



MODELAGEM DE CROSS-SELLING NO MARKETING DE RELACIONAMENTO POR MÁQUINAS DE VETOR DE SUPORTE

CROSS-SELLING MODELING IN THE MARKETING OF RELATIONSHIP BY
VECTORS SUPPORT MACHINES

ANDERSON GUIMARÃES DE PINHO

RESUMO

Analistas de marketing estão frequentemente em busca de novas maneiras de se vender mais produtos e serviços para os mesmos clientes. É de grande debate em processos KDD (Knowledge Discovery Database) a discussão sobre modelagem de cross-selling, ou venda cruzada, pela aplicação de tradicionais técnicas de inteligência computacional e estatísticas como Redes Neurais e Regressão Logística. O presente artigo abordará a técnica conhecida como Máquinas de Vetor de Suporte (MVS) para resolução de problemas desta natureza em contrapartida as demais já utilizadas. No presente artigo, utilizaremos uma base de dados real de uma empresa vendedora de motos a pessoas físicas como aplicação da técnica. MVS tratase de um tipo especial de redes neurais alimentadas adiante, e representa real vantagem às outras técnicas, pois permite ao analista o domínio na determinação das superfícies de separação entre grupos, em nosso caso, de interessados e não interessados na compra de outra moto. Esta predição é interessante do ponto de vista de marketing, pois permite a antecipação pela empresa de um desejo do cliente, viabilizando uma ação de marketing pró ativa para incremento de receita através da venda focada.

Palavras-chave: Venda cruzada. Máquinas de vetor de suporte. Customer relationship management. Redes neurais artificiais.

ABSTRACT

Analysts of marketing are often looking for new ways to sell more products and services for the same customers. In processes of KDD (Knowledge Discovery Database) when discussing on modeling of cross-selling, researchers often apply traditional techniques of intelligence computing and statistics as neural networks and logistic regression. This article discusses the technique known as Support Vector Machines (SVM) for resolution of such problems. In this article, we use a database of a real business problem, when decided to apply this technique. SVM is a special type of feed forward neural networks, and represents real advantage to other techniques, it allows the analyst to determine the area of separation between groups, in our case, interested and not interested in buying another motorcycle. It allows the anticipation of a desire from the client, enabling an action of pro-active marketing to increase revenue through a focus sale.

Keywords: Cross-sell. Support vector machines. Customer relationship management. Artificial neural networks.

INTRODUÇÃO

Manfred, apud Page et al. afirmaram que as empresas podem crescer de três maneiras: (1^a) atraindo novos consumidores; (2) perdendo menos clientes; (3) fazendo mais negócios com os existentes. É consenso geral no marketing de relacionamento que é mais rentável focar recursos nos clientes existentes para o aumento da participação da carteira (mais negócios), do que em penetração de mercado alvo (novos consumidores). Sendo assim, a perda de menos clientes passa a ser consequente de um círculo virtuoso entre antecipação de desejos e satisfação das necessidades. Desta forma, devemos focar recursos iniciais de marketing no aumento da rentabilização dos clientes, ou seja, fazendo-se mais negócios com eles.

Para o aumento da rentabilidade do cliente, ações de *venda cruzada* ou *venda superior* podem ser elaboradas. A venda superior – em inglês, “up-selling” – acontece quando há a venda de um produto/serviço superior em termos de receita líquida para a empresa, *substituto* ao atual do cliente. Já venda cruzada, ou “cross-selling”, acontece através da venda *adicional* de um produto/serviço ao cliente da empresa. E particularmente, será foco da estratégia principal de mineração deste estudo.

Neste âmbito, para auxílio e criação de uma oferta de cross-selling, as principais perguntas feitas a um analista de banco de dados de marketing envolvido na descoberta de conhecimento, ou mineração de dados, são: Que produto oferecer a quem? E quando oferecer?

A principal idéia em marketing de relacionamento e mineração de dados é que dados contém informação valiosa para tomada de decisão empresarial. Mineração de dados é como qualquer ferramenta: não é suficiente somente entendê-la como funciona, é necessário saber como iremos utilizá-la. Desta forma, para ser efetivo, a mineração de dados deve ocorrer num contexto que permita a organização mudar o comportamento do cliente como resultado daquilo que ela aprende sobre ele (Berry, p.7).

Existem diversas técnicas disponíveis no mercado sobre mineração de dados. No capítulo 2, faremos um resumo das principais técnicas utilizadas tipicamente em problemas de crossselling. Em 3 apresentaremos nossa metodologia de análise, não tão usual, porém com eficácia a ser comprovada, denominada Máquinas de Vetor de Suporte (MVS), ou Support Vector Machines. Desta forma, resultados da metodologia utilizada nos permitirão concluirmos sobre a quem ofertar o produto, e quando ofertá-lo.

MOTIVAÇÃO DO ESTUDO

No contexto de cross-selling, fala-se muito sobre propensão a compra de um produto. Desta forma, clientes são classificados de acordo com suas chances de comprarem determinado produto. Dentre as técnicas mais utilizadas para modelos de propensão, podemos destacar as seguintes: Regressão Logística, Redes Neurais, e Árvores de Decisão.

Podemos citar o trabalho de Weng em [9] na montagem de uma coleção de árvores de decisão pela metodologia denominada “Randow Forest”, solucionando um problema de venda cruzada financeiro. Ou ainda Park em [10], na comparação das três técnicas mencionadas no parágrafo anterior, para solução de um problema de venda cruzada no mercado de saúde.

Neste artigo, trataremos da técnica de Máquina de Vetor de Suporte. Acreditamos ser uma alternativa vantajosa as tradicionais técnicas utilizadas pelos seguintes motivos: (1) A precisão na superfície de decisão que separa as classes é um parâmetro especificado pelo

usuário; (2) Diferentes máquinas de aprendizagem fornecem diferentes representações de regularidades estatísticas intrínsecas aos dados de treinamento; (3) É garantido que se encontre um extremo global da superfície de erro, onde o erro se refere à diferença entre a resposta desejada e a saída da máquina de vetor de suporte, entre outros.

Como veremos em 4, o padrão de modelagem neste artigo trata-se de um interesse a compra, e não propriamente a compra diretamente. Mesmo assim, veremos que os resultados obtidos viabilizarão uma ação do tipo venda cruzada, sujeita a possivelmente a conversão em vendas. Como qualquer algoritmo ou modelo de mineração de dados, após a implementação e mensuração dos primeiros resultados, parâmetros de uma máquina de vetor de suporte podem ser ajustados para uma melhoria no desempenho obtido. Seguiremos com um breve detalhamento da metodologia em foco.

METODOLOGIA DE ANÁLISE

MÁQUINAS DE VETOR DE SUPORTE

É bastante freqüente a discussão a cerca de redes neurais artificiais sobre o problema conseqüente à excessiva duração da fase de aprendizado denominado “overfitting”. Isto é uma conseqüência dos algoritmos de aprendizados por minimização do erro, na tentativa de se selecionar o melhor modelo. Proposta por Vapnik (1995), as Máquinas de Vetor de Suporte (MVS) estão ganhando popularidade tendo em vista múltiplas características e performances em relação às redes neurais tradicionais (Gun, 1998). O objetivo seria produzir um classificador que trabalhasse bem em exemplos não vistos, ou seja, apresenta-se boa generalização.

MVS são redes do tipo alimentada a diante (feed-forward), ou seja, o sinal da cada entrada se propaga para frente, e não há retro-alimentação. Assim como perceptrons de múltiplas camadas ou redes de função de base radial, MVS podem ser usadas para classificação de padrões e regressão.

Com fins de exemplificação, vamos supor um conjunto de padrões separáveis em duas classes para um problema típico de classificação. O objetivo principal de uma MVS seria encontrar um hiperplano através de uma superfície de decisão, de tal forma que a margem de separação entre padrões positivos e negativos fosse máxima.

Uma MVS apresentaria esta propriedade desejável fundamentada na teoria de aprendizagem estatística. O objetivo em modelagem estatística é selecionar um modelo de um espaço hipotético, o qual é o mais próximo possível (com alguma medida de erro) de uma função traçada no espaço de características alvo. Erros podem acontecer de duas formas:

(1) Aproximação do erro, como conseqüência do espaço hipotético ser menor que o espaço alvo, e neste caso a função traçada estaria fora do espaço hipotético. Seria uma ruim escolha do modelo de espaço que resultaria num alto erro de aproximação, e é referida como erro na aproximação do modelo.

(2) Erro de estimação, relacionado a procedimentos de aprendizado resultantes em um modelo não ótimo de um espaço hipotético.

A formulação de máquinas de vetor de suporte envolve a minimização do risco estrutural (SEM em inglês). Este princípio indutivo é baseado no fato de que a taxa de erro de uma MVS sobre o grupo de teste (generalização) é limitada pela soma da taxa de erro do treinamento e

por um termo que depende da dimensão de Vapnik-Chervnenkis, ou V-C (Haykin, p.349). Supondo nosso exemplo de padrões separáveis, uma MVS apresentaria zero para o erro de generalização, e minimizaria o termo V-C.

Um conceito central ao aprendizado de MVS é o núcleo do produto interno (ou função Kernel) entre um vetor de suporte X_i e outro X retirado do espaço de entrada. Vetores de suporte consistem de um subconjunto de pontos retirados do conjunto de treinamento. Dependendo de como este núcleo interno é gerado, podemos criar diferentes máquinas de aprendizagem por vetor de suporte, caracterizadas por superfícies de decisão não lineares próprias. Em geral, podemos usar o algoritmo de aprendizagem por vetor de suporte para criarmos três tipos de máquinas de aprendizagem: (1) máquinas polinomiais; (2) de função de base radial; (3) perceptrons de duas camadas. Neste artigo veremos somente a aplicação das duas primeiras.

Não detalharemos por completo toda a teoria por detrás de uma MVS, por ser bastante exaustiva a literatura tendo em vistas as propostas deste artigo. Para estudos adicionais recomendamos [5], [6]. A seguir, uma breve discussão sobre o processo de otimização de uma MVS.

PROJETO ÓTIMO DE UMA MVS

A construção de uma superfície de decisão *não linear* do espaço de entrada, cuja imagem no espaço de características é linear, pode ser encontrada pela maximização da seguinte função objetivo:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (1)$$

Dado um conjunto de amostra de treinamento $\{(x_i, d_i)\}_{i=1}^N$, onde x_i é o padrão de entrada para o i -ésimo exemplo e d_i é a resposta desejada correspondente, a resolução do problema em (1) consiste em encontrar os multiplicadores de Lagrange $\{\alpha_i\}_{i=1}^N$ a que maximizam a função de objetivo, sujeito as seguintes restrições:

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (2)$$

$$0 \leq \alpha_i \leq C \text{ para } i = 1, 2, \dots, N \quad (3)$$

Onde C é um parâmetro positivo especificado pelo usuário. Havendo determinado os multiplicadores de Lagrange ótimos, representados por $\alpha_{a,i}$, podemos calcular o vetor de peso ótimo W_0 por:

$$w_0 = \sum_{i=1}^N \alpha_{o,i} d_i \varphi(x_i) \quad (4)$$

w_0 conecta o espaço de características ao espaço de saída, reconhecendo que $(\cdot)_{ij}$ (imagem induzida no espaço de características devido a x_i) desempenha o papel de entrada para o vetor de peso w . A restrição (2) surge da otimização do lagrangiano $Q(a)$ em relação ao bias $b=w_0$ para $(\cdot)_{ij} = 1$ (note que a primeira componente de w_0 representa o bias ótimo b_0). Desta forma, para as classes de padrões com notações 1 e -1 podem ser classificadas da seguinte forma:

$$\begin{aligned} \text{se } w^T_0 x_i + b_0 \geq +1 \text{ então } d'_i &= +1 \\ \text{se } w^T_0 x_i + b_0 \leq -1 \text{ então } d'_i &= -1 \end{aligned} \quad (5)$$

A função em (1) é dada pela expansão do núcleo do produto interno $K(X, X_i)$. Abaixo apresentaremos três núcleos de produto interno para construção de superfícies de decisão (linear ou não) no espaço de entrada, dentro os quais particularmente utilizaremos neste trabalho o polinomial, e o de base radial. Desta forma, $(\cdot)_{ij} K(X, X_i)$ será igual a:

$$\text{Polinomial} \rightarrow (X^T X_i + 1)^p \quad (6)$$

$$\text{Radial} \rightarrow \exp\left(-\frac{1}{2\sigma^2} \|X - X_i\|^2\right) \quad (7)$$

$$\text{Perceptron} \rightarrow \tanh(\beta_0 X^T X_i + \beta_1) \quad (8)$$

Da mesma forma que C, p, σ, β_0 e β_1 são parâmetros definidos pelo usuário. Seguiremos o artigo com o detalhamento de um caso para aplicação da metodologia.

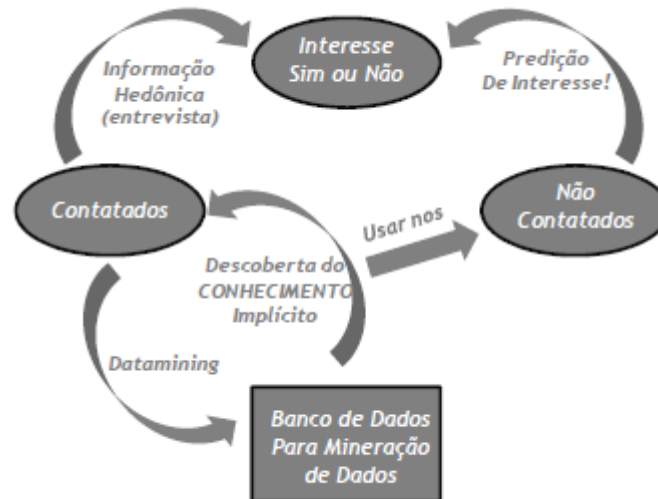
CASO CONCESSIONÁRIA DE MOTOS

Analistas de marketing estão sempre preocupados em antecipar e despertar os desejos de seus consumidores. Com o objetivo de obter a opinião dos seus clientes sobre a satisfação e novas expectativas sobre produtos e serviços vendidos – informação geralmente não encontrada em bancos de dados empresariais ao nível de cliente – uma pesquisa de mercado do tipo quantitativa foi conduzida com 406 de seus clientes.

Dentre muitos aspectos abordados pela pesquisa, uma pergunta em particular nos chamou a atenção, e assim motivou o desenvolvimento deste estudo. A pergunta “O Sr.(a) tem algum interesse em comprar outra moto conosco nos próximos 12 meses?”. Como mencionado no primeiro parágrafo, esta pergunta foi feita a apenas 406 clientes. A empresa apresenta outros 9.778 clientes em sua base de dados. Seria interessante, para os analistas, saberem quais destes também apresentam interesse em comprar outra moto. Isto seria motivo para criarem

uma ação especial de marketing, com o objetivo de transformar este *interesse* na compra, em um *desejo* de compra. Isto anteciparia lucros futuros, evitaria perdas para a concorrência, e, sobretudo, tornaria cada um de seus clientes mais rentáveis.

Figura 1: Estratégia de Descoberta



Fonte: Própria Autoria

No entanto, a pergunta individual para mais de 9.000 clientes sobre o interesse de compra de outra moto é totalmente desmotivada. Se o conhecimento implícito no comportamento de o comportamento de interesse em outra moto (informação hedônica), o mesmo conhecimento poderia ser aplicado para todos os outros 9.000, predizendo seus interesses, sem que nenhuma pergunta fosse feita pessoalmente.

Este mapeamento do conhecimento somente poderia ser feito através de técnicas de mineração de dados, dentre as quais as apresentadas aqui são grandes promissoras. Anteriormente, podemos ver um esquema dos objetivos detalhados neste capítulo.

SELEÇÃO DE VARIÁVEIS

Os gestores da área de marketing da concessionária dispõem de um banco de dados com informações gerenciais sobre cada um de seus clientes. Muitas informações são contidas de natureza cadastral, demográfica, geográfica, ou ainda, transacional.

Informações do tipo transacional, através da análise estatística, oferecem a prática do *business intelligence* (BI). Dados brutos subsidiam a tomada de decisão nos escalões médios e altos da empresa (Tarsila, p.8). Já o *datamining* diz respeito à procura de novos padrões de vinculação entre as variáveis registradas, subsidiando conhecimento novo e útil, e neste caso, de natureza transacional.

Nosso trabalho começa com a mineração de dados de múltiplas fontes, a maioria transacional, para mapear uma informação hedônica, o interesse na compra de uma moto. Desde já, muitas variáveis são disponibilizadas *ao nível de cliente* (cada um representando uma linha), representadas na tabela 1.

Tabela 1: Variáveis para Análise de Seleção

Variáveis	Descrições
txtNomeLojaMaisFrequencia	Nome da loja mais frequente.
flgClienteClube	Participa de algum clube.
flgEmailValido	Possui e-mail válido.
flgFormatoValidoEndereco	Possui endereço válido.
flgMotosUlt12Meses	Comprou moto nos últimos 12 meses.
flgPecasAcessBoutUlt12Meses	Comprou peças/acess./bout. últimos 12 meses.
flgServicosUlt12Meses	Utilizou algum serviço nos últimos 12 meses.
flgTemMotoB	Possui moto do tipo B.
flgTemMotoC	Possui moto do tipo C.
flgTemMotoA	Possui moto do tipo A.
flgTemMotoD	Possui moto do tipo D.
flgTemMotoE	Possui moto do tipo E.
flgTemMotoF	Possui moto do tipo F.
flgTemMotoG	Possui moto do tipo G.
numQtdePecasAcessBout	Qtde total de peças/acess./bout. comprados.
numQtdePecasAcessBoutUlt12Meses	Qtde de peças/acess./bout. últimos 12 meses.
numQtdeServicos	Quantidade total de serviços.
numQtdeServicosUlt12Meses	Qtde total de serviços nos últimos 12 meses.
numTempoInatividadeDias	Tempo desde a última compra em dias.
numTempoMedioEntreComprasDias	Tempo médio entre compras.
numTempoRelacDias	Tempo desde a primeira compra em dias.
qtdTotalMotos	Quantidade total de motos compradas.
txtNomeLojaMaisAntiga	Loja com compra mais antiga.
txtNomeLojaMaisGastou	Loja onde mais gastou.
txtNomeLojaMaisRecente	Loja com compra mais recente.
txtSegmentoCliente	Segmento cliente segundo critérios da empresa.
txtSexo	Sexo do cliente.
vlrPagoServicos	Valor pago total em serviços.
vlrPecasAcessBout	Valor pago total em peças/acess./bout..
vlrTotalMotos	Valor pago total em motos.
qtdMotosUlt12Meses	Qtde total de motos nos últimos 12 meses.

Fonte: Própria Autoria

Contínuas

Para seleção das variáveis contínuas, utilizou-se a estatística de teste t-student. Este teste é muito utilizado quando estamos interessados em comparar médias de duas populações 10 distintas, a partir de amostras independentes. Em nosso caso, estamos interessados em saber se as médias de uma variável contínua em análise diferem significativamente entre a população de interessados e não interessados.

Para aplicação do teste em questão, é necessária que seja verificada a normalidade de distribuição da variável de interesse, entre grupos. Caso esta normalidade não seja verificada, outro teste seria mais apropriado como o de Wilcoxon, mais conhecido como Mann-Whitney (Bussab, pág. 366). Para o nosso problema, assumiu-se que as variáveis seguissem distribuição normal. Acredita-se que mesmo a análise técnica desta premissa poderia nos levar a uma conclusão ruidosa, pois dispomos de poucos casos para comparação – 287 com interesse e 119 sem interesse –, e muitas vezes apresentando valores nulos.

Sendo assim, as hipóteses a serem testadas ficam:

$$\begin{aligned}
 H_0 &: \mu_{Sim} = \mu_{N\tilde{a}o} \\
 H_1 &: \mu_{Sim} \neq \mu_{N\tilde{a}o}
 \end{aligned}
 \tag{6}$$

E a estatística de teste é dada pela seguinte fórmula:

$$t_{observado} = \frac{|\mu_{Sim} - \mu_{N\tilde{a}o}|}{\sqrt{\frac{\sigma_{Sim}^2}{n_{Sim}} + \frac{\sigma_{N\tilde{a}o}^2}{n_{N\tilde{a}o}}}}
 \tag{7}$$

Compara-se o $t_{observado}$ com o $t_{tabelado}$ da distribuição t-student com (n_1+n_2-2) graus de liberdade, e nível de significância de α . Em caso de $t_{observado} > t_{tabelado}$, rejeita-se H_0 ao nível de confiança de $1-\alpha$. E sendo assim, assumi-se que as médias entre grupos para variável de interesse diferem significativamente uma da outra, devendo ser incluída na modelagem.

Tabela 2: Variáveis Contínuas Seleccionadas

Variáveis	t-Student Observado	p-Valor	Resultado do Teste
numQtdePecasAcessBoutUlt12Meses	0,692	0,489	F
numQtdeServicosUlt12Meses	0,452	0,651	F
numQtdeServicos	2,730	0,007	S
numTempoMedioEntreComprasDias	-1,415	0,158	F
qtdTotalMotos	3,630	0,000	S
vlrPagoServicos	2,150	0,032	S
vlrTotalMotos	3,062	0,002	S
numQtdePecasAcessBout	1,267	0,206	F
qtdMotosUlt12Meses	-0,598	0,550	F

Fonte: Própria Autoria

Resultados anteriormente refletem as variáveis mais significativas ao nível de significância de 5%, ou seja, p-valor é menor que α . Optamos por critérios empíricos sobre a inclusão de variáveis na coluna “Resultado” preenchida com “F”, onde o do p-valor é maior que α ou seja, resultado não significativo. Acreditamos que embora contrariando o resultado do teste t, estas variáveis são importantes para diferenciar o cliente interessado ou não na compra de uma moto.

Catagóricas

Para seleção de variáveis categóricas, utilizou-se a estatística de teste X_2 (Qui-Quadrado). A estatística X_2 é definida como uma medida de associação (ou dependência) entre duas variáveis qualitativas. Em nosso problema, se pelo teste X_2 assume-se que uma variável categórica é dependente da variável que determinada o classe de interesse, diz-se que esta primeira variável é útil para modelar o valor da segunda, e neste caso, inclui-se ela na análise.

A aplicação do teste é feita através de tabelas de contingências.

Suponha que temos uma tabela de contingência de dimensão r por s , dado pelo cruzamento de duas variáveis qualitativas X e Y , classificadas em r categorias A_1, \dots, A_r para X e s categorias B_1, \dots, B_s para Y . Defini-se a estatística de teste X^2 de Pearson como sendo:

$$X^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} \quad (8)$$

Onde n_{ij} é o número de elemento observados na célula ij da tabela de contingência, e n_{ij}^* o valor esperado da célula ij dado pela fórmula abaixo:

$$n_{ij}^* = \frac{\text{total linha } i \text{ X total coluna } j}{\text{total observações}} \quad (9)$$

Dado as seguintes hipóteses a serem testadas:

$$\begin{aligned} H_0 &: \text{Variável } Y \text{ depende de } X \\ H_1 &: \text{Variável } Y \text{ não depende de } X \end{aligned} \quad (10)$$

Se aceita H_0 ao nível de significância α , caso o valor observado em (3) seja maior que o X^2_{tabelado} com $(r - 1) \times (s - 1)$ graus de liberdade e $(1 - \alpha)$ de confiança. Mais detalhes sobre o teste X^2 , recomendamos Bussab (p.87,187-188). Da mesma forma, para o teste t-Student recomenda-se Bussab (p.361-381).

Tabela 3: Variáveis Categóricas Seleccionadas

Variáveis	Qui-Square Observado	p-Valor	Resultado do Teste
*txtNomeLojaMaisFrequencia	8,857	0,012	S
flgTemMotoA	4,230	0,040	S
txtSegmentoCliente	10,943	0,001	S
flgEmailValido	2,372	0,124	F

Fonte: Própria Autoria

Ao mesmo nível de significância (5%), as variáveis anteriormente foram seleccionadas. Por critérios empíricos também, optamos pela inclusão da variável flgEmailVálido - coluna "Resultado" preenchida com "F" -, onde o do p-valor é maior que α .

TRANSFORMAÇÃO DE VARIÁVEIS

Transformações dos dados são requeridas para uma melhor adequação dos métodos aqui discutidos e melhoria no desempenho. Em [3], por exemplo, vemos que se a transformação for bem empregada em variáveis contínuas, problemas como não normalidade, não linearidade, ou não homocedasticidade podem ser corrigidos.

Utilizamos três tipos de transformação nos dados. A primeira dela trata-se na *representação*

por bits. Receberam esta transformação, variáveis tipicamente categóricas (variáveis que eram contínuas foram antes transformadas em classes para depois serem assim representadas). Em bits, uma variável com n classes/categorias era representada no mínimo a dígitos binários, onde $n \leq 2^a$.

A segunda trata-se da *normalização por faixa de medição*, dados são transformados na faixa de {0,1} segundo (4). Variáveis associadas a tempo receberam esta transformação.

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (11)$$

Pela normalização padrão, dados eram normalizados segundo a função em (5). Variáveis associadas à quantidade e valor receberam esta normalização.

$$x'_i = \frac{x_i - \mu_i}{\sigma_i} \quad (12)$$

A tabela abaixo oferece um resumo sobre as variáveis finais selecionadas, e suas transformadas.

Tabela 4: Variáveis Finais de Modelagem

Variáveis	Natureza Original	Transformação Utilizada	Variáveis Codificadas
numQtdePecasAcessBoutUlt12Meses	Contínua	Bits	X1, X2
numQtdeServicosUlt12Meses	Contínua	Bits	X3, X4
txtNomeLojaMaisFrequencia	Categórica	Bits	X5, X6
flgTemMotoHD	Categórica	Bits	X7
numQtdeServicos	Contínua	Padrão	X8
numTempoMedioEntreComprasDias	Contínua	Faixa Medição	X9
qtdTotalMotos	Contínua	Padrão	X10
txtSegmentoCliente	Categórica	Bits	X11
vlrPagoServicos	Contínua	Padrão	X12
vlrTotalMotos	Contínua	Padrão	X13
flgEmailValido	Categórica	Bits	X14
numQtdePecasAcessBout	Contínua	Padrão	X15
qtdMotosUlt12Meses	Contínua	Padrão	X16
Tem Interesse na Compra	Categórica	--	Y

Fonte: Própria Autoria

Outliers presentes foram considerados na análise, uma vez que não representam valores incorretos, e sim uma discrepância existente entre alguns clientes. Por outro lado, caso optássemos pela exclusão de alguma informação da amostra, poderíamos estar perdendo informação valiosa, uma vez que dispomos de poucos casos para modelagem.

DETALHAMENTO DO EXPERIMENTO

Para as funções kernel em (6) e (7), parâmetros foram variados conforme quadro abaixo. Lembre-se que C é um parâmetro definido pelo usuário, para cada uma das classes de interesse, e ρ e σ – parâmetros específicos do kernel polinomial, e radial, respectivamente –

definidos por d.

Tabela 5: Parâmetros

kernel	d	net.c=[c1 c2]
linear	d=1:1:1	c1=10:10:100 ; c2=10:10:100
RBF	d=0.001 ; 0.01 ; 0.1:0.1:0.5 ; 1:1:10 ; 100 ; 1000	c1=10:10:100 ; c2=10:10:100
poly	d=1:1:10	c1=10:10:100 ; c2=10:10:100

Fonte: Própria Autoria

Notações do tipo A:B:C, significam que o parâmetro em foco, variava de A até C, com acréscimo intervalares de B. Para o kernel polinomial, por exemplo, temos um total de 1000 15 (10x10x10) possíveis MVSs a serem analisadas. No linear teremos um total de 100 (10x10) possíveis MVS, e no RBF 1900 (19x10x10).

Decidiu-se também variar diferentes tipos de banco de dados. Em um momento, testaremos todas as variações na tabela 5 para o banco de dados conforma tabela 4 na codificação apresentada, e outra com a transformada de bits em ordinal. Num segundo momento, testaremos a codificação apresentada, e a de bits em ordinal, para somente as variáveis dados como significativas pelos testes apresentados, estando assim livre de interpretação dos analistas.

Ao final, teremos um total de 12 experimentos (3 funções kernel x 4 bancos de dados). Para cada experimento, utilizaremos todos os 406 clientes respondentes da pesquisa de mercado, particionados entre grupos de treinamento (80%) e teste (20%). Em cada experimento, será necessária a escolha de uma MVSs que assim fosse “campeã” entre todas verificadas. A seguir, apresentaremos qual foi o critério de escolha entre MVSs por experimento.

CRITÉRIO DE ESCOLHA ENTRE DIFERENTES MÁQUINAS DE VETOR DE SUPORTE

A forma mais usual de se avaliar a qualidade de um modelo em problemas de classificação é feita em termos de acurácia e abrangência. Acurácia como o % de acerto dado pelo modelo na classe objetivada pelo problema (em nosso caso clientes com interesse em outra moto), e abrangência como o % de cobertura da classe objetivada na base utilizada.

Tabela 6: Espaços de Ocorrências na Modelagem do Interesse em Outra Moto

Classes		Classificada		
		Sim	Não	Total
Observada	Sim	A	B	(A+B)
	Não	C	D	(C+D)
	Total	(A+C)	(B+D)	(A+B+C+D)

Fonte: Própria Autoria

A modelagem de um problema que classifique corretamente clientes com interesse na compra de outra moto ou não, apresenta 4 possíveis ocorrências em virtude de seu histórico passado

observado, conforme tabela 6.

Onde A, B, C, e D são números inteiros, dados pelas clientes pertencentes a estas categorias. Defini-se então a acurácia de uma máquina de vetor de suporte como i:

$$Ac_i = \frac{A_i}{(A_i + C_i)} \quad (13)$$

E abrangência como:

$$Ab_i = \frac{A_i}{(A_i + B_i)} \quad (14)$$

Suponham que a hipótese nula, H_0 , de um modelo estatístico seja: assumir que todos os clientes interessados na compra de uma nova moto. Podemos definir dois tipos de erros encontrados em testes de hipóteses estatísticos (Bussab, p.323). O primeiro, erro do tipo 1, a probabilidade de eu rejeitar H_0 dado que ela é verdadeira, ou seja, a probabilidade de eu assumir que o cliente não é interessado, dado que ele será um interessado. O segundo, erro do tipo 2, seria a probabilidade de eu aceitar H_0 , dado que H_0 é falsa. Podemos definir então ambos os erros, para uma MVS, da seguinte forma:

$$ET1_i = \frac{B_i}{(A_i + B_i)} \quad (15)$$

$$ET2_i = \frac{C_i}{(A_i + C_i)} \quad (16)$$

Sendo assim podemos dizer que uma boa solução para o problema seria aquela que maximizasse a Ac_i e Ab_i , e minimizassem os $ET1_i$ e $ET2_i$. Reparem que $Ac_i = (1 - ET2_i)$, e $Ab_i = (1 - ET1_i)$. Reparem também que estamos lidando com múltiplos objetivos na avaliação de uma MVS. Para problemas desta natureza, podemos utilizar uma combinação de múltiplos objetivos numa única função f , de tal forma a buscarmos modelos que a maximizem ou minimizem, dependendo de suas características. Para o nosso problema, utilizou-se como objetivo a maximização da função abaixo.

$$f(Ac_i, Ab_i, ET1_i, ET2_i) = \frac{(A_i + D_i)}{(C_i + B_i)} \quad (17)$$

Percebam que a maximização da função acima atende aos quatro objetivos aqui detalhados: maximizar acurácia e abrangência, e minimizar erros do tipo 1 e 2.

RESULTADOS OBTIDOS

Resultados para as variações dos parâmetros e banco de dados apresentados no capítulo 5 podem ser verificadas anteriormente. Lembrando que apresentamos somente os resultados por experimento cujo f verificado é máximo. Observa-se que utilizando a codificação de bits para ordinal em todas as variáveis, função kernel do tipo RBF, $C1=60$, $C2=20$, e $d=0.3$, gerou o

melhor resultado entre experimentos. Neste caso, resultado da função f igual a 2,69. Como a amostra de treinamento encontrava-se ponderada entre interessados (50%) e não interessados (50%), pode-se dizer que os ganhos foram significantes em relação a um modelo aleatório. Mesmo muitas de outras melhores configurações deixaram a desejar, pois ou mostraram ser eficientes na classe de interessados, ou na classe de não interessados.

Tabela 7: Melhores Máquinas, Resultados para o Grupo de Treinamento

Grupo de Variáveis	Resultados						
	Parametrização				% Acerto Treinamento		
	Kernel	f (max)	net.c (max)	d (max)	Sim	Não	Total
Todas Variáveis	Linear	1,29	[20 20]	qq	65,26	69,47	67,37
	RBF	1,82	[90 90]	1	88,42	90,53	89,47
	Poly	1,4	[70 70]	2	93,68	93,68	93,68
Somente Significativas	Linear	1,4	[40 20]	qq	94,74	12,63	53,68
	RBF	1,4	[30 30]	0.1	92,63	87,37	90,00
	Poly	1,82	[10 50]	4	20,00	88,42	54,21
Todas (bits para ordinal)	Linear	1,53	[100 80]	qq	81,05	53,68	67,37
	RBF	2,69	[60 20]	0.3	100,00	84,21	92,11
	Poly	1,67	[40 100]	2	96,84	46,32	71,58
Significativas (bits para ordinal)	Linear	1,18	[10 20]	qq	31,58	88,42	60,00
	RBF	1,53	[80 50]	0.2	94,74	82,11	88,42
	Poly	1,67	[70 80]	3	53,68	47,37	50,53

Fonte: Própria Autoria

Tabela 8: Melhores Máquinas, Resultados para o Grupo de Teste

Grupo de Variáveis	Resultados						
	Parametrização				% Acerto Teste		
	Kernel	f (max)	net.c (max)	d (max)	Sim	Não	Total
Todas Variáveis	Linear	1,29	[20 20]	qq	50,00	62,50	56,25
	RBF	1,82	[90 90]	1	70,83	58,33	64,58
	Poly	1,4	[70 70]	2	70,83	45,83	58,33
Somente Significativas	Linear	1,4	[40 20]	qq	100,00	16,67	58,33
	RBF	1,4	[30 30]	0.1	66,67	50,00	58,33
	Poly	1,82	[10 50]	4	37,50	91,67	64,58
Todas (bits para ordinal)	Linear	1,53	[100 80]	qq	66,67	54,17	60,42
	RBF	2,69	[60 20]	0.3	79,17	66,67	72,92
	Poly	1,67	[40 100]	2	87,50	37,50	62,50
Significativas (bits para ordinal)	Linear	1,18	[10 20]	qq	25,00	83,33	54,17
	RBF	1,53	[80 50]	0.2	70,83	50,00	60,42
	Poly	1,67	[70 80]	3	62,50	62,50	62,50

Fonte: Própria Autoria

CONCLUSÕES E PRÓXIMOS PASSOS

Há de se concluir que o problema em questão é de difícil modelagem tendo em vista os resultados obtidos, e os resultados esperados pela metodologia existente por detrás das

máquinas de vetor de suporte. A rede escolhida com base na função f , entre experimentos, foi a única que resultou num percentual de acerto no grupo de teste maior que 70%.

É necessária a comparação de outros modelos, especialmente neurais, para validação da eficácia em se utilizar máquinas de vetor de suporte. Todavia, máquinas de vetor de suporte apresentam grande vantagem estrutural, pois permitem ao critérios/parâmetros de seleção estrutural para otimização das superfícies de separação.

Do ponto de vista gerencial, a implementação de uma MVS nos processos internos da empresa garantirá uma lista continua de potenciais clientes interessados na compra de novas motos, à medida que estes adquirissem este perfil. Dependendo das verbas dispostas para impactar por ações de marketing potenciais clientes interessados, outras configurações de MVS podem ser utilizadas, visando aumentar ou diminuir os erros de acerto em uma determinada classe.

Para os 9817 não entrevistados, verificamos que 55% apresentam uma predição positiva de interesse na compra de outra moto.

REFERÊNCIAS BIBLIOGRÁFICAS

BUSSAB, Wilton de O.; MORETTIN, Pedro A.. **Estatística Básica**. 5ed. São Paulo: Saraiva, 2006.

TARSILA, G. B. T.. **Análise de Basket e Regressão Logística para o Aumento da Rentabilização de uma Loja de Vestuário Feminino**. Monografia de graduação do curso de Ciências Estatísticas, ENCE: Rio de Janeiro, 2008.

PACHECO, Marco A. C.; VELLASCO, Marley B. R.; LAZO, Juan G.. **Notas de Aula do Curso Sistemas Inteligentes Aplicados**. Pontifícia Universidade Católica, Rio de Janeiro,RJ, 2008.

HAYKIN, Simon. **Redes Neurais, Princípios e Práticas**. Paulo Matins Engel. 2ed. Porto Alegre: Bookman, 2001.

GUNN, Steve R.. **Support Vector Machines for Classification and Regression**.University of Southampton, Faculty of Engineering, Science and Mathematics, 1998.

CHAVES, Adriana C. F.. **Notas de Aula do Curso Support Vector Machines**. Pontifícia Universidade Católica, Rio de Janeiro, RJ, 2008.

BERRY, Michael; LINOFF, Gordon. **Data Mining Techniques for Marketing, Sales and Customer Relationship Management**. John Wiley & Sons, Indianapolis, Indiana, 2004.

MAFREDI, Vanilde; MENDES, Jorge. **Curso de Extensão: Implementação e Gestão deCRM – Customer Relationship Management**. Material de Aula, Pontifícia Universidade Católica do Rio de Janeiro: Rio de Janeiro, 2006.

PARK, Kwan; DONOHUE, Steve. **Multistage Cross-Sell Model of Employers inthe Financial Industry**. Sugi 28 Conference, Whashington State, USA, 2003.

WENG, Shifeng. **Random Forest for the Cross-selling Business Problem**. PAKDD2006 data mining competition, Nanyang Technological University, Singapore.