

MODELOS DE CREDIT E BEHAVIOR SCORING: MODELAGEM E PRINCIPAIS INDICADORES DE PERFORMANCE

Joyce Maceno da Rocha (PIBIC/CNPq/FA/Uem), Daniele Cristina Tita Granzotto (Orientador), e-mail: dctgranzotto@uem.br.

Universidade Estadual de Maringá / Centro de Ciências Exatas e da Terra/Maringá, PR.

Área e subárea do conhecimento conforme tabela do [CNPq/CAPES](#): 10203001 – Probabilidade e Estatística Aplicadas

Palavras-chave: Modelagem estatística, regressão logística, indicadores de performance

Resumo:

Os modelos de *credit* e *behavior scoring* são modelos amplamente usados na classificação durante o processo de aprovação e acompanhamento do risco. Esses modelos são adotados pelas instituições financeiras quando solicitado crédito a mesma, onde precisam decidir sobre sua concessão ou não, pautados no risco que este cliente apresenta para a instituição. Sendo assim, este projeto tem por objetivo estudar os modelos de regressão logística simples e múltipla, e acompanhar desde sua validação por meio de indicadores de performance: índice de estabilidade populacional (IEP), Kolmogorov Smirnov (KS), coeficiente de Gini e acurácia.

Introdução

Diante de uma solicitação de crédito as instituições financeiras precisam decidir sobre sua concessão ou não, pautados no risco que este cliente apresenta para a instituição. Nesta difícil decisão de concessão ou não de crédito ao consumidor, a estatística apresenta ferramentas, amplamente difundidas, que auxiliam na tomada de decisão, modelos de classificação de risco, modelos estocásticos de risco de crédito e modelos de risco de portfólio.

Dentre os modelos supracitados, os modelos de classificação de risco são os mais aplicados na mensurabilidade do risco de tomadores em transações individuais ou de uma carteira de crédito como um todo. Este modelos atribuem uma medida (pontuação), que representa a expectativa de risco de *default*, geralmente expressa na forma de uma classificação de risco *rating* ou de pontuação *score*, que permita identificar o quão próximo o consumidor está de dois grupos: “bom” que é provável cumprir com suas obrigações financeiras ou um grupo de “mau”, cujo pedido deve ser negado devido à alta probabilidade de faltar com seus compromissos na instituição financeira.

Pensando nesta grande área de aplicação bancária, o presente projeto tem por objetivo estudar os modelos de regressão logística simples e múltiplo, a fim de aplicá-lo a dados reais. Além de modelar uma base de dados bancária, também temos o interesse de estudar os indicadores de performance desta classe de modelos a fim de monitorá-los. Para acompanhamento e validação do modelo construído utilizamos os seguintes indicadores de performance: índice de estabilidade populacional (IEP), Kolmogorov Smirnov (KS), coeficiente de Gini e acurácia.

Materiais e métodos

A regressão logística, ou análise logit, foi desenvolvida pelo estatístico David Cox em 1958 e é uma técnica de análise multivariada, apropriada para as situações nas quais a variável dependente é categórica e assume um entre dois resultados possíveis (binária). Apesar de ser o modelo mais simples e amplamente utilizado, o modelo de regressão logística binária tem extensões para mais de dois níveis da variável dependente:

- item quando se trata de respostas categóricas com mais de dois valores uma opção é o uso da regressão logística multinomial;
- caso as respostas sejam várias categorias, ordenadas, podemos modelar via regressão logística ordinal.

Independente do tipo, o objetivo da regressão logística é gerar uma função matemática cuja resposta permita estabelecer a probabilidade de uma observação pertencer a um grupo previamente determinado, em razão do comportamento de um conjunto de variáveis independentes. Ou seja, matematicamente temos

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

sendo, $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ os coeficientes a serem estimados.

A base de dados disponível para o trabalho refere-se a clientes pessoa jurídica, tomadoras de crédito com investimentos acima de R\$50 mil. Apesar de considerada pequena, a carteira tem em torno de 8 mil clientes, esta é capaz de movimentar mensalmente mais de R\$ 2 bilhões. Neste cenário é redundante afirmar que maus clientes impactariam de maneira direta os lucros e prejuízos sofridos pela instituição. O estudo de performance, foi realizado por um período de doze meses antecedentes ao ponto de observação. Após terem sido acompanhados, os clientes em estudo foram reclassificados.

Assim, as seguintes variáveis foram utilizadas no processo de modelagem:

- Percentual da linha de crédito utilizado;
- Máximo de dias em atraso nos últimos 12 meses para o produto;
- Tempo de relacionamento com o banco;
- Índice de utilização da conta garantida flex;

- Índice médio de utilização dos últimos 12 meses e média de dias que utilizou o cheque especial no mesmo período;
- Percentual saldo médio devedor e credor dos últimos 12 meses;
- Percentual saldo médio credor e faturamento;
- Restritivos abertos;
- Restritivos no último ano;
- Evolução do saldo total pelo saldo de curto prazo

Resultados e Discussão

Os seguintes resultados foram obtidos a partir do modelo construído.

Tabela 1 - Estimativas dos parâmetros do modelo

Variáveis em Estudo	Parâmetro Estimado
Percentual de linha de crédito utilizado	
Até 50%	0,61
Acima de 50%	
Máximo Dias de atraso nos últimos 12 meses	
Sem atraso	0,40
Até 15 dias	
15 dias ou mais	-0,74
Tempo no Banco	
Até 1 ano	0,02
De 1 a 2 anos	0,09
De 2 a 4 anos	0,18
De 4 a 8 anos	0,34
De 8 a 16 anos	0,57
➤ 16 anos	1,03
missing	
IU Conta Garantida Flex	
Até 50%	-0,18
Mais de 95%	
Missing ou de 50 a 95%	041
Média de IU nos últimos 12 meses com média de dias de utilização do cheque nos últimos 12 meses	
IU acima de 90% e 10 ou mais dias de utilização do cheque	-0,90
IU acima de 80%	-0,58
IU de 20 a 80% ou Sem limite	
IU de 5 a 20% ou IU menor que 5% com 10 ou mais dias de utilização	0,37
IU até 5% com menos de 10 dias de utilização	1,13
Saldo médio devedor/saldo médio credor ult 6 meses	
Até 50% ou sem saldo devedor	
Acima de 50%	
Saldo credor/ faturamento	
0%	-0,53
≤ 0,000104	-0,41
De 0,000104 a ≥ 0,006652	
➤ 0,006652	0,21
Restritivos abertos	
Sem Restritivo atual nos últimos 12 mese	

CCF ou REFIN ou mais de 2 restritivos abertos	-0,58
Restritivos no Último ano	
Possui CCF e REFIN no último ano	-1,31
Possui CCF e REFIN mais ação ou protesto ou restr. interno	-0,69
Não possui os restritivos	
Evolução do saldo total/ saldo curto prazo	
Sem Bacen nos 12 meses	
Ev sldtot<0 ev sldCP<0	-0,16
Ev sldtot>0 ev sldCP<0	-0,33
Ev sldCP>0	-0,49

Para o modelo apresentado acima o índice de estabilidade populacional se manteve entre 10 e 20% ao longo dos próximos 12 meses. Ainda observamos um KS para todos os meses superiores a 70%, como se espera para essa classe de modelos (behavior scoring).

Conclusões

Através do cálculo dos indicadores de performance é possível tomar decisões visando encontrar um ponto ótimo, minimizando o risco tomado ao credor do crédito e com isso, maximizando a rentabilidade da empresa (banco).

Instituições financeiras no geral, utilizam combinações entre os status dos indicadores, para uma possível ação.

Agradecimentos

Agradeço ao CNPq e Fundação Araucária por esse projeto.

Referências

ANDERSON, Raymond. **The credit scoring toolkit: theory and practice for retail credit risk management and decision automation**. Oxford University Press, 2007.

HAIR, Joseph F. et al. **Multivariate data analysis**. Upper Saddle River, NJ: Prentice hall, 1998.

HOSMER, David W.; LEMESHOW, Stanley. **Applied regression analysis**. New York, John Willey, 1989.

HOSMER, David W.; LEMESHOW, Stanley. Special topics. **Applied Logistic Regression, Second Edition**, p. 260-351, 2000.

MAYS, Elizabeth (Ed.). **Handbook of credit scoring**. Global Professional Publishing, 2001.[7] THOMAS, Lyn; CROOK, Jonathan; EDELMAN, David. **Credit scoring and its applications**. Siam, 2017.