

UM ESTUDO SOBRE A ÁRVORE DE DECISÃO E SUA IMPORTÂNCIA NA HABILIDADE DE APRENDIZADO

Paola Guarisso Crepaldi¹, Renato Nogueira Perez Avila², Jaqueline Paula M. de Oliveira³ Paulo Ricardo Rodrigues⁴, Renan Lucio Martins⁵

RESUMO

A utilização da árvore de decisão oferece uma maneira simples para o acesso ao conhecimento, é também um ótimo meio de minerar classes e várias outras informações extremamente úteis que são extraídas em valores de atributos de conjuntos de dados. Existem ótimos serviços no qual a adoção deste tipo de recurso em ambientes de trabalho. A primeira facilidade é a construção baseada na classificação de conjuntos de treinamento, porém, o maior serviço que a árvore de decisão oferece são suas habilidades de aprendizado. Assim, quando o treinamento for completamente passado, a árvore de decisão é alimentada.

Palavras Chave: Árvore de Decisão, Aprendizado, Administração.

ABSTRACT

The use of the decision tree offers a simple way for the access to the knowledge, is also an excellent one half of minerar classrooms and several other extremely useful information that are extracted in values of attributes of data sets. Excellent services in which exist the adoption of this type of resource in work environments. The first easiness is the construction based on the classification of sets of training, however, the biggest service that the decision tree offers is its abilities of learning. Thus, when the training will be completely past, the decision tree is fed.

Keywords: Tree of Decision, Learning, Administration.

¹ Professora Especialista pela PUC- Pontificia Universidade Católica de Londrina em Comércio Exterior e Logística Internacional e Graduada pela Unopar - Universidade Norte do Paraná em Ciências Contábeis, ² Professor, Graduado em Tecnologia em Processamento de Dados, Graduado em Licenciatura Plena em Informática, Especialista em Ciência da Computação e Mestre em Telecomunicações, Doutorando em Ciência da Educação, docente de vários cursos de Graduação da INESUL, ³ ⁴ ⁵ Graduando em Ciências Contábeis – Faculdade Integrada Inesul Londrina

INTRODUÇÃO

Neste artigo iremos abordar um assunto não muito conhecido na área de graduação, iremos falar sobre Arvore de Decisão. As árvores de decisão são representações simples do conhecimento, e um meio eficiente de construir classificadores que predizem ou revelam classes ou informações úteis baseadas nos valores de atributos de um conjunto de dados.

O mais interessante sobre o programa de árvores de decisão não é a sua construção a partir de classificação de um conjunto de treinamento, e sim a sua habilidade de aprendizado. Quando o treinamento é finalizado, é possível alimentar sua árvore de decisão construída a partir de exemplos com novos casos a fim de classificá-los.

O objetivo desse artigo é mostrar que existem vários tipos de arvores de decisão em todas as áreas que envolvam o ramo empresarial, pois o ambiente empresarial requer decisões rápidas e precisas, para tanto existem vários métodos de auxílio e a arvore de decisão é uma delas. Temos arvores de decisão na área financeira, contas a pagar e a receber, compras, ativo, passivo, provisões, etc. Existem inúmeros modos de arvores de decisões, basta saber onde vai ser feita uma.

O processo de tomada de decisão dentro das organizações modernas demanda um profissional que conheça o ambiente em que a empresa está inserida e também esteja ciente das suas constantes e freqüentes mudanças. Por decisão, entende-se a escolha que alguém realiza entre, no mínimo, duas alternativas possíveis, utilizando o meio que julga ser o melhor disponível para atingir um determinado objetivo (CORRAR; THEÓPHILO, 2004).

A utilidade do estudo será analisado um balanço simplificado e a partir da posição inicial da empresa far-se-ão simulações dos trajetos que a empresa poderá seguir e quais os resultados de cada um deles. Assim esse estudo será focado apenas nas decisões financeiras que envolvem investimentos e aplicações de recursos, apesar de entender-se que este procedimento possa ser aplicado em outros tipos de decisões empresariais, como as decisões operacionais. Serão descritos os cenários e suas probabilidades de ocorrência, desenvolvendo se, com isso, as alternativas de decisão para atingira meta proposta.

Uma arvore de decisão utiliza uma estratégia de *dividir para-conquistar*: Um problema complexo é decomposto em sub-problemas mais simples. Recursivamente a mesma estratégia é aplicada a cada subproblema. A capacidade de discriminação de uma arvore vem da: Divisão do

espaço definido pelos atributos em subespaços. A cada sub-espaço é associada uma classe. Crescente interesse CART (Breiman, Friedman, et.al.) C4. 5 (Quinlan) Splus, Statistica, SPSS. (João Gama, 2002)

O que é uma Arvore de Decisão?

Árvores de Decisão são métodos de classificação de dados no contexto da chamada Mineração de Dados (*Data Mining*). Podem ser usadas em conjunto com a tecnologia de indução de regras, mas são as únicas a apresentar os resultados hierarquicamente (com priorização). Nelas, o atributo mais importante é apresentado na árvore como o primeiro nó, e os atributos menos relevantes são mostradas nos nós subsequentes. A vantagem principal das Árvores de Decisão é a **tomada de decisões** levando em consideração os atributos mais relevantes, além de compreensíveis para a maioria das pessoas. Ao escolher e apresentar os atributos em ordem de importância, as Árvores de Decisão permitem aos usuários conhecer quais fatores mais influenciam os seus trabalhos.

De acordo com RAGSDALE (2001), a árvore de decisão é composta por nós (representados por círculos e quadrados) interconectados por ramos (representados por linhas). Um nó quadrado é chamado de nó de decisão porque representa uma decisão. Ramos emergindo do nó de decisão representa as diferentes alternativas para uma decisão particular.

Segundo Garcia (2000), as Árvores de Decisão também consistem de: nodos (nós), que representam os atributos, e de arcos (ramos), provenientes desses nodos e que recebem os valores possíveis para esses atributos (cada ramo descendente corresponde a um possível valor desse atributo). Nas árvores existem nodos folha (folha da árvore), que representam as diferentes classes de um conjunto de treinamento, ou seja, cada folha está associada a uma classe. Cada percurso na árvore (da raiz à folha) corresponde a uma regra de classificação.

Cada nó de decisão contém um teste num atributo, cada ramo descendente corresponde a um possível valor deste atributo, cada folha está associada a uma classe, cada percurso na arvore (da raiz à folha) corresponde a uma regra de classificação, cada folha corresponde a uma região, hiper-retângulo, a intersecção dos hiper-retângulos é vazio e a união dos hiper-retângulos é o espaço completo. (João Gama, 2002)

Árvores de decisão – Exemplo 1:

Uma companhia enfrenta um problema relacionado com um produto (Px) desenvolvido por um dos seus laboratórios de pesquisa. Têm que decidir se prosseguem com o teste de mercado deste produto ou se simplesmente descontinuam o seu desenvolvimento. Foi estimado que o teste de mercado custará 100 mil euros. A experiência indica que apenas 30% dos produtos passam no teste de mercado.

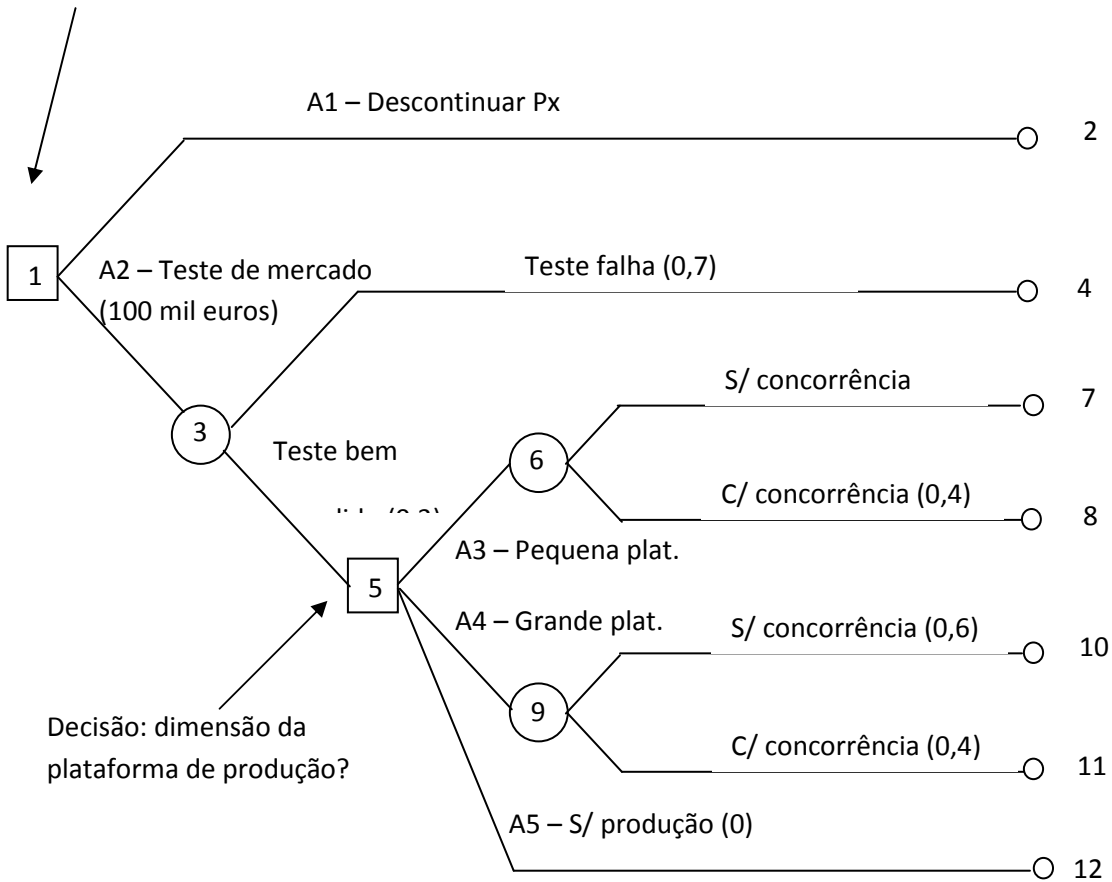
Se Px passar no teste de mercado, a companhia terá que enfrentar uma nova decisão relacionada com as dimensões da plataforma de produção do produto. Uma pequena plataforma custa 150 mil euros a construir e permite a produção de 2000 unidades/ano, enquanto uma plataforma maior custa 250 mil euros e permite a produção de 4000 unidades/ano.

O departamento de mercado estimou que existe 40% de probabilidade da concorrência responder com um produto similar e que o preço por unidade vendida será o seguinte (em euros):

	Plataforma grande	Plataforma pequena
Concorrência responde	20	35
Concorrência não responde	50	65

Assumindo que a vida de mercado para o produto Px está estimada em 7 anos e que os custos de funcionamento das plataformas é de 50 mil euros/ano, deve a companhia seguir em frente com o teste de mercado de Px?

Decisão: teste de mercado?



Passo 1

Para cada caminho entre o nó inicial e cada um dos nós terminais vamos calcular o lucro associado (da esquerda para a direita). Valores em milhares de euros.

	Nó 2	Nó 4	Nó 7	Nó 8	Nó 10	Nó 11	Nó 12
Benefícios	0	0	910	490	1400	560	0
Custos	0	100	600	600	700	700	100
Lucros	0	-100	310	-110	700	-140	-100

Passo 2

Vamos agora considerar as probabilidades introduzidas no problema, trabalhando da direita para a esquerda da árvore de decisão.

Nó 6

Ramo para 7 – probabilidade 0,6 e lucro total 310

Ramo para 8 – probabilidade 0,4 e lucro total -110

Valor monetário esperado:

$$0,6 * (310) + 0,4 * (-110) = 142 \text{ mil euros}$$

Nó 9

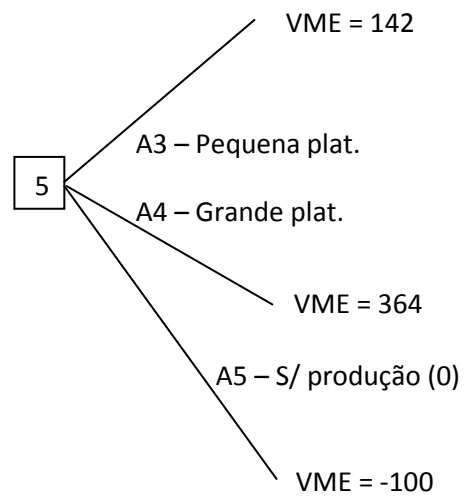
Ramo para 10 – probabilidade 0,6 e lucro total 700

Ramo para 11 – probabilidade 0,4 e lucro total -140

Valor monetário esperado:

$$0,6 * (700) + 0,4 * (-140) = 364 \text{ mil euros}$$

Substituindo os nós de mudança pelos correspondentes VME:



Assim, no nó 5 temos 3 alternativas possíveis:

1 – Construir uma plataforma pequena VME = 142

2 – Construir uma plataforma grande VME = 364

3 – Não construir nenhuma plataforma VME = -100

A alternativa 2 é a melhor. Vamos repetir este processo para o nó de decisão 1 considerando apenas esta alternativa.

Nó 3

Ramo para 4 – probabilidade 0,7 e lucro total -100

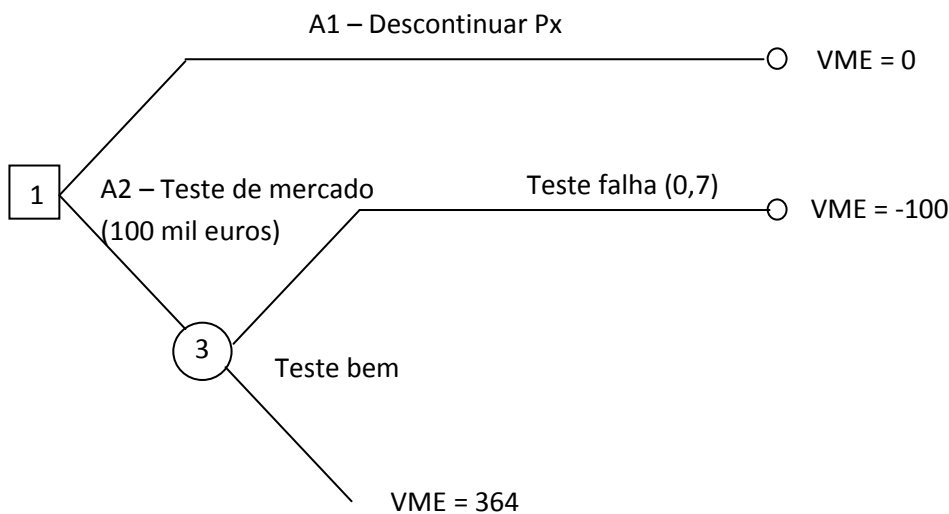
Ramo para 5 – probabilidade 0,3 e lucro total 364

Valor monetário esperado:

$$0,7 * (-100) + 0,3 * (364) = 39,2 \text{ mil euros}$$

Assim, no nó de decisão 1 temos duas alternativas possíveis:

- 1- Descontinuar o produto VME = 0
- 2- 2- Proceder ao teste de mercado VME = 39,2



A árvore de decisão serve para representar um processo de decisão no tempo em que o decisor deve efetuar uma decisão. Para Gomes *et al.* (2002), as etapas da técnica para o auxílio da tomada de decisão são definidas como: a) definição do tema; b) definição do objetivo, metas e submetas; c) construção da árvore de decisão; d) revisão da árvore de decisão; e) encerramento.

Árvores de decisão - Exemplo Dois:

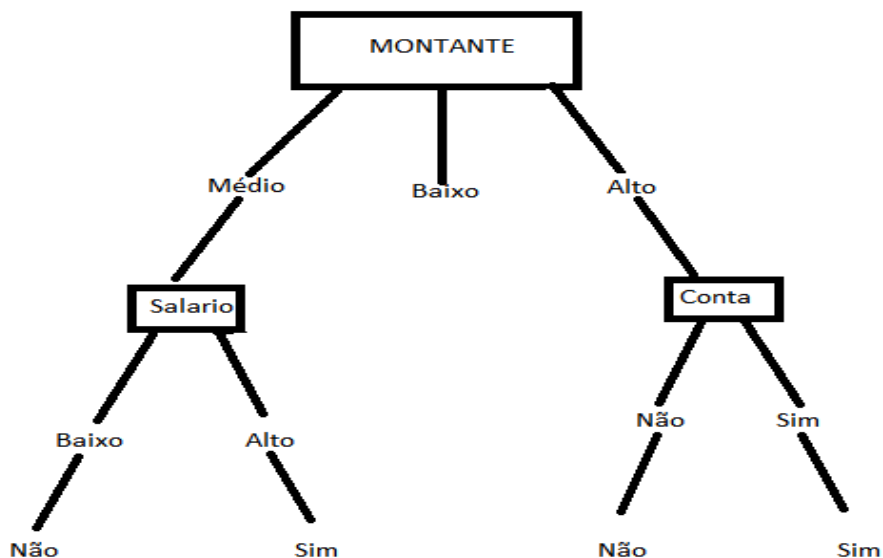
Quinlan (1993) desenvolveu uma técnica que permitiu o uso da representação do conhecimento por meio das Árvores de Decisão. Sua contribuição consistiu na elaboração de um algoritmo chamado ID3 que, juntamente com suas evoluções (ID4, ID6, C 4.5, See 5), é uma ferramenta adequada ao uso da referida técnica.

As Árvores de Decisão são representações simples do conhecimento e um meio eficiente de construir classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados (GARCIA, 2000).

A figura abaixo apresenta um exemplo de Árvore de Decisão, na qual constam dados que relatam as condições para uma pessoa receber um empréstimo. Nesse caso existem duas possíveis classes: Sim (receber empréstimo) e Não (não receber empréstimo). Os atributos são montante, salário e conta.

O atributo Montante pode assumir os Valores de Médio, Alto ou Baixo; o atributo Salário pode assumir Valor Baixo ou Valor Alto; e o atributo Conta pode ser Sim ou Não. Alguns dados são exemplos da classe Sim, ou seja, os requisitos exigidos por um banco a uma pessoa para a concessão de um empréstimo são satisfatoriamente preenchidos. Outros são da classe Não, isto é, os requisitos exigidos não são plenamente satisfeitos. A classificação, nesse caso, resulta numa estrutura de árvore, que pode ser usada para todos os objetos do conjunto (BRADZIL, 1999).

Figura : Exemplo de uma Árvore de Decisão Simples



É possível derivar regras de uma Árvore de Decisão, com o intuito de facilitar a leitura e a compreensão por parte do usuário. Assim, as Árvores de Decisão podem ser representadas como conjuntos de regras do tipo **se-então** (*if-then*). As regras são escritas considerando o trajeto do nó raiz até uma folha da árvore. Árvores de Decisão e Regras de Classificação são métodos geralmente utilizados em conjunto. Devido ao fato de as Árvores de Decisão tenderem em crescer muito, como mostram algumas aplicações, elas são muitas vezes substituídas pelas regras. Isso acontece em virtude dessas últimas poderem ser facilmente modularizadas. Uma regra pode ser compreendida sem que haja a necessidade de referenciar-se a outras (INGARGIOLA, 1996).

Com base na Árvore de Decisão apresentada na figura acima, pode-se exemplificar a derivação de regras. Dois exemplos de regras obtidas a partir dessa árvore são mostrados a seguir:

- **Se** montante = médio **e** salário = baixo **então** classe = não.
- **Se** montante = médio **e** salário = alto **então** classe = sim.

Muitos são os algoritmos de classificação que utilizam a representação sob o formato de Árvores de Decisão (WITTEN e FRANK, 2000). O algoritmo ID3, cuja criação se baseou em sistemas de inferência e em conceitos de sistemas de aprendizagem, foi um dos primeiros de Árvore de Decisão. Logo após, foram elaborados diversos outros algoritmos, sendo os mais conhecidos o C4.5, o CART (*Classification and Regression Trees*) e o CHAID (*Chi Square Automatic Interaction Detection*), além de outros (GARCIA, 2000).

A utilização de Árvores de Decisão apresenta as seguintes vantagens: não assumem nenhuma distribuição particular para os dados; as características ou os atributos podem ser categóricos (qualitativos) ou numéricos (quantitativos); pode-se construir modelos para qualquer função desde que o número de exemplos de treinamento seja suficiente; elevado grau de compreensão. Após a construção de uma Árvore de Decisão, é importante avaliá-la por meio da utilização de dados que não tenham sido usados no treinamento. Essa estratégia permite estimar como a árvore generaliza os dados e adapta-se a novas situações, além de determinar a proporção de erros e acertos ocorridos na construção da árvore (BRADZIL, 1999).

Confronto entre Árvore de decisão e Maxver

Maxver é a técnica de classificação supervisionada mais popular para tratamento de dados de satélites. Este método é baseado no princípio de que a classificação errada de um pixel particular não tem mais significado do que a classificação incorreta de qualquer outro pixel na imagem (Richards, 1993). O usuário determina a significância nos erros de atributos especificados para uma classe em comparação a outras.

A eficácia do Maxver depende, principalmente, de uma precisão razoável da estimativa do vetor médio e da matriz de covariância de toda classe espectral. Isso depende da quantidade de pixels incluídos nas amostras de treinamento. O resultado do Maxver é tanto melhor quanto maior o número de pixels numa amostra de treinamento para implementá-los na matriz de covariância (Previdelli, 2004).

Nas classificações em que se usa o algoritmo Maxver, cada classe é determinada por um modelo de classes gaussiano e multivariado, capaz de descrever a distribuição das classes no espaço de atributos. Este modelo leva em conta as extensões de dados em direções espectrais determinadas. No espaço RGB, elipsóides concêntricas de equi-probabilidade, cuja localização, forma e tamanho refletem a média, variância e covariância da classe, determinam a probabilidade do pixel pertencer à classe.

Crósta (1993) considera que o método Maxver deve ser aplicado quando o analista conhece bem a imagem a ser classificada, para que possa definir classes que sejam representativas.

Árvore de decisão

Uma árvore de decisão tem a função de particionar recursivamente um conjunto de treinamento, até que cada subconjunto obtido deste particionamento contenha casos de uma única classe. Para atingir esta meta, a técnica de árvores de decisão examina e compara a distribuição de classes durante a construção da árvore. O resultado obtido, após a construção de uma árvore de decisão, são dados organizados de maneira compacta, que são utilizados para classificar novos casos (Holsheimer & Siebes, 1994; Brazdil, 1999).

A árvore de decisão não assume nenhum modelo estatístico *a priori*, sendo a divisão do espaço de atributos feita de acordo com as amostras provenientes do treinamento. Assim, no espaço RGB, as classes são separadas por planos que as delimitam.

Maxver × Árvore de Decisão

Neste trabalho, destacamos a hipótese de que a classificação por árvore de decisão pode obter um melhor resultado que a classificação por Maxver, para os casos em que os pixels das classes não são “bem comportados”, isto é, não obedecem a uma distribuição gaussiana, como nos casos em que existem duas respostas espectrais distintas para uma mesma classe ou a “fronteira” das classes se emaranham no espaço de atributo. Algumas possíveis configurações, num plano bidimensional, que poderiam ser melhor classificadas pela árvore de decisão do que por Maxver são esquematizadas abaixo (**Figura 1**):

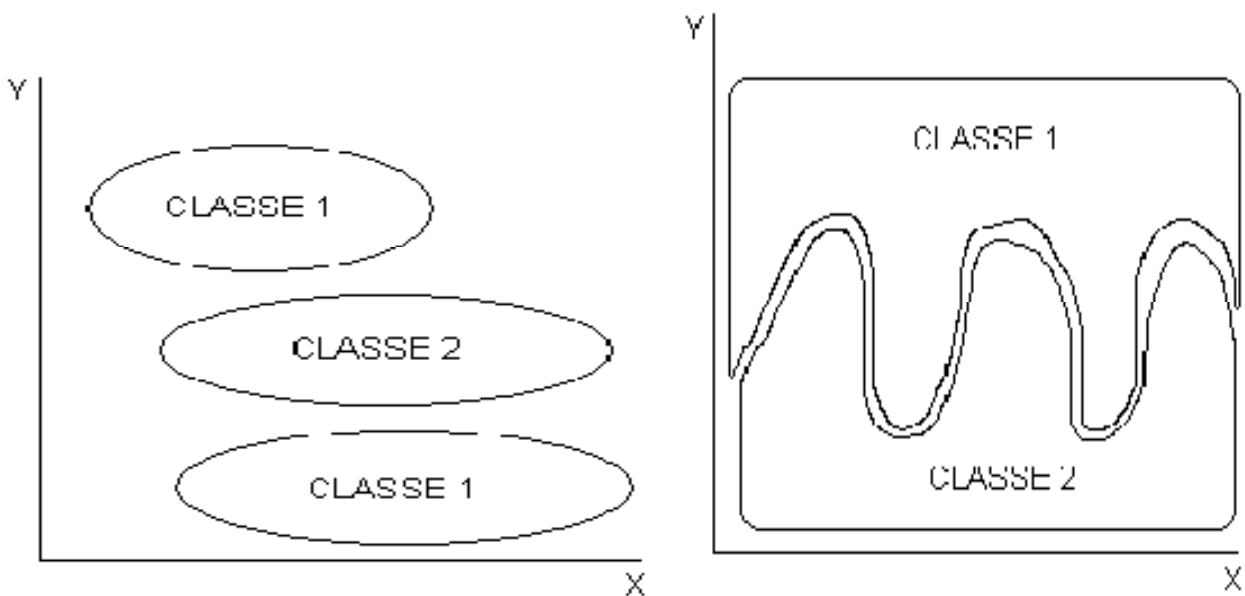


Figura 1: Configurações onde árvore de decisão classifica com melhor acuracidade que Maxver.

IMG

O sistema IMG consiste de:

- a) Biblioteca-IMG: um conjunto de classes, rotinas e “operadores sobrecarregados” para processamento de imagens (PI) escrito em C++;

b) Programas-IMG: um conjunto de programas para PI escritos utilizando as rotinas acima.

A finalidade do Sistema IMG é fornecer meios para testar rapidamente novos algoritmos de PI. A sua interface com o usuário é pouco amigável e os comandos são dados na janela de prompt do DOS (comandos de linha).

O algoritmo de árvore de decisão da biblioteca do sistema foi desenvolvido para trabalhar com imagens tridimensionais, através do modelo de cores RGB, que pode ser encontrada em qualquer tecnologia que utilize o modelo aditivo. No processamento de imagens, através desta biblioteca, os componentes da imagem digital colorida é fracionado nos três canais de cores. Portanto, neste sistema, as imagens de ensorimento remoto podem ser trabalhadas em 3 bandas, uma em cada canal.

CONSIDERAÇÕES FINAIS

A árvore de decisão é muito importante para a tomada de uma decisão, demonstra elevado grau de comprometimento. É usada muito em empresas, entretanto também pode-se utilizá-la em outros lugares, como na vida particular.

Portanto pode-se concluir que a árvore de decisão esta englobada em todo âmbito empresarial. O ambiente empresarial requer decisões rápidas e precisas, para tanto existem vários métodos de auxílio. Neste estudo, é considerada a integração de dois desses métodos: a análise de balanços e o modelo de árvores de decisão. Para demonstrar a forma de utilização do ferramental conceitual foi desenvolvido um estudo de simulação numérica considerando o aspecto financeiro de uma empresa hipotética.

A partir das demonstrações contábeis desenvolvidas apenas com o intuito de provar que mesmo uma empresa rentável pode ser otimizada, foram obtidos índices relativos à situação da empresa, sendo esses dados transferidos para o gráfico de árvores de decisão, através do qual foram identificados os possíveis caminhos alternativos para que a empresa pudesse cumprir determinados objetivos estabelecidos e, ainda, melhorá-los.

A aplicação do modelo de árvore de decisão na análise dos índices econômico-financeiros facilita o entendimento, uma vez que o administrador trabalha com informações resumidas em indicadores gerenciais e com recursos gráficos. Essa ferramenta, além disso, indica o aspecto “pró-ativo” da empresa, pois estimula a busca continua de melhorias para a sua posição.

REFERÊNCIAS

GODET, Michel. Manual de prospectiva e estratégica: da antecipação à ação. Lisboa: Dom Quixote, 1993.

GUERREIRO, Reinaldo. Modelo Conceitual de Sistema de Informação de Gestão Econômica: Uma contribuição à teoria da comunicação da contabilidade. Tese de doutoramento. Departamento de Contabilidade FEA-USP. São Paulo: USP, 1989.

LACHTERMACHER, Gerson. Pesquisa operacional na tomada de decisões: modelagem em Excel. Rio de Janeiro: Campus, 2002.

BRAZDIL, P. Construção de Modelos de Decisão a partir de Dados. Disponível por WWW em: <http://www.ncc.up.pt/~pbrazdil/Ensino/ML/DecTrees.html>, 1999.

BREIMAN, L.; FRIEDMAN, J.H.; OLSHEN, R.A. **Stone, Classification And Regression Trees**, Wadsworth, 1984. R.Adm., São Paulo, v.40, n.3, p.225-234, jul./ago./set. 2005

CARVALHO, A.P. de L.F. de. *Redes neurais artificiais*. 2000. Disponível em: <http://www.icmc.sc.usp.br/~andre/neural1.html>

GARCIA, S.C. *O uso de árvores de decisão na descoberta de conhecimento na área da saúde*. In: SEMANA ACADÊMICA, 2000. Rio Grande do Sul: Universidade Federal do Rio Grandedo Sul, 2000.

INGARGIOLA, G. *Building classification models: ID3 and C4.5*. 1996. Disponível em: <http://www.cis.temple.edu/~ingargio>

QUINLAN, J.C. *C4.5: programs for machine learning*. San Mateo: Morgan Kaufmann, 1993. 302p.

http://www.ccog.up.pt/~jgama/Aulas_ECD/arv.pdf

<http://www.revistasusp.sibi.usp.br/pdf/rege/v15n3/v15n3a7.pdf>

<http://ascot1.dominiotemporario.com/doc/A9.pdf>

http://www.congressosp.fipecafi.org/artigos12004/an_resumo.asp?cod_trabalho=424

<http://www.congressosp.fipecafi.org/artigos12004/424.pdf>

<http://aimotion.blogspot.com/2009/04/artigo-introducao-arvores-de-decisao.html>

<http://www.liaad.up.pt/~jgama/Bdc/arv.pdf>