principais riscos é o risco de inadimplência.

Saber escolher bem para quem liberar crédito e o quanto liberar é essencial para que as instituições financeiras obtenham bons retornos dos créditos cedidos. Para que esta decisão seja tomada, existem diversos fatores a serem analisados e o grande volume de pessoas e empresas buscando por crédito, torna inviável que estas decisões sejam tomadas de forma manual. O histórico das instituições com os clientes faz com que seja possível entender e agrupar clientes em perfis semelhantes, e estes dados são utilizados na criação de modelos que preveem se um cliente será ou não uma boa escolha para a instituição que está analisando o crédito a ser cedido. Melhores modelos de crédito se tornam diferenciais para as instituições, ajudando-as a maximizar os lucros.

O uso de modelos estatísticos trás mais agilidade e confiança nas decisões tomadas, pois levam em consideração o histórico e informações de outros clientes, ao invés de somente a visão subjetiva dos analistas. No entanto, desde as mais simples as mais sofisticadas técnicas de análise de crédito trazem alguma incerteza e os clientes selecionados podem não pagar o valor combinado, ou pagar parcialmente.

A partir de momento em que os clientes ficam inadimplentes, o novo desafio é traçar estratégias para recuperar os valores. Os devedores tem perfis diferentes e a forma de abordá-los na cobrança da dívida pode impactar no pagamento. Entender o comportamento destes devedores é essencial para que a estratégia adequada seja utilizada. Para alguns clientes, uma cobrança feita muito cedo pode fazer com que um cliente que iria pagar dívida deixe e pagar, para outros esta cobrança feita mais cedo poderia estimular o pagamento. A negativação de uma dívida, tras gastos para as instituições, e muitas vezes a instituição poderia evitar o gasto ao esperar mais um curto período para o pagamento.

O desafio é saber qual estratégia tomar em cada cliente, e novamente, fazer de forma manual não seria a melhor estratégia, tanto pelo grande volume quanto pela visão subjetiva. Novamente vemos a necessidade e a importância de criar modelos que ajudem a prever a probabilidade de um devedor quitar sua dívida ou parte dele e/ou em quanto tempo isso aconteceria. Com estes modelos e o estudo de como tratar cada perfil de devedor, as empresas podem maximizar a recuperação do crédito.

Não há muito conteúdo disponível acerca de modelos de recuperação de crédito no Brasil, os dados são de difícil acesso, fazendo com o que o conhecimento dos modelos desenvolvidos fiquem restritos as empresas que os desenvolveram. A exploração destes modelos é importante e tem potencial para trazer grandes retornos as instituições que os utilizam, por isso, vimos a importância de explorar e desenvolver as técnicas de recuperação através de modelos estatísticos e de aprendizado de máquina.

2 Contextualização e Revisão teórica

2.1 Collection Score

2.2 Regressão Logística

Umas das técnicas mais conhecidas, quando estamos falando em classificação binária é a regressão logística.

2.3 Floresta aleatória

2.4 SVM - Suport Verctor Machine

2.5 Análise discriminate

2.6 Avaliação de modelos de classificação

2.7 Avaliação de modelos de tempo **** Melhorar esse nome

3 Simulação dos dados

Os dados utilizados para o ajuste de modelos de *collection Score* costuma levar em consideração diversos tipos de informação relacionados á

A Anexo