

# Árvore de Decisão Aplicada em Base de Dados de Sistemas Industriais Automatizados

Iangelles Sad Melo<sup>1</sup>, Michelli Marlane da Silva<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação - Universidade Presidente Antônio Carlos (UNIPAC)  
Rua Palma Bageto Viol S/N - Barbacena - MG - Brasil

{rabies\_55, michelli\_marlane}@hotmail.com

**Abstract.** *Currently the production processes are capable of generating a very large volume of data, stored in these databases that can reach the house of terabytes, known as Data Warehouse. This paper discusses the process of KDD (Knowledge Discovery in Database Knowledge Discovery in Database), comprised mainly by Data Mining (Data Mining) applying the technique Decision Tree in a production database, generated by a industrial automated system. Aiming the extraction of this knowledge base.*

**Resumo.** *Atualmente os processos produtivos são capazes de gerar um volume de dados muito grande, estes armazenados em bancos de dados que podem alcançar a casa de terabytes, denominados como Data Warehouse. O presente trabalho aborda o processo de KDD (Knowledge Discovery in Database Descoberta de Conhecimento em Banco de Dados), compreendido principalmente pelo Data Mining (Mineração de Dados) aplicando a técnica de Árvore de Decisão em uma base de dados produção, gerada por um sistema industrial automatizado. Tendo como objetivo a extração do conhecimento desta base.*

## 1. Introdução

De acordo com Pierre Levy (Levy, 1996) o conhecimento é o processo pelo qual a realidade se reflete e se reproduz no pensamento humano, com o propósito de alcançar a verdade no seu valor como prova para se afirmar que se conhece alguma coisa, ou seja, uma atitude embasada num conceito que obtém um resultado previamente esperado. O processo de geração do conhecimento está mais ligado a transmissão deste entre indivíduos do que a geração pelo próprio.

O processo de gestão do conhecimento demanda um número constante de decisões acerca das atividades presentes em uma organização, tendo como base para essas decisões o conhecimento adquirido com a experiência profissional, impactando diretamente em toda a cadeia produtiva da organização. Para auxiliar nesse gerenciamento, ferramentas computacionais geradoras de novos conhecimentos são essenciais, como exemplo o processo de *KDD (Knowledge Discovery in Database Descoberta de Conhecimento em Bancos de Dados)*. No processo de KDD está compreendida uma etapa denominada *Data Mining (Mineração de Dados)*. Nesta etapa são utilizadas técnicas para extração dos padrões de conhecimento. Uma destas técnicas, a Árvore de Decisão, será utilizada neste trabalho.

## 2. Motivação

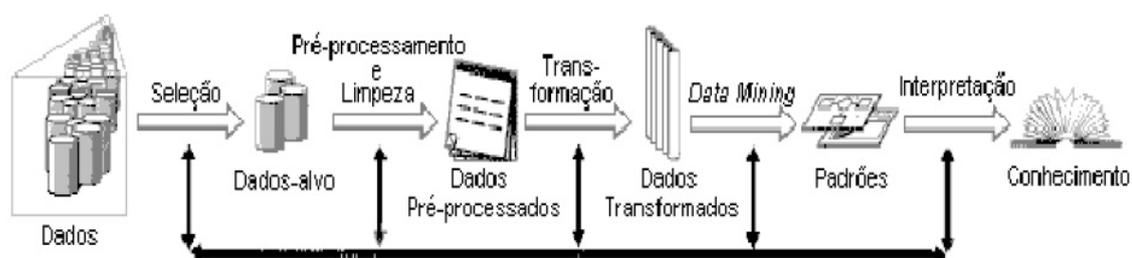
Um exemplo para o emprego da descoberta de conhecimento, são os processos produtivos automatizados onde o volume de informações geradas é muito grande. Estas informações geralmente são armazenadas em bancos que muitas das vezes ficam na casa de *terabytes*. A automatização de um processo geralmente tem como consequência a redução de efetivo nas empresas, centralizando a operação dos equipamentos em uma única pessoa. Fazendo com que ocorrências significativas passem despercebidas, e as devidas atitudes não sejam tomadas, podendo acarretar em danos a estrutura física e/ou ao processo da empresa.

Atualmente as plantas industriais automatizadas podem atingir altos níveis de complexidade podendo ter em alguns casos até 15000 pontos de comunicação (troca de dados entre sinais de campo e controladores, denominados na leitura técnica e comercial de *tags*). Sistemas de automação demasiadamente complexos dificilmente podem ser supervisionados de uma forma eficiente por um único operador (Kaskurewicz et al., 1997).

## 3. Metodologia

Para a construção da base de conhecimentos do sistema será utilizado o processo de Descoberta do Conhecimento em Bases de Dados (*KDD - Knowledge Discovery in Databases*). O processo de Descoberta do Conhecimento em Bases de Dados é composto por várias etapas, que são: seleção, pré-processamento e limpeza, transformação, *data mining* e interpretação dos resultados ([AVI1998]).

A figura 1 mostra o processo de descoberta de conhecimento em banco de dados (KDD).



**Figura 1. Processo de descoberta de conhecimento em banco de dados KDD.**  
Fonte (Gimenes, 2000)

Como sugere Kamber (Kamber, 2001) KDD corresponde a todo processo tendo *Data Mining* como uma de suas etapas. A seguir descrição de cada etapa segundo Kamber:

- Limpeza de dados: remoção de dados inconsistentes;
- Integração de dados: fase em que os dados pesquisados provindos de vários lugares são combinados;
- Seleção de dados: fase onde dados relevantes para análise são recuperados das bases de dados;
- Transformação de dados: fase onde os dados são transformados em um formato apropriado para a mineração de dados;

- *Data Mining*: fase onde métodos inteligentes são aplicados para a extração de padrões de dados;
- Validação de padrões: identificação dos padrões de representação do conhecimento realmente interessante, baseado em questões de interesse apresentadas previamente;
- Apresentação do conhecimento: onde a visualização e técnicas de representação do conhecimento são usadas para apresentar o conhecimento minerado para o usuário final.

Como pode-se perceber, as etapas de KDD descritas por Gímenes e Kamber são relativamente parecidas, na maioria das vezes iguais, porém nomeadas e interpostas de maneiras diferentes.

#### 4. *Data Mining*

*Data Mining* é uma tecnologia que emergiu da intersecção de três áreas: estatística clássica, inteligência artificial e aprendizado de máquina, sendo a primeira a mais antiga delas.

O *Data Mining* é uma ferramenta usada para buscar e extrair informações úteis de grandes volumes de dados, em um processo chamado descoberta de conhecimento em banco de dados (KDD). Inclui tarefas conhecidas, como extração de conhecimento, arqueologia de dados, dragagem de dados e colheita de dados. Todas essas atividades são realizadas automaticamente e permitem uma rápida descoberta, mesmo por quem não é programador. Os dados muitas vezes estão soterrados em grandes bancos de dados, *datawarehouse*, documento de texto, informações e conhecimentos reunidos ao longo de muitos anos. Para Kimball (Kimball, 1996) *DataWarehouse* define-se assim: Conjunto de ferramentas e técnicas de projeto, que quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos permitirá que planejem e construam um *DataWarehouse*..

A figura 2 a seguir, proposta por DalAlba (DalAlba, 1999), mostra algumas das principais características comparativas entre banco de dados operacionais e *DataWarehouse*.

Características	Bancos de dados Operacionais	<i>DataWarehouse</i>
Objetivo	Operações diárias do negócio	Analisar o negócio
Uso	Operacional	Informativo
Unidade de trabalho	Inclusão, alteração, exclusão.	Carga e consulta
Tipo de usuário	Operadores	Comunidade gerencial
Condições dos dados	Dados operacionais	Dados Analíticos
Volume	Megabytes – gigabytes	Gigabytes – terabytes
Histórico	60 a 90 dias	5 a 10 anos
Estrutura	Estática	Variável
Manutenção desejada	Mínima	Constante
Acesso a registros	Dezenas	Milhares
Atualização	Continua (tempo real)	Periódica (em <i>batch</i> )
Integridade	Transação	A cada atualização
Número de índices	Poucos/simples	Muitos/complexos
Intenção dos índices	Localizar um registro	Aperfeiçoar consultas

**Figura 2. Comparação entre Banco de Dados Operacionais e *DataWarehouse***  
 Fonte (DalAlba, 1999)

## 5. Tarefas Realizadas por Técnicas de Mineração de Dados

Denominando-se uma tarefa como um tipo de problema de descoberta de conhecimento a ser solucionado, nestas tarefas podem ser aplicadas técnicas de mineração de dados. Essas tarefas são apresentadas de forma resumida na figura 3 (Dias, 2001).

Tarefa	Descrição	Exemplos
Classificação	Constrói um modelo de algum tipo que possa ser aplicado a dados não classificados a fim de categorizá-los em classes, o objetivo é descobrir um relacionamento entre um atributo meta (cujo valor será previsto) e um conjunto de atributos de previsão	Classificar pedidos de crédito Esclarecer pedidos de seguros fraudulentos Identificar a melhor forma de tratamento de um paciente
Estimativa (ou Regressão)	Usada para definir um valor para alguma variável contínua desconhecida	Estimar o número de filhos ou a renda total de uma família Estimar o valor em tempo de vida de um cliente Estimar a probabilidade de que um paciente morrerá baseando-se nos resultados de diagnósticos médicos Prever a demanda de um consumidor para um novo produto
Associação	Usada para determinar quais itens tendem a ser adquiridos juntos em uma mesma transação	Determinar que produtos costumam ser colocados juntos em um carrinho de supermercado
Segmentação (ou Clustering)	Processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos	Agrupar clientes por região do país Agrupar clientes com comportamento de compra similar Agrupar seções de usuários Web para prever comportamento futuro de usuário
Sumarização	Envolve métodos para encontrar uma descrição compacta para um subconjunto de dados	Tabular o significado e desvios padrão para todos os itens de dados Derivar regras de síntese

**Figura 3. Tarefas realizadas por técnicas de mineração de dados.**  
Fonte (Dias, 2001)

A figura 4, segundo Dias (Dias, 2002), traz as técnicas de mineração de dados normalmente utilizadas associadas às tarefas, conforme descritas a cima.

Técnica	Descrição	Tarefas	Exemplos
Descoberta de Regras de Associação	Estabelece uma correlação estatística entre atributos de dados e conjuntos de dados	Associação	Apriori, AprioriTid, AprioriHybrid, AIS, SETM (Agrawal e Srikant, 1994) e DHP (Chen <i>et al.</i> , 1996).
Árvores de Decisão	Hierarquização dos dados, baseada em estágios de decisão (nós) e na separação de classes e subconjuntos	Classificação Regressão	CART, CHAID, C5.0, Quest (Two Crows, 1999); ID-3 (Chen <i>et al.</i> , 1996); SLIQ (Metha <i>et al.</i> , 1996); SPRINT (Shafer <i>et al.</i> , 1996).
Raciocínio Baseado em Casos ou MBR	Baseado no método do vizinho mais próximo, combina e compara atributos para estabelecer hierarquia de semelhança	Classificação Segmentação	BIRCH (Zhang <i>et al.</i> , 1996); CLARANS (Chen <i>et al.</i> , 1996); CLIQUE (Agrawal <i>et al.</i> , 1998).
Algoritmos Genéticos	Métodos gerais de busca e otimização, inspirados na Teoria da Evolução, onde a cada nova geração, soluções melhores têm mais chance de ter "descendentes"	Classificação Segmentação	Algoritmo Genético Simples (Goldberg, 1989); Genitor, CHC (Whitley, 1993); Algoritmo de Hillis (Hillis, 1997); GA-Nuggets (Freitas, 1999); GA-PVMINER (Araújo <i>et al.</i> , 1999).
Redes Neurais Artificiais	Modelos inspirados na fisiologia do cérebro, onde o conhecimento é fruto do mapa das conexões neuronais e dos pesos dessas conexões	Classificação Segmentação	Perceptron, Rede MLP, Redes de Kohonen, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB (Azevedo, 2000), (Braga <i>et al.</i> , 2000), (Haykin, 2001)

**Figura 4. Técnicas de mineração de dados.**Fonte (Dias, 2002)

A mineração de dados pode ser aplicada de duas formas: como um processo de verificação e como um processo de descoberta (Groth, 1998). No processo de verificação o usuário sugere uma hipótese e tenta prová-la com utilização de técnicas de análise estatística e multidimensional. Já no processo de descoberta não há suposição antecipada e utiliza técnicas como árvore de decisão, redes neurais, algoritmos genéticos.

Segundo Harrison (Harrison, 1998), a escolha da técnica de mineração de dados dependerá da tarefa específica a ser executada e dos dados disponíveis para análise. Berry e Linoff (1997) sugerem que a seleção das técnicas de mineração de dados deve ser dividida em dois passos:

- Traduzir o problema de negócio a ser resolvido em séries de tarefas de mineração de dados;
- Compreender a natureza dos dados disponíveis em termos de conteúdo e tipos de campos de dados e estrutura das relações ente os registros.

Na definição da técnica de mineração de dados, podem ser considerados os seguintes parâmetros:

- Tipo de problema de descoberta de conhecimento a ser solucionado;
- Característica dos dados;
- Forma de aplicação da mineração de dados;
- Disponibilidade de ferramenta de mineração de dados.

## 6. Árvore de Decisão

A árvore de decisão consiste de uma hierarquia de nós internos e externos que são conectados por ramos. O nó interno, também conhecido como nó decisório ou nó intermediário, é a unidade de tomada de decisão que avalia através de teste lógico qual será o próximo nó descendente ou filho. Em contraste, um nó externo (não tem nó descendente), também conhecido como folha ou nó terminal, está associado a um rótulo ou a um valor.

Em geral, o procedimento de uma árvore de decisão é o seguinte: apresenta-se um conjunto de dados ao nó inicial (ou nó raiz que também é um nó interno) da árvore; dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos e este procedimento é repetido até que um nó terminal é alcançado. A repetição deste procedimento caracteriza a recursividade da árvore de decisão.

Na figura 5 [MACHADO] pode-se visualizar um modelo de árvore de decisão assim como na figura 6 [MACHADO] seu pseudocódigo.

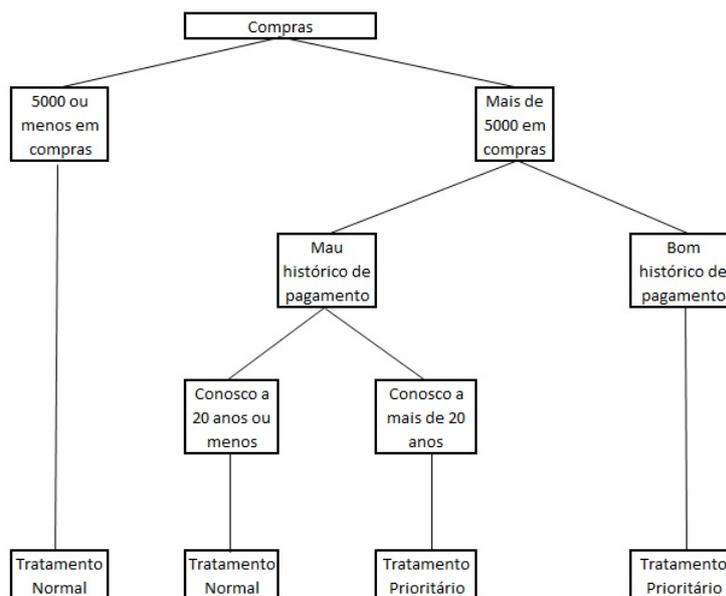


Figura 5. Modelo de árvore de decisão.

```

Se Compras_Cliente > 5 000
  Se Hist_Pagamento = "Bom"
  Então
    Tratamento_Cliente = "Prioritário"
  Senão
    Se Idade_Cliente > 20
    Então
      Tratamento_Cliente = "Prioritário"
    Senão
      Tratamento_Cliente = "Normal"
  Senão
    Tratamento_Cliente = "Normal"

```

**Figura 6. Pseudocódigo.**

## 7. Aplicação das Fases

### 7.1. Base de dados

A definição da base de dados é essencial, como a técnica definida para este trabalho é a árvore de decisão, a base necessitaria ter seus dados não classificados mas, possíveis de serem categorizados. A base de dados inicial era composta pelos seguintes campos:

- Data: campo que armazenava o dia referente aos registros;
- Base Úmida: registro do volume de material ainda úmido;
- Umidade: percentual de umidade encontrado no material;
- Base Seca: registro do volume de material, já aplicado o fator de umidade;
- Base Úmida Combustível: registro do volume de combustível ainda úmido;
- Umidade: percentual de umidade encontrado no combustível;
- Orçado kWh/t: valor orçado de consumo de energia;
- Combustível: quantidade de combustível utilizado, já aplicado o fator de umidade;
- Tempo de Funcionamento: total de horas de funcionamento dentro do período de 24h;
- Orçado Rendimento: valor de t/h orçado;
- Realizado Rendimento: valor de t/h alcançado;
- Acumulado: valor produção acumulado.

### 7.2. Limpeza e Seleção de Dados

Para início, foi necessária a aplicação da limpeza dos dados bem como seleção dos mesmos. Assim dados inconsistentes e valores zerados foram expurgados. Em seguida foi feita a definição dos atributos, excluindo aqueles que não apresentavam dados irrelevantes como valores orçados, acúmulo de produção.

Entrando na fase de transformação dos dados, fez-se necessário o estudo um estudo mais aprofundado da base para uma melhor dragagem dos dados. Em consequência mais alguns dados foram retirados e uma classe foi criada para identificação de parâmetros.

### 7.3. Transformação dos Dados

Com a base já moldada, nesta permaneceu os seguintes atributos: base úmida do material, umidade do material, base seca do material, base úmida do combustível, base seca do combustível, umidade de combustível, horas de funcionamento. A classe criada, baseou-se no volume de produção horária, onde, valores a cima de 90 foram classificados como classe "A", entre 90 e 80 classe "B" e abaixo de 80 classe "C". Esta base continha num total de 327 instâncias.

### 7.4. Mineração dos Dados

Para etapa de mineração dos dados a técnica de árvores de decisão foi aplicada com o uso do software *Knime* versão 2.0.2. Este utiliza o algoritmo C4.5, mostrado com um pseudocódigo conforme figura 7 [BARBOSA AND CARNEIRO AND TAVARES] seguir.

```
1   if CRITERIOPARADA(exemplos)
2       ESCOLHECLASSE(exemplos)
3   else
4       melhor = ESCOLHEATRIBUTO(subAtributos; exemplos)
5       arvore =nova arvore com nó raiz= melhor
6       particao = ESCOLHEPARTICAO(melhor)
7       while particao
8           exp =elementos de exemplos com melhor= p
9           subAvr = INDUCAOCARTEC4.5(exp; subA melhor)
10          ADICIONARAMOARVORE(p; subAvr)
11
12  PODAARVORE(arvore)
```

Figura 7. Pseudocódigo do algoritmo C4.5.

Onde:

- EscolheAtributo(): faz uma busca gulosa, selecionando a característica que maximiza a divisão dos dados por meio de entropia <sup>1</sup>;
- EscolherPartição(): o algoritmo atribui a cada valor do atributo um ramo próprio;
- CriterioParada() o algoritmo só para de dividir se cada folha contém casos de uma única classe, ou até não ter como particionar mais porque os dois casos têm os mesmos valores para cada atributo, mas pertencem a classes diferentes;
- EscolhaClasse() o algoritmo atribui ao nó terminal a classe mais provável dentro dos exemplos;
- PodaArvore() o algoritmo usa a poda baseada no erro, ele permite utilizar o próprio conjunto de treino para efetuar a poda da árvore e tem a vantagem de não obrigar a separação do conjunto de treino, em conjunto de treino e conjunto de teste.

<sup>1</sup>A entropia caracteriza a impureza dos dados, num conjunto de dados, ela caracteriza a falta de homogeneidade dos dados de entrada em relação a sua classificação.

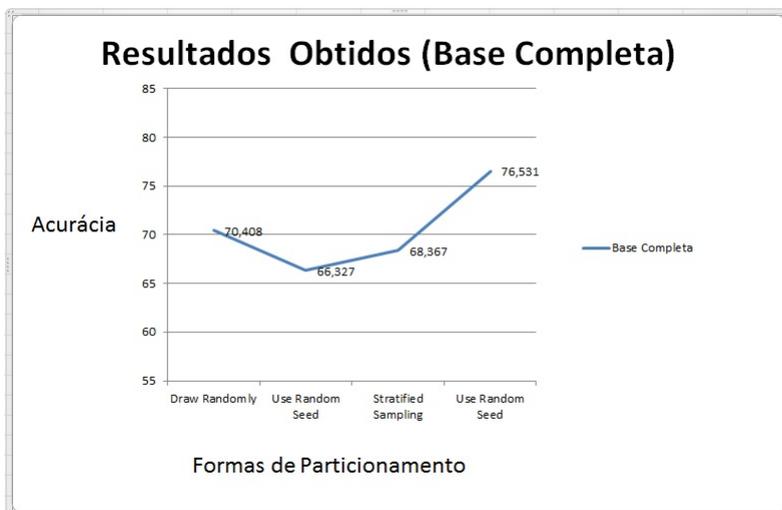
## 8. Testes

Para os testes realizados com a base de dados citada anteriormente, o parâmetros de treinamento da árvore de decisão foi de 70% dos registros. Para encontrar a melhor acurácia, a variação no número de registros por nó foi necessária, onde após vários testes os valores definidos foram 1, 2, 15 e 50 por poderem expressar de forma significativa a variação, conforme as tabelas a seguir.

Nas tabelas abaixo, a linha em negrito demonstram os melhores resultados, onde a quantidade mínima de registros por nó foi de valor 1. Nesta linha fica claro o maior percentual de acurácia nos testes realizados.

Os resultados obtidos a partir da base completa são apresentados abaixo na tabela e no seu gráfico respectivamente.

Base Completa				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>70,408</b>	<b>66,327</b>	<b>68,367</b>	<b>76,531</b>
2	71,429	62,327	65,306	74,49
15	60,204	52,041	61,224	67,347
50	55,102	54,082	53,061	57,143



A partir do resultado anterior, iniciaram-se testes para encontrar os registros que diretamente impactariam no percentual de acurácia.

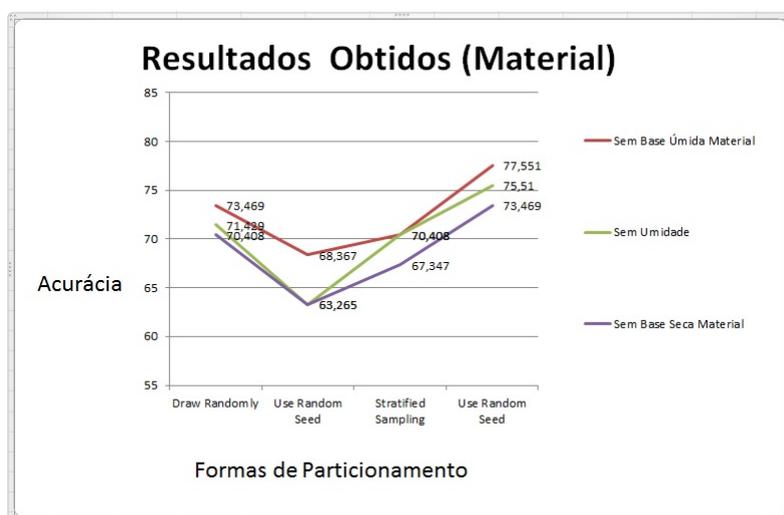
Para isso, simplesmente pela ordem que os registros encontravam-se na base, foram sendo tirados um a um e, posteriormente, gerado o resultado novamente. Após colhido os resultados o registro era retornado a base para retirada do seguinte.

As três primeiras colunas eram referentes ao material, onde podemos visualizar os resultados obtidos, conforme abaixo.

Sem Base Úmida Material				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>73,469</b>	<b>68,367</b>	<b>70,408</b>	<b>77,551</b>
2	67,347	62,245	69,388	73,469
15	53,061	52,041	67,347	67,347
50	53,061	54,082	56,347	57,143

Sem Umidade				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>71,429</b>	<b>63,265</b>	<b>70,408</b>	<b>75,51</b>
2	66,327	62,245	69,388	73,469
15	68,367	48,98	61,224	64,286
50	56,122	51,02	51,02	57,143

Sem Base Seca Material				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>70,408</b>	<b>63,265</b>	<b>67,347</b>	<b>73,469</b>
2	67,347	64,286	58,163	75,51
15	58,163	51,02	53,061	66,327
50	57,143	56,122	58,163	58,163



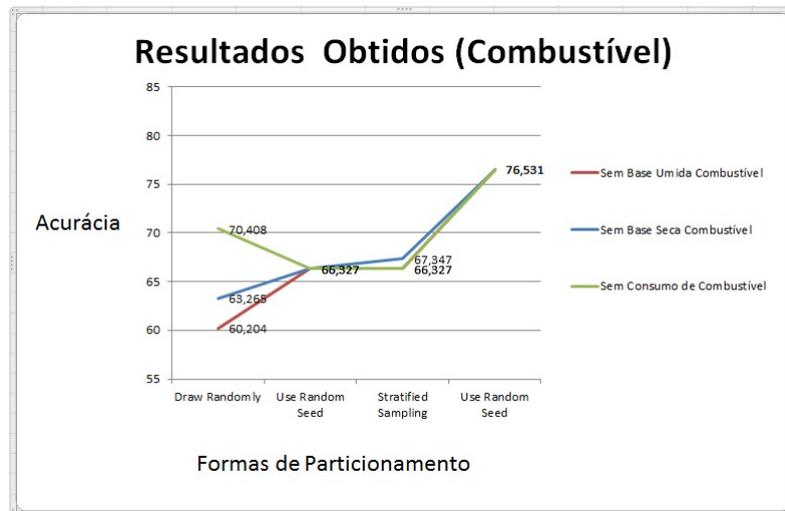
Com estes resultados, verifica-se que a retirada dos registros de Base Úmida Material e Umidade, a média dos percentuais de acurácia forma os melhores neste grupo.

O mesmo procedimento foi aplicado aos registros relativos ao combustível e seus resultados podem ser acompanhados a seguir.

Sem Base Umida Combustível				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>60,204</b>	<b>66,327</b>	<b>66,327</b>	<b>76,531</b>
2	60,204	62,245	63,265	74,490
15	57,143	52,041	66,327	67,347
50	57,143	54,082	60,204	57,143

Sem Base Seca Combustível				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>63,265</b>	<b>66,327</b>	<b>67,347</b>	<b>76,531</b>
2	64,286	62,245	68,367	74,490
15	52,041	52,041	62,245	67,347
50	44,898	54,082	53,061	57,143

Sem Consumo de Combustível				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>70,408</b>	<b>66,327</b>	<b>66,327</b>	<b>76,531</b>
2	71,429	62,245	67,347	74,490
15	64,286	52,041	62,245	67,347
50	56,122	54,082	53,061	57,143



Já nesta etapa, os resultados não foram tão bons ou significativos, pois a variação da acurácia foi baixa.

Observando e comparando os resultados obtidos até aqui, percebe-se que a extração dos valores relativos ao material poderia trazer uma melhor acurácia. Assim, foram retirados os registros que melhor percentual de acurácia demonstraram, porém ao menos um dado de volume de material na produção tem que permanecer.

Como os melhores resultados foram na retirada da Base Úmida do Material e Umidade, repetiu-se o procedimento anterior, mas agora, extraindo os dois registros simultaneamente. O que nos resulta nos seguintes dados:

Sem Base Úmida Material e sem Umidade				
Qtde Mínima de Reg. Por Nó	Draw Randomly	Use Random Seed	Stratified Sampling	Use Random Seed
	%	%	%	%
<b>1</b>	<b>78,571</b>	<b>81,633</b>	<b>76,531</b>	<b>66,327</b>
2	74,490	73,469	75,51	62,245
15	66,327	57,143	65,306	58,163
50	50,000	55,102	56,122	47,959



O gráfico final mostra o comparativo de acurácia quando gerados os resultados com a base completa e com a retirada dos registros de Base Úmida Material e Umidade. É notável a melhoria do resultado, mostrando que sem os valores dos registros retirados o aprendizado da árvore de decisão foi melhor.

Por fim, foi buscada uma associação do resultado às práticas reais vividas. Para tal, a citação da colaboradora Kelly Pereira Melo (Supervisora de Gestão de Produção):

”...o cimento é um produto fabricado a partir das misturas de matérias primas, cujas umidades são perdidas ao longo do processo. Como exemplo podemos citar a escória que passa um método de secagem antes de adicionada as demais. E é a partir da informação dos volumes sem umidade que se calcula a produção de cimento.”

## 9. Conclusão

Com os resultados, pode-se notar que os valores referentes à Base Úmida do Material e Umidade não são significativos tanto para análise do processo quanto para os testes realizados neste trabalho e refletiam diretamente no aprendizado da árvore de decisão. Após estes terem sido retirados e mantida as configurações como para todos os outros registros, obtendo resultados pertinentes ao esperado e buscado pelos nichos de produção similares ao utilizado neste trabalho.

Mas é importante o conhecimento da base de dados, ou seja, seu conteúdo deve ser interpretado e entendido, para que se possa fazer uma boa preparação da mesma e assim poder-se definir qual a melhor ferramenta a ser utilizada. Com isso um resultado consistente e fundado na realidade dos dados será obtido.

## Referências

- CARVALHO, FLÁVIA PEREIRA DE AND JUNIOR, AGUINALDO FAGUNDES AND SILVEIRA, JORGE GUEDES AND AZAMBUJA, MARCELO CUNHA DE, *KDD - NMS: Um Sistema de Descoberta de Conhecimento e Mineração em Bases de Dados de Sistemas de Gerência de Redes*, 2003. Disponível em: [http://www.rnp.br/\\_arquivo/wrnp2/2003/oscf01a.pdf](http://www.rnp.br/_arquivo/wrnp2/2003/oscf01a.pdf). Acesso em: 23/03/2012.
- Descoberta do conhecimento (KDD)*. Disponível em: <http://sites.google.com/site/mineracaodedados1b/descoberta-do-conhecimento-kdd>. Acesso em: 23/03/2012.
- DANTAS, ERIC ROMMEL G. AND JÚNIOR, JOSÉ CARLOS ALMEIDA PATRÍCIO AND LIMA, DANIEL SILVA DE AND AZEVEDO, RYAN RIBEIRO DE, *O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões*. Disponível em: [http://www.aedb.br/seget/artigos08/331\\_331\\_Artigo\\_SEGET\\_EJDR\\_Versao\\_Final\\_010808.pdf](http://www.aedb.br/seget/artigos08/331_331_Artigo_SEGET_EJDR_Versao_Final_010808.pdf). Acesso em: 25/03/2012.
- TURBAN, MCLEAN WETHERBE., *Tecnologia da Informação para Gestão, Bookman*, 3ª edição, págs. 344, 345, 408.
- SISTEMAS ESPECIALISTAS - Data Mining (DM): alguns conceitos*. Disponível em: <http://palhares2.flu.angelfire.com/uniceub/ia/iat004.pdf>. Acesso em: 26/03/2012.
- SELLITTO, MIGUEL AFONSO, *Inteligência Artificial: uma aplicação em uma indústria de processo contínuo*. Disponível em: <http://www.scielo.br/pdf/gp/v9n3/14574.pdf>. Acesso em: 28/04/2012.
- CARDOSO, FÁTIMA, *A inteligência do Chip: softwares que imitam o raciocínio humano*. Disponível em: <http://super.abril.com.br/tecnologia/inteligencia-chip-sofwarees-imitam-raciocinio-humano-440972.shtml>. Acesso em: 01/05/2012.
- BARRELLA, WAGNER DÄUMICHEN, *Bases Conceituais para o Desenvolvimento de um Sistema Especialista para Indústrias, Alemanha: Instituto Tecnológico Brasil*, 1998.
- DIAS, MARIA MADALENA, *A inteligência do Chip: softwares que imitam o raciocínio humano*. Disponível em: <http://super.abril.com.br/tecnologia/inteligencia-chip-sofwarees-imitam-raciocinio-humano-440972.shtml>. Acesso em: 01/05/2012.
- BARBOSA, JULIANA MOREIRA AND CARNEIRO, TIAGO GARCIA DE SENNA AND TAVARES, ANDREA IABRUDI, *Métodos de Classificação por Árvores de Decisão Disciplina de Projeto e Análise de Algoritmos*. Disponível em: <http://www.decom.ufop.br/menotti/pa111/files/PCC104-111-ars-11.1-JulianaMoreiraBarbosa.pdf>. Acesso em: 26/11/2012.
- MACHADO, RAFAEL GOMES, *Análise e Projecto de Sistemas*. Disponível em: [www.esenweiseu.net/recursos](http://www.esenweiseu.net/recursos). Acesso em: 27/11/2012.