

MODELO INTEGRADO PARA ANÁLISE DO DESLIGAMENTO DE EMPREGADOS: UM ESTUDO DE CASO

Felipe de Moraes Oliveira

Thiago de Melo Rezende

PETROBRAS

Av. República do Chile 65, Rio de Janeiro – RJ. CEP 20031-912

felipe.moraes@petrobras.com.br

thiagorezende@petrobras.com.br

Celso F. Araujo

Rafael de O. Valle dos Santos

Ricardo Vitor J. C. Vasconcelos

PETROBRAS

Av. Nilo Peçanha 151, Centro, Rio de Janeiro - RJ. CEP 20020-100

celsoaf@petrobras.com.br

rvsantos@petrobras.com.br

ricardovitor@petrobras.com.br

RESUMO

Em um cenário de economia emergente, onde as grandes empresas têm sua demanda por recursos humanos aumentada, é natural que os profissionais da área gestão de pessoas intensifiquem o uso de modelos e ferramentas científicas de análise que os ajudem nas atividades de planejamento e tomada de decisão. Tomando como exemplo uma grande empresa do mercado brasileiro, este trabalho exemplifica como é possível empregar modelos de previsão e de mineração de dados de maneira integrada, visando analisar a dinâmica dos desligamentos de empregados. O uso destes modelos é particularmente útil para os profissionais ligados à gestão de efetivo.

PALAVRAS CHAVE. Recursos humanos. Modelos de previsão. Mineração de dados.

Área principal: OA - Outras Aplicações em PO.

ABSTRACT

On an emerging economy scenario, where major enterprises have their demand for human resources raised, it is natural for people management professionals to intensify the use of scientific models and analysis tools to help them on planning and decision-taking activities. Taking a leading company from the Brazilian market as an example, this work shows how it is possible to employ forecasting and data mining models in an integrated fashion, aiming to analyze employees resignation dynamics. The use of these models is particularly useful for professionals related to human resources management.

KEYWORDS. Human resources. Forecasting. Data mining.

Main area: OA - Outras Aplicações em PO.

1. Introdução

De uma forma geral, as empresas têm como um de seus principais desafios a gestão adequada de seu efetivo. No caso daquelas que possuem um número considerável de empregados, isso pode se tornar crítico, principalmente se forem considerados os inconvenientes que a rotatividade na força de trabalho pode implicar.

Assim sendo, é de se esperar que grandes empresas façam estudos para prever a demanda futura por empregados, permitindo, assim, a organização de seus processos de seleção, contratação e treinamento. Logo, para dar suporte à previsão de efetivo futuro, é interessante que a empresa possua mecanismos de previsão da taxa de desligamentos de seus empregados. Tal taxa poderia, eventualmente, ser segregada por carreira ou por área da empresa.

Considerando que os desligamentos podem ser motivados por aposentadoria, falecimento, demissão, pedido do colaborador ou qualquer outra razão, existe uma demanda por parte dos gestores de RH para que, além da previsão quantitativa da taxa de desligamento, seja definido o perfil dos empregados que se desligam. O conhecimento deste perfil pode nortear políticas de retenção e de contratação.

Neste trabalho considera-se o exemplo de uma grande empresa do mercado nacional, com efetivo na casa dos milhares. Algumas características específicas desta empresa, como a contratação apenas por concurso público e a intensa concorrência do mercado pelos profissionais, indicam a importância de se dispor de mecanismos adequados para a previsão de desligamentos.

O objetivo aqui é descrever o desenvolvimento de um modelo integrado para análise quantitativa e qualitativa do desligamento de empregados, baseado em modelos estatísticos simples (Lapponi, 2005) e de mineração de dados (Witten e Frank, 2005). A estratégia usada para construção deste modelo respeitou as características da gestão de RH da empresa estudada e contemplou a análise de duas carreiras distintas, tendo como base o histórico dos desligamentos ocorridos entre os anos de 2005 e 2010 (total de seis anos).

2. Metodologia

Neste trabalho, são exploradas duas técnicas consideradas complementares: estatística básica e mineração de dados. Enquanto a estatística é um ramo consolidado das ciências exatas, a disciplina conhecida como *mineração de dados* vem ganhando espaço, como se pode ver pela multiplicidade de fontes *on-line* sobre o assunto (Homepage “*The Data Mine*”) e pelo aumento na disponibilidade de software específico, gratuito ou não (Homepage do WEKA; Homepage do SAS).

A metodologia usada no desenvolvimento do modelo integrado para análise de desligamentos foi fundamentada na disponibilidade dos dados de *efetivo* e de *desligamentos*. A base de dados utilizada refere-se ao período 2005-2010 e registra o efetivo total e os respectivos desligamentos por idade, tempo de empresa, sexo, carreira e área de atuação.

Considerando que o modelo integrado foi desenvolvido sob a premissa de que apenas as variáveis intrínsecas à empresa deveriam ser consideradas como explicativas para a previsão das taxas de desligamento¹, a estratégia utilizada no seu desenvolvimento pode ser assim resumida:

1. Segregação dos dados entre empregados ativos e desligados por carreira;
2. Previsão quantitativa das taxas de desligamento por carreira, através de *médias móveis* com intervalo de 3 anos;
3. Obtenção do perfil qualitativo dos empregados desligados a partir do algoritmo C4.5;
4. Integração dos resultados quantitativos e qualitativos, para uso dos gestores de RH.

¹ Um exemplo de variável exógena que se poderia considerar, se estivesse disponível: índice mensal de oferta de emprego no mercado, variando por carreira.

2.1. Médias móveis

Dado o reduzido intervalo histórico (6 anos), considerou-se o uso do método de *médias móveis* (Lapponi, 2005) como o mais adequado para a obtenção da previsão das taxas de desligamento. Com séries mais longas, outros métodos poderiam ser testados, e.g., amortecimento exponencial.

No método de médias móveis com janela L , a previsão da taxa em $t+1$, calculada no instante t , é a média aritmética dos valores de taxa observados em $t, t-1, \dots, t-L+1$. Neste trabalho por exemplo, foram consideradas médias móveis com janela deslizante de 3 (três) anos.

2.2. Algoritmo C4.5

Para determinação do perfil dos empregados desligados, optou-se pelo uso do algoritmo C4.5, que atua como um classificador estatístico por árvore de decisão (Quinlan, 1993; Witten e Frank, 2005; Homepage da *Rulequest*, 2011). Este algoritmo, criado por Ross Quinlan, é considerado um dos “10 mais” na disciplina de *mineração de dados* (Wu e Kumar, 2009).

Dado um conjunto de elementos (registros) com propriedades conhecidas (atributos), o C4.5 usa as propriedades dos elementos para classificá-los em categorias pré-determinadas. Por exemplo, considere-se um conjunto *Pessoas*, onde cada elemento, i.e. cada *pessoa*, contém as propriedades *Estado_Civil*, *Filhos*, *Sexo*, *Renda* e *Gasto_Alimentação*; uma possível aplicação do algoritmo seria classificar os elementos de *Pessoas* por faixa de *Gasto_Alimentação*. Em um algoritmo baseado em árvores de decisão, cada classificação é definida por um conjunto de regras SE-ENTÃO; os *antecedentes* (parte SE das regras) conteriam um ou mais dos atributos *Estado_Civil*, *Filhos*, *Sexo* e *Renda*; os *conseqüentes* (parte ENTÃO das regras) indicariam a faixa (discreta) do atributo *Gasto_Alimentação* na qual uma determinada *pessoa* se encaixa. Os passos básicos do C4.5 estão descritos no quadro abaixo (Quadro 1):

→ Considere um conjunto de *treinamento* S , com o atributo objetivo assumindo um número qualquer de rótulos nominais (classes).

1. Se todas as instâncias em S tiverem o mesmo valor de atributo objetivo, crie um *nó-folha* com esse valor e pare a execução; senão, **escolha um atributo explicativo A** para ramificar a árvore (crie um novo *nó-decisório*);
2. Divida as instâncias de S em subconjuntos S_1, S_2, \dots, S_n , de acordo com os valores (ou faixas de valores) de A .
3. Aplique o algoritmo recursivamente para cada um dos conjuntos S_i .

COMO ESCOLHER O ATRIBUTO DE RAMIFICAÇÃO (heurística de seleção):

4. Para cada atributo candidato, calcula-se uma grandeza chamada Entropia;
5. Escolhe-se o atributo que formar o ramo com a menor soma ponderada de Entropias (maior Ganho de Informação);

→ Um atributo candidato é um atributo que ainda não tenha tido sua capacidade de partição esgotada.

Quadro 1 – Funcionamento do C4.5

Para demonstrar o uso do C4.5 aplicado a análise de desligamentos, utiliza-se um banco de dados fictício, exibindo 21 empregados desligados recentemente (Tabela 1). A tabela lista, para cada desligado, tempos de permanência na empresa (em anos), idades e expectativas de aposentadoria (em anos) quando da admissão. Os tempos esperados para aposentadoria são estimados da seguinte maneira:

Seja TVP = tempo de vinculação previdenciária, quando da admissão; então:

1. Se Sexo = masculino, então Expectativa = $\max(35 - TVP, 0)$;

2. Se Sexo = feminino, então Expectativa = $\max(30 - TVP, 0)$.

Tabela 1 – Exemplo de uso do C4.5. Banco de dados fictício com 21 registros.

| Tempo de permanência | Sexo | Expectativa de aposentadoria na admissão | Idade na admissão |
|----------------------|-----------|--|-------------------|
| 2 | masculino | 34 | 20 |
| 23 | masculino | 23 | 32 |
| 8 | masculino | 29 | 25 |
| 32 | masculino | 34 | 19 |
| 26 | masculino | 26 | 28 |
| 35 | masculino | 35 | 20 |
| 26 | masculino | 26 | 28 |
| 6 | feminino | 17 | 38 |
| 30 | feminino | 30 | 20 |
| 31 | masculino | 33 | 21 |
| 5 | masculino | 27 | 27 |
| 3 | masculino | 33 | 21 |
| 30 | masculino | 30 | 23 |
| 32 | masculino | 32 | 21 |
| 31 | masculino | 31 | 23 |
| 30 | masculino | 30 | 27 |
| 3 | feminino | 23 | 29 |
| 20 | masculino | 20 | 37 |
| 30 | masculino | 30 | 26 |
| 1 | masculino | 31 | 21 |
| 2 | feminino | 22 | 34 |

Para aplicação do C4.5, o primeiro passo a ser dado é o estabelecimento das classes a serem consideradas no estudo. Neste trabalho, o interesse é classificar os desligados por *tempo de permanência na companhia*; assim, esta variável, chamada de *variável objetivo*, deve ser segregada (*discretizada*) em um número finito de classes. Como exemplo, sugere-se uma segregação do tempo de permanência em 4 classes:

1. Até 5 anos;
2. De 6 a 20 anos;
3. De 21 a 29 anos;
4. 30 ou mais anos.

A Figura 1 mostra uma *árvore de decisão* gerada pelo C4.5, quando aplicado sobre o banco de dados do exemplo, utilizando o tempo de permanência como variável objetivo.

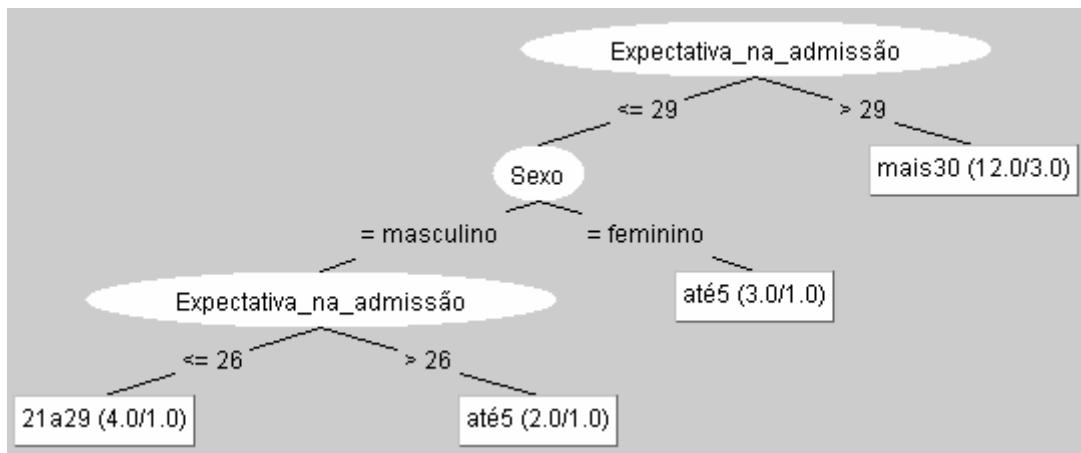


Figura 1 – Saída do algoritmo C4.5 (exemplo fictício).

A leitura das folhas da árvore dá-se da seguinte maneira: “*classe (total de elementos classificados / elementos erroneamente classificados)*”. Por exemplo: “*mais30 (12.0/3.0)*” significa que para a faixa de permanência maior ou igual a 30 anos, foram classificados 12 indivíduos sendo que, destes 12, houve 3 indivíduos classificados erroneamente (75% de acerto). O índice de acerto geral da árvore mostrada ficou em 62%.

3. Experimentos

Como citado anteriormente, os modelos estatísticos e de mineração de dados descritos neste trabalho foram testados com dados reais, colhidos de uma grande empresa brasileira. Os experimentos de teste foram estruturados da seguinte forma:

1. Definição de duas carreiras para realização do estudo: A e B²;
2. Levantamento do número de desligamentos por carreira, por mês, entre 2005 e 2010;
3. Levantamento do efetivo total, por carreira, por mês, entre 2005 e 2010. Considerando a grande quantidade de dados que representa este período, a gestão de RH pôde disponibilizar apenas o efetivo total, por carreira, dos meses de *dezembro* entre 2005 e 2010.
4. Cálculo da *taxa de desligamento*, por *carreira* e por *ano*. Para cada carreira, a taxa de desligamento é o quociente entre o número de desligamentos observados em um dado ano e o efetivo medido em dezembro daquele mesmo ano.
5. Projeção dos desligamentos, usando média móvel com intervalo de 3 anos. Seguem os pares histórico – projeção:
 - a. Histórico 2005/06/07 – projeção para 2008;
 - b. Histórico 2006/07/08 – projeção para 2009;
 - c. Histórico 2007/08/09 – projeção para 2010;
 - d. Histórico 2008/09/10 – projeção para 2011.
6. Obtenção do MAE (Mean Absolute Error – Erro Absoluto Médio) para as projeções;
7. Obtenção dos *perfis de empregados que se desligaram* utilizando o algoritmo C4.5.

3.1. Resultados: previsão de taxas

Os cálculos de média móvel foram realizados com planilha eletrônica MS-Excel.

O gráfico a seguir (Figura 2) apresenta o resultado das projeções das taxas de desligamento, em pontos percentuais, para a carreira A:

². Os nomes das carreiras selecionadas foram omitidos por confidencialidade.

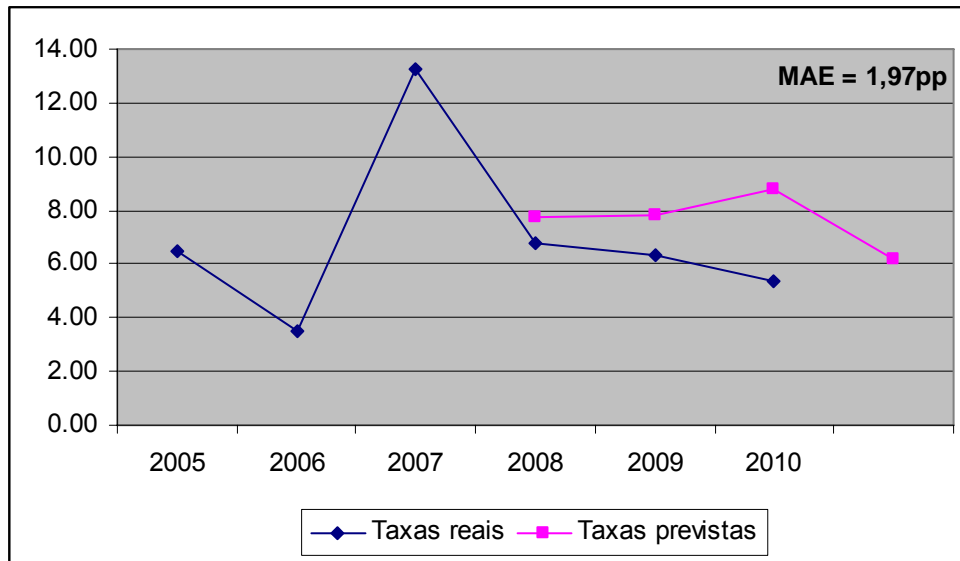


Figura 2 - Taxas de desligamento: previstas x realizadas no período 2005-2010 (carreira A).

O gráfico na , apresenta o resultado das projeções das taxas de desligamento, em pontos percentuais, para a carreira B:

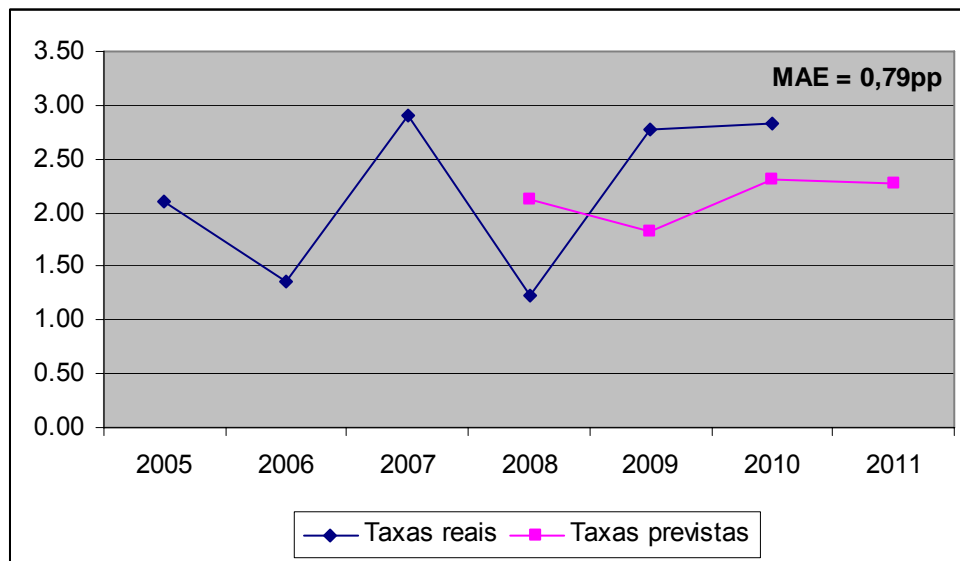


Figura 3 - Taxas de desligamento: previstas x realizadas no período 2005-2010 (carreira B).

3.2. Resultados: perfis de desligamento

Para os experimentos relacionados a obtenção de perfis, foi utilizada a implementação do C4.5 existente no software WEKA (Homepage do WEKA, 2011).

A partir do histórico de empregados desligados, compreendendo o período de 2005 a 2010, as seguintes informações foram levantadas:

1. Sexo;
2. Data de admissão;
3. Tempo de vinculação previdenciária;

4. Idade;
5. Motivo do desligamento;
6. Último cargo ocupado;
7. Última área em que trabalhava.

Com base nestas informações, realizou-se a análise do perfil dos empregados desligados das duas carreiras selecionadas, A e B. Para tanto, as seguintes variáveis foram consideradas: *Idade na Admissão*, *Expectativa de Aposentadoria na Admissão* e *Tempo de Permanência na Empresa*. Como exemplificado na seção 2.2, os perfis foram classificados de acordo com diferentes faixas de *Tempo de Permanência*.

Primeiramente, foi feita a análise para a carreira A, que apresentou 228 empregados desligados no período estudado (6 anos). A análise foi feita considerando dois diferentes conjuntos de perfis, de acordo com o tempo de permanência na companhia. Os resultados estão nos gráficos a seguir (Figura 4 e Figura 5).

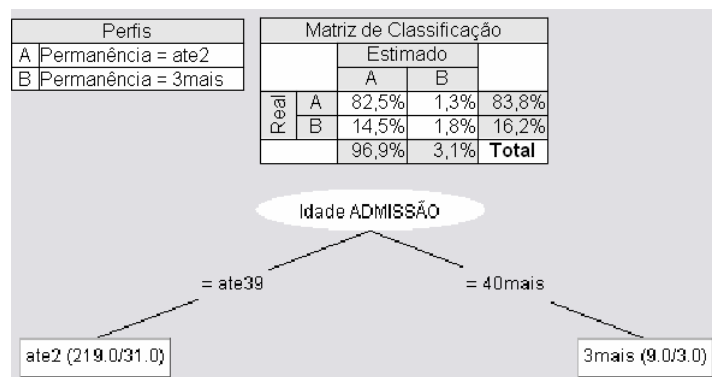


Figura 4 - Saída do algoritmo considerando duas faixas de permanência: até e inclusive 2 anos completos; 3 ou mais anos completos de trabalho na empresa.

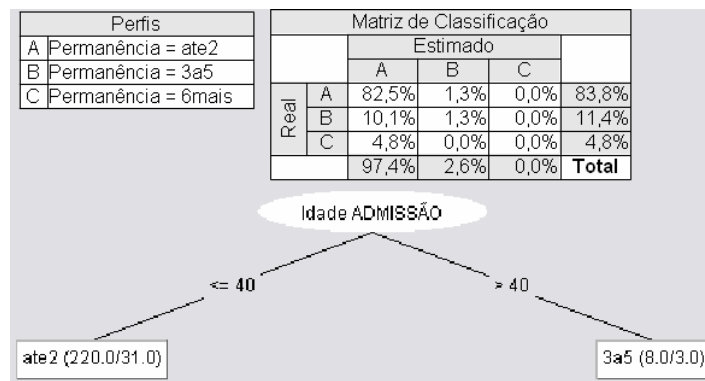


Figura 5 - Saída do algoritmo considerando três faixas de permanência: até e inclusive 2 anos completos; entre 3 e 5 anos completos; 6 ou mais anos completos de trabalho na empresa.

Em seguida, foi feita análise semelhante para carreira B, com 280 empregados desligados no período (6 anos). Os resultados estão exibidos logo abaixo (Figura 6 e Figura 7).

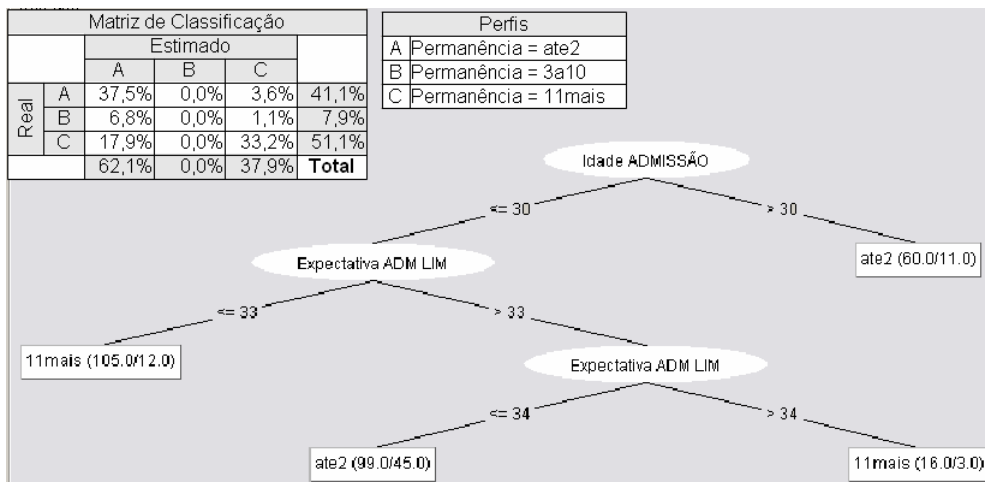


Figura 6 - Saída do algoritmo considerando três faixas de permanência: até e inclusive 2 anos completos; entre 3 e 10 anos completos; 11 ou mais anos completos de trabalho na empresa.

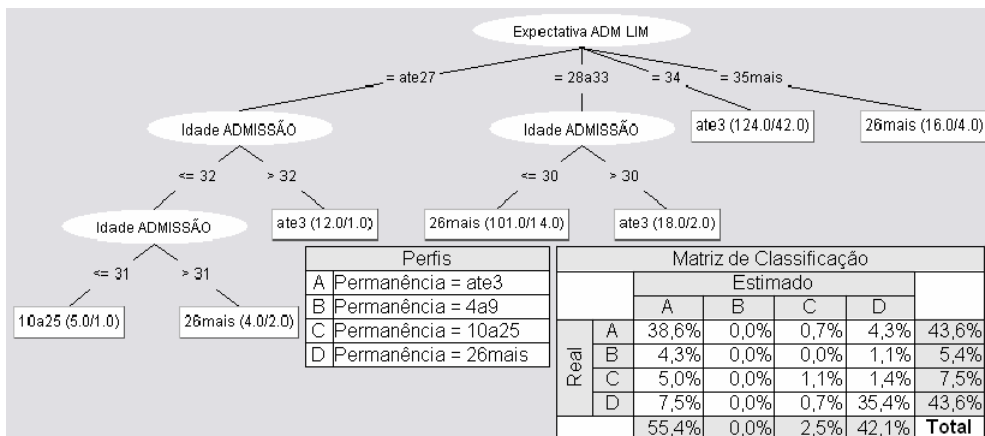


Figura 7 - Saída do algoritmo considerando quatro faixas de permanência: até e inclusive 3 anos completos; entre 4 e 9 anos completos; entre 10 e 25 anos completos; 26 ou mais anos completos de trabalho na empresa.

4. Integração dos modelos

Os resultados apresentados na seção anterior devem ser usados de maneira integrada, funcionando como ferramenta de apoio à decisão. Nas próximas subseções exemplifica-se o procedimento de integração proposto, por carreira.

4.1. Carreira A

A taxa de desligamentos projetada (por média móvel) para os empregados da carreira A, no ano de 2011, foi de 6,17% (MAE de 1,97pp). Assim, o número previsto de empregados desligados em 2011 é igual a 48, i.e., 782 (nº total de empregados em jan 2011) $\times 6,17\% = 48$. Utilizando as informações das árvores de decisão, é possível segregar estes resultados por faixas de tempo de permanência na companhia. Assim temos os seguintes resultados:

- Para árvore de decisão com 2 faixas:
 - Até 2 anos: $82,5\%$ dos desligamentos = $48 \times 0,825 = 40$ empregados
 - Acima de 2 anos de permanência = $48 \times 0,018 = 1$ empregado
- Para árvore de decisão com 3 faixas:
 - Até 2 anos: $82,5\%$ dos desligamentos = $48 \times 0,825 = 40$ empregados
 - Entre 3 e 5 anos: 1 empregado

- Acima de 6: 0 empregados

Obs.: A diferença entre o total de desligamentos (48) e o somatório por faixas (41 e 41) está relacionada com a incerteza do modelo (erros de 15%).

4.2. Carreira B

Para os empregados da carreira B, a taxa de desligamento projetada para 2011 é 2,28% (MAE = 0,79pp). Assim, o número previsto de empregados a serem desligados em 2011 é igual a 57, i.e., 2508 (nº total de empregados em jan 2011) * 2,28% = 57. Utilizando as informações das árvores de decisão, é possível segregar estes resultados por faixas de *tempo de permanência* na companhia. Assim temos os seguintes resultados:

- Para 3 faixas:
 - Até 2 anos: 37,5% dos desligamentos = $57 * 0,375 = 21$ empregados
 - Entre 3 e 10 anos: 0% = 0 empregados
 - Acima de 11 anos: 33,2% = $57 * 0,332 = 19$ empregados
- Para 4 faixas:
 - Até 3 anos: 38,6% dos desligamentos = $57 * 0,386 = 22$ empregados
 - Entre 4 e 9 anos: 0% = 0 empregados
 - Entre 10 e 25 anos: 1,1% = $57 * 0,01 = 1$ empregado
 - Acima de 26 anos: 35,4% = $57 * 0,354 = 20$ empregados

Obs.: A diferença entre o total de desligamentos (57) e o somatório por faixas (40 ou 43) está relacionada com a incerteza do modelo (erros de 30% e 25% respectivamente).

4. Conclusão

Um dos maiores desafios na gestão de pessoas nos dias de hoje é a atração e retenção de talentos em um mercado altamente competitivo. Portanto, é de suma importância que as empresas conheçam o perfil dos empregados existentes em seu efetivo, bem como o comportamento da rotatividade deles no mercado em que atuam.

Considerando a necessidade de se obter um modelo de projeção de desligamentos, com base nas variáveis intrínsecas à empresa em questão, pode-se concluir que as técnicas de mineração de dados aqui utilizadas contribuíram significativamente para a identificação das principais características que influenciam no desligamento dos empregados.

Como trabalho futuro, pode-se sugerir que o modelo integrado aqui desenvolvido seja aplicado às demais carreiras da mesma empresa. Desta forma, seria possível a identificação de agrupamentos de carreiras com comportamentos semelhantes (*clusters*), relevantes para identificação de áreas consideradas críticas, potenciais alvos de programas de atração e retenção de talentos.

Referências

Homepage da Rulequest [WWW], <http://www.rulequest.com/index.html>, acessada em abril/2011.

Homepage The Data Mine [WWW], <http://www.the-data-mine.com/>, acessada em abril/2011.

Homepage do SAS [WWW], <http://www.sas.com/offices/latinamerica/brazil/>, acessada em abril/2011.

Homepage do WEKA [WWW], <http://www.cs.waikato.ac.nz/ml/weka/index.html>, acessada em abril/2011.

Lapponi, J. C., *Estatística usando Excel (4ª edição)*, Campus, 2005.

Quinlan, J. R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

Witten, I. H. e Frank, E., *Data mining: practical machine learning tools and techniques (2nd edition)*, Morgan Kaufmann, 2005.

Wu, X. e Kumar, V., *The top ten algorithms in data mining*, Chapman & Hall/CRC, 2009.
<http://www.the-data-mine.com/>