

5 Modelagem e análise dos resultados

Neste capítulo serão detalhados os resultados da modelagem realizada por meio de árvores de decisão, a escolha final do modelo das variáveis que melhor definem o perfil de clientes propensos a aceitar ofertas de uso de dados via campanhas de *cross-selling* e as implicações dos resultados obtidos.

5.1. Caracterização da amostra

Todas as informações utilizadas na modelagem foram levantadas a partir da base de dados da operadora de telefonia móvel estudada. Nos próximos parágrafos serão apresentadas características da amostra final segundo algumas estatísticas descritivas essenciais.

Os dados levantados apresentam clientes, em sua maioria, com bastante tempo de relacionamento com a empresa. Pode ser observado na Figura 4.2 que, 51% da amostra possui acesso de voz há mais de três anos na empresa.

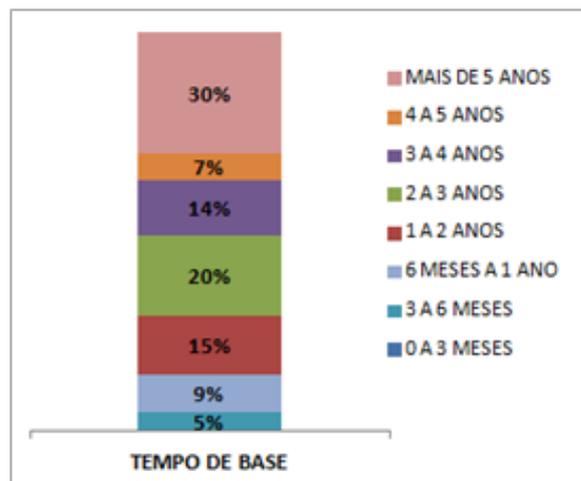


Figura 4.2: Tempo de Base
Fonte: Própria

Embora uma grande parte dos clientes na base total da empresa sejam usuários de iPhone, no grupo de clientes analisados, estes usuários representam apenas 11% do total. Isso pode ser explicado pelo filtro de “não usuários de internet” feito previamente, o que leva a crer que usuários de iPhone geralmente são usuários de internet.

A Figura 4.3 apresenta a distribuição da base por tipo de *smartphone*. Nela é possível identificar que a maioria dos selecionados, 74%, são usuários de aparelhos Samsung, Nokia ou LG.

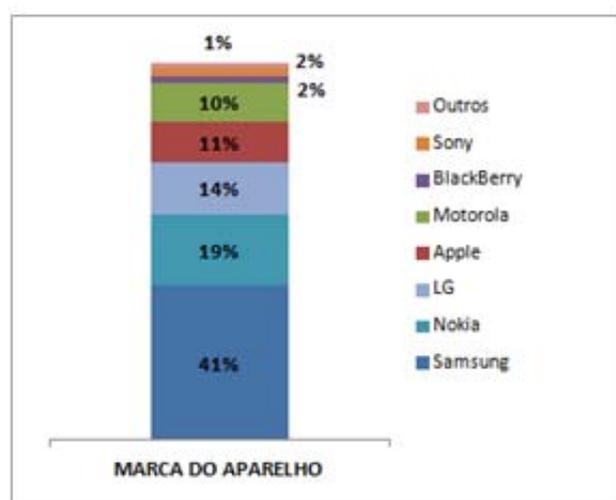


Figura 4.3: Marca do Aparelho
Fonte: Própria

A segmentação dos clientes na empresa analisada leva em consideração seu gasto total mensal, onde os clientes que gastam mais são classificados como clientes diamante, seguidos pelos clientes ouro, prata e bronze.

Como pode ser observado na Figura 4.4, dentre os clientes do grupo analisado, 84% estão nas categorias prata e bronze, indicando que não são, em sua maioria, clientes de alto valor. Assim como ocorre na marca de aparelho, devido ao filtro de não usuários de internet no celular, podemos inferir que clientes de alto valor já são usuários de internet.

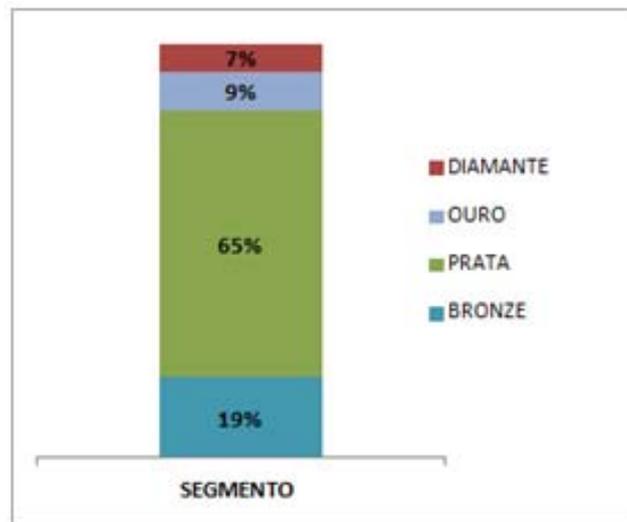


Figura 4.4: Segmento
Fonte: Própria

Na Figura 4.5, nota-se que apenas 8% dos clientes analisados possuem planos de voz acima de 100 minutos. Como a quantidade de minutos utilizados está diretamente relacionada ao gasto mensal do cliente, esta frequência é coerente com a classificação dos segmentos vista anteriormente.

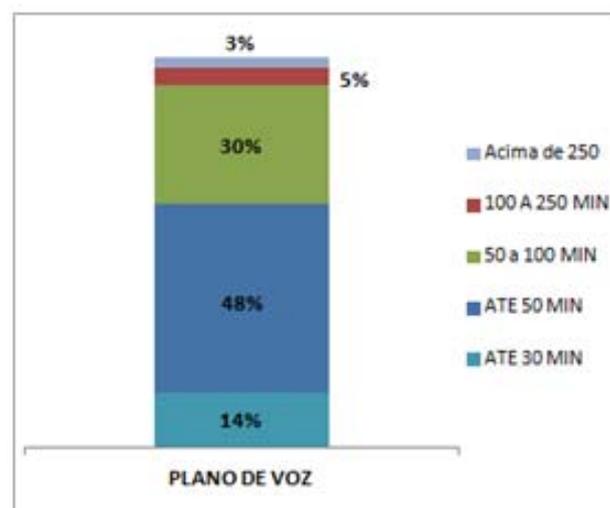


Figura 4.5: Plano de Voz
Fonte: Própria

5.2. Modelagem

No início da modelagem, três técnicas de *data mining* foram testadas para identificar que atributos são mais importantes na predição dos clientes mais propensos à contratação do serviço de internet para o celular: Regressão Logística, Redes Neurais e Árvores de Decisão. Entre estas, a modelagem por árvores de decisão foi escolhida como técnica a ser aplicada neste estudo devido à natureza dos dados de entrada e por apresentar com maior clareza quais variáveis participantes do modelo são relevantes, fornecendo também a ordem de importância de tais variáveis para a predição.

Uma variável alvo binária foi necessária para estimar a Árvore de Decisão. Essa variável foi criada a partir de uma campanha real de *cross-selling* na operadora estudada, onde o valor 0 (zero) representa os clientes que aceitaram a oferta e 1 (um) os clientes que não ativaram a oferta. Lembra-se que a oferta era relativa à assinatura de um plano de internet para uso no celular.

Para a construção da árvore de decisão, foram consideradas inicialmente todas as variáveis disponíveis. Entretanto, dada à natureza desta técnica desta modelagem, onde cada uma das variáveis é classificada de acordo com a importância, apenas as variáveis mais relevantes são consideradas no resultado final. As outras variáveis que foram incluídas no modelo inicialmente não contribuíram significativamente para a predição.

Depois da otimização dos parâmetros, um modelo final foi obtido, com uma predição satisfatória da aceitação do serviço de internet para o celular. A árvore de decisão estimada foi capaz de escolher, dentre todas as variáveis iniciais que poderiam ser úteis, as mais importantes para a predição.

Para que fosse possível generalizar os resultados, foi necessário treinar o modelo primeiro para depois fazer sua validação. A amostra foi repartida em dois subconjuntos mutuamente exclusivos onde o primeiro, que continha 70% da amostra, foi utilizado para o treinamento do modelo. Os 30% restantes (nunca antes vistos pelo modelo) foram usados na validação do modelo estimado. Este processo é feito para determinar o quanto o modelo é preciso na prática, ou seja, como seria o seu desempenho para um novo conjunto de dados. Assim, é possível garantir que o modelo realmente possui capacidade de generalização.

5.3. Ajuste da árvore de decisão

Um *lift chart* representa graficamente quanto o modelo melhora os resultados obtidos. Através do *lift chart* é possível comparar os resultados das diferentes técnicas aplicadas, para utilizar no modelo final apenas aquelas que trazem os melhores resultados (LIMA, HOR-MEYLL, FERREIRA, 2012).

Neste estudo, com o modelo final, foi possível obter um *lift* de 1,85, conforme demonstrado na Figura 5.1. Esse *lift* indica que, com a utilização do modelo, se abordados somente os 10% de clientes classificados como mais propensos a aceitar uma oferta de internet móvel, seria possível melhorar em 1,85 vezes a efetividade da campanha. Ou seja, se o retorno inicial da campanha de *cross-selling* foi de 1,05%, com a utilização do modelo, esse retorno poderia passar a ser de 1,94%, quase o dobro.



Figura 5.1: *Lift Chart* do Modelo Final
Fonte: Própria

De forma geral, os modelos não conseguem acertar em todas as classificações que serão feitas no futuro. Uma aproximação do erro de classificação pode ser obtida utilizando os resultados obtidos no treino do modelo. Depois de criado o modelo, os resultados obtidos no treino são confrontados com os resultados obtidos na validação. O resultado desta análise é a Matriz de Confusão.

A **matriz de confusão** quantifica quantos exemplos da base de dados utilizada seriam classificados corretamente pelo modelo construído e quantos exemplos não seriam classificados corretamente pelo modelo (LIMA, HORMEYLL, FERREIRA, 2012).

A matriz de confusão obtida após o ajuste do modelo está apresentada na Tabela 5.1. Nela, identificamos que o modelo foi capaz de prever corretamente 88% dos casos em que o retorno era negativo, ou seja, que o cliente não aceitou a oferta. Isso significa que o modelo acerta muito em prever quem não vai aceitar a oferta. Os benefícios deste cenário são:

- Menor desgaste da base, na medida em que clientes que não aceitariam a oferta não serão abordados.
- Redução de custo, já que a abordagem será mais efetiva.

É possível verificar também que o modelo foi capaz de prever corretamente somente 33% dos casos em que o retorno era positivo, ou seja, que o cliente aceitou a oferta. Isso significa que o modelo não acerta tanto em prever quem vai aceitar a oferta. O complicador deste cenário é que a empresa pode deixar de abordar clientes que seriam propensos à aceitação da oferta.

		PREDITO	
		0	1
REAL	0	88%	12%
	1	67%	33%

Tabela 5.1: Matriz de Confusão do Modelo Final

Fonte: Própria

O resultado da matriz de confusão pode ser visualizado mais facilmente no Gráfico de Acerto da Figura 5.2 abaixo.

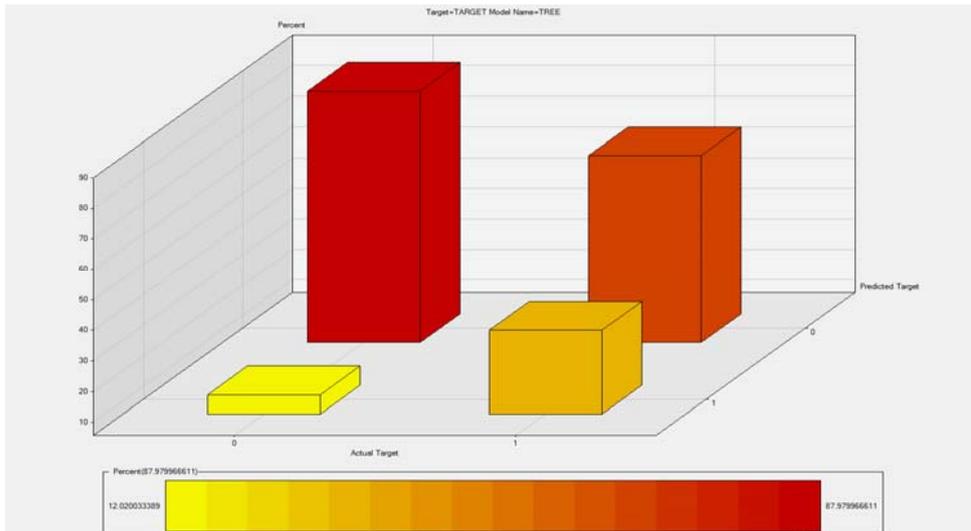


Figura 5.2: Gráfico de Acerto

Fonte: Própria

5.4. Estrutura final da árvore de decisão

De acordo com o algoritmo utilizado pela árvore de decisão, a variável que contribui mais para a distinção de duas classes é alocada na primeira divisão. Seguindo a ordem decrescente de importância, as variáveis subsequentes que contribuem para a divisão aparecem nos demais níveis, até que nenhuma outra variável seja útil para distinguir duas classes.

Na estrutura final da árvore de decisão, observa-se que a variável mais relevante foi a marca do aparelho utilizada pelo cliente, seguida da quantidade de minutos que o cliente utiliza em chamadas recebidas e do seu tempo de relacionamento com a operadora.

O diagrama completo desta modelagem é apresentado na Figura 5.3. A influência de cada uma das variáveis selecionadas para a estrutura final da árvore na definição da variável alvo será discutida na sessão seguinte.

Figura 5.3: Diagrama da Árvore de Decisão

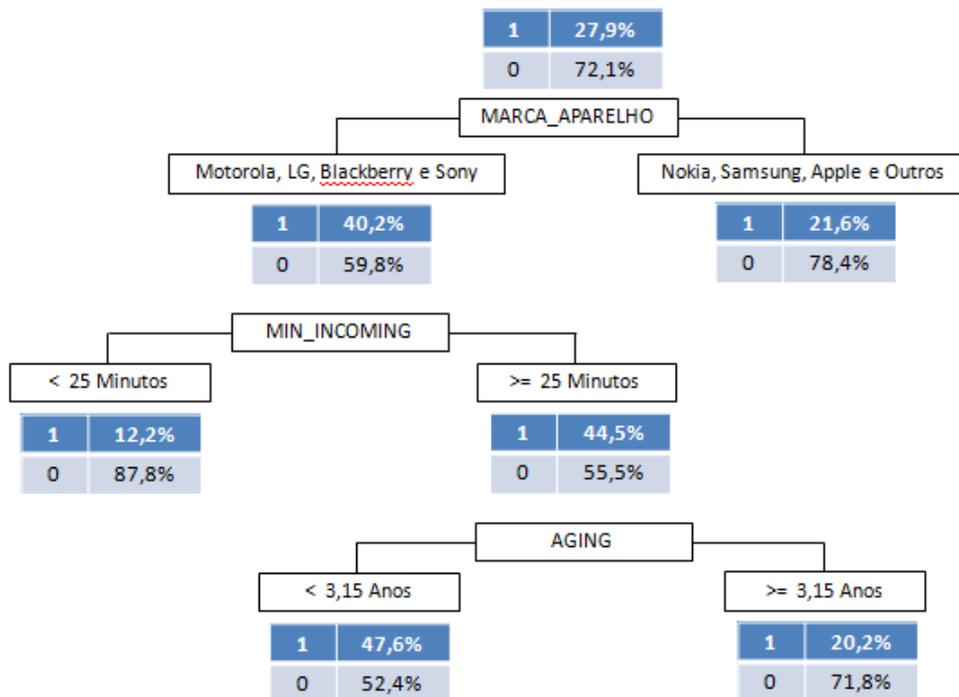


Figura 5.3: Diagrama da Árvore de Decisão
Fonte: Própria

5.5. Análise dos resultados

A variável mais importante para determinar a propensão de clientes a aceitar ofertas de uso de dados no celular foi a marca do aparelho que o cliente possui no momento do contato. De acordo com a primeira divisão, tem-se que, se forem abordados os usuários de *smartphones* Motorola, LG, Blackberry e Sony, as chances de o cliente aceitar a oferta de internet para uso no celular aumentam de 27,9% para 40,2%. Por outro lado, se forem abordados os usuários de *smartphones* Nokia, Samsung, Apple e Outros, as chances de o cliente aceitar a oferta caem de 27,9% para 21,6%.

Como visto anteriormente, na seleção do grupo de análise, foi feito um filtro prévio de não usuários de internet no celular. Assim, há indícios de que os usuários de *smartphones* Nokia, Samsung, Apple e Outros, que ainda não usam a internet no celular, de fato não têm interesse no serviço. Portanto, não há necessidade de abordá-los com ofertas desse tipo. É possível que sejam clientes mais idosos, não familiarizados com a tecnologia, ou mesmo clientes que usam a internet no celular apenas via wi-fi.

A segunda variável mais importante foi a quantidade de minutos utilizados em chamadas recebidas. Esta divisão mostra que, para clientes com aparelhos Motorola, LG, Blackberry e Sony, se um indivíduo fala mais de 25 minutos em chamadas recebidas, as chances dele aceitar a oferta aumentam de 40,2% para 44,5%. Por outro lado, se um cliente fala menos de 25 minutos em chamadas recebidas, as chances de aceitar a oferta caem consideravelmente, de 40,2% para 12,2%.

A quantidade de chamadas recebidas pode estar relacionada à utilização do celular como um todo, com outros tipos de chamadas, outros tipos de serviços e utilidades, e até mesmo com a aceitação do canal de comunicação utilizado para a oferta (*Smart Message*). Assim, é provável que, quanto maior a quantidade de chamadas recebidas, maior a utilização do celular e maior a propensão do cliente em conhecer e utilizar novos serviços que antes eram desconhecidos seja pelo serviço em si ou por uma oferta da operadora. Por exemplo, estes clientes poderiam desconhecer que a operadora possui planos de internet para uso no celular com preço acessível.

A terceira variável mais importante foi o tempo de relacionamento do cliente com a operadora. Esta divisão mostra que, para clientes com aparelhos Motorola, LG, Blackberry e Sony e que usam mais do que 25 minutos em chamadas recebidas, se um indivíduo tem mais de 3,15 anos de relacionamento com a operadora, as chances dele aceitar a oferta caem de 44,5% para 20,2%. Por outro lado, se um cliente tiver menos de 3,15 anos de relacionamento com a operadora, as chances de aceitar a oferta aumentam de 44,5% para 47,6%.

O tempo médio de relacionamento entre o cliente e a operadora pode estar relacionado à idade do cliente. Se isto for uma realidade, podemos assumir que clientes com mais idade não são o público alvo para este tipo de oferta. Essa suposição poderia ser confirmada se tivessem sido disponibilizados mais dados demográficos (sexo, idade, renda, escolaridade, etc) do grupo de clientes analisados.

Na seleção inicial das variáveis foram levantados alguns dados que poderiam ser relevantes para o modelo, mas que não apareceram na estrutura final da árvore, especialmente os contatos do cliente com o *call center* da operadora. Esperava-se que, quanto menos contatos o cliente tivesse com o *call center*, menos motivos ele teria de insatisfação e possivelmente estaria mais aberto a

aceitar ofertas de *cross-selling*. Entretanto, o modelo mostrou que esta não é uma variável suficientemente relevante. É possível que, se as interações tivessem sido levantadas em um período maior de tempo, os resultados fossem diferentes.

O mesmo aconteceu com informações relacionadas ao gasto do cliente com a operadora como o segmento e o plano de voz. Talvez essas não fossem as variáveis ideais para estimar a faixa de gastos do cliente na operadora. Talvez fosse melhor utilizar dados de faturamento total, que corresponde ao comprometimento mensal do cliente de fato.

5.6.

Aplicação do modelo proposto ao caso estudado

Conforme apresentado anteriormente, os meios de comunicação mais utilizados pela operadora para ações de *cross-selling* hoje são o SMS e o *Smart Message*.

Na ação de estímulo à contratação do serviço de internet para uso no celular, o canal escolhido foi o *Smart Message*. Isso porque, para ações que envolvem a contratação do serviço no momento do recebimento da comunicação, o *Smart Message* é um canal mais eficiente, já que, para retirar a mensagem da tela do celular, o cliente precisa dar um comando de aceite ou recusa. No caso do SMS, o cliente pode visualizar a mensagem, deixar para entrar em contato com a operadora posteriormente e esquecer ou desistir.

Porém, apesar de ser mais eficiente, o canal *Smart Message* exige a contratação de um fornecedor externo, já que não é uma plataforma própria da operadora em questão. Nas entrevistas com funcionários, foi dito que a operadora tem um limite de 300.000 envios diários estabelecido por contrato com o fornecedor. Ou seja, com uma única ação de *cross-selling* (cujos dados foram analisados nesta pesquisa), a operadora utilizou metade de sua capacidade diária de envios, já que foram abordados 150.000 clientes.

O contrato determina ainda que a remuneração do fornecedor externo deve corresponder a 20% da receita líquida gerada pela primeira fatura de cada cliente após a contratação do serviço. Mas ainda que a operadora tenha esse custo por ativação, quanto mais ativações melhor.

Com a modelagem proposta foi possível identificar que, abordando apenas 10% dos 150.000 clientes iniciais, seria possível obter uma efetividade quase duas vezes maior na ação de *cross-selling*. Isso porque seriam abordados apenas os clientes mais propensos a aceitar a oferta de internet móvel. Dessa forma, a operadora usaria apenas 5% de sua capacidade diária de envios, ao invés de 50%, e poderia fazer outras campanhas de *cross-selling* em paralelo para aumentar sua receita. Estimulo a aquisição de outros tipos de serviços como canais de música, educação ou saúde, por exemplo.

Este resultado confirma o exposto na literatura analisada, especialmente os estudos de caso de Ahn et al (2011) e Li, Sun e Montgomery (2011), que também obtiveram melhora de performance significativa nas ações de *cross-selling* com a utilização de modelos preditivos.

Em ações futuras, os resultados do modelo proposto poderiam ser facilmente aplicados. Basta que a operadora passe a usar filtros adicionais na geração do *mailing* para as ações de *cross-selling* de estímulo ao uso de serviço de internet no celular, além dos que já são aplicados. Estes novos filtros corresponderiam às variáveis classificadas como relevantes pela árvore de decisão. Ou seja, nas próximas campanhas, deveriam ser selecionados apenas clientes que atendessem aos seguintes critérios:

- Usuários de *smartphones* Motorola, LG, Blackberry e Sony
- Que recebam utilizem mais de 25 minutos em ligações recebidas por mês
- Que possuam um plano de voz pós-pago há menos de 3 anos na operadora