

Modelagem de Previsão



Ronald Poon Affat. FIA FSA MAAA CFA

VP e Diretor – América do Sul

Agosto de 2014

O que há em comum entre essas empresas?

Google



amazon.com[®]

facebook



NETFLIX

Resposta

- QUANTIDADE ENORME DE DADOS / Mineração de dados
- Acesso a computação de alta velocidade
- Técnicas e software avançados para o desenvolvimento de modelos
- Ambiente de forte concorrência, em que até os pequenos diferenciais produzem grandes ganhos de participação de mercado (market share)
- Todas utilizam a Modelagem de Previsão (MP)

Estudo de caso: Netflix



Contextualização

- A Netflix oferece filmes e programas de TV, entre outros, mediante uma taxa mensal.
- Os assinantes têm uma "fila" de conteúdos a que desejam assistir.
- Podem avaliar, com um sistema de 5 estrelas, os filmes e programas que veem.
- A Netflix oferece recomendações personalizadas utilizando o Cinematch.
- A Netflix observou uma grande relação entre a retenção de clientes e a quantidade de filmes que os assinantes veem e avaliam positivamente.

Objetivos comerciais

- As recomendações de conteúdos são parte essencial da estratégia da empresa.
- A Netflix fez até um "concurso" para **aumentar a precisão dessas recomendações**.
- Para a empresa, esse sistema das recomendações representa **"um dos nossos ativos mais valiosos"**.

Estudo de caso: Netflix (cont.)



O Concurso

- Base de dados disponibilizada ao público com mais de 100 milhões de avaliações de assinantes
- O banco de dados era muito simples, com muito poucas variáveis (usuário, título do filme, ano de lançamento, avaliação e data dela). Os participantes podiam acrescentar outros dados a sua análise.
- É oferecido um **prêmio** de US\$ 1 milhão a quem consiga desenvolver um modelo (de previsão das avaliações futuras dos usuários) **que supere o Cinematch em mais de 10%**.
- O modelo e o método dos algoritmos vencedores, em várias etapas do concurso, **devem ser disponibilizados** ao público.
- **No primeiro ano**, mais de 20.000 equipes de 150 países foram inscritas e mais de 2.600 delas apresentação predições. (“PhDs trabalhando por cerca de um dólar por hora”)
- **A solução vencedora** foi uma **combinação de 107 modelos distintos de várias equipes**.

— Cerca de 2.000 horas de trabalho

— **75% da melhoria deveu-se a uma combinação de três modelos**

Beisebol

- Em 1999, Billy Beane (administrador do Oakland Athletics) desenvolveu uma utilização inovadora para a mineração de dados.
- Um time sem muitos recursos
 - Em 12º lugar (entre 14 equipes) em termos de folha de pagamento
 - Como competir contra aos times mais ricos?
- Beane contratou um grande especialista para analisar estatísticas.
- Assim, conseguiu contratar jogadores excelentes, mas pouco valorizados no mercado.
- 12 meses depois de Beane assumir o time, o Oakland estava em 2º lugar!

Conclusão

- Beane *quantificou* o desempenho dos jogadores.
 - Não de maneira perfeita, mas melhor que os outros
- Conclusão:
 - Fique atento para os campos em que é preciso um especialista para tomar decisões baseadas em uma síntese criteriosa de informações quantificáveis em várias dimensões.
 - Isso não se parece com subscrição de seguros?
 - Com um Modelagem de Previsão , talvez dê para ganhar o jogo??

Exemplos de Modelagem de Previsão

1. Previsão do tempo
2. Definição de locais onde ocorrerão crimes
3. Projeções da procura por certos produtos em supermercados
4. Redução das taxas de evasão escolar `drop out rates`
5. Definição de locais para abertura de novas lojas do McDonald's
6. Oportunidades de venda cruzada
7. Filtros para determinar se uma certa email é spam
8. Busca de doadores potenciais para universidades, instituições de caridade e os politicos
9. Setor de telecomunicações – projeções de demanda e previsão de churn
10. Mecanismo de busca Google
11. Estimativas de público nos estádios
12. Atribuição de pontuação de crédito
13. Previsão de congestionamentos de tráfego
14. Busca de clientes que possam estar grávidas (Target)

História real

Target

Target coupons inform father of daughter's pregnancy

By **Kevin Hoffman** Mon., Feb. 20 2012 at 7:30 AM

5 Comments



Categories: **Target**



192



30



28



1



5

A Minneapolis man discovered that his daughter was pregnant after Target started sending her coupons for baby clothes and cribs, according to **an eye-opening New York Times story**.

The story chronicles how Target builds complex profiles of its customers, assigning each of them a "Guest ID" and studying their purchase choices to predict future sales opportunities. Pregnancy presents a pointed opportunity to change buying behavior, so Target's algorithms are particularly good at identifying new customers by the second trimester.



Target knows when you are pregnant.

A formula da gravidez



O que é Modelagem de Previsão ?

A modelagem preditiva é um processo, não um produto!

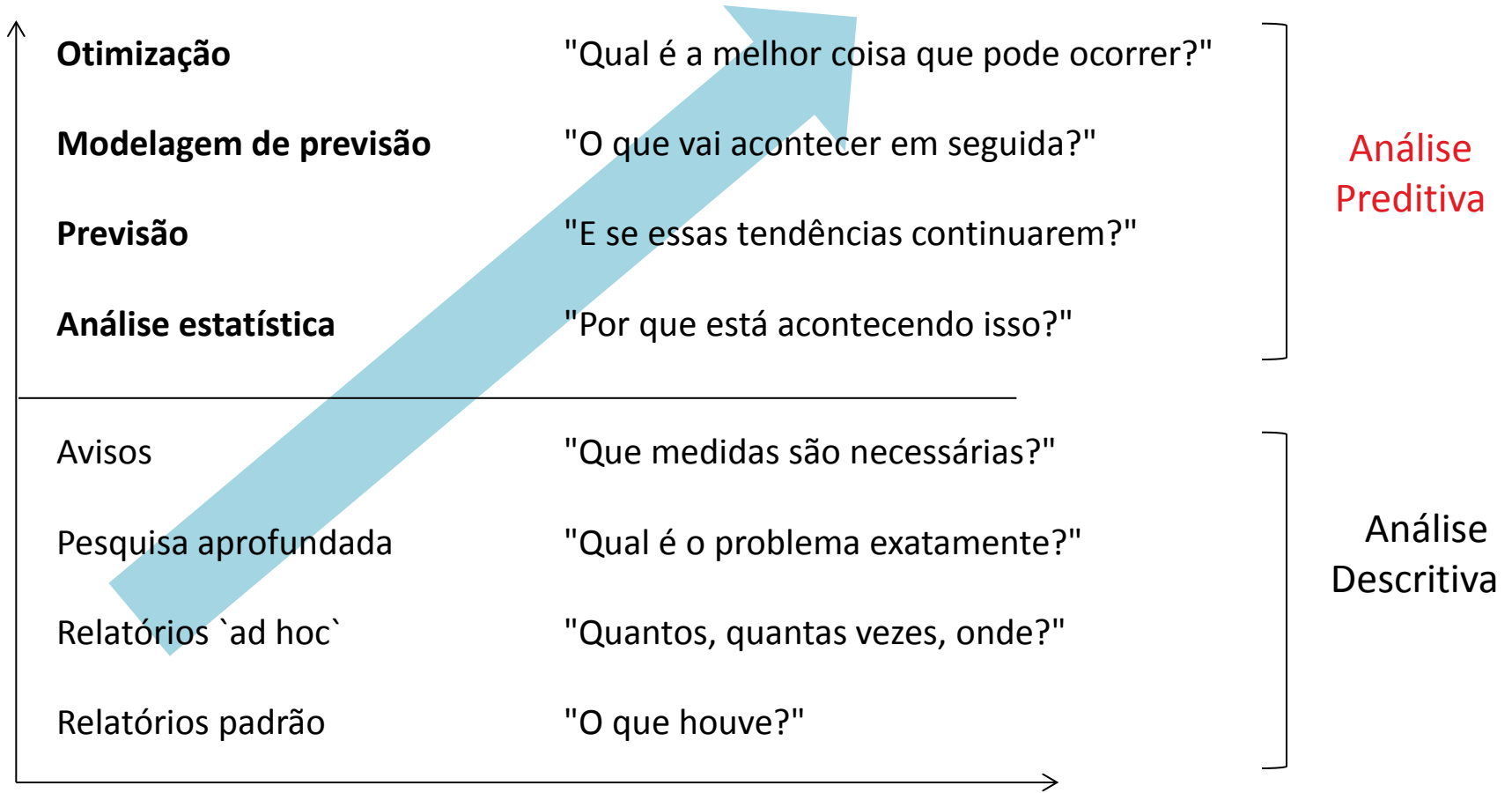
É essencial dispor de bons dados (variáveis e resultados) para desenvolver um modelo que faça sentido.

Um processo pelo qual fatos atuais ou passados são utilizados para fazer previsões sobre acontecimentos ou comportamentos futuros

As "previsões", em geral, são feitas com base em modelos estatísticos avançados.

Normalmente, os modelos buscam determinar a probabilidade de um resultado futuro (acontecimentos e comportamentos).

Análise Descritiva X Análise Preditiva



Dados e análise avançada

Por que a "Modelagem de Previsão" está sendo tão bem aceita?

Oferta



Demanda

Esses resultados são oriundos de talento, intenção, ferramentas e dados

- **Mix de ciência de dados e visão de negócio, conhecimento técnico**

-36% das empresas possuem time dedicado para *insights* de dados

- **Qualidade, dados consistentes armazenados com fácil acesso**

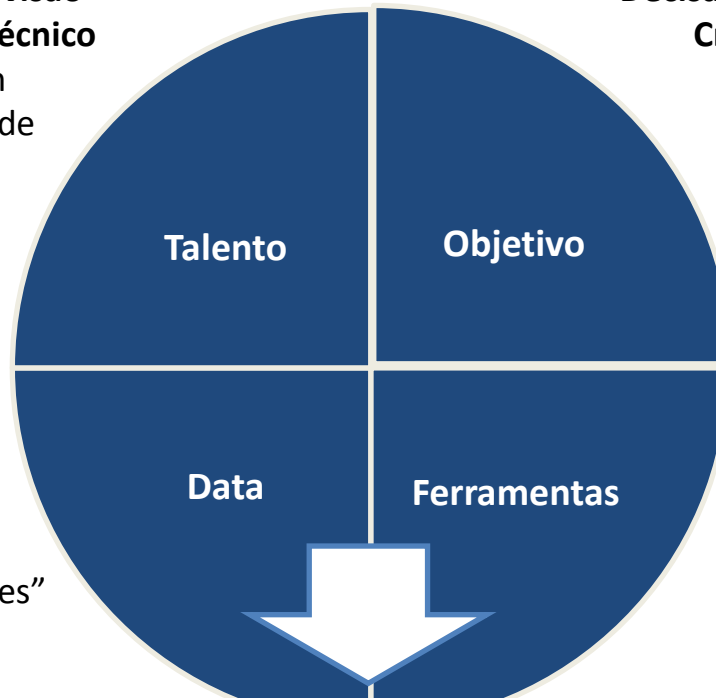
-Apenas 19% possuem “alta qualidade, dados consistentes” em sua organização

- **Decisão para ter orientação por dados; Criação de estruturas, processos e incentivos para apoiar a tomada de decisão por dados**

-23% têm uma estratégia clara para a incorporação efetiva de estrutura de análise de dados

- **Ferramentas “estado da arte” como Hadoop, NoSQL, HPCC ou algoritmos automatizados**

-38% estão usando ferramentas “estado da arte”



Apenas 4% das empresas atendem a todos esses critérios

4 principais tipos de oportunidades com Big Data

1

Melhorar processos internos



Usa mídia social e algoritmos complexos para identificar influenciadores da marca sobre temas importantes



Usa análise para encontrar relações inesperadas que melhoram as vendas (ex, cerveja ao lado de fraldas)

2

Melhorar produtos ou serviços existentes



Maximiza o engajamento pela escolha da hora ideal para enviar mensagens em nome de seus clientes



Usa passes eletrônicos para monitorar visitantes em seu parque de diversões e otimizar a experiência do parque

3

Construir novas ofertas



Integra dados meteorológicos em tempo real com dados de clientes internos para o varejo, seguros e CPG



Usa análise para monitorar o comportamento do condutor e proativamente selecionar sua base de clientes

4

Transformar o modelo de negócios



Detecta fraude telefônica em tempo real usando análise de áudio e uma base de dados de "fraudadores"



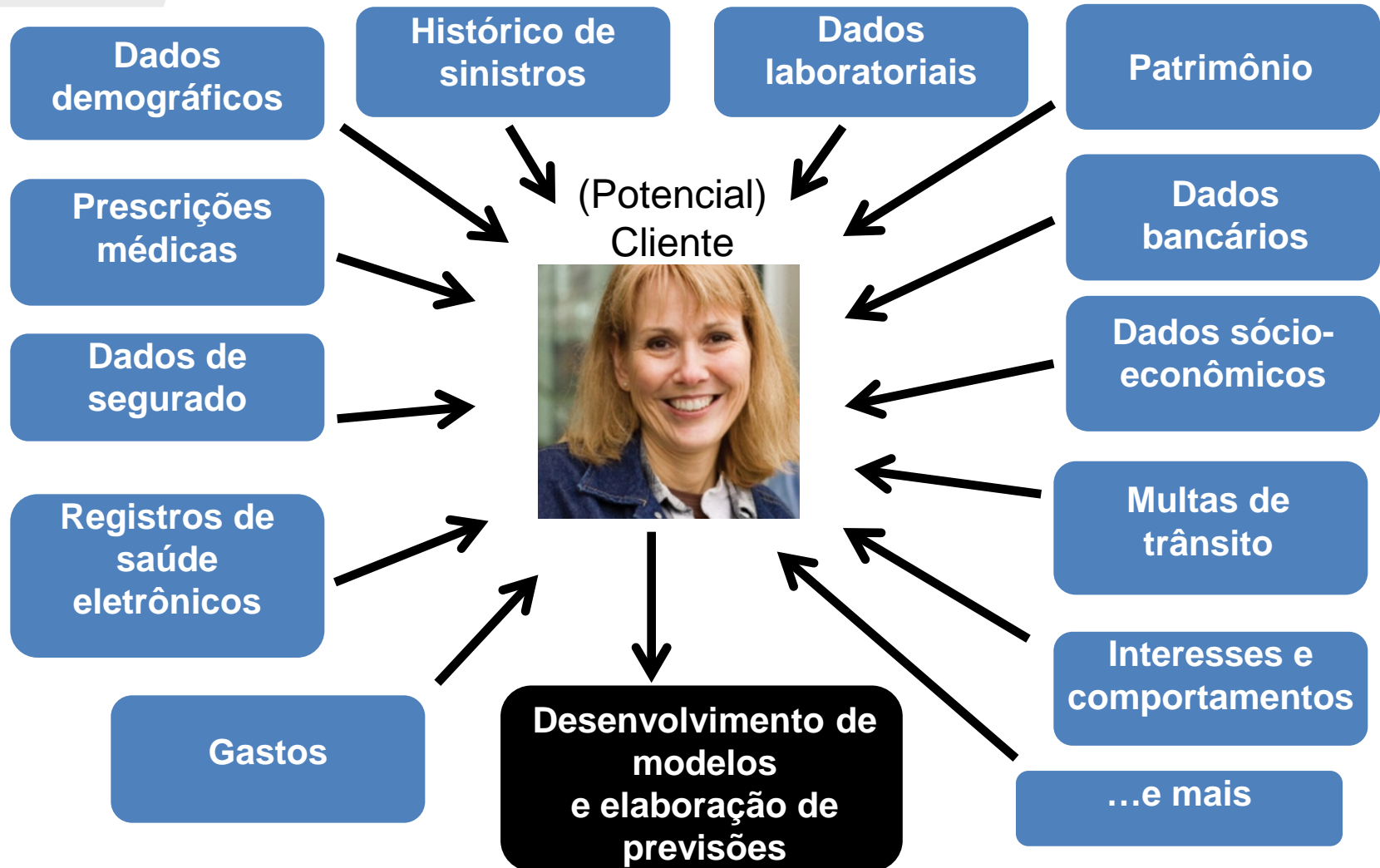
Usa análise preditiva para direcionar o cuidado preventivo, tornando-se rentável para atender pacientes "doentes"

Aplicação no setor de seguros



Optimização do valor dos dados disponíveis

Comportamental/Titular da apólice/Dados da apólice/Territorial





Muitas aplicações possíveis...

Subscrição

- Identificação dos melhores níveis de risco
- Decisões rápidas e coerentes
- Priorização de casos
- Redução da quantidade de apólices não-emitidas

Sinistros

- Previsão da frequência de sinistros
- Estimativas da frequência de sinistros
- Priorização de recursos
- Identificação de sinistros com maior probabilidade de fraude/rescisão

Precificação / Reservas

- Precificação mais adequada
- Incorporação mais eficiente de correlações
- Premissas baseadas em fórmulas
- Computação de variações de reserva

Análise de sinistralidade

- Identificação mais precisa de fatores indutores de sinistro
- Trabalho com dados pouco confiáveis
- Criação de tabelas das premissas

Vendas e Marketing

- Campanhas eficientes
- Recomendações
- Modelos para recrutamento de corretores
- Monitoramento de corretores

Gestão de apólices vigentes

- Segmentação de clientes
- Previsão de cancelamentos e planejamento de estratégias de retenção
- Modelos baseados em Customer Lifetime Value

Projetos de Análise Preditiva do RGA no Mundo

Reino Unido:

- Configuração da base (mortalidade, morbilidade e cancelamentos)
- Modelo de precificação baseado no CEP
- Aperfeiçoamento da análise de experiência
- Previsão de tabagismo
- Subscrição preditiva com de agências de crédito

Europa:

- Subscrição preditiva com dados de "bancassurance"

África do Sul:

- Análise da taxa de rescisão de seguros em grupo por invalidez permanente

Uma série de aplicações para análise de dados em todo o mundo.



Índia:

- Previsão de sinistros fraudulentos

Austrália:

- Subscrição preditiva / venda cruzada com base em dados de "bancassurance"

EUA:

- Substituição do preço de seguros em grupo por invalidez permanente
- Segmentação de clientes
- Segmentação de riscos geográficos
- Premisas de cancelamentos
- Modelo de subscrição preditiva / venda cruzada com dados de não-vida
- Caducidade da cobertura ao final de apolice

Ásia:

- Subscrição preditiva com base em dados de "bancassurance"
- Aprimoramento da segmentação de preço
- Disposição de compra

Visão geral do processo de modelagem preditiva

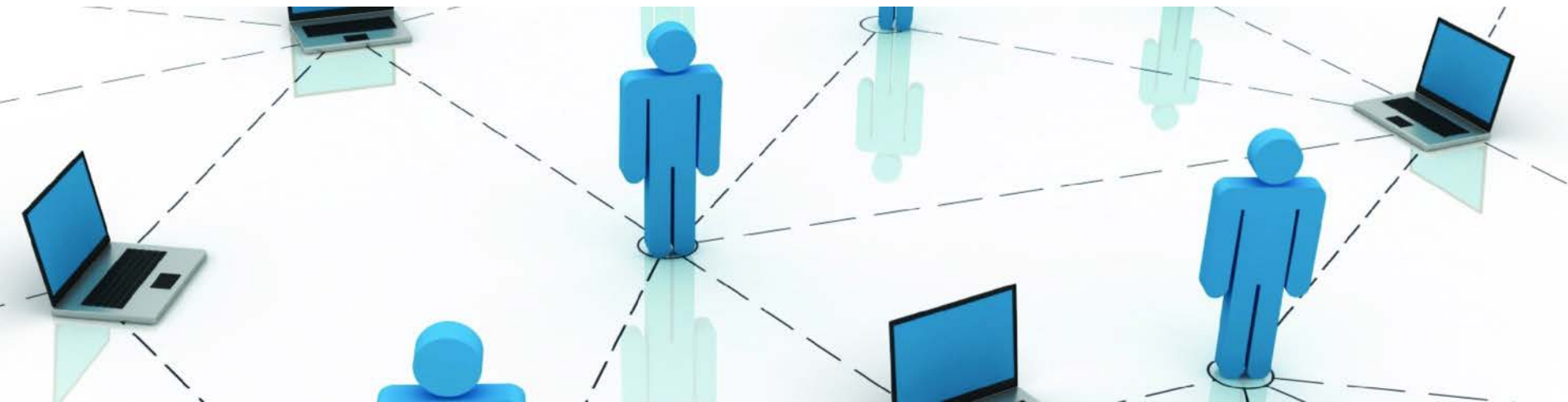
1. Definição da finalidade do modelo

2. Coleta e preparação dos dados

3. Desenvolvimento dos modelos

4. Interpretação e aplicação dos modelos

5. Acompanhamento dos resultados e atualizações



Actuaries do it with.... models



Técnicas de desenvolvimento de modelos

- Supervisionado: ***previsão de uma variável-alvo***
 - Regressão → MLG (GLM)
 - Redes neurais
 - CART: Classification And Regression Trees (Árvores de Regressão e Classificação)
- Não-supervisionado: ***sem variável-alvo***
 - Agrupamento (clustering)
 - Componentes Principais (redução de dimensionalidade)

... análise avançada

Pesquisa operacional

- Simulação
- Optimização
- Recozimento simulado (Simulated Annealing)

Elaboração de previsões

- Transformadas de Fourier
- Transformadas wavelets

- Agrupamento k-means
- Algoritmos genéticos
- Teoria dos grafos

• Redes neurais

• Análise harmônica

- **Regressão linear, logística, MLG, MARS**
- Análise de séries temporais
- Redes Bayesianas

- Análise de Links
- Árvores de decisão
- Floresta aleatória
- Máquinas de vetores de suporte

- Método de Monte Carlo
- Componentes principais

Mineração de dados

• Visualização

- Confiabilidade/análise de sobrevivência
- Análise de Variância Univariada (ANOVA)
- Análise de Variância Multivariada (MANOVA)
- Correlação
- Análise fatorial

- Consulta de dados
- OLAP
- Tabulação cruzada

BI

• SQL

Estatística

Comentários finais a respeito de modelos...

- É comum dividir os dados aleatoriamente em 70% para a aplicação do modelo e 30% para validação;
- Escolha a técnica de desenvolvimento de modelos mais adequada para o problema;
- Selecione, preferencialmente, os dados e as variáveis consideradas mais preditivas;
- Após o desenvolvimento, a adequação do modelo deve ser testada com base na métrica de "critério informativo";
- Quanto mais variáveis e dimensões, maior a precisão do modelo e
- Os resultados do modelo devem traduzir-se em aplicações "mensuráveis" em termos de negócios.

Trabalho com Estudos de Caso



Subscrição: Substituição da seleção médica pela pontuação



Estudo de caso 1: Segmentação do risco

Estudo do caso de empresa asiática, mas típico de todo o mundo

Os objetivos do banco/seguradora eram:

- Crescimento das vendas pelo canal de "bancassurance"
- Venda de produtos voltados à proteção de clientes de bancos com emissão garantida ou emissão simplificada e mínimo impacto no preço
- Aprimoramento da experiência do cliente:
 - ✓ Agilização do processo de emissão de apólices
 - ✓ Simplificação do processo de subscrição para clientes com maior probabilidade de serem considerados padrão
- Redução das despesas de aquisição
- Máximo aproveitamento do valor dos dados internos

Estudo de caso 1: Segmentação do risco

A subscrição preditiva exige dois conjuntos de variáveis...

Preditoras

Exemplos de um banco/seguradora:

- Idade
- Filial
- Ativos sob gestão
- Segmento do cliente
- Nacionalidade

Outras possíveis fontes:

- Outras apólices de seguro
- Dados de marketing
- Agências de classificação de risco de crédito



Resposta

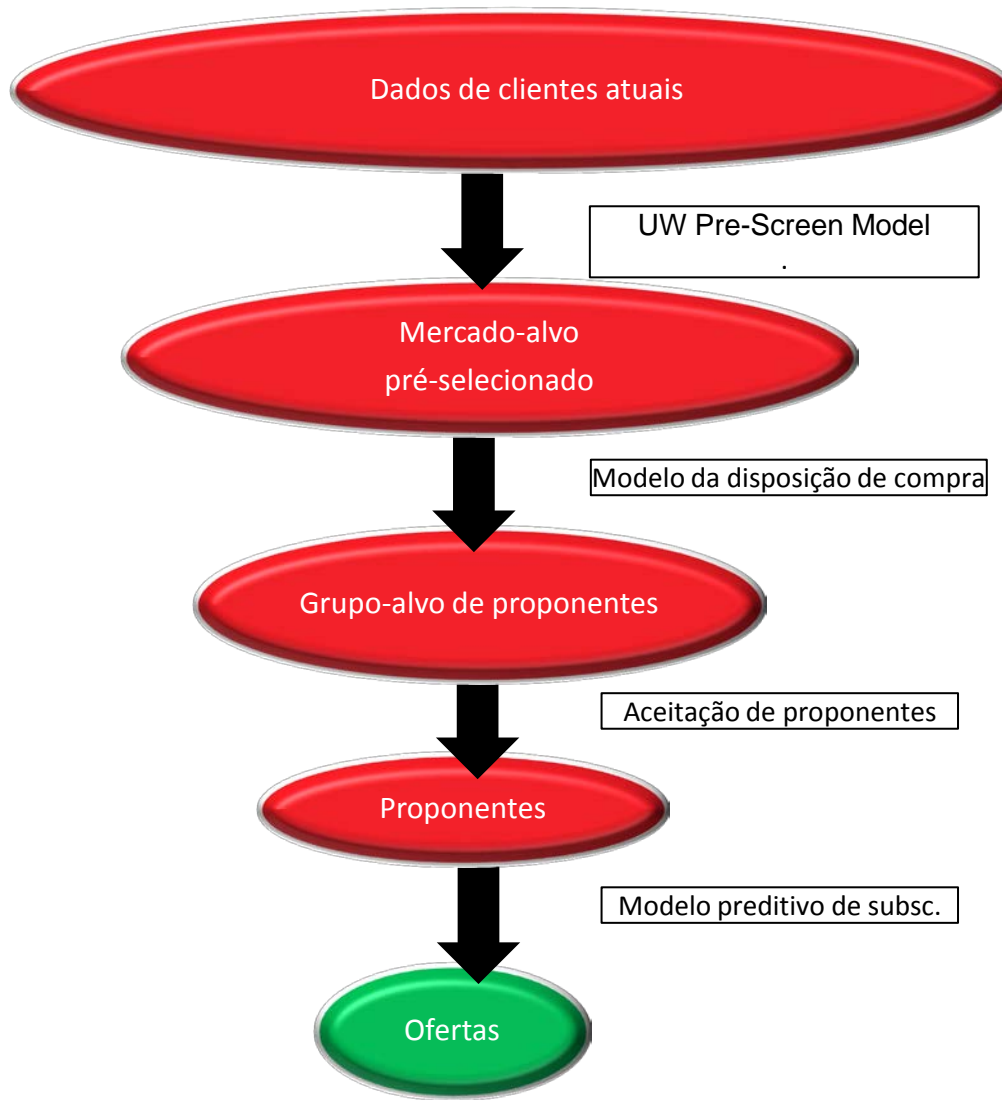
Decisão seguindo os procedimentos normais de subscrição:

- Risco padrão
- Classificado
- Recusado

Fora dos padrões

Estas fontes devem estar vinculadas às mesmas vidas

Criação de oportunidades com a Análise Preditiva

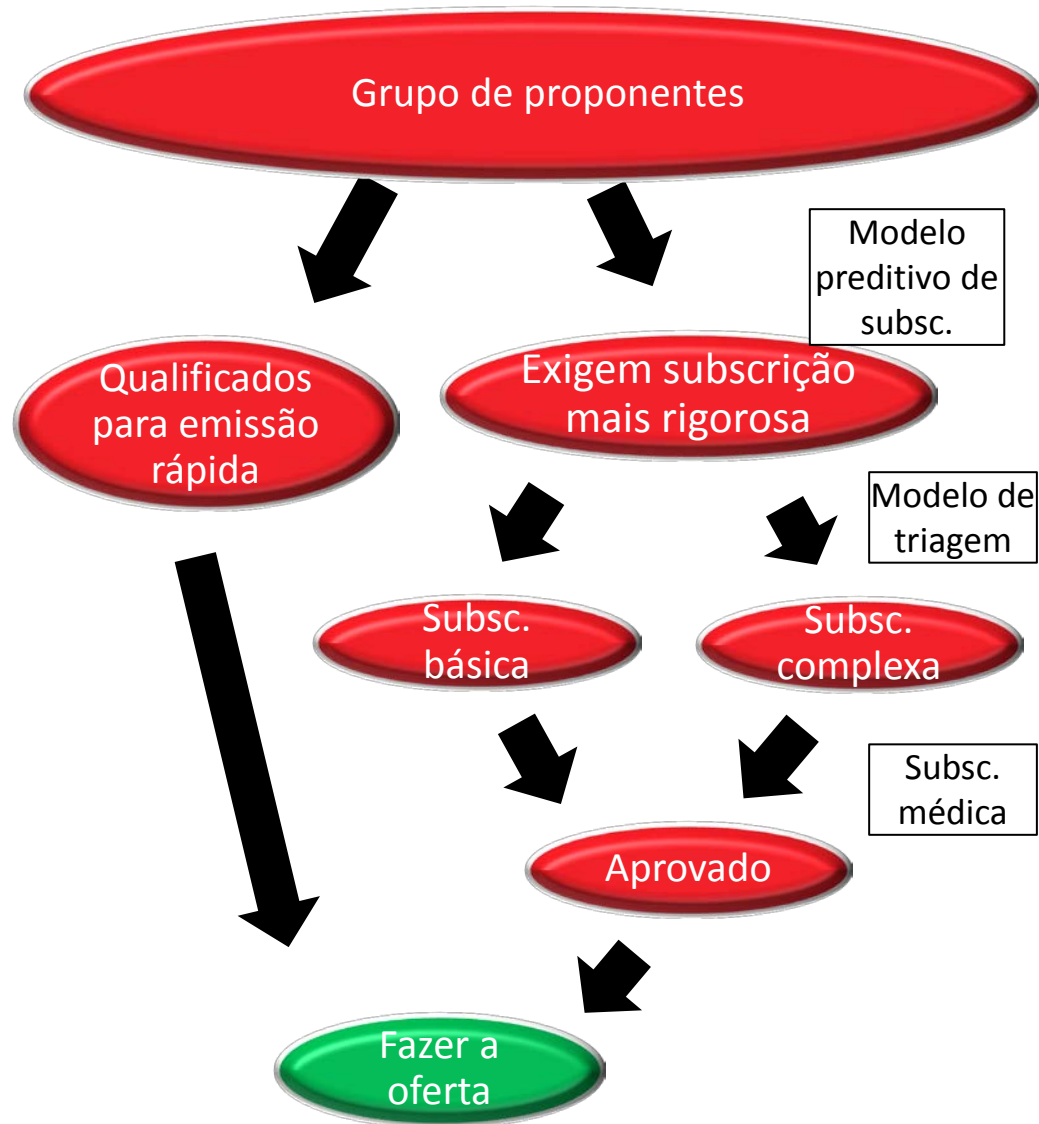


Dados existentes podem ser utilizados para identificar clientes que ofereçam oportunidades de venda cruzada

- Foco das ações de marketing, redução do tempo e custos de subscrição e combate à antisseleção por meio de uma eliminação rápida e barata dos riscos mais importantes no início do processo
- Identificação, com base no perfil de consumo, de clientes com maior probabilidade de comprar o produto novo
- Os programas mais eficazes terão taxas de aceitação muito mais altas que as abordagens de marketing convencionais.
- O projeto do modelo contemplará, de forma equilibrada, os objetivos de colocação, mortalidade e preço.
- A implantação do modelo pode ser totalmente baseada em dados ou combinada com métodos de subscrição convencionais.
- A triagem inicial deve reduzir o número de recusas e ofertas sem propostas.

Criação de oportunidades com a Modelagem de Previsão

O processo de subscrição pode ser aprimorado, de forma a otimizar o valor dos dados eletrônicos utilizados no procedimento.

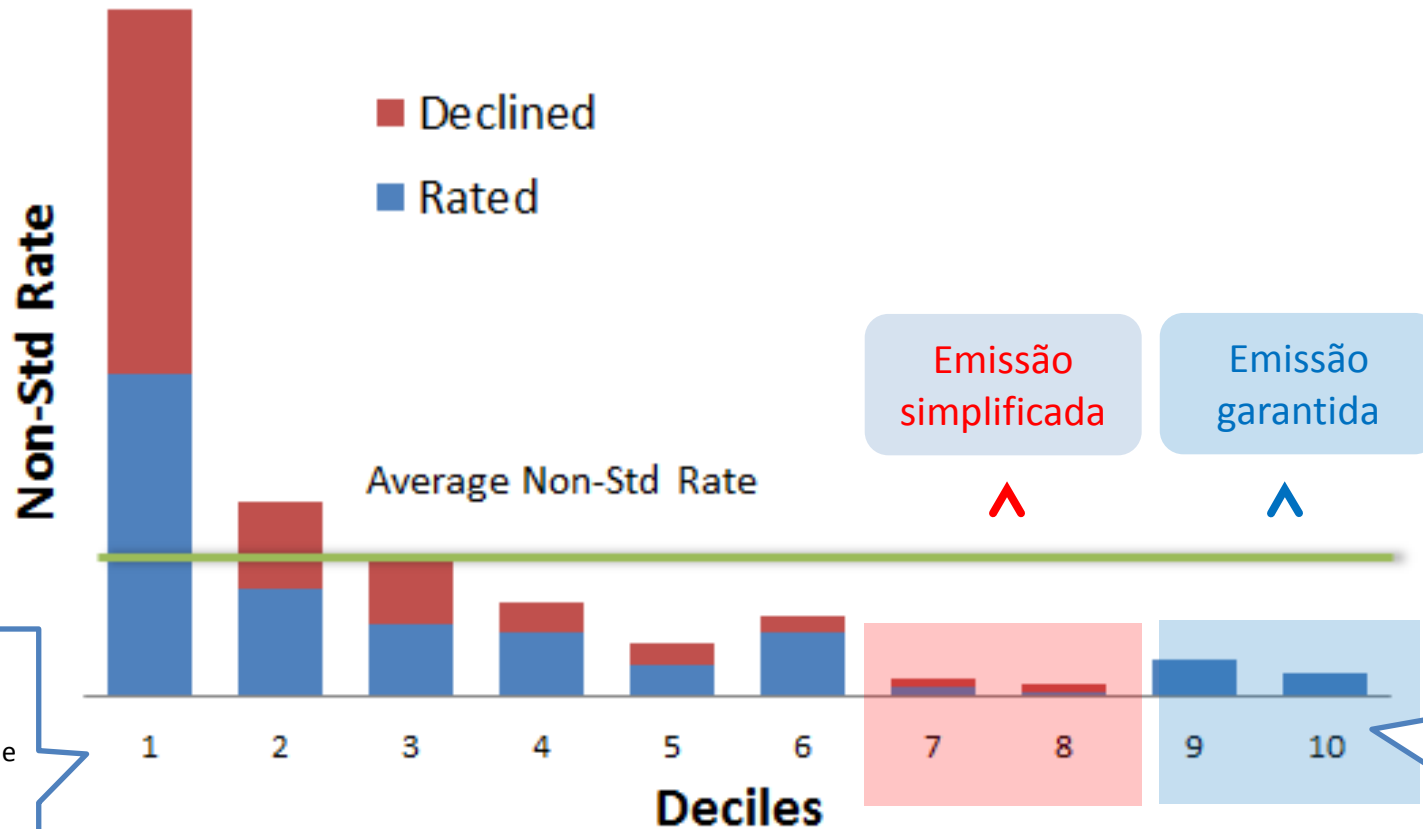


- Utilização da Análise Preditiva para permitir a emissão rápida, nos casos dos menores riscos, sem subscrição integral.
- Redução do tempo entre a proposta e a oferta para melhorar as taxas de colocação nos casos de risco mais baixo
- Subsc. médica integral para proponentes que não se qualificam à emissão rápida.
- Podem-se dividir os casos de acordo com o nível de complexidade.
- Possibilidade de aplicação de um modelo ou da subsc. convencional

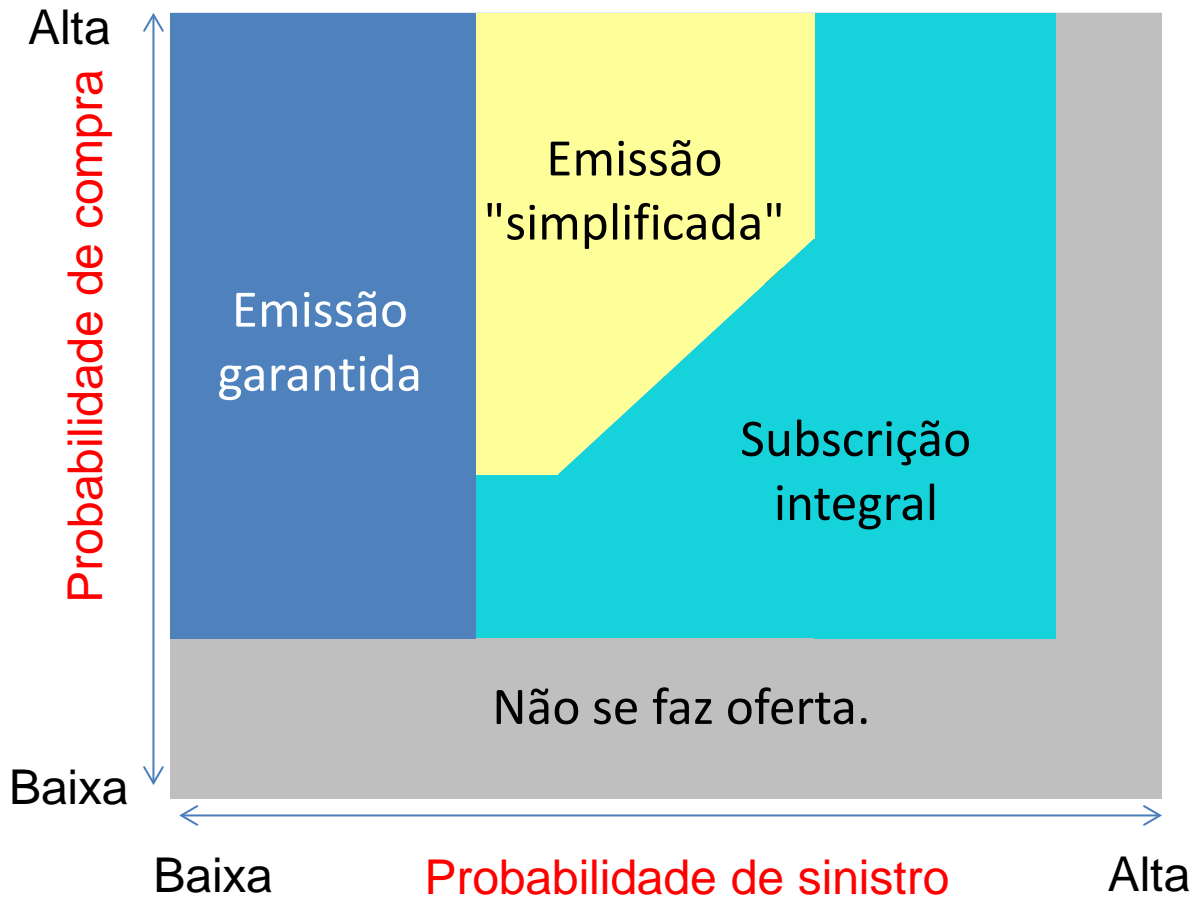
Estudo de caso 1: Modelo de subscrição pré-aprovada

Modelo de seleção de clientes a partir da subscrição preditiva

Lift Plot for Validation Results



Estudo de caso 1: Segmentação do risco



- Cada modelo retorna uma pontuação de "probabilidade".
- Salvaguardas contras os maiores riscos e antisseleção
- Emissão acelerada para os clientes com a menor probabilidade de sinistros
- Subsc. integral nos casos de riscos mais elevados
- Processo mais simplificado para os clientes na faixa intermediária
- Os limiares e a estrutura podem variar com a idade, importância segurada total, tipo de produto, canal de distribuição etc.

* A área não é proporcional à verdadeira quantidade de clientes.

Estudo de caso 2: Segmentação do risco

Risco de viagens internacionais:

- As viagens internacionais aumentaram muito nos últimos 50 anos, o que repercutiu na área de seguros de vida.
- Grande discrepância nas taxas de mortalidade e morbidade de países diferentes

Objetivos

- Avaliar o risco de viagens internacionais e residência no exterior
- Fornecer uma base de comparação uniforme para todos os países do mundo
- Conclusões baseadas em fatos e dados, em vez de opiniões e ideias pré-concebidas



Estudo de caso 2: Segmentação do risco

Coleta de dados de 205 países/regiões

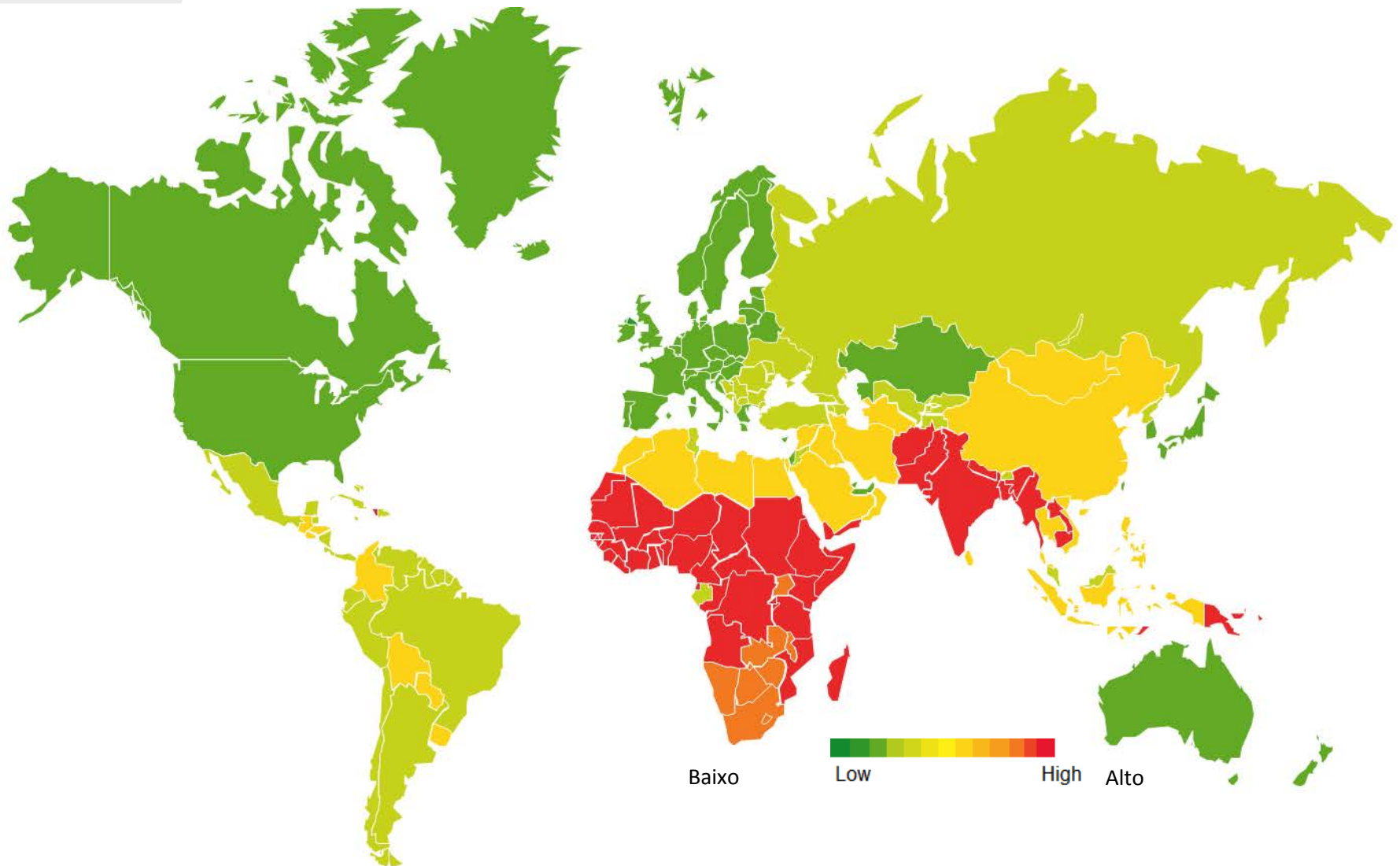
Os campos de dados foram ponderados de maneira adequada:

- expectativa de vida, mortalidade materna, mortalidade infantil, crianças abaixo do peso ideal, obesidade dos adultos, prevalência do HIV, mortalidade por doenças contagiosas, número de médicos, densidade médica, saneamento básico, água potável, leitos hospitalares, tráfego, homicídios, conflitos militares, mortalidade de estrangeiros, dióxido de carbono, acidentes de trabalho, concentração de partículas em suspensão, número de internautas, uso de telefones celulares, malha viária, PIB per capita, corrupção, taxa de escolaridade e Índice de Gini

Dados de várias fontes:

- CIA, OMS, Fórum Econômico Mundial, Banco Mundial, ONU, Center for Systemic Peace, Departamento de Estado dos EUA, pueblo.gsa.gov, Elsevier, Transparência Internacional

Estudo de caso 2: Segmentação do risco



Resumo

- Manutenção da qualidade de dados;
- É necessário que a cultura da empresa aceite processos decisórios baseados em dados;
- Pilote o projeto e divulgue os resultados para conseguir apoio (buy in);
- Com um processo bem conduzido, os dados da empresa podem se tornar um diferencial;
- Um modelo eficiente pode resultar em regras de atuação que gerem ganhos significativos;
- Nunca é tarde demais para começar a reunir os dados certos;
- Alguns cálculos serão necessários...
- ... Mas a RGA pode ajudar!

É hora de fazer a primeira previsão:

Qual desses dois fez o papel do estatístico em "O Homem que Mudou o Jogo"?





<http://www.analytics-for-insurance.com/>

Analytics for Insurance Europe 2014

2-day Conference & Exhibition | October 6-7, 2014 | Marriott Regents Park, London

organised by



[Overview](#) [Agenda](#) [Speakers](#) [Sponsor](#) [Partners](#) [Attendees](#) [Venue](#) [Download Brochure](#) [Register Now](#)

Embed Analytics into Operations to Reduce Costs and Improve Pricing





80 people already registered



65% end users



Pan-European focus



10º Congresso Brasileiro & Ibero-Americano de Atuária



Ronald Poon Affat FIA FSA MAAA CFA

VP e Diretor: RGA Re – América do Sul e o
Caribe

rpoonaffat@rgare.com

