

Projeto Supervisionado

Caio Almasan de Moura

ra: 095620

Índice

1. Introdução	
2. Principal Projeto: Modelo de Score	
2.1. <i>Objetivo</i>	pg 3
2.2. <i>Agentes Envolvidos</i>	pg 3
2.3. <i>Contextualização</i>	pg 3
2.4. <i>Conceitos</i>	pg 4
2.5. <i>Cronograma</i>	pg 5
3. Escopo para Modelagem	
3.1. <i>Linhas de Créditos Consideradas</i>	pg 5
3.2. <i>Poder Preditivo do Modelo</i>	pg 5
4. Análise Univariada	pg 6
5. Análise Bivariada	pg 6
6. Análise Multivariada	
6.1. <i>O que é Análise multivariada?</i>	pg 6
6.2. <i>Regressão Logística</i>	pg 7
7. Validação de modelo	
7.1. <i>O teste de K-S para duas amostras</i>	pg 7
7.2. <i>O teste do risco relativo</i>	pg 7
8. Estabilidade do modelo final	
8.1. <i>Índice de Estabilidade Populacional</i>	pg 8
8.2. <i>Teste KS1</i>	pg 8
9. Conclusão	pg 8

1. Introdução

Estágio em Modelagem de Risco de Crédito

Intituição: Banco Panamericano

Cargo horária: 18 horas semanais (6 horas por dia, 3 dias por semana)

Principais atividades: Análise de bases no EXCEL e SPSS. calculo de índices (FPD, FPD2, M3OVER30, PDD entre outros), manutenção dos modelos de crédito para todos os produtos (Veículos, Crédito Pessoal, Crédito Imobiliário, Consórcio, CDC), frequente acompanhamento das noticias macroeconomicas que influencia diretamente na linha de crédito.

2. Principal Projeto: Modelo de Score

2.1. Objetivo

Descrever o procedimento de desenvolvimento do modelo de concessão de crédito para pessoa física na linha de crédito motos do Banco Panamericano.

2.2. Agentes Envolvidos

Área de Gerência Executiva de Modelagem do Banco Panamericano.

2.3. Contextualização

O modelo desenvolvido tem a finalidade de estimar a probabilidade de inadimplência nas operações de concessão para pessoa física, utilizando horizonte de 9 meses com marcação de atraso no final do período.

Este relatório fornece detalhes do desenvolvimento, das análises e dos resultados obtidos.

Na primeira etapa do projeto foram definidas as necessidades e os objetivos do projeto. O segundo passo é caracterizado pela análise do conceito, construção das variáveis, consolidação das bases, definição da técnica a ser utilizada e estimação dos modelos. Para esses modelos a técnica utilizada foi regressão logística.

O relatório está organizado da seguinte forma:

2.4. Conceitos

Com o objetivo de definir o conceito de inadimplência que seria utilizado na modelagem, foram comparadas as taxas de inadimplência geradas por vários cenários: M6OVER60, M6OVER90, M9OVER60 e M9OVER90.

Para maior entendimento, segue abaixo definições e conceitos que serão usados durante a análise:

-Out-of-time: Amostra de safras com período posterior ao usado para o desenvolvimento e validação do modelo.

-Out-of-sample: Amostra de uma safra atual.

-KS (*Kolmogorov-Smirnov*): Estatística usada para a comparação de distribuição entre duas amostras.

-Coeficiente de ROC: Estatística que analisa a capacidade preditiva de um modelo de classificação, baseado nas distribuições acumuladas de bons e maus, para cada faixa de escore.

-Gini: Estatística calculada com base no coeficiente de ROC que analisa a capacidade preditiva de um modelo de classificação.

-PSI (*Population Stability Index*): Estatística usada para verificar possíveis distorções no perfil da população em relação à base de referência. É baseada na proporção de indivíduos em cada categoria.

-KS1: Estatística também usada para verificar possíveis distorções no perfil da população em relação à base de referência. É baseada na distribuição acumulada de ambas as populações.

Escore: Pontuação atribuída pelo modelo de crédito conforme o perfil do cliente e da proposta.

2.5. Cronograma

- Entrega da base de dados: 09/04/2011

- Entrega da base com marcação de atraso de mercado, com conceito de Bancos/Financeiras, feito pela SERASA Experian: 02/05/2011
- Entrega dos resultados do desenvolvimento do modelo: 25/05/2011
- Entrega da documentação contendo as técnicas e procedimentos utilizados nas fases de desenvolvimento, teste e validação do modelo: 31/05/2011
- Implantação do modelo: a ser definido.

3. Escopo para a Modelagem

3.1. Linhas de Crédito consideradas

A linha de crédito considerada para desenvolvimento do modelo foi veículos motos, exceto produtos de transferência de dívida, leasing e para pessoa jurídica.

3.2. Poder preditivo do modelo

Os índices utilizados para avaliar o poder preditivo do modelo foram KS2, Coeficiente de ROC, Coeficiente de Gini, PSI e KS1.

4. Análise Univariada

O que é a análise univariada?

Utilizamos a análise univariada para estudar a distribuição de frequência de cada variável separadamente. A partir desta podemos identificar valores inconsistentes, missings, outliers, investigar sua origem e decidir como tratá-los de modo que não comprometam a validade e adequabilidade dos modelos estatísticos.

5. Análise Bivariada

O que é a *análise bivariada*?

O interesse maior na análise bivariada é analisar a relação entre duas variáveis, ou seja, a relação da variável preditora com a variável 'status' que caracteriza o cliente.

Os objetivos dessa análise são:

- Analisar o potencial discriminador de uma variável preditora;
- Analisar e refinar a categorização das variáveis;
- Identificar comportamentos estranhos ou inesperados de uma variável;
- Identificar correlações entre as variáveis preditoras.

6. Análise Multivariada

6.1. O que é a *análise multivariada*?

Conjunto de métodos que permite análise simultânea de duas ou mais variáveis, levando em consideração as correlações existentes, que permitem inferências sobre o conjunto de variáveis em um nível de significância conhecido.

6.2. *Regressão logística*

A técnica utilizada no desenvolvimento dos modelos foi regressão logística. Além de apresentar um custo computacional razoavelmente baixo, seus parâmetros são facilmente interpretáveis. Esse método consiste em estimar a probabilidade através da combinação linear das características preditivas.

7. Validação do Modelo

7.1. O teste de K-S para duas amostras

O KS é uma das medidas de avaliação de performance mais utilizadas no mercado e mede a capacidade do escore de distinguir os “bons” e “maus” pagadores. Quanto maior for a estatística de KS, maior será a separação entre os clientes “bons” e “maus”. O valor encontrado é a máxima diferença entre as distribuições acumuladas de “bons” e “maus”.

7.2. O teste do risco relativo

O risco relativo é uma medida de chance dentro de cada classe. Representa o quociente entre a proporção de bons e a proporção de maus na categoria. É uma medida ponderada pelo total em relação ao status do cliente (bom/mau).

Quanto maior for o risco relativo, maior será a chance de encontrar clientes bons em relação aos clientes maus dentro de cada classe.

8. Estabilidade do Modelo Final

8.1. Índice de Estabilidade Populacional

O ÍEP/PSI é utilizado para comparar as distribuições de variáveis categorizadas.

O cálculo é feito através da fórmula seguinte:

$$IEP = (r_i - m_i) \times \ln \left(\frac{r_i}{m_i} \right)$$

Na qual r_i é a proporção de indivíduos da categoria i da amostra de referência e m_i é a proporção de indivíduos da categoria i da amostra de monitoramento.

8.2. Teste KS1

O KS1 compara distribuições de entrada de propostas observadas no desenvolvimento com a entrada de propostas obtidas em safras mais recentes, ou seja, é a máxima diferença entre as distribuições acumuladas da base de desenvolvimento e da base mensal de monitoramento.

O cálculo é feito através da fórmula seguinte:

$$KS1 = Máx|F_{r_i} - F_{m_i}| \quad (8.2.1)$$

Em que F_r é a distribuição acumulada da amostra de referência (desenvolvimento) e F_m é a distribuição acumulada da amostra mensal de monitoramento.

9. Conclusão

Baseados em estudos macroeconômicos, estáticos, análise de perfis e com histórico de nossas bases internas, criamos um novo Modelo de Score que já está implantado em nosso sistema. Apesar de estar em vigor há pouco tempo (um mês), alguns estudos já indicam a redução da zona cinzenta e uma melhor distribuição e performance entre os Bons e os Maus clientes. A estimativa é que, em 6 meses, nossos índices de inadimplência sejam reduzidos em até 30%, sem afetar a produtividade do banco, ou até com possível aumento da produção.