CENTRO UNIVERSITÁRIO FEEVALE

DAIANA PEREIRA DOS SANTOS

MINERAÇÃO EM NOTAS FISCAIS DE ENTRADA DE UMA EMPRESA CALÇADISTA

Novo Hamburgo, junho de 2008.

DAIANA PEREIRA DOS SANTOS

MINERAÇÃO EM NOTAS FISCAIS DE ENTRADA DE EMPRESA CALÇADISTA

Centro Universitário Feevale
Instituto de Ciências Exatas e Tecnológicas
Curso de Ciência da Computação
Trabalho de Conclusão de Curso

Professor Orientador: Juliano Varella de Carvalho

Novo Hamburgo, junho de 2008.

AGRADECIMENTOS

Gostaria de agradecer a todos os que, de alguma maneira, contribuíram para a realização desse trabalho de conclusão, em especial:

Aos amigos e às pessoas que convivem comigo diariamente, minha gratidão, pelo apoio emocional - nos períodos mais difíceis do trabalho. Agradeço ao professor Juliano por ter orientado este trabalho com sabedoria e destreza.

Enfim, agradeço ao meu marido Edson pelo apoio e paciência e principalmente a Deus que está me proporcionando esta oportunidade.

RESUMO

Esse trabalho apresenta o processo de descoberta do conhecimento com ênfase na fase de mineração ou *Data Mining* (DM). Será relatado as etapas do processo de KDD - *Knowledge discovery in databases*, as técnicas de mineração e seus principais classificadores: árvore de decisão e redes neurais. Com base em alguns artigos que fazem uso da tecnologia de DM, procurou-se mostrar de forma abrangente aquilo que pode ser realizado a partir do processo de mineração. Como estudo de caso, será apresentada a aplicação do processo de mineração em uma base de dados específica de uma empresa calçadista, com auxílio da ferramenta Weka e seus algoritmos. O objetivo deste estudo de caso é descobrir conhecimento novo e transformá-lo em ações que tragam melhorias para a área de suprimentos da empresa.

Palavras-chave: *Data Mining*, Weka, Fornecedores, Nota fiscal, Conhecimento.

ABSTRACT

This research shows the process of knowledge discovery with emphasis in the fase of mining or Data Mining. Will be related the stage of the process of KDD – Knowledge discovery in databases, the techniques of mining and its mains classifiers: decision tree and neural network. With the base on some articles that do using of the technology of DM, found it to show in the comprehensive form that can be realized from a process of mining. As a case, this research will show the application of the process of mining in the a specified database of shoe's factory with the help of the Weka application and yours algorithms. The main of this case is discovery new knowledge and change it to actions that meaning benefits to the supply chain.

Key words: Data Mining, Weka, Supplied, Invoice, knowledge.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ARFF	*Attribute Relation File Format*
DM	*Data Mining*
EDI	*Eletronic Data Interchange*
KDD	*Knowledge Discovery in Databases*
RNA	Redes Neurais Artificiais
SERASA	Centralização de Serviços dos Bancos S. A.
SGBD	Sistema gerenciador de banco de dados
SPC	Serviço de Proteção ao Crédito
WEKA	*Waikato Environment for Knowledge Analysis*

SUMÁRIO

INTRODUÇÃO

Empresas de grande porte e que possuem muitas operações transacionais, geram uma enorme quantidade de informações que ficam registradas em seus bancos de dados. Estas informações não fornecem conhecimento a "olho nu", pois se encontram em sua forma bruta e precisam ser lapidadas. Conhecimento novo, descoberta de novas tendências e relações entre processos e produtos, são as pedras preciosas que as empresas procuram, mas que muitas vezes não conseguem encontrar, por não terem acesso às ferramentas adequadas.

A exploração do D*ata Warehouse*[1] das empresas através de análise dos dados, permite aos usuários extrair informações úteis e que podem auxiliar na tomada de decisões (LAUDON, 1999) . Esta análise de dados pode ser feita através de processamento analítico *on-line* (OLAP - O*n-Line Analytical Processing*) e mineração de dados (DM-*Data Mining).*

Harrison (1998) define DM como um processo de exploração e análise de uma grande massa de dados, fazendo uso de meios automáticos ou semi-automáticos, objetivando a descoberta de padrões e regras significativas. Para se aplicar o DM em uma base de dados, a mesma precisa passar por um pré-processamento onde devem ser apontadas as fontes internas e externas, gerando subconjuntos dos dados que sofrerão o processo de mineração. (QUONIAN et al,2001)

Um bom exemplo do uso de *Data Mining* é em redes de supermercados que guardam as informações referentes a seus clientes e suas compras a fim de montar perfis de consumo. Através das técnicas de mineração, neste caso, obtêm-se padrões que podem auxiliar o supermercado a descobrir quais produtos foram vendidos em conjuntos ou itens que devem

---

[1] *Data Warehouse* segundo Laudon (1999), são bancos de dados que armazenam dados atuais e históricos de interesse potencial para os gerentes de toda a empresa, estes dados são padronizados e consolidados de forma que possam ser usados para análises gerenciais e tomada de decisão.

estar próximos nas prateleiras. Outro exemplo do uso de DM são as dicas que aparecem em sites de comércio eletrônico quando o consumidor escolhe um determinado produto para comprar. Estas dicas mostram quais foram os outros produtos adquiridos por pessoas que também compraram o produto que está sendo selecionado pelo consumidor. (X ESCOLA,2002)

A motivação deste trabalho consiste em explorar a base de dados de uma empresa calçadista, mais especificamente a base onde são registradas as notas fiscais de entrada originadas de transações de compra de suprimentos. Com base nas técnicas de DM e uso de ferramentas gratuitas, busca-se encontrar informações importantes e que podem ajudar na tomada de decisões da gerência de suprimentos, apontar alguma ineficiência no processo ou ainda relacionar fatos ocultos que podem agregar conhecimento à equipe.

Este trabalho iniciará com um embasamento teórico referente as etapas do processo de descoberta de conhecimento em bases de dados (KDD – *Kownledge Discovery in Databases*) e em seguida sobre o processo de mineração, suas tarefas e alguns classificadores. No capítulo 2 será relatado três aplicações que fizeram uso das técnicas de *Data Mining,* o capítulo 3 mostra o estudo de caso fazendo analogia as etapas do KDD continuando no capítulo 4 com ênfase nas etapas de mineração, avaliação e consolidação do conhecimento.

DESCOBERTA DE CONHECIMENTO

Definição

A forma normal de se obter conhecimento dos registros de dados é através da análise e interpretação de um especialista. Mas como analisar grandes massas de dados e obter informações importantes em curtos espaços de tempo? Eis uma dúvida que pode ser respondida através do processo de descoberta de conhecimento em banco de dados, denominado KDD - *Knowledge discovery in databases*.

Como o próprio nome já define, KDD consiste em um processo que exige pesquisa em um conjunto de dados a fim de obterem-se freqüências, combinações e modelos que sejam novidade, que tenham valor agregado, e que possam ser aplicados na prática e totalmente inteligíveis para o ser humano. (FAYYAD,1996)

Como todo processo, este também é composto por várias etapas que contemplam duas fases distintas: preparação e mineração dos dados. Na primeira fase deverá haver a escolha da técnica de mineração a ser usada com base no problema e na base de dados envolvida, em seguida é feita uma limpeza nos dados através de um pré-processamento e por fim os dados sofrem uma transformação para facilitar seu uso na próxima fase. Na segunda fase inicia-se o processo de mineração onde deverá ser escolhido o algoritmo a ser usado. Este processo gerará relatórios que serão analisados por especialistas (VIANA, 2004).

Etapas do KDD

O processo de descoberta do conhecimento não se resume apenas a analisar grandes bases de dados e obter padrões, é necessário trabalhar com estas informações e transformá-las

em retorno empresarial (CARVALHO,2005). Incorporar os processos de DM e usá-los de forma efetiva pode tornar uma empresa reativa em pro ativa (HARRISON,1998).

Para que os padrões sejam descobertos, é necessário que os dados estejam simplificados a fim de que se possa descartar aquilo que é específico e, privilegiar aquilo que é genérico (NAVEGA,2002).

Dentro do processo total de KDD, se gasta 60% do tempo na preparação dos dados e apenas 10% no processo de mineração (CABENA,1998). Pelo fato do KDD ser um processo, sugere que existam vários passos. Estes passos são considerados interativos porque a maioria requer avaliação e tomada de decisão por parte do usuário e, iterativos porque é possível, de qualquer etapa, voltar para alguma etapa anterior até que se chegue ao resultado esperado. A figura 1.1 mostra os passos do KDD que serão detalhados nas próximas seções deste capítulo.

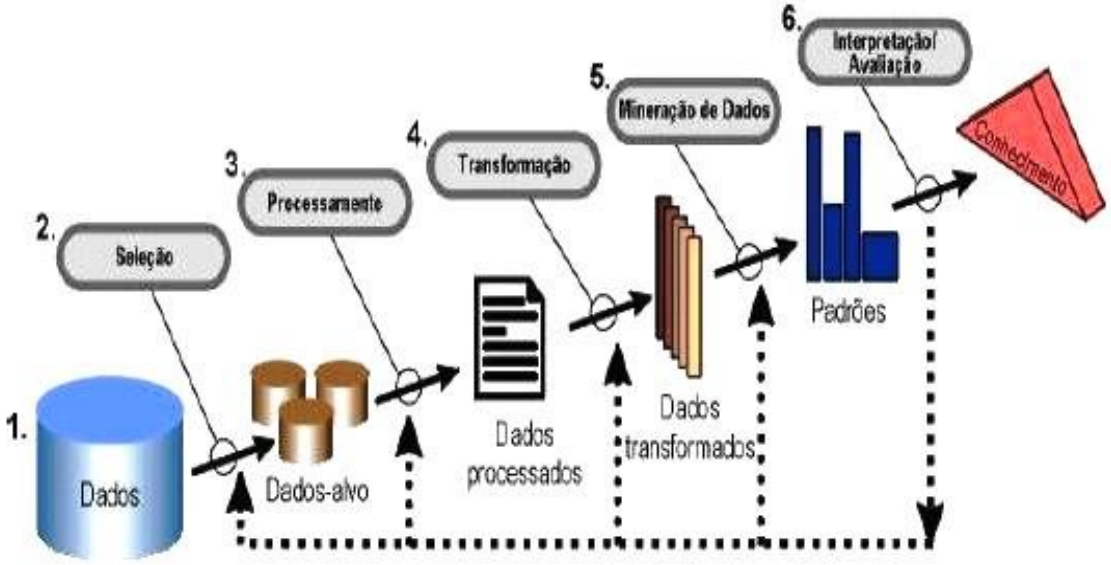

Figura 1.1 – Etapas do processo de KDD

Compreensão do domínio da aplicação

Este passo é comum no início de qualquer projeto significativo. O requisito mínimo é que os problemas e os objetivos do negócio sejam percebidos e tenham potencial para retornar valores válidos.

Faz-se necessário nesta fase, a colaboração do analista de negócios que possui o domínio do conhecimento e do analista de dados que pode começar a traduzir os objetivos do projeto para a aplicação posterior do DM (CABENA,1998).

Seleção de um conjunto de dados alvo

O ponto forte deste processo é identificar a origem dos dados e extrair aqueles que são necessários para uma análise preliminar na preparação para o processo de DM. Nesta fase, a análise dos dados já pode começar a focar nos algoritmos de DM que mais combinarão com o negócio (CABENA,1998).

Pré-processamento de limpeza dos dados

Esta etapa tem como principal função assegurar a qualidade dos dados selecionados. Possibilita que haja uma familiarização dos dados permitindo que se saiba procurar onde está o real conhecimento durante a fase de DM.

Esta fase é a mais problemática dentro de processo de preparação dos dados, pois os dados mais operacionais nunca foram modelados ou capturados para propósitos de DM, por isso possuem muitos ruídos e distorções que precisam ser removidos (CABENA,1998).

Transformação

Nesta fase, os dados pré-processados são transformados para que se possa criar um modelo de dados analítico. A acurácia e validade do resultado final vão depender de como o analista de dados decidiu estruturar e apresentar as entradas que, servirão de base para a escolha de um algoritmo específico de DM.

A transformação dos dados incide em converter valores para formatos padrões, reduzir o número total de variáveis usadas ou categorizá-las de acordo com seu valor (CABENA,1998).

Mineração

Esta etapa do processo consiste basicamente em escolher a tarefa apropriada de DM, seus respectivos algoritmos e executá-los em cima da base escolhida e pré-processada pelas etapas anteriores. A mineração está intimamente conectada com a etapa posterior que é a análise dos resultados, pois é com base nesta análise que se descobre se os objetivos foram

alcançados. Se for necessário, esta etapa pode ser executada quantas vezes for preciso até que se chegue ao resultado satisfatório.

Os algoritmos são escolhidos de acordo com um conjunto de fatores que incluem o real objetivo do negócio, habilidade em manipular qualquer tipo de dados, saídas de fácil compreensão, escalabilidade e algum nível de familiaridade.

A execução dos algoritmos requer o acompanhamento de um analista de dados e até algumas vezes, de um administrador de bancos de dados caso seja necessário alguma alteração no modelo. O que for definido nesta etapa irá variar de acordo com o tipo de aplicação que está sendo desenvolvida.

Interpretação/Avaliação

O processo de mineração dos dados não faria sentido se os padrões descobertos não fossem avaliados a fim de se descobrir se as causas dos problemas foram solucionadas ou se o objetivo da empresa foi alcançado (CARVALHO,2005). Esta etapa deve ser executada pelo analista de dados e pelo analista de negócios. Eventualmente poderá ser contatado também o executivo responsável para fornecer esclarecimentos sobre as descobertas encontradas, relacionando-as aos objetivos do negócio a fim de validá-las (CABENA,1998).

Consolidação do conhecimento descoberto

Esta etapa fecha o ciclo da descoberta do conhecimento e é responsável por colocar em ação todo o conhecimento adquirido durante as etapas anteriores. O papel do analista de negócios e do executivo responsável é tornar convincente todas as descobertas e encontrar a melhor maneira de explorá-las. Caso se chegue à conclusão de que as descobertas não foram suficientes para resolver o problema apontado no início do processo, pode-se aplicar todo o ciclo novamente (CABENA,1998).

Mineração de dados

O *Data Mining* ou Mineração de Dados, consiste na exploração e análise de grandes quantidades de dados através de meios automáticos ou semi-automáticos, para descobrir padrões e regras significativas (HARRISON,1998). Apesar de ser mais usado em processos

comerciais, o *Data Mining* vem ganhando grandes proporções à medida que as empresas começam a produzir grande quantidade de dados e sentem a necessidade de armazená-los.

A acirrada competitividade entre as empresas, faz com que passem a direcionar seus produtos e serviços a públicos específicos. A descoberta dos perfis destes clientes e suas preferências pode ser obtida através do processo de mineração. Outros fatores que estão possibilitando a disseminação do *Data Mining* é a queda de preço das tecnologias de processamento paralelo e a disponibilidade de softwares de DM no mercado (HARRISON,1998).

O DM é uma tecnologia que trabalha em uma área interdisciplinar pois suas técnicas usam conhecimento de outras áreas conforme mostra figura 1.2.

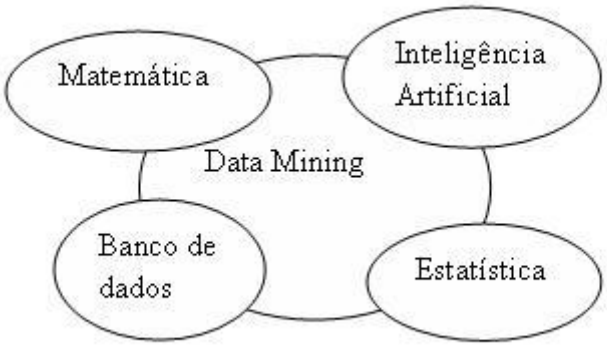

Figura 1.2 – Áreas que fornecem conhecimento às técnicas de DM

As tarefas da Mineração de Dados

Através de estudos feitos na área, definiu-se que o processo de mineração de dados faz uso de diversas tarefas e técnicas, onde as tarefas são classes de problemas e as técnicas são grupos de soluções. Uma tarefa pode ser solucionada por mais de uma técnica e algumas técnicas, podem solucionar tarefas diferentes. Nesta seção será discutido as principais tarefas da mineração de dados.

Classificação

É a tarefa de *Data Mining* mais usada por se parecer muito com a compreensão que o ser humano tem do ambiente no qual está inserido (HARRISON,1998). Esta tarefa consiste

em criar classes de objetos ou dados, que comparados com outros apresentam semelhanças, indicando possíveis agrupamentos.

Esta tarefa consiste em categorizar objetos de acordo com suas características e agrupá-los em conjuntos de classes predefinidas, por este motivo pode ser considerada como preditiva. Ela faz uso de resultados discretos e sua principal técnica é a árvore de classificação (VIANNA, 2004). Por exemplo, uma população pode ser dividida em classes sociais (baixa, média ou alta), sexo (masculino ou feminino), idade e escolaridade. Analisando cada indivíduo e classificando-o de acordo com suas características, pode-se descobrir, por exemplo, em que classe social se encontra o maior número de indivíduos do sexo masculino com 2º grau completo estando na faixa etária de 20 a 30 anos.

Estimativa/Previsão

Com base nos valores de variáveis conhecidas, a estimativa tem a função de estipular valores de variáveis desconhecidas. Faz uso de valores contínuos e também é considerada uma tarefa preditiva já que existem informações definidas anteriormente que servirão de base para a aplicação da tarefa. Por exemplo, pode-se estimar as chances de um paciente se recuperar de determinada doença avaliando um conjunto de diagnósticos, sua idade e condições financeiras.

Já a tarefa de previsão, funciona basicamente da mesma maneira, porém as variáveis futuras são conhecidas. Pode-se citar como exemplo o faturamento de uma empresa, que tem previsão de faturar 40 milhões de reais no mês de dezembro do ano corrente, com base no faturamento que teve no ano passado neste mesmo período.

Associação/Descrição

O processo de DM pode também simplesmente, ter a função de descrever a maneira como determinados dados foram produzidos a fim de aumentar o conhecimento das pessoas. Uma boa descrição gera uma boa explicação, ou pelo menos sugere onde se pode começar a procurar por ela.

Através desta tarefa descritiva, é possível obter padrões que apontam relacionamentos entre itens armazenados criando uma associação entre eles. Por exemplo, analisando as transações de compra de um determinado grupo de clientes, pode-se concluir que determinados produtos sempre ocorrem juntos na mesma compra. Esta informação pode servir de apoio ao gerente do supermercado quando ele for elaborar um novo *layout* de

prateleiras ou mesmo na montagem de um catálogo de produtos, itens casados podem ficar próximos fisicamente ou participarem de promoções.

Segmentação

Esta tarefa consiste em agrupar elementos de uma população heterogênea em pequenos grupos ou *clusters* mais homogêneos. Em contradição a tarefa de classificação, na segmentação, não se faz necessário a existência de classes predefinidas. Nada precisa ser dito ao sistema, o próprio algoritmo descobre os objetos semelhantes e faz o agrupamento, sem ter a necessidade de interação do usuário. Como exemplo, pode-se citar a carteira de clientes de uma grande loja, onde analisando cada cliente, é possível agrupá-los de acordo com seus hábitos de consumo e usar esta informação em uma outra tarefa de DM.

Classificadores

Nesta seção será discutido sobre duas técnicas de mineração usadas pela tarefa de classificação: árvores de decisão e redes neurais. Estas duas técnicas se baseiam em aprendizado supervisionado, cujo processo de criação automática de um modelo de classificação tem origem em um conjunto de registros chamado de conjunto de treinamento (CABENA,1998).

Árvores de decisão

Consiste em um algoritmo de classificação que usa um formato representado por modelos SE-ENTÃO considerado de fácil compreensão, pois as regras ficam bem explícitas possibilitando que especialistas avaliem os resultados e identifiquem atributos chave. É considerado um algoritmo supervisionado pelo fato da necessidade de se saber com antecedência as classes dos registros usadas no conjunto de treinamento (VIANNA,2004) .

Uma árvore de decisão é composta por um conjunto de nós que são conectados através de ramificações, estes nós se dividem em três tipos conforme mostra a figura 1.3.

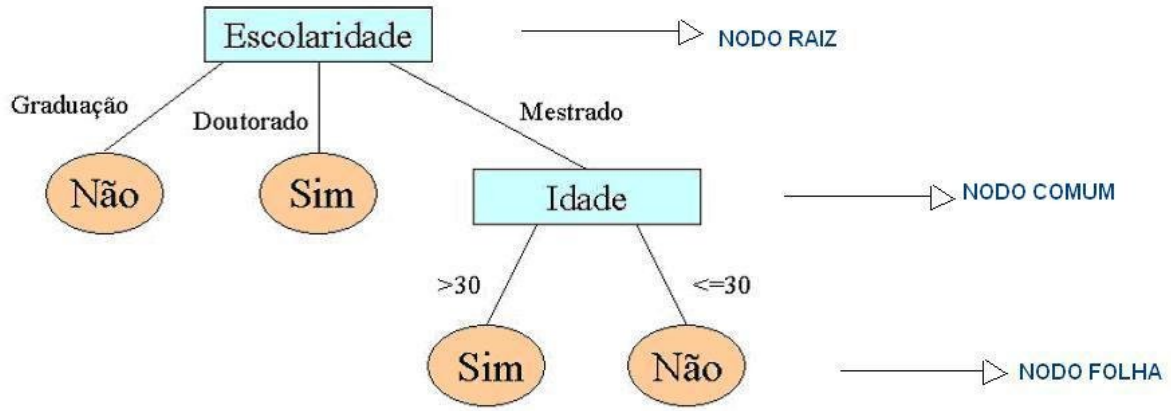

Figura 1.3 – Exemplo de árvore de decisão com nodos identificados. (GONÇALVES, s.a)

Nodo Raiz: nodo que inicia a árvore;

Nodos Comuns: divide um atributo e gera novas ramificações;

Nodo Folha: possui as informações de classificação do algoritmo.

Durante a fase de treinamento do algoritmo é possível parametrizar o número máximo de níveis possíveis na árvore. É candidato a ser um nó de uma árvore, o atributo que melhor classificar os dados, e os escolhidos são chamados de atributos divisores ou atributos teste (PICHILIANI,2006).

Seguem quatro passos do funcionamento do algoritmo para a criação de uma árvore de decisão:

Geração do nó raiz: cria-se um nó raiz contendo as probabilidades calculadas de cada valor do atributo de classificação.

Encontrar nós a serem divididos: os candidatos são os nós que não são folhas e que ainda não possuem divisões. Também não podem classificar a amostra totalmente, ou seja, não pode haver classes calculadas com 100% de probabilidade de classificar sua amostra.

Divisão de nó: deve-se escolher um atributo que melhor classifica os dados. Em seguida considerar o atributo que mais gera nós folha e os que menos geram nós que podem ser divididos.

Criação do nó: a partir do atributo escolhido cria-se o nó e suas ramificações com todos os possíveis valores do atributo.

Um novo processo de análise é disparado sobre os nós gerados e o algoritmo volta para o processo de encontrar nós a serem divididos. O algoritmo pára sua execução quando somente encontrar nós com probabilidade de classificação igual a 100% (PICHILIANI,2006).

Redes Neurais

São consideradas as técnicas mais comuns usadas pelos processos de DM e consistem na combinação de neurônios interligados por sinapses cujo algoritmo tem a função de simular o cérebro humano. Adquirem aprendizado através de um conjunto de dados de treinamento e possuem uma particularidade que os diferenciam dos demais; eles podem gerar saídas equivalentes às entradas, que não existiam durante a fase de treinamento (LUDWIG, 2007).

Uma rede Neural representa muito bem um conjunto de dados que sofrem modificações com o passar do tempo, pois pode ser projetada para ter seus pesos sinápticos alterados em tempo real (LUDWIG, 2007). Outra vantagem da RNA é que elas detectam padrões nos dados de forma semelhante ao pensamento humano (HARRISON,1998).

É considerada uma desvantagem da RNA, o fato de não se saber por que uma rede chegou a um determinado resultado. É necessário introduzir dados para validação a fim de se obter o erro médio quadrático cuja análise, indicará a veracidade do resultado (LUDWIG, 2007).

Uma RNA é formada por uma ou mais camadas que podem ser compostas por um ou mais neurônios (LUDWIG,2007), e estas camadas estão dispostas conforme mostra figura 1.4.

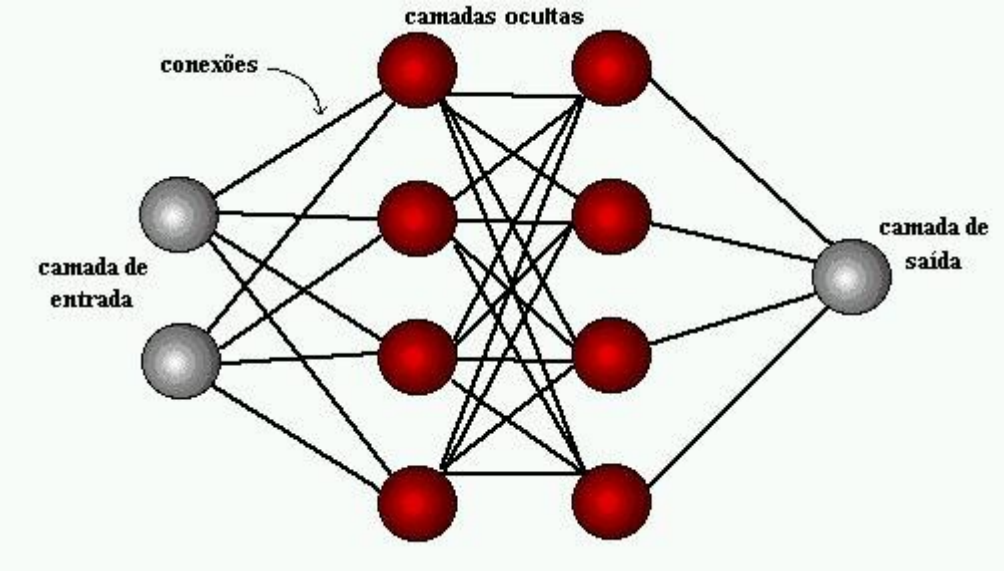

Figura 1.4 – Arquitetura de um RNA (CARVALHO, s.a.).

Camadas de entrada: não possui nenhum neurônio, é composta por nós em quantidade equivalente ao número de sinais de entrada da rede. Nesta camada não há processamento, apenas a representação dos dados de entrada e sua distribuição para cada neurônio da camada seguinte.

Camadas ocultas: composta por pelo menos um neurônio que tem a função de possibilitar que a rede extraia estatística e represente problemas que na mesma linha não podem ser separados.

Camadas de saída: possui um número de neurônios equivalente ao número de sinais de saída da rede.

Não existe uma formalização em relação à arquitetura de uma RNA, a quantidade de neurônios ou camadas pode ser alterada de acordo com a necessidade do projetista (LUDWIG, 2007).

O funcionamento de uma RNA requer um treinamento que é adquirido através de algoritmos de aprendizado. A escolha do algoritmo está intimamente ligada à forma como a estrutura da RNA foi definida, que pode variar de acordo com as conexões entre as camadas.

Pode-se dar de duas maneiras o processo de aprendizado de uma RNA:

Supervisionado: trabalha-se com um conjunto de pares de dados de entrada e saída desejada. Dados os conjuntos de entrada, a RNA retorna um conjunto de valores de saída desejado. Caso haja muita diferença, os pesos sinápticos[2] e os níveis de bias[3] são ajustados até que esta diferença seja minimizada.

Não-supervisionado: a rede recebe treinamento apenas com base em valores de entrada. Ela classifica os dados pelo reconhecimento de padrões por meio de processos chamados competição e cooperação entre os neurônios.

[2] Pesos sinápticos: grandeza que representa a intensidade de ligação entre os neurônios (LUDWIG, 2007).

[3] Níveis de Bias: possibilita que um neurônio apresente saída não nula ainda que todas as entradas sejam nulas (LUDWIG, 2007).

APLICAÇÕES DE DATA MINING

Com base no estudo feito em alguns artigos que utilizam *Data Mining*, foram selecionados 3 (três) com o intuito de dar uma visão mais abrangente daquilo que pode ser realizado a partir das técnicas de mineração. Com base nisso, pode-se constatar que o DM não é somente usado para fins comerciais podendo ser aplicado em qualquer base de dados que forneça informação suficiente para extração de conhecimento. O primeiro artigo usa DM para classificar o fruto Pequi de acordo com sua quantidade de caroços, o artigo seguinte tem como base um catálogo de teses francesas que possuem o Brasil como assunto e que através dos processos de DM conseguiu descobrir padrões e conhecimento novo e por fim, o último artigo usa as técnicas de DM para classificar uma base de clientes em adimplentes e inadimplentes com o objetivo de auxiliar uma empresa na hora da concessão de crédito.

Utilização de algoritmos simbólicos para identificação do número de caroços do fruto Pequi (OLIVEIRA et al, 2002)

O Pequi é uma fruta popular do cerrado brasileiro e que possui muito valor econômico para a região. Seu caroço possui muitos espinhos o que pode ocasionar ferimentos nas gengivas se a fruta não for degustada de forma correta. Seu uso na culinária é bastante apreciado, pois possui aroma e sabor marcantes e peculiares.

Este artigo se propõe a trabalhar com um conjunto de dados formado pelas características do fruto pequi a fim de classificá-lo, com base em suas dimensões, em frutos com um, dois ou três caroços.

Como ferramenta de *Data mining* os autores optaram por usar o *software Weka* por várias razões e entre elas o fato dele se encontrar disponível na internet, fácil instalação e implementação em Java, o que deixa o sistema apto a funcionar em diversos ambientes. Esta ferramenta possui dois algoritmos de classificação J48.J48 e o J48.PART, mas nessa pesquisa

foi usado somente o primeiro, cujas particularidades podem ser vistas em (GONCHOROSKI,2007).

Para a montagem da base de dados, foram colhidos e analisados diversos frutos, os quais geraram 215 registros que foram divididos em três classes de acordo com os valores de um conjunto de atributos. A distribuição dos registros em cada classe ficou assim:

- Classe 1: 70 registros
- Classe 2: 75 registros
- Classe 3: 70 registros

Cada registro possui seis atributos e estará associado a uma classe conforme mostra a tabela 2.1. Os valores possíveis classificados como contínuos, indicam a medida variável de cada fruto.

Tabela 2.1 – Atributos da base de dados (OLIVEIRA et al, 2002)

Nome do Atributo	Descrição	Valores Possíveis
Comp (mm)	Indica o comprimento do fruto medido de uma ponta a outra	Contínuo
Larg (mm)	Indica a largura do fruto	Contínuo
Espes (mm)	Indica a espessura do fruto	Contínuo
Deq (mm)	Indica o diâmetro equivalente	Contínuo
Vol (g/ml)	Indica o volume	Contínuo
Esferic	Indica a esfericidade	Contínuo
Classe	Identifica o número de caroços do fruto	1, 2 e 3

Esta base de dados foi exportada para um arquivo “.txt” usando o ponto e vírgula como separador das colunas. A ferramenta *Weka* se encarregou de fazer a divisão dos registros em conjuntos de treinamento (70%) e teste (30%) segundo o paradigma de estimativa de erro *Holdout*, que é um dos mais usados. Os registros foram distribuídos nestes conjuntos de forma aleatória.

Os algoritmos foram executados separadamente e em seguida associados a métodos de meta-aprendizagem (*bagging* e *boosting*) com objetivos de encontrar melhores resultados conforme mostra tabela a seguir:

Tabela 2.2 – Taxa de desempenho dos algoritmos (OLIVEIRA et al, 2002)

Algoritmo	Métodos	% Erro	% Acertos
J48.PART		26	74
J48.PART Algoritmo	Bagging Métodos	20 % Erro	80 % Acertos
J48.PART	Boosting	16	84
J48.J48		23	77
J48.J48	Bagging	19	81
J48.J48	Boosting	14	86

Os autores constataram que o melhor resultado refere-se ao método *Boosting* associado ao algoritmo J48.J48, que obteve o menor percentual de erro e o maior de acertos. Com base neste baixo nível de erros, as hipóteses geradas por esta pesquisa foram consideradas válidas e de grande consistência na classificação do fruto pequi em relação ao seu número de caroços.

Este artigo permitiu aquisição de aprendizado em relação ao fruto Pequi, e uma maior habilidade em se trabalhar com a ferramenta de *Data Mining: Weka*, que terá suas características relatadas no próximo capítulo deste trabalho.

Inteligência obtida pela aplicação de *Data Mining* em base de teses francesas sobre o Brasil (QUONIAN,2001)

Usando como estudo de caso um catálogo de teses francesas chamado *Doc Théses*, este artigo tem por objetivo demonstrar as técnicas de DM focando as teses que tiveram o Brasil como assunto da pesquisa, totalizando 1355 teses. Nesta amostra estão incluídas também teses defendidas por brasileiros na França referente ao período de 1969 a 1999.

Estando a base de dados pronta para o processo de mineração, foi aplicado o software *Dataview* responsável por extrair indicadores de tendências. Este software está alicerçado em métodos bibliométricos[4] e tem por objetivo transformar dados em inteligência a fim de gerar elementos para análise estatística. Dois aspectos devem ser levados em

[4]Bibliometria – Aplicação de métodos estatísticos ou matemáticos sobre um conjunto de referências bibliográficas (QUONIAN, apud ROSTAING,1996).

consideração no processo decisório: o valor e a validade da informação, pois irão influenciar decisivamente em todo o processo de descoberta do conhecimento.

O *Dataview* usa como método de mensuração, as ocorrências com que cada elemento bibliográfico aparece classificando-o em estado primário, estado condensado ou co-ocorrência. A figura 2.1 mostra em que momento o software *Dataview* participa do processo.

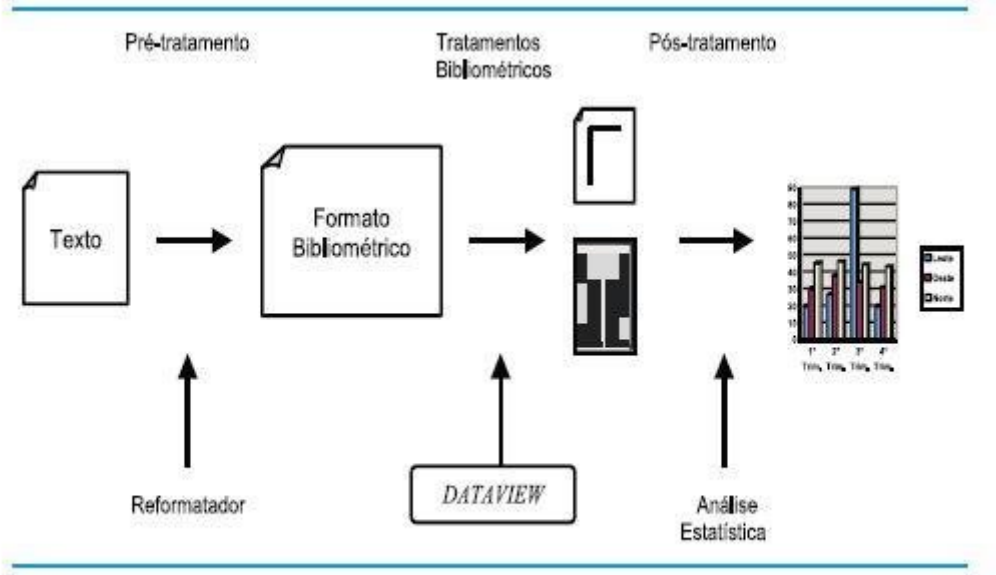

Figura 2.1 – Posição do *Dataview* em um estudo bibliométrico (QUONIAN, 2001)

A bibliometria é composta por 3 leis básicas:

1 – Lei de Bradford: baseia-se no comportamento de ocorrências repetitivas dentro de uma área específica do conhecimento;

2 – Lei de Lotka: baseia-se na contribuição que cada autor fornece para o progresso da ciência;

3 – Lei de Zipf: chamada lei quantitativa da atividade humana e subdivide-se em primeira e segunda lei de Zipf. A primeira leva em consideração o número de vezes que as palavras aparecem no texto e na segunda, estabelece-se quais delas aparecem com menos freqüência.

Este artigo levou em consideração a curva Zipf que se encontra dividida em 3 zonas diferentes conforme mostra figura 2.2.

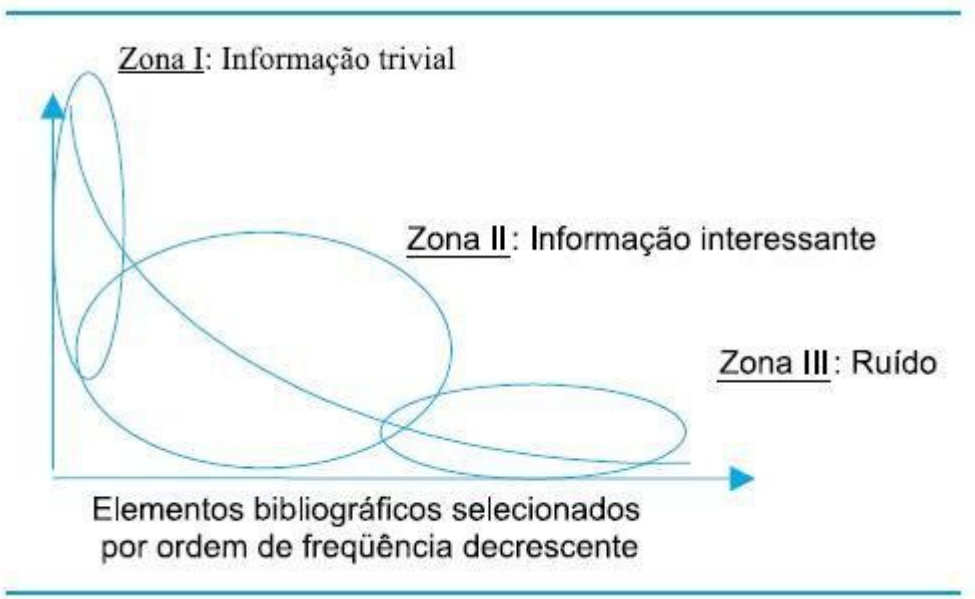

Figura 2.2 – Zonas de distribuição (QUONIAN, 2001)

Zona I – temas centrais da análise bibliométrica;

Zona II – temas periféricos e informações com potencial de inovação;

Zona III – conceitos são avaliados e classificados em emergentes ou ruídos.

Apenas as zonas I e II foram consideradas nesta pesquisa. Conforme mostra a figura 2.3, um terço do total das teses que tinham o Brasil como assunto da pesquisa ou como país origem do pesquisador, se refere ás áreas de economia, sociologia e ciências tecnológicas – zona I.

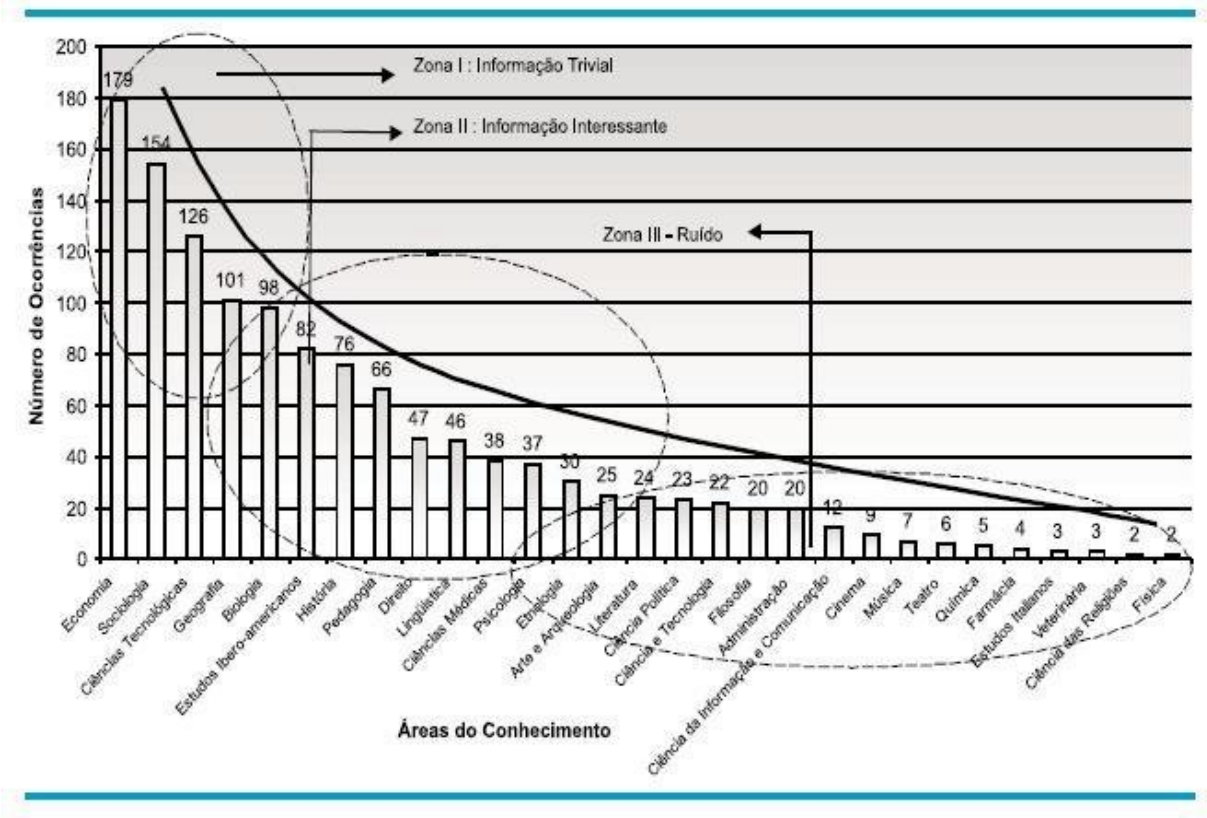

Figura 2.3 – Ocorrência por área de conhecimento (QUONIAN,2001)

Pertencentes a zona II, foi encontrado a pedagogia, ciências médicas, estudos libero-americanos e história que, são consideradas áreas emergentes.

Na análise feita sobre os orientadores brasileiros, pode-se destacar Fréderic Mauro na área de História onde foi o professor que mais orientou teses de 1969 a 1999, o que corresponde a 25% do total de teses orientadas dentro do seu grupo. Esta pesquisa originou outra informação importante: Paris é a cidade onde se encontram o maior número de teses defendidas, em torno de 62% do total.

Dentro do período analisado, pode-se também constatar quais os anos em que houve maior interesse das pessoas em escrever teses sobre determinado assunto. Por exemplo, nos anos 70 aumentou o interesse pela área de Direito, o que se atribui às circunstâncias políticas vividas pelo Brasil nesta época.

Pode-se concluir que a aplicação dos processos de *Data Mining* foi reveladora já que mostrou as áreas de maior interesse, o orientador que mais teses defendeu, as cidades que abrigam o maior número de teses e o período com que tudo isso ocorreu com maior e menor freqüência.

Mineração de dados: um estudo de caso de concessão de crédito explorando o software *Weka* (OLIVEIRA,2005)

A carteira de clientes de uma empresa pode trazer muitas informações importantes e dentre elas, características que permitem avaliar se um cliente é merecedor ou não de uma concessão de crédito. Pelo fato das organizações terem dificuldades em avaliar seus atuais e novos clientes, este artigo tem por objetivo avaliar uma base de clientes ativos através das técnicas de DM e exploração da ferramenta *Weka*, a fim de fornecer informações suficientes para a tomada de decisão.

A empresa selecionada para este estudo possui hoje uma avaliação manual, lenta e imprecisa de seus clientes, o que aumenta a probabilidade de erros. A idéia é confrontar através de regras de classificação, os dados dos clientes já existentes, a fim de classificá-los em adimplente e inadimplente e usar estas mesmas regras para classificar também o perfil dos novos clientes.

A organização elegeu 29 atributos que ela considera relevante para a concessão do crédito, sendo que 9 atributos são iguais, porém são provenientes de 3 empresas diferentes, onde o cliente já possui um relacionamento. São eles:

- Tempo de relação com o fornecedor;

- Forma de pagamento;

- Valor médio mensal das compras;

- Valor da maior compra;

- Data da maior compra;

- Valor da última compra;

- Data da última compra;

- Pontualidade;

- Conceito: ótimo, bom, ruim ou inativo.

O vigésimo oitavo atributo seria uma consulta ao SPC (Serviço de proteção ao crédito) e SERASA (centralização dos serviços de Bancos S.A.) para investigar a situação do cliente e o último, a opinião de quem estaria preenchendo o cadastro que, com base nestas informações, concederia ou não o crédito ao cliente.

A seqüência deste estudo se deu em 3 etapas conforme descritas a seguir, onde na primeira foram realizados testes com diversos algoritmos que, com base nos resultados e sua confiabilidade, partiu-se para a segunda e terceira etapas onde foram executados os treinamentos, classificação e inclusão de novos clientes.

Primeira etapa: foram usados 100 clientes e aplicou-se 5 algoritmos (*ConjunctiveRule, DecisionTable, DecisionStump, J48, ADTree*). A figura 2.4 mostra a eficiência destes algoritmos com destaque para o *ADTree* e *DecisionTable* que obtiveram a melhor performance em relação aos demais.

	% classificados corretamente	% classificados incorretamente	Adimplente correto	Adimplente incorreto	Inadimplente correto	Inadimplente incorreto
Conjunctive Rule	86	14	65	2	21	12
Decision Table	91	9	65	2	26	7
Decision Stump	86	14	65	2	21	12
J48	83	17	66	1	18	16
ADTree	95	5	67	0	28	5

Figura 2.4 – Tabela comparativa do uso de algoritmos: base 100 clientes (OLIVEIRA, 2005)

Segunda etapa: foram usados os mesmos cinco algoritmos da etapa anterior, agora utilizando uma base de 80 clientes para treinamento e 20 para testes. A figura 2.5 mostra a eficiência dos algoritmos com ênfase para os algoritmos *DecisionTable, ConjunctiveRule, DecisionStump* e *J48.*

	% classificados corretamente	% classificados incorretamente	Adimplente correto	Adimplente incorreto	Inadimplente correto	Inadimplente incorreto
Conjunctive Rule	70	30	8	4	6	2
Decision Table	80	20	11	1	5	3
Decision Stump	70	30	8	4	6	2
J48	70	30	12	0	2	6
ADTree	50	50	8	4	2	6

Figura 2.5 – Tabela comparativa uso algoritmos: base 80 clientes treinamento e 20 para testes (OLIVEIRA,2005)

Os algoritmos selecionados para a etapa seguinte foram o *DecisionTable* por sua confiabilidade e o *J48* por ser o mais conhecido nas literaturas.

Terceira etapa: nesta etapa foi simulada a inclusão de um novo cliente. A fim de comprovar a utilidade e viabilidade da ferramenta *Weka*, foram utilizados para estes testes os algoritmos escolhidos na etapa anterior, no qual tiveram seu comportamento avaliado usando uma base composta por 100 clientes e outra com apenas 1 (cliente teste). Foram realizados 7 testes de inclusão de um novo cliente a fim de classificá-lo (adimplente ou inadimplente) com base nos demais clientes já cadastrados e classificados. A figura 2.6 mostra a eficiência destes dois algoritmos. Nota-se que o grau de confiabilidade foi bastante grande mesmo quando houve a tentativa de corrompê-los.

		% classificados corretamente	% classificados incorretamente	Adimplente correto	Adimplente incorreto	Inadimplente correto	Inadimplente incorreto
Decision Table	Teste 1	100	0	0	0	1	0
	Teste 2	0	100	0	1	0	0
	Teste 3	100	0	1	0	0	0
	Teste 4	0	100	0	0	0	1
J48	Teste 5	100	0	1	0	0	0
	Teste 6	0	100	0	1	0	0
	Teste 7	100	0	1	0	0	0

Figura 2.6 – Tabela comparativa no uso de algoritmos para a inclusão de um novo cliente. (OLIVEIRA,2005)

Pode-se concluir que através das técnicas de *Data Mining*, usando a ferramenta *Weka* para a criação de regras de classificação, foi possível analisar e classificar os clientes desta empresa, em adimplentes ou inadimplentes. Esta análise não só auxilia na tomada de decisão como também fornece agilidade e confiabilidade na difícil tarefa de conceder crédito.

No próximo capítulo será relatado o funcionamento do estudo de caso estipulado para este trabalho. Haverá a apresentação do ambiente atual contemplando as principais tabelas do banco de dados que serão utilizadas, suas características e finalidade. Em seguida será explicado o motivo pelo qual se optou em usar a ferramenta *Weka* e uma breve descrição de suas características e funcionamento. Para finalizar, será descrito uma proposta de como será aplicado na base de dados de uma empresa calçadista, o processo de mineração e quais objetivos se desejam atingir.

PROPOSTA DE *DATA MINING* EM BASE DE NOTAS FISCAIS DE EMPRESA CALÇADISTA

Este capítulo tem o propósito de definir o ambiente nos quais os processos de *Data Mining* serão aplicados, como isso será feito, quem são as pessoas que serão diretamente atingidas por este processo e por fim, o que se deseja descobrir e como este conhecimento agregará valor à empresa. A empresa escolhida para o projeto é a Calçados Azaléia S. A. localizada na cidade de Parobé, onde a aluna exerce a função de analista de sistemas da área de suprimentos. A medida que os passos forem sendo executados, será feito analogia com as etapas do KDD mencionadas no capítulo 1 deste trabalho.

Ambiente atual

A Calçados Azaléia S. A. trabalha com banco de dados *Oracle* desde 1994, onde iniciou a migração do ambiente *Mainframe* para um servidor de banco de dados relacional. Dona de marcas como Azaléia, Dijean, OLK e Olympikus, esta empresa tem uma produção diária de cerca de 120.000 pares e é considerada uma das maiores fabricantes de calçados femininos da América Latina. Recentemente esta empresa foi comprada pelo grupo Vulcabrás S.A.

Por ser uma empresa de grande porte que executa diariamente milhares de transações que vão desde a compra da matéria-prima até o faturamento do calçado para o cliente, ela gera milhares de registros em seu banco de dados. Este foi um dos motivos pelo qual a aluna optou em executar o estudo de caso deste trabalho nesta empresa, já que a finalidade das técnicas de *Data Mining* é justamente essa, varrer uma imensidão de dados para encontrar padrões e obter conhecimento novo. Esta fase inicial do trabalho refere-se a etapa do KDD - Compreensão do domínio da aplicação.

Pela experiência que a aluna possui como analista de sistemas da área de suprimentos, a base de dados onde fica registrada as notas fiscais de entrada na empresa, foi eleita como objeto de estudo. As notas fiscais de entrada são documentos que se referem à compra de suprimentos pertencentes ou não ao processo produtivo da empresa. Esta base é formada por tabelas do SGBD *Oracle*, e será usada então para a aplicação das técnicas de *Data Mining*.

As tabelas principais deste modelo se chamam NOTAS_FISCAIS_RECEBIMENTOS e ITENS_NOTAS_RECEBIMENTOS. Nestas tabelas estão armazenadas todos os documentos de entrada referentes aos mais diversos itens que vão desde a matéria-prima para fazer o calçado, passando por máquinas e equipamentos, até a compra de produtos de higiene e limpeza. Os dados existentes nestas tabelas se referem ao ano atual + ano anterior e se localizam no banco de dados principal da empresa chamado Oraaza1. No anexo 2 pode ser verificado a descrição destas duas tabelas. Todo início de ano é feito um salvamento dos dados anteriores a um ano em um outro banco de dados chamado Oracd, este banco de dados é usado especificamente para guardar dados históricos, feito isso, os dados do banco de produção são então excluídos. Os dados históricos são armazenados a fim de atender auditorias internas e prestações de contas à fiscalização.

Outras tabelas também serão usadas para complementar este trabalho, pois algumas informações eleitas para constar neste estudo não se encontram nas tabelas de notas, mas sim em tabelas relacionadas a elas conforme mostra o modelo de dados exibido na figura 3.1.

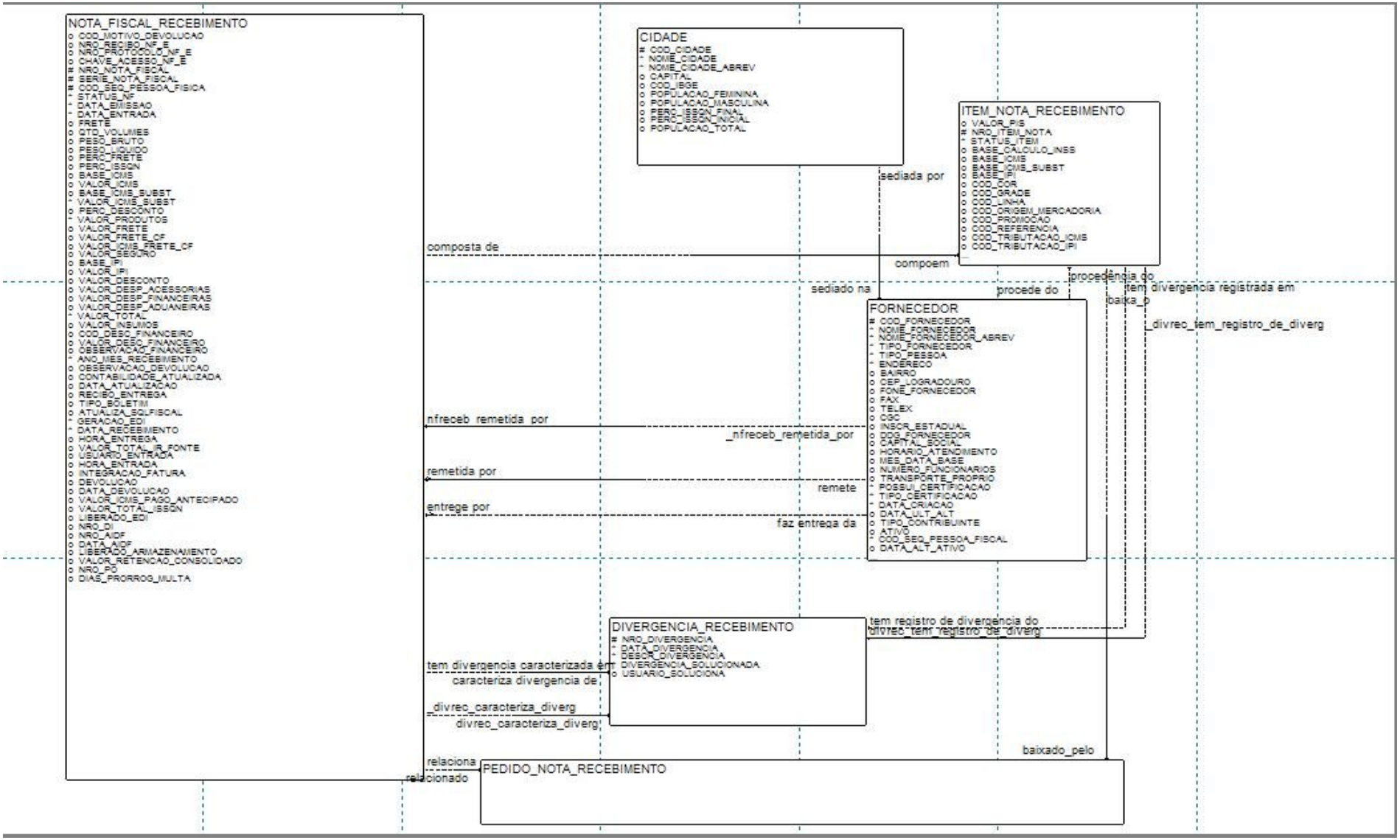

Figura 3.1 – Modelo ER das tabelas utilizadas neste estudo de caso

Segue abaixo a lista das demais tabelas que serão usadas neste estudo:

- FORNECEDORES: cadastro de todas as pessoas físicas e jurídicas para o qual a empresa emite ou recebe notas fiscais;

- EMPRESAS: cadastro de todas as empresas do grupo Azaléia;

- DEVOLUCOES_RECEBIMENTOS: tabela onde ficam registradas todas as devoluções que a empresa fez a um fornecedor seja ele uma empresa do grupo ou um terceiro;

- DIVERGENCIAS_RECEBIMENTOS: tabela que recebe todas as divergências existentes entre a nota fiscal de entrada e o pedido de compra atendido por esta nota;

- PEDIDOS_NOTAS_RECEBIMENTOS: relação dos pedidos no qual o item da nota fiscal de entrada está atendendo;

- CIDADES: cadastro de cidades.

A área de suprimentos da Azaléia é composta por vários sistemas, dentre eles pode-se citar o faturamento de materiais, recebimento, estoque, compras, preços, EDI[5] de fornecedores, logística, abastecimento e planejamento. Para a realização deste trabalho, alguns desses sistemas precisaram ser acionados a fim de se buscar informação, para que o processo de mineração fosse executado da forma mais eficiente possível.

Descobrindo Conhecimento com a Ferramenta Weka

Neste trabalho foi utilizado o software *WEKA* para a aplicação dos algoritmos de *Data Mining* na base de dados selecionada. Este software reúne uma série de características favoráveis a sua utilização, quais sejam:

- Gratuito: software livre (e também possui código aberto) disponível para *download* no próprio site da ferramenta (www.cs.waikato.ac.nz/ml/weka);
- Interface amigável: instalação simples e de fácil utilização;
- Escrito em Java: permite ser executado por diversos sistemas operacionais (Windows, Linux, Macintosh);

[5] EDI: processo de troca eletrônica de dados (GRUPO GOL,2007).

- Disponibilidade: qualquer pessoa pode baixá-lo e executá-lo sem restrições;
- Documentação *on-line*;
- Seus pacotes podem ser "plugados" em qualquer *software* a ser desenvolvido, pois chamadas aos algoritmos existentes são realizadas facilmente;
- Muitos trabalhos já utilizam esta ferramenta em seus processos. Como referência podemos citar (OLIVEIRA et al,2002) e (OLIVEIRA,2005).

Esta ferramenta foi desenvolvida pela Universidade de Waikato na Nova Zelândia e consiste em um conjunto de algoritmos de aprendizado de máquina que fornecem condições para extração do conhecimento (WITTEN, 2000).

É possível com esta ferramenta executar alguns passos do processo de mineração em cima de um conjunto de dados, sem precisar escrever uma só linha de código. Porém o processo de mineração exige que os dados estejam pré-processados e gravados dentro de um arquivo ".arff" (*attribute relation file format*). Para transformar os dados extraídos do banco de dados para este tipo de arquivo, os passos a seguir devem ser executados:

1 – Exportar os dados para um arquivo com a extensão CSV, cujo delimitador dos dados deve ser a vírgula;

2 – Abrir o arquivo como texto simples e salvar com a extensão ARFF;

3 – Rotular o conteúdo do arquivo para que a ferramenta saiba o que significa cada campo:

@relation – usado para rotular o conjunto de dados

@attribute – usado para rotular os atributos

@data – usado para identificar os dados

4 – Salvar o arquivo como texto sem formatação.

Segue abaixo exemplo de um arquivo gerado pelo passo 1 (figura 3.2) e em seguida um arquivo formatado pelos demais passos descritos acima (figura 3.3).

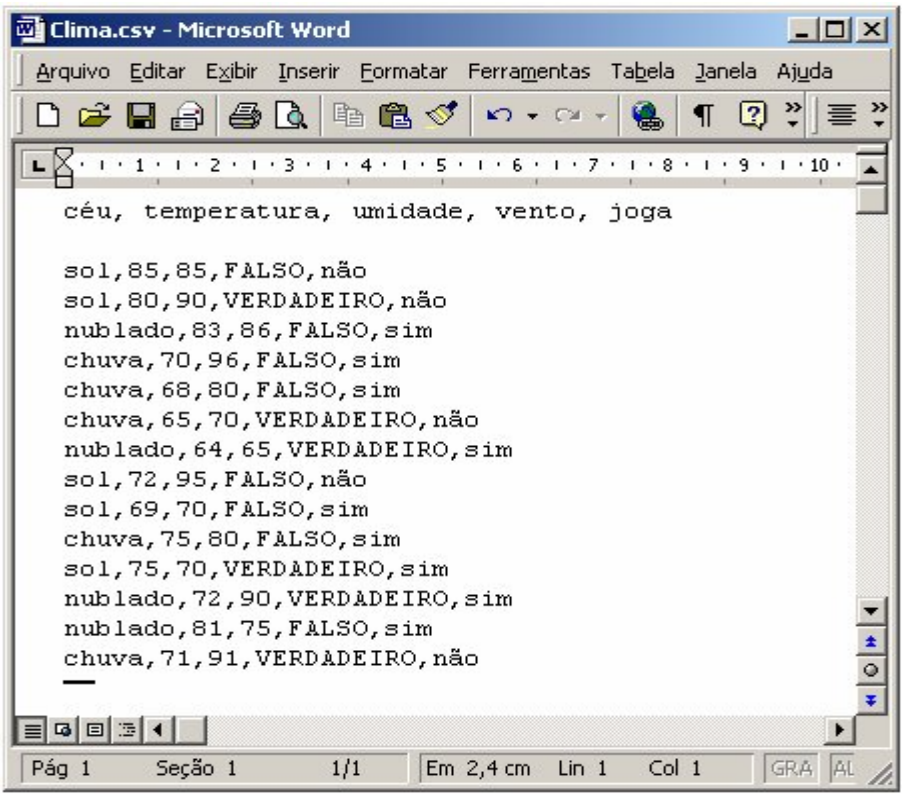

Figura 3.2 – Atributos separados por vírgula

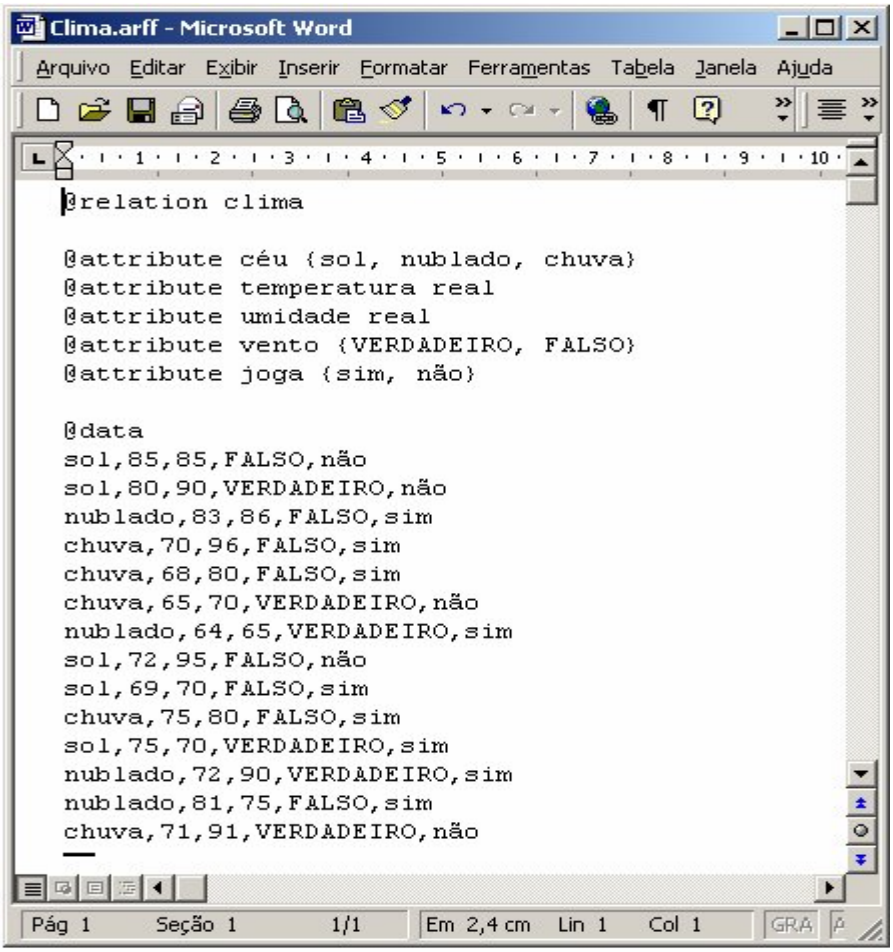

Figura 3.3 – Arquivo no formato ARFF

Após estes passos o arquivo está pronto para ser analisado pela ferramenta.

A ferramenta *Weka* é formada por um conjunto de pacotes: *atributte selection, classifiers, clustering, association rules* e *estimators*. Cada pacote é formado por vários algoritmos que possuem funções específicas de acordo com as tarefas de *Data Mining* mencionadas no capítulo 1 deste trabalho.

Neste trabalho usou-se especificamente o algoritmo de classificação: J48.J48 e o algoritmo de agrupamento: *SimpleK-Means.*

Proposta de uso de KDD no Setor de suprimentos de Empresa Calçadista

O objetivo deste trabalho é garimpar a base de dados de uma empresa calçadista mais especificamente a base onde fica registrada toda entrada de notas fiscais referente à compra de suprimentos e afins para a produção de calçados e áreas de apoio.

A idéia é descobrir qual o perfil dos fornecedores que mais geram divergências entre o pedido de compra e a entrega da mercadoria com base em padrões, freqüências e conhecimento novo que as técnicas de *Data Mining* são capazes de fornecer.

Pretende-se utilizar as técnicas de agrupamento e classificação com o auxílio da ferramenta *Weka* e seus algoritmos para obtenção deste conhecimento.

Estas informações poderão ser usadas pela área de suprimentos da empresa para diminuir o número de divergências atuando na sua causa. Dependendo dos resultados, poderão ocorrer mudanças nas negociações com os fornecedores, ajustes no sistema e até mesmo o surgimento de propostas de automatizar processos que ainda estejam sendo executados manualmente.

Durante todo o processo de pesquisa esteve á disposição da aluna um usuário da área de materiais desta empresa que exerce a função de gerente de importação e assessoria de compras, ele possui conhecimento dos dados e do negócio da empresa. A aluna pôde contar também com o apoio do gerente de informática que possui conhecimento da área e tem grande experiência como analista de dados.

Estudo e definição da modelagem de dados

Neste momento do trabalho, inicia-se a etapa de KDD - Seleção de um conjunto de dados alvo. Procurou-se encontrar um atributo foco nos dados que registram todas as

movimentações feitas por diversos tipos de materiais dentro da empresa. Porém, através do estudo em artigos que fizeram uso de *Data Mining* e a aplicação deste processo relatado em Gonchoroski (2007), chegou-se a conclusão de que esta base não seria ideal para os objetivos nos quais este trabalho se propõe. Constatou-se que, para a aplicação de *Data Mining* precisavam-se ter dados que indicassem uma ação, uma descoberta de perfis ou algo que hoje na empresa é visto como um processo que pode ser melhorado através do estudo das causas que o geram.

Foi providenciado então um estudo em cima da base disponível no banco de dados da empresa. Onde através de um contato com o gerente da área de suprimentos, obteve-se um relatório gerado por uma empresa de consultoria externa chamada Accenture. Esta empresa analisou toda a área de suprimentos da Calçados Azaléia no período de julho a setembro de 2005 com o objetivo de qualificar os serviços prestados apontando melhorias nos processos. Este relatório apresenta comparações entre a situação atual da área, suas deficiências e o que pode ser melhorado a partir de pesquisas feitas no banco de dados da empresa, entrevistas a usuários chaves e acompanhamento de processos. Além da parte descritiva, este relatório mostra também através de gráficos, tabelas e tópicos, o perfil da área de compras apontando em seguida as oportunidades e recomendações.

Com base então neste relatório foi identificado um ponto onde a aplicação de *Data Mining* poderia ser bastante útil: a avaliação de fornecedores. Hoje a empresa não tem um processo específico para avaliar seus fornecedores, apenas conta com um *ranking* onde somente é possível relatar os maiores fornecedores com base no valor faturado. Pretende-se, desta forma, utilizar-se de outros atributos para se obter uma forma mais eficiente e realista de avaliá-los.

A partir desta premissa, a aluna passou a pesquisar no sistema quais atributos existentes poderiam ser considerados como atributos válidos no estudo de caso. Em reunião com um dos gerentes da área de suprimentos foram definidos os seguintes itens:

- **Divergência**: Indica se a entrada da mercadoria na empresa está condizente com o que fora solicitado na ordem de compra. Podem ocorrer divergências de prazo, preço, quantidade, condição de entrega e outros. A geração de uma divergência pode ser bastante impactante na cadeia de suprimentos, por exemplo, uma entrega atrasada pode ocasionar parada no processo produtivo pela falta do produto encomendado.

- **EDI (*Eletronic Data Interchange*)**: Indica se o fornecedor possui sistema automatizado de recepção da ordem de compra e envio da nota fiscal e, indica se a nota fiscal de entrada a ser pesquisada foi recebida eletronicamente. Pois é possível que, mesmo que o

fornecedor tenha EDI, a nota de entrada possa não ter vindo eletronicamente ocasionada por falha no processo. Este atributo é considerado bastante importante pela empresa que está trabalhando a vários anos para que todos seus fornecedores se adequem ao processo, pois este traz mais agilidade, rapidez e segurança no tráfego de informações. A Calçados Azaléia é pioneira na implementação do EDI entre as empresas do Vale dos Sinos (GRUPO GOL, 2007). Automatizou seu almoxarifado no ano de 2001 e junto com ele implementou a troca eletrônica de dados com alguns fornecedores voluntários. A partir dessa parceria, outras empresas da região também aderiram ao projeto e então surgiu a necessidade de padronização das mensagens trafegadas, o que originou o Grupo de Otimização Logística – GOL. Maiores detalhes sobre este grupo podem ser verificadas no site www.gol.org.br.

- **Região**: atributo que idenficará a partir do endereço comercial do fornecedor qual é a sua região. Ela pode ser classificada em: Sul, Sudeste, Norte, Nordeste,Centro-Oeste e Exterior (fornecedores de fora do país).

- **Tipo de fornecedor**: este atributo indica qual o tipo de material que o fornecedor comercializa. Pode ser classificado em: ‘matéria-prima’, ‘manutenção e construção’, ‘transporte’, ‘importação’, ‘diversos’ e ‘máquinas e equipamentos’.

- **Devolução**: indicará se a nota fiscal avaliada teve ou não devolução para o fornecedor. Não existe no sistema o motivo pelo qual a mercadoria foi devolvida, mas na maioria dos casos é por defeito e má qualidade no produto.

- **Pedido de compra**: indica se a nota está atendendo um pedido de compra ou não. Entende-se que tudo que um fornecedor fatura para a Azaléia deveria estar apontando para um pedido de compra criado no sistema, pois o registro da compra é entendido como um item de confiança tanto para a empresa em relação à área de suprimentos quanto ao fornecedor que pode comprovar a solicitação. Sem o registro da compra pode-se abrir espaço para roubos e desvios de material.

Passando a executar a etapa do KDD – Pré-processamento de limpeza dos dados, nesta pesquisa foram desconsideradas as empresas do grupo Azaléia que, pelo fato de emitirem notas fiscais para suas coligadas, estas entram no sistema de recebimento da empresa destino considerando que a empresa origem é o fornecedor da nota. Considerá-las distorceria os dados selecionados já que este tipo de nota fiscal tem características diferenciadas das demais notas recebidas de terceiros. Por exemplo, notas emitidas entre empresas do grupo não tem EDI, porque o próprio sistema se encarrega de gerar a nota no recebimento da empresa destino, que por sua vez não gera divergência já que esse tipo de nota não tem pedido de compra. O pedido de compra somente é utilizado para compra de terceiros.

Serão consideradas somente notas fiscais que se referem a compra de materiais tanto do mercado interno quanto do externo, pois a base também é composta por notas de devolução, notas sem valor comercial, notas de transferências (já excluídas no item anterior) e notas de retornos (empréstimos, demonstração, conserto etc.).

O levantamento dos dados deu origem a *query* da figura 3.4, onde neste exemplo selecionam-se apenas dados do ano de 2008 cujas notas já estejam atualizadas (recebidas e armazenadas: status_nf = 5):

```
COLUMN EDI FORMAT A3
COLUMN REGIAO FORMAT A15
COLUMN NF_EDI FORMAT A6
COLUMN DIVER FORMAT A5
COLUMN NF_PED FORMAT A5
COLUMN DEVOL FORMAT A5
COLUMN TIPO_FORNECEDOR FORMAT A25
COLUMN CLASSIFICACAO FORMAT A10
COLUMN VALOR_NF FORMAT 999999
COLUMN QTDE_ITENS_NF FORMAT 999999
SELECT /*+ CHOOSE */
F.POSSUI_EDI EDI,
CRC.RV_MEANING TIPO_FORNECEDOR,
RETORNA_REGIAO(C.UF) REGIAO,
NFR.GERACAO_EDI NF_EDI,
RETORNA_TEMPO_FORNECEDOR(F.DATA_CRIACAO) CLASSIFICACAO,
RETORNA_DIVERGENCIA(NFR.COD_EMPRESA, NFR.NRO_NOTA_FISCAL, NFR.SERIE_NOTA_FISCAL, NFR.COD_SEQ_PESSOA_FISICA ) DIVER,
RETORNA_PEDIDO(NFR.COD_EMPRESA, NFR.NRO_NOTA_FISCAL, NFR.SERIE_NOTA_FISCAL, NFR.COD_SEQ_PESSOA_FISICA ) NF_PED,
RETORNA_DEVOLUCAO(NFR.COD_EMPRESA, NFR.NRO_NOTA_FISCAL, NFR.SERIE_NOTA_FISCAL, NFR.COD_SEQ_PESSOA_FISICA ) DEVOL,
NFR.VALOR_TOTAL VALOR_NF,
RETORNA_QTDE_ITENS(NFR.COD_EMPRESA, NFR.NRO_NOTA_FISCAL, NFR.SERIE_NOTA_FISCAL, NFR.COD_SEQ_PESSOA_FISICA)
QTDE_ITENS_NF
FROM CIDADES C,
     FORNECEDORES F,
     CG_REF_CODES CRC,
     NOTAS_FISCAIS_RECEBIMENTOS NFR
WHERE C.COD_CIDADE = F.COD_CIDADE
  AND F.COD_FORNECEDOR = NFR.COD_FORNECEDOR
  AND CRC.RV_LOW_VALUE = F.TIPO_FORNECEDOR
  AND NFR.DATA_ATUALIZACAO > '01-JAN-2008'
  AND NFR.STATUS_NF = 5
  AND NFR.IDENTIF_NOTA_FISCAL IN ( 1,2,3,7)
  AND CRC.RV_DOMAIN = 'TIPO_FORNECEDOR'
  AND NOT EXISTS ( SELECT 1 FROM PARAMETROS_ESTOQUES PE WHERE PE.COD_FORNECEDOR_CIA = NFR.COD_FORNECEDOR )
/
```

Figura 3.4 – Query executada para selecionar dados

Para executar a *query* acima, foi necessário se conectar com o usuário, senha e banco em *software* de conexão específica para o qual o banco de dados utilizado suporta. Neste exemplo, está se usando um banco de dados *Oracle*, por isso a ferramenta utilizada para rodar a *query* foi o SqlPlus. Após sua execução, a *query* gerou o arquivo texto exibido na figura 3.5:

QUERY1.TXT

EDI	TIPO_FORNECEDOR	REGIAO	NF_EDI	CLASSIFICA	DIVER	NF_PE	DEVOL	VALOR_NF	QTDE_ITENS_NF
N	MATERIA-PRIMA	NORTE	N	ANTIGO	S	S	N	580	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1131	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1131	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1131	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1131	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1131	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	2272	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1136	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1136	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1136	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	2272	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1131	1
N	DIVERSOS	NORTE	N	ANTIGO	S	S	N	1136	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103530	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	420	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	S	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	52080	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	103950	1
N	MATERIA-PRIMA	NORDESTE	N	NOVO	N	S	N	51870	1

Figura 3.5 – Arquivo texto gerado a partir da *query* de pesquisa.

O próximo passo foi abrir este arquivo texto em planilha Excel e salvá-lo como tipo CSV (separado por vírgula). O passo seguinte consistiu em abrir o arquivo “.csv” em editor de texto normal. Nota-se que os dados estão separados por vírgula, pré-requisito do conteúdo do arquivo a ser importado pelo *Weka*. Em seguida, o arquivo foi salvo com a extensão “.arff”. Passou-se então a rotular o conteúdo do arquivo conforme descrito no capítulo 3.2 deste trabalho e exibido na figura 3.6:

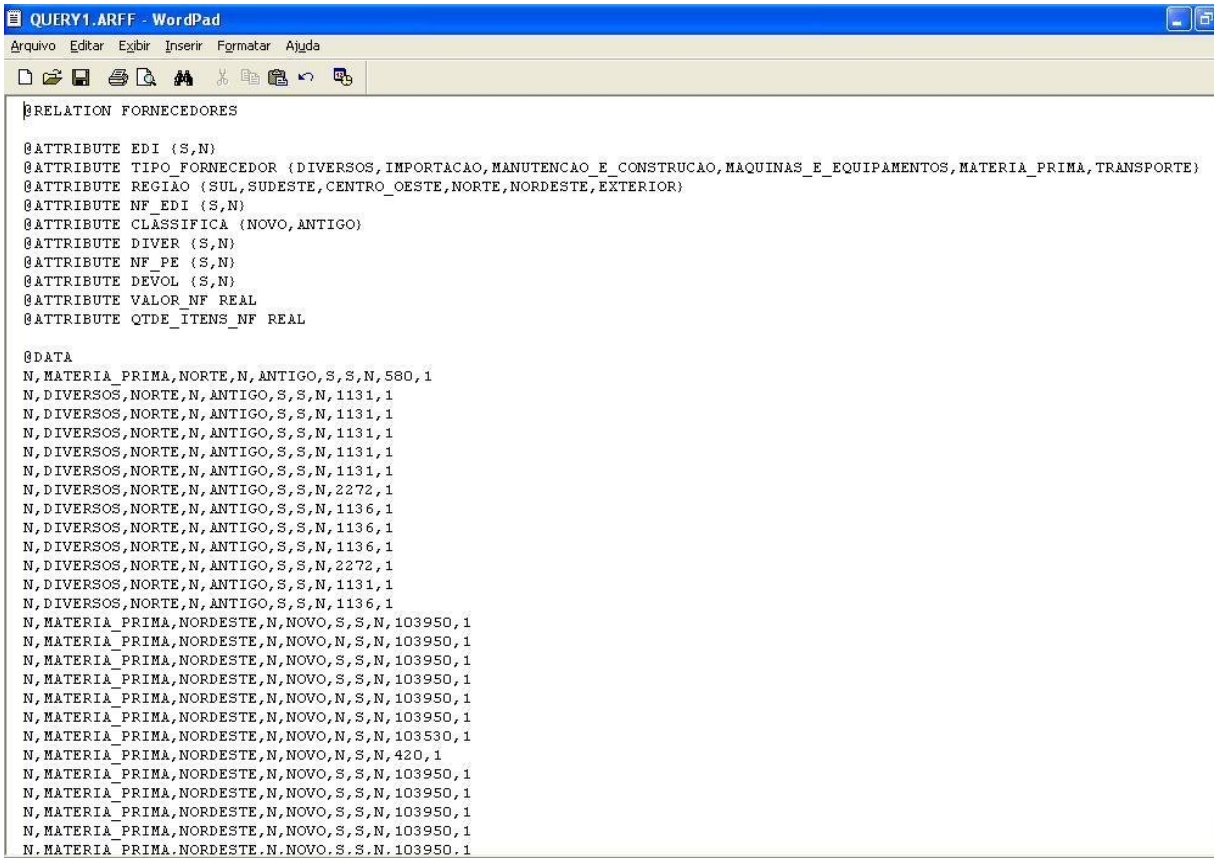

Figura 3.6 – Arquivo no formato arff (conteúdo rotulado)

A partir deste formato o arquivo ficou pronto para ser importado pelo *software Weka*. O processo de transformar os dados em um arquivo padrão refere-se a etapa de KDD – Transformação.

A *query* trabalhada neste momento foi a mesma apresentada acima. Entretanto, selecionou-se um intervalo maior de dados que englobou o ano de 2007 e 2008 até o mês de abril. O resultado desta *query* gerou um arquivo com a extensão ARFF de aproximadamente 160.000 registros que, devido ao seu tamanho não pode ser processado pelas configurações padrões do software *Weka*, ocasionando um erro de falta de memória conforme mostra figura 3.7.

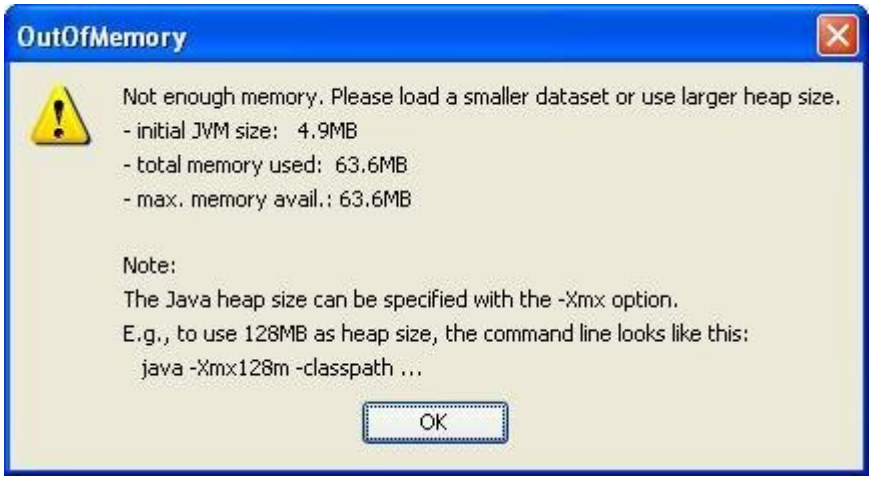

Figura 3.7 – Erro na manipulação de arquivos grandes pelo Weka

Para solucionar este problema, a memória alocada para a máquina virtual Java precisou ser aumentada através do comando a seguir, digitado no *prompt* do MS-Dos:

Java –Xmx512m –classpath weka.jar weka.gui.explorer.Explorer.

Este comando aumentou a memória alocada de 4.9 MB para 512 MB permitindo então a manipulação de arquivos com grandes quantidades de registros.

Aplicou-se então o algoritmo J48 em cima deste arquivo através do software *Weka*. O uso deste algoritmo pode ser visto também em (GONCHOROSKI,2007) e (OLIVEIRA,2002). A partir deste ponto começa-se a utilizar as etapas de KDD chamadas de Mineração e Interpretação/Avaliação, onde os dados são minerados e avaliados logo em seguida a fim de se verificar se a mineração está atingindo o objetivo proposto.

Na próxima seção foi feito um agrupamento dos dados pelo fato de que a árvore gerada pelo algoritmo J48 ficou muita extensa e de difícil compreensão. O algoritmo *SimpleK-Means* foi escolhido para executar esta tarefa.

Agrupamento dos dados

Analisando o anexo I, nota-se que existem muitos dados e possibilidades, fazendo com que o software gere uma árvore com muitos nodos e folhas, causando ao analisador muita confusão e dificuldade na extração do conhecimento.

Com a finalidade de gerar uma árvore menor, optou-se então por aplicar o processo de agrupamento ou *clustering*, a fim de se descobrir o grupo de registros com a maior similaridade. O processo de agrupamento consiste em colocar dentro do mesmo conjunto,

elementos que possuem maior similaridade entre si do que com os elementos dos outros conjuntos. (OCHI, 2008)

Através da ferramenta *Weka* em conjunto com o algoritmo *SimpleK-Means,* foram realizadas várias análises utilizando a mesma base de dados conforme descrição a seguir. Optou-se em utilizar este algoritmo pelo fato dele ser bastante utilizado no meio acadêmico como pode ser visto em (GONCHOROSKI,2007), (SANTOS,2005) e (SCHWERTNER,s.a).

Análise I

Nesta análise optou-se em ignorar um atributo de cada vez a fim de identificar algum que pudesse causar um impacto maior ou menor nos agrupamentos. A figura 3.8 mostra a análise gerada pelo *Weka* ao ser ignorado o campo: EDI.

```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -N 10 -S 10
Relation:     NOTAS_FISCAIS
Instances:    164093
Attributes:   10
              TIPO_FORNECEDOR
              REGIAO
              NF_EDI
              CLASSIFICA
              DIVER
              NF_PE
              DEVOL
              VALOR_NF
              QTDE_ITENS_NF
Ignored:
              EDI
Test mode:    evaluate on training data

=== Model and evaluation on training set ===


kMeans
======
```

```
Number of iterations: 10
Within cluster sum of squared errors: 103859.99201788545

Cluster centroids:

Cluster 0
        Mean/Mode:  DIVERSOS SUL N ANTIGO N N N 14422.7304     1.9739
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
276603.6775     4.378
Cluster 1
        Mean/Mode:  MANUTENCAO_E_CONSTRUCAO NORDESTE N ANTIGO S S N  5408.8221     2.6287
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
19275.7131     3.4531
Cluster 2
        Mean/Mode:  MATERIA_PRIMA SUL S ANTIGO S S N  3298.5382     1.6182
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
8159.1621     0.9028
Cluster 3
        Mean/Mode:  MATERIA_PRIMA SUL S ANTIGO S S N  4465.7288    21.0204
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
6522.3064     7.5241
Cluster 4
        Mean/Mode:  MATERIA_PRIMA SUL S ANTIGO N S N  2543.0693    19.2068
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
3433.9496     5.9231
Cluster 5
        Mean/Mode:  MATERIA_PRIMA SUL S ANTIGO N S N  1487.2808     3.4722
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
5939.7739     2.8758
Cluster 6
        Mean/Mode:  MANUTENCAO_E_CONSTRUCAO SUL N ANTIGO S S N  4617.2488     3.4451
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
23652.2403     6.73
Cluster 7
        Mean/Mode:  DIVERSOS SUL S ANTIGO N N N  2120.8676     5.746
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
8257.2743     3.9215
Cluster 8
        Mean/Mode:  MATERIA_PRIMA NORDESTE S ANTIGO S S N  6272.0769     5.0215
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
10930.7712     4.9578
Cluster 9
        Mean/Mode:  MATERIA_PRIMA SUL S ANTIGO S S N  3774.4993     8.1063
        Std Devs:   N/A         N/A         N/A         N/A         N/A         N/A         N/A
7432.5366     2.6363

Clustered Instances

 0       27813 ( 17%)
 1       11900 (  7%)
```

```
2       38484 ( 23%) +
3        4421 (  3%)
4       13529 (  8%)
5       29188 ( 18%)
6       17257 ( 11%)
7        2130 (  1%) -
8       10538 (  6%)
9        8833 (  5%)
```

Figura 3.8 – análise gerada pelo software *Weka* usando algoritmo *SimpleK-Means*

A letra "K" do algoritmo *SimpleK-Means* significa que se pode parametrizar um número K de *clusters* possíveis a ser gerado em cada processamento. Neste estudo especificamente este número foi parametrizado em K = 10 para todas as análises, já que um número menor geraria *clusters* com dados fora da realidade, pelo fato de terem sido agrupados pela similaridade. E um número maior de clusters, geraria informações muito espalhadas, e os dados estariam pulverizados em vários *clusters* dificultando o entendimento. Foram feitos vários testes com diversos números para o K e chegou-se a conclusão de que a parametrização em K =10 *clusters* atenderia nesta situação a necessidade apontada no início desta seção. A escolha do número 10 se deu pelo fato de ser um número intermediário que facilitou o entendimento e propiciou um número considerável de agrupamentos.

Acompanhando o relatório descrito na figura 3.8, encontra-se nas primeiras linhas: a parametrização referente a quantidade de *clusters* gerados, o nome do conjunto de dados, a quantidade de registros processados, os atributos a serem considerados nesta análise e o atributo ignorado.

Em seguida, o relatório mostra como ficou a distribuição dos atributos em cada *cluster*. Note que o objetivo do agrupamento é escolher os dados por afinidade, logo os valores de cada atributo mencionados neste relatório, se referem aos atributos que são a maioria em cada *cluster*.

E por fim este relatório faz uma síntese de cada *cluster* indicando o percentual de registros agrupados.

Esta análise foi executada 10 vezes, onde em cada uma delas um atributo diferente foi ignorado. Após estas análises terem sido concluídas, foram verificados alguns pontos:

- Em todas as análises realizadas nota-se um conjunto de dados que se destaca. Para os *clusters* que agruparam o maior número de registros, os conteúdos listados abaixo na tabela 3.1 foram encontrados na maioria dos casos.

Tabela 3.1 – conteúdo dos *clusters* que agruparam mais registros

ATRIBUTO	VALOR
EDI	S
TIPO_FORNECEDOR	MATÉRIA-PRIMA
REGIÃO	SUL
NF_EDI	S
CLASSIFICA	ANTIGO
DIVER	S
NF_PE	S
DEVOLUÇÃO	N

- Os campos “valor_nf” e “qtde_itens_nf” foram retirados em análises futuras, pois pulverizam os registros prejudicando seu agrupamento, dada a diversidade de valores existentes. Na figura 3.9, pode ser verificado o conteúdo dos clusters 93 e 95 destacados em negrito. Estes dois *clusters* poderiam ter sido agrupados se os campos referentes ao valor da nota e quantidade de itens não estivessem sendo contemplados, já que o restante dos atributos desses dois *clusters* possui os mesmos conteúdos.

```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -N 100 -S 10
Relation:     NOTAS_FISCAIS
Instances:    164093
Attributes:   10
              EDI
              TIPO_FORNECEDOR
              REGIAO
```

```
              NF_EDI
              CLASSIFICA
              DIVER
              NF_PE
              DEVOL
              VALOR_NF
              QTDE_ITENS_NF
Test mode:    evaluate on training data
=== Model and evaluation on training set ===
kMeans
======
Number of iterations: 40
Within cluster sum of squared errors: 35095.03134333702

Cluster centroids:

Cluster 0
         Mean/Mode:  N DIVERSOS NORDESTE N ANTIGO N N N   5810.2109      1.4547
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          34228.5667      0.8334
Cluster 1
         Mean/Mode:  S DIVERSOS NORDESTE N ANTIGO S S N   8867.451       7.6807
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          21460.0979      8.9197
Cluster 2
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO S S N    316.8371      1
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A           271.4936      0
Cluster 3
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO S S N   1574.7694      2
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          1653.618       0
Cluster 4
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO N S N    967.6471     10
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          1531.3238      0
Cluster 5
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO N S N    719.5039      7
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          1351.8781      0
.
.
.
.
Cluster 88
         Mean/Mode:  S MATERIA_PRIMA SUL N ANTIGO N N N   2780.9702      2.7535
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          8662.5787      3.2724
Cluster 89
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO S S N   5658.1103     34.2051
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
```

```
N/A         N/A           4176.7658      5.9474
Cluster 90
         Mean/Mode:  N DIVERSOS NORDESTE N ANTIGO N N N   7632.8378     10.3923
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          31747.3063      4.5027
Cluster 91
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO N S N    709.6593      4
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A           1825.7058      0
Cluster 92
         Mean/Mode:  S DIVERSOS SUL S ANTIGO N S N   1810.6747      9.6938
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A           8387.9945      6.0306
Cluster 93
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO N S N    690.8705      9
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A           1667.7609      0
Cluster 94
         Mean/Mode:  N MATERIA_PRIMA SUL S ANTIGO N S N   2333.8889      1.2698
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A           4255.6818      0.6012
Cluster 95
         Mean/Mode:  S MATERIA_PRIMA SUL S ANTIGO N S N    914.7274     12
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A            720.628       0
Cluster 96
         Mean/Mode:  N MANUTENCAO_E_CONSTRUCAO SUL N ANTIGO S S N   2356.7319      1.1656
         Std Devs:    N/A           N/A           N/A           N/A           N/A           N/A
N/A         N/A          13288.4655      0.3718

Clustered Instances

 0        5357 (  3%)
 1         592 (  0%)
 2        5960 (  4%)
 3        4181 (  3%)
 4        1074 (  1%)
 5         762 (  0%)
.
.
.
88        2284 (  1%)
89         390 (  0%)
90         339 (  0%)
91        2997 (  2%)
92         787 (  0%)
93         695 (  0%)
94          63 (  0%)
95        1152 (  1%)
96        4088 (  2%)
```

Figura 3.9 – Agrupamento parametrizado em 100 clusters

Nota-se que mesmo parametrizando a geração dos dados em K = 100 *clusters*, o algoritmo utilizou apenas 97 para fazer o agrupamento dos dados, ou seja, o algoritmo conseguiu agrupar os dados em um número menor de clusters do que fora estipulado.

No próxima seção será feita uma nova análise dos dados selecionados, ignorando então os campos referentes ao valor da nota e quantidade dos itens.

Análise II

Partindo-se da mesma *query* utilizada na seção anterior, foram ignorados então os atributos "valor_nf" e "qtde_itens_nf" e aplicadas as mesmas regras de agrupamento: ignorando um atributo de cada vez. Com base nestas verificações montou-se a tabela 3.2.

Tabela 3.2 – Resumo da análise de agrupamento

Análise	EDI	TIPO	REGIÃO	NF_EDI	CLASSIFICAÇÃO	DIVER	NF PED	DEVOL	%	Cluster
1	S	Matéria-prima	Sul	S	Antigo	S	S	N	30	2
						N			27	3
2	X	Matéria-prima	Sul	S	Antigo	S	S	N	32	2
	X					N			26	3
3	S	X	Sul	S	Antigo	S	S	N	32	2
		X				N			28	3
4	S	Matéria-prima	X	S	Antigo	S	S	N	33	2
			X			N			25	3
5	S	Matéria-prima	Sul	X	Antigo	N	S	N	28	3
				X		S			27	2
6	S	Matéria-prima	Sul	S	X	S	S	N	30	2
					X	N			27	3
7	S	Matéria-prima	Sul	S	Antigo	X	S	N	55	2
						X			12	3
8	S	Matéria-prima	Sul	S	Antigo	S	X	N	30	2
						N	X		25	3
9	S	Matéria-prima	Sul	S	Antigo	S	S	X	30	2
						N		X	27	3

Nesta tabela estão registrados os dois clusters que mais agruparam registros em cada análise com seus respectivos atributos, conteúdos, percentuais de agrupamento e número do

cluster selecionado. Os campos marcados com ‘X’, indicam o campo que foi ignorado na análise corrente. Conforme já havia sido constatado na pesquisa anterior, os *clusters* com o maior número de registros agrupados, possuem valores muito similares, onde cada registro, na maioria dos casos, se diferenciou dos demais apenas pelo campo “Diver”. Nota-se que este campo, ao ser ignorado (análise 7 da tabela 3.2), propiciou um maior agrupamento dos registros em relação aos demais campos ignorados em outras análises.

Com base neste ponto de referência, optou-se então em gerar um arquivo Arff contendo somente os registros equivalentes ao *cluster* de maior índice de agrupamento. Pois entende-se que este *cluster* gerará uma árvore menor o que facilitará sua análise e obtenção de conhecimento. Este novo arquivo foi gerado a partir do *software Weka*, na pasta *cluster*, executando os seguintes procedimentos: após o processamento dos dados, com o botão direito clicou-se no relatório existente no *box* “*result list*” e em seguida foi escolhida a opção “*visualize cluster assignments*” conforme mostra figura 3.10.

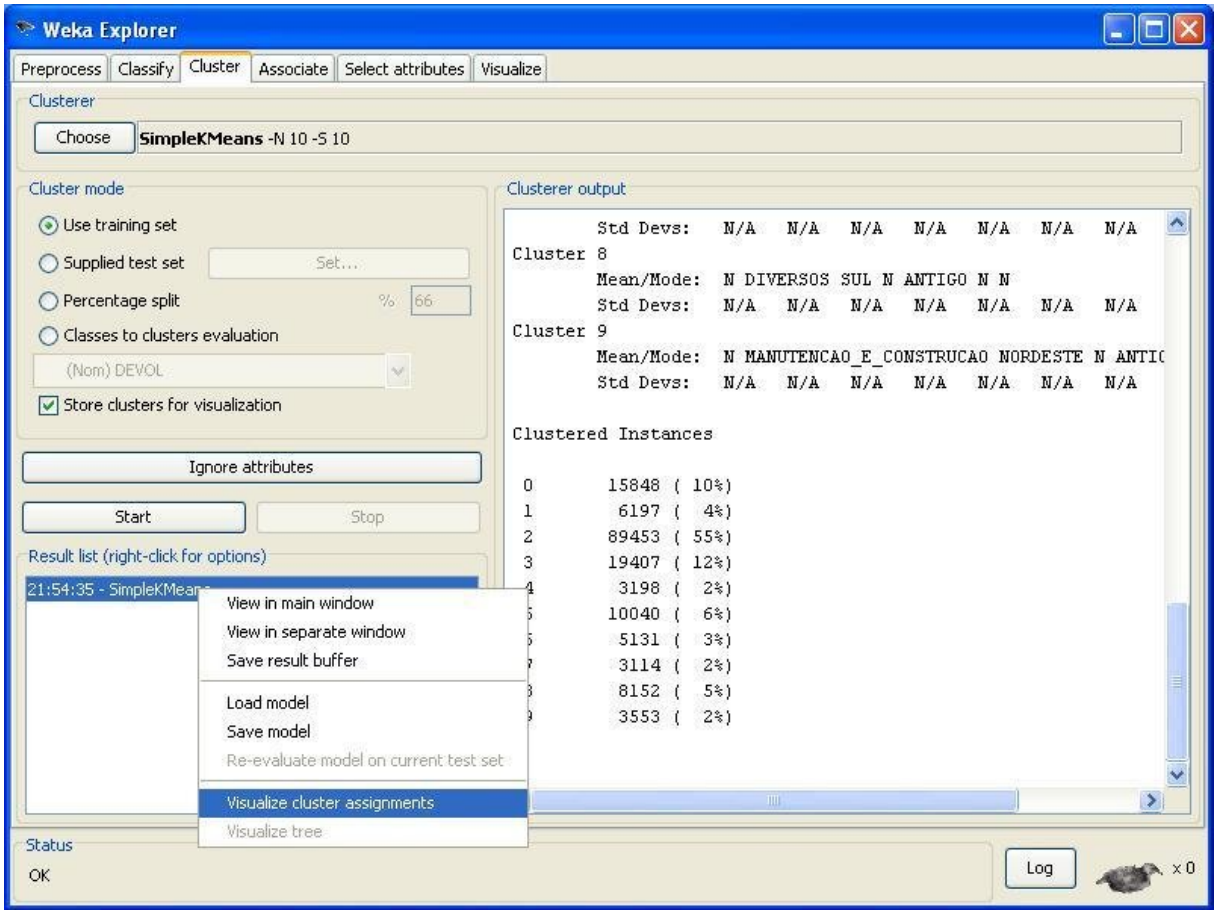

Figura 3.10 – Tela clusterização

Este procedimento gerou um arquivo com a extensão “.Arff” com todos os registros processados indicando para cada um em qual *cluster* o mesmo foi classificado. Após, foi

necessário separar, com ajuda do *software* Microsoft Excel, somente os registros pertencentes ao *cluster*2, *cluster* que, em nossa análise, alcançou o maior número de registros similares.

Este arquivo ficou com 89453 registros cujo conteúdo dos campos é similar aos conteúdos mencionados na tabela 3.2 para a análise 7 e *cluster* 2.

Entende-se através das análises feitas até agora que o atributo alvo seja o campo “Diver” pelo fato dele causar grande impacto nas análises, principalmente ao ser ignorado na análise 7 da tabela 3.2. Nota-se que o algoritmo conseguiu agrupar 55% dos registros em um único cluster (cluster 2).

Outro fator também descoberto, é que este campo é o único que tem seu conteúdo diferenciado entre os dois maiores clusters de cada análise. Olhando para a realidade da empresa, entende-se que estudar os perfis dos fornecedores que geram divergências pode ser um fator auxiliador no trabalho que já vem sendo executado na busca da redução do número de divergências. Acredita-se que revelando tais perfis, pode-se iniciar um trabalho mais focado em um grupo específico de fornecedores. Por exemplo, se o número de divergências tende a ser menor em fornecedores que trabalham com EDI, revela-se então mais um motivo para buscar junto aos fornecedores, a implementação da troca eletrônica de dados .

No próximo capítulo será visto a aplicação do algoritmo J48 nesta base de dados gerada pelo processo de agrupamento.

APLICAÇÃO DO ALGORITMO J48

O algoritmo J48.J48, mais popular algoritmo da Weka, é responsável por montar um modelo de árvore de decisão baseado em um conjunto de dados de treinamento (OLIVEIRA,2002).

Partindo então dos dados gerados conforme descrito no capítulo anterior, aplicou-se então o algoritmo J48. Esta classificação gerou o relatório visualizado na figura 4.1.

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     NOTAS_FISCAIS_CLUSTER2
Instances:    89453
Attributes:   8
              EDI
              TIPO_FORNECEDOR
              REGIAO
              NF_EDI
              CLASSIFICA
              DIVER
              NF_PE
              DEVOL
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree
------------------

REGIAO = SUL
|   EDI = S
|   |   NF_EDI = S
|   |   |   NF_PE = S
|   |   |   |   TIPO_FORNECEDOR = DIVERSOS: N (5523.0/2332.0)
|   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (5941.0/2441.0)
|   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (20.0/5.0)
```

```
|   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   CLASSIFICA = NOVO: S (68.0/23.0)
|   |   |   |   |   CLASSIFICA = ANTIGO
|   |   |   |   |   |   DEVOL = S: S (1112.0/445.0)
|   |   |   |   |   |   DEVOL = N: N (58495.0/27407.0)
|   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   NF_PE = N: S (1482.0/426.0)
|   |   NF_EDI = N
|   |   |   NF_PE = S: S (5087.0/1168.0)
|   |   |   NF_PE = N
|   |   |   |   DEVOL = S: S (16.0/1.0)
|   |   |   |   DEVOL = N
|   |   |   |   |   CLASSIFICA = NOVO: S (14.0/1.0)
|   |   |   |   |   CLASSIFICA = ANTIGO: N (1680.0/516.0)
|   EDI = N: S (590.0/63.0)
REGIAO = SUDESTE: S (9425.0/843.0)
REGIAO = CENTRO_OESTE: S (0.0)
REGIAO = NORTE: S (0.0)
REGIAO = NORDESTE: S (0.0)
REGIAO = EXTERIOR: S (0.0)

Number of Leaves  :       19

Size of the tree :        29

Time taken to build model: 3.61 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances       53782               60.1232 %
Incorrectly Classified Instances     35671               39.8768 %
Kappa statistic                          0.2335
Mean absolute error                      0.4477
Root mean squared error                  0.4731
Relative absolute error                 90.2132 %
Root relative squared error             94.9806 %
Total Number of Instances            89453

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0.377     0.133      0.772     0.377     0.507    S
  0.867     0.623      0.539     0.867     0.665    N

=== Confusion Matrix ===

     a     b   <-- classified as
 18339 30255 |     a = S
  5416 35443 |     b = N
```

Figura 4.1 – Relatório gerado pelo processamento do algoritmo J48

Analisando este relatório pode ser verificado que foram usados todos os dados como base para o treinamento. Em seguida, encontra-se a própria árvore gerada pelo algoritmo em sua forma linear, o número de folhas e o tamanho da árvore. Pode ser verificado também o tempo de construção do modelo, um resumo estatístico dessa árvore e por fim a matriz de confusão.

Note no resumo que em torno de 60% dos dados foram classificados de forma correta contra 40% de forma incorreta, o que nos revela uma árvore ruim pelo fato do número de classificações incorretas, ser muito expressivo.

Na figura 4.2 pode-se verificar a árvore gerada em sua forma *top-down.*

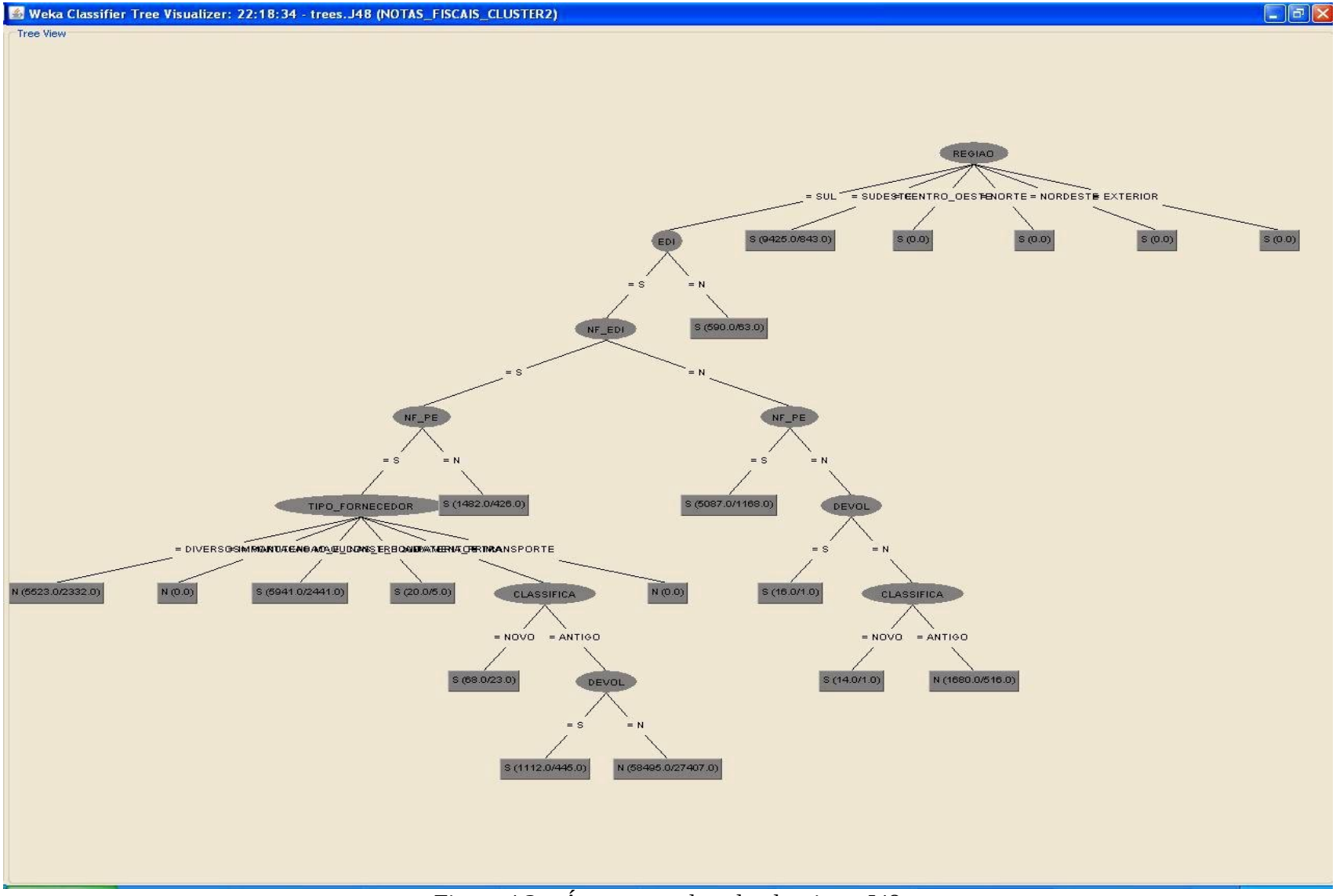

Figura 4.2 – Árvore gerada pelo algoritmo J48

Mesmo com um pré-processamento dos dados através do agrupamento, pôde-se verificar ainda uma árvore muito pulverizada, onde os dados encontram-se espalhados demais dificultando a análise. Foram executadas outras tentativas de gerar uma árvore mais confiável, mas não houve sucesso. Nos casos onde campos de forma aleatória foram ignorados, os percentuais de erro somente aumentaram. Também não houve redução dos erros quando utilizou-se o modo de teste e treinamento *cross-validation* (que separa os registros em número N de *folders* e usa um para treinamento e os demais para teste e assim sucessivamente até que todos os *folders* são treinados).

Apesar de atingir um nível considerável de erros, foi possível extrair algum conhecimento da árvore gerada conforme descrito abaixo e ilustrado na tabela 4.1:

Tabela 4.1 – Resultado obtidos a partir de uma árvore de decisão

%	Região	Tipo	Novo	EDI	OC	NF EDI	Devolução	Divergência
91.05	Sudeste							Sim
89.32	Sul			Não				Sim
71.26	Sul			Sim	Não	Sim		Sim
77.04	Sul			Sim	Sim	Não		Sim
66.18	Sul	Matéria-prima	Sim	Sim	Sim	Sim		Sim
57.78	Sul	Diversos		Sim	Sim	Sim		Não
69.29	Sul		Não	Sim	Não	Não	Não	Não

- 91.056% das notas emitidas de fornecedores da região Sudeste, **geram divergência**;

- 89.32% das notas emitidas de fornecedores da região Sul que não possuem EDI, **geram divergência**;

- 71.26% das notas emitidas de fornecedores da região Sul que possuem EDI cuja nota foi enviada eletronicamente sem pedido de compra, **geram divergência**;

- 77.04% das notas emitidas de fornecedores da região Sul que possuem EDI cuja nota não foi enviada eletronicamente, mas tem pedido de compra associado, **geram divergência**;

- 66.18% das notas emitidas por novos fornecedores de matéria-prima da região Sul que tem EDI, cuja nota fiscal está associada a um pedido de compra e foi enviada eletronicamente, **geram divergência**;

- Fornecedores da região Sul classificados como Diversos que possuem EDI, cuja nota foi enviada eletronicamente e está associada a um pedido de compra, apenas 57.78% **não geram divergência;**

- 69.29% das notas emitidas por fornecedores antigos da região Sul que possuem EDI mas que a nota não entrou por EDI, não tem pedido de compra e não houve devolução, **não geram divergência.**

Os percentuais acima calculados levaram em consideração o número de registros classificados em cada nodo terminal cujo cálculo é definido pela divisão do número de registros incorretos pelo número de registros classificados, multiplicados por 100 e o resultado subtraído de 100. Por exemplo, no primeiro item citado acima o cálculo ficou assim: 100 – (843/9425*100)=91.056.

Com base nas descobertas citadas acima pode-se elencar alguns itens para se trabalhar no projeto da redução de divergências, no qual a empresa já vem executando alguns procedimentos. Esta é a etapa final do KDD – Consolidação do conhecimento descoberto, onde tenta se colocar em prática, o conhecimento adquirido na fase de mineração:

- O fato de um fornecedor não ter implementado a troca eletrônica de dados é um dos causadores da divergência, então pode-se reforçar junto aos fornecedores que ainda não aderiram ao projeto que o façam o quanto antes. O projeto de convencer os fornecedores a implementarem o EDI já está em andamento na empresa já faz alguns anos, pode-se então incluir na lista de benefícios, o fato de haver uma redução no número de divergências.

- Outro ponto a ser trabalhado são os novos fornecedores que mesmo tendo EDI ainda geram as notas com muitas divergências. Uma sugestão é incluir no contrato com estes fornecedores a exigência de que a nota deve vir com as

mesmas informações enviadas no pedido de compra que, entende-se ser as informações negociadas anteriormente entre as partes, ou mesmo pensar em um plano de punição com multas ou prorrogações automáticas do prazo de pagamento.

- Sugere-se também fazer um trabalho de redução de divergências por regiões que, facilitado pela geografia, realizar um *Workshop* explicativo a fim de que os fornecedores tomem conhecimento do real impacto que causa dentro da empresa o fato da nota fiscal vir com divergências em relação ao pedido de compra.

O processo de DM revelou alguns detalhes e curiosidades sobre os perfis dos fornecedores que eram desconhecidos e que poderão ser usados tanto no projeto de avaliação de fornecedores quanto no projeto de redução das divergências. O projeto de avaliação de fornecedores ainda não foi iniciado, mas o de reduzir as divergências está em andamento e já poderá contar com os dados aqui descobertos.

CONCLUSÃO

Pode-se afirmar que gerar dados apenas com a finalidade de registrar as transações executadas por uma empresa, pode indicar que a empresa esteja deixando de ganhar tanto em valores monetários quanto na eficiência dos processos. As técnicas de *Data Mining* mostraram que é possível resgatar conhecimento em cima de grandes massas de dados e usar este conhecimento para alavancar os negócios de uma empresa, ou repensar estratégias de relacionamento com seus clientes e fornecedores. Nota-se também que por mais tecnologias que sejam usadas para automatizar processos, a avaliação de um especialista, tanto no decorrer do processo quanto nos resultados, continua sendo imprescindível.

Acredita-se que aplicar as técnicas de descoberta de conhecimento na base de dados de Calçados Azaléia S. A. foi um processo de aprendizado tanto para a aluna quanto para os usuários envolvidos que até então, tinham o *Data Mining* apenas como um conceito. O processo revelou algumas informações novas que poderão ser usadas no projeto de redução do número de divergências entre o pedido de compra e a nota fiscal. Como trabalhos futuros sugere-se implementar o projeto de avaliação de fornecedores, que parametrizados com seu grau de avaliação, pode-se novamente aplicar as técnicas de DM e descobrir novos perfis levando em consideração os já revelados por este trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

CABENA, Peter et al. **Discovering Data Mining from concept to implementation**. New Jersey, USA: Prentice Hall Ptr, 1998.

CARVALHO, Luis Alfredo Vidal de. **DataMining** : A Mineração de Dados no Marketing , Medicina, Economia, Engenharia e Administração. Rio de Janeiro: Ciência Moderna, 2005.

CARVALHO, Prof. André Ponce de Leon F. de. *Redes Neurais Artificiais*. S.a. Disponível em: <http://www.icmc.usp.br/~andre/research/neural/index.htm>. Acesso em 23 out. 2007.

FAYYAD, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic. *From Data Mining to Knowledge Discovery in Databases.* 1996. Disponível em: <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf.> Acesso em 02 out. 2007.

GRUPO GOL. *Introdução ao EDI*. Downloads, 2007. Disponível em: <www.gol.org.br/downloads/guia_edi.pdf>. Acesso em 21 Nov. 2007.

GONÇALVES, Eduardo Corrêa. Extração de árvores de decisão com a ferramenta de Data Mining Weka. Disponível em <http://www.devmedia.com.br/articles/viewcomp.asp?comp=3388> Acesso em 16 out. 2007.

GONCHOROSKI, Sidinei Pereira. Utilização das técnicas de KDD em um Call Center Ativo. Novo Hamburgo:2007. 125 p.

GUARDA, Álvaro. *Aprendizado de máquina: árvore de decisão Indutiva*. S.a. Disponível em: <http://www.decom.ufop.br/prof/guarda/CIC250/ArvoreDecisaoIndutiva.pdf> Acesso em 02 out. 2007.

HARRISON, Thomas H. **Intranet Data Warehouse.** São Paulo: Berkeley Brasil, 1998.

LUDWIG Jr., Oswaldo; Montgomey, Eduard. **Redes Neurais** : Fundamentos e aplicações com programas em C. Rio de Janeiro : Editora Ciência Moderna Ltda, 2007.

OCHI, Luiz Satoru; Dias, Carlos Rodrigo. Soares, Stênio S. Furtado.*Clusterização em Mineração de Dados*. 2008. Disponível em <http://www.sbc.org.br/bibliotecadigital/download.php?paper=37>. Acesso em 29 de abril de 2008.

OLIVEIRA, Fernando Luiz de et al. Utilização de algoritmos simbólicos para a identificação do número de caroços do fruto Pequi, In: IV Encontro de estudantes de informática do estado de Tocantins, 2002, Palmas. **Encontro de Estudantes de Informática do Tocantins – Ecoinfo**. Palmas, 2002. p. 34-43.

OLIVEIRA, Alessandra Marchiori de; Smiderle, Andréia. Mineração de dados: um estudo de caso de concessão de crédito explorando o software Weka. **Revista de Informática Mater Dei,** Pato Branco, Paraná : v.2, n.2, p.17-22, 2005.

ORALLO, José Hernández; Ramirez, Cèsar Ferri. Práctica de Minéria de datos: Introducción el Weka. Curso de doctorado extracción automática de conocimiento am bases de datos e Ingeniería del software - Universitat Politècnica de València, Março, 2006.

PICHILIANI, Mauro. *Data Mining na prática: Árvores de Decisão*. 27 de novembro de 2006. Disponível em: <http://www.imasters.com.br/artigo/5130/sql_server/data_mining_na_pratica_arvores_de_decisao/> . Acesso em 14 out. 2007.

QUONIAN, Luc; Tarapanoff, Kira;Júnior, Rogério Henrique de Araújo; Álvares, Lílian. *Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil* . 2001. Disponível em <http://www.scielo.br/pdf/ci/v30n2/6208.pdf>. Acesso em agosto de 2007.

SANTOS, David Moises Barreto dos. Seleção de Modelos de Classificação através de Heurísticas. Campina Grande: Julho 2005. 99p.

SCHWERTNER, Marco Antônio; Rigo, Sandro José; Oliveira, José Palazzo M. de. Mineração de uso em sistema de informação na Web. São Leopoldo. S.a. 10p.

WITTEN, Ian H.; Franck, Eibe. **Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations**. Hamilton, New Zealand: Morgan Kaufmann Publishers, 2000. p. 265-320.

VIANA, Reinaldo. Mineração de dados : Introdução e Aplicações. **Sql Magazine**, Rio de Janeiro : v.10, n.1, p.16-25, 2004.

X ESCOLA de Informática da SBC-Sul. **ERI 2002**. 13 a 17 maio, 2002. Caxias do Sul, Criciúma, Cascavel: s.e.

ANEXO I

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

NF_PE = S
|   QTDE_ITENS_NF <= 5
|   |   VALOR_NF <= 549
|   |   |   QTDE_ITENS_NF <= 3
|   |   |   |   VALOR_NF <= 53
|   |   |   |   |   NF_EDI = S
|   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: N (1630.0/365.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   |   |   |   |   VALOR_NF <= 20: S (160.0/26.0)
|   |   |   |   |   |   |   |   VALOR_NF > 20: N (179.0/69.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 1: S (85.0/25.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (1.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   REGIAO = SUL
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   |   |   |   |   |   VALOR_NF <= 22: N (1230.0/474.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 22
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 30: S (299.0/99.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 30
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 34
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 32
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 31: N (39.0/17.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 31: S (23.0/7.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 32: N (73.0/18.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 34
|   |   |   |   |   |   |   |   |   |   |   |   EDI = S: S (593.0/246.0)
|   |   |   |   |   |   |   |   |   |   |   |   EDI = N: N (11.0/4.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 1
|   |   |   |   |   |   |   |   |   CLASSIFICA = NOVO: S (2.0)
|   |   |   |   |   |   |   |   |   CLASSIFICA = ANTIGO: N (1485.0/414.0)
|   |   |   |   |   |   |   REGIAO = SUDESTE: S (9.0/2.0)
```

```
|   |   |   |   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   |   |   |   REGIAO = NORTE: N (0.0)
|   |   |   |   |   |   |   REGIAO = NORDESTE: S (419.0/118.0)
|   |   |   |   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   |   NF_EDI = N
|   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: S (866.0/311.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: S (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 2: S (986.0/425.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 2
|   |   |   |   |   |   |   |   REGIAO = SUL: N (25.0/9.0)
|   |   |   |   |   |   |   |   REGIAO = SUDESTE: N (0.0)
|   |   |   |   |   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   |   |   |   |   REGIAO = NORTE: N (0.0)
|   |   |   |   |   |   |   |   REGIAO = NORDESTE
|   |   |   |   |   |   |   |   |   VALOR_NF <= 23: S (2.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 23: N (3.0/1.0)
|   |   |   |   |   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: N (24.0/9.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 2
|   |   |   |   |   |   |   |   EDI = S: S (378.0/134.0)
|   |   |   |   |   |   |   |   EDI = N
|   |   |   |   |   |   |   |   |   VALOR_NF <= 33: S (34.0/7.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 33
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 39: N (19.0/4.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 39: S (32.0/5.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 2: N (24.0/10.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (1.0)
|   |   |   |   VALOR_NF > 53
|   |   |   |   |   REGIAO = SUL
|   |   |   |   |   |   NF_EDI = S
|   |   |   |   |   |   |   VALOR_NF <= 223
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: N (596.0/242.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: S (0.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   |   |   |   |   |   |   DEVOL = S: S (19.0/8.0)
|   |   |   |   |   |   |   |   |   DEVOL = N
|   |   |   |   |   |   |   |   |   |   CLASSIFICA = NOVO
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 1: N (12.0/4.0)
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 1
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 118: N (3.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 118: S (8.0/1.0)
|   |   |   |   |   |   |   |   |   |   CLASSIFICA = ANTIGO: N (994.0/477.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (2.0/1.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 174
|   |   |   |   |   |   |   |   |   |   |   DEVOL = S: N (19.0/6.0)
|   |   |   |   |   |   |   |   |   |   |   DEVOL = N: S (1975.0/723.0)
```

```
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 174
|   |   |   |   |   |   |   |   |   |   |   EDI = S
|   |   |   |   |   |   |   |   |   |   |   |   DEVOL = S
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 209: S (6.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 209: N (3.0)
|   |   |   |   |   |   |   |   |   |   |   |   DEVOL = N: N (734.0/361.0)
|   |   |   |   |   |   |   |   |   |   |   EDI = N: S (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 1
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 68: N (227.0/82.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 68
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 218
|   |   |   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   DEVOL = S: N (20.0/7.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   DEVOL = N: S (734.0/326.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 92: N (116.0/37.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 92: S (300.0/140.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 218
|   |   |   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 2: N (40.0/7.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 220: N (4.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 220: S (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: S (0.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 223
|   |   |   |   |   |   |   |   |   |   CLASSIFICA = NOVO: N (57.0/26.0)
|   |   |   |   |   |   |   |   |   |   CLASSIFICA = ANTIGO: S (5619.0/2075.0)
|   |   |   |   |   |   |   NF_EDI = N: S (7409.0/2497.0)
|   |   |   |   |   |   REGIAO = SUDESTE: S (2020.0/363.0)
|   |   |   |   |   |   REGIAO = CENTRO_OESTE: S (0.0)
|   |   |   |   |   |   REGIAO = NORTE: S (72.0/20.0)
|   |   |   |   |   |   REGIAO = NORDESTE
|   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS
|   |   |   |   |   |   |   |   DEVOL = S
|   |   |   |   |   |   |   |   |   VALOR_NF <= 159: N (3.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 159: S (5.0/1.0)
|   |   |   |   |   |   |   |   DEVOL = N
|   |   |   |   |   |   |   |   |   EDI = S
|   |   |   |   |   |   |   |   |   |   CLASSIFICA = NOVO: N (73.0/26.0)
|   |   |   |   |   |   |   |   |   |   CLASSIFICA = ANTIGO
|   |   |   |   |   |   |   |   |   |   |   NF_EDI = S: S (455.0/190.0)
|   |   |   |   |   |   |   |   |   |   |   NF_EDI = N: N (111.0/49.0)
|   |   |   |   |   |   |   |   |   EDI = N: S (1262.0/516.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: S (0.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (1515.0/556.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (60.0/14.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   |   EDI = S: S (2586.0/771.0)
|   |   |   |   |   |   |   |   EDI = N: N (48.0/22.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: S (0.0)
|   |   |   |   |   |   REGIAO = EXTERIOR: S (0.0)
|   |   |   QTDE_ITENS_NF > 3
```

```
|   |   |   |   NF_EDI = S
|   |   |   |   |   REGIAO = SUL
|   |   |   |   |   |   VALOR_NF <= 165
|   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: N (409.0/38.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   |   |   |   |   |   VALOR_NF <= 122: S (32.0/6.0)
|   |   |   |   |   |   |   |   VALOR_NF > 122: N (12.0/4.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: N (1.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA: N (1947.0/398.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   |   |   VALOR_NF > 165
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 4
|   |   |   |   |   |   |   |   EDI = S
|   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: S (31.0/11.0)
|   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: N (100.0/46.0)
|   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (1.0)
|   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA: N (1153.0/364.0)
|   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   |   |   |   |   EDI = N: S (4.0/1.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 4
|   |   |   |   |   |   |   |   VALOR_NF <= 346
|   |   |   |   |   |   |   |   |   DEVOL = S: S (5.0)
|   |   |   |   |   |   |   |   |   DEVOL = N
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: S (12.0/6.0)
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S
(20.0/5.0)
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (1.0)
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA: N (196.0/64.0)
|   |   |   |   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   |   |   |   |   VALOR_NF > 346: S (167.0/49.0)
|   |   |   |   |   REGIAO = SUDESTE: S (109.0/9.0)
|   |   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   |   REGIAO = NORTE: N (0.0)
|   |   |   |   |   REGIAO = NORDESTE
|   |   |   |   |   |   CLASSIFICA = NOVO
|   |   |   |   |   |   |   VALOR_NF <= 119: S (6.0/1.0)
|   |   |   |   |   |   |   VALOR_NF > 119: N (34.0/14.0)
|   |   |   |   |   |   CLASSIFICA = ANTIGO: S (156.0/65.0)
|   |   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   |   |   NF_EDI = N: S (976.0/307.0)
|   |   VALOR_NF > 549: S (51602.0/11089.0)
|   QTDE_ITENS_NF > 5
|   |   REGIAO = SUL
|   |   |   NF_EDI = S
|   |   |   |   VALOR_NF <= 4198
|   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS
|   |   |   |   |   |   VALOR_NF <= 199: N (639.0/82.0)
|   |   |   |   |   |   VALOR_NF > 199
```

```
|   |   |   |   |   |   |   DEVOL = S: N (13.0/4.0)
|   |   |   |   |   |   |   DEVOL = N: S (424.0/93.0)
|   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (572.0/178.0)
|   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (11.0/1.0)
|   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   VALOR_NF <= 761: N (8601.0/1606.0)
|   |   |   |   |   |   VALOR_NF > 761
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 14
|   |   |   |   |   |   |   |   VALOR_NF <= 1940
|   |   |   |   |   |   |   |   |   DEVOL = S
|   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 11: S (73.0/18.0)
|   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 11
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 917: N (6.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 917: S (14.0/5.0)
|   |   |   |   |   |   |   |   |   DEVOL = N: N (3487.0/1129.0)
|   |   |   |   |   |   |   |   VALOR_NF > 1940
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 11: S (1146.0/442.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 11
|   |   |   |   |   |   |   |   |   |   DEVOL = S: S (22.0/4.0)
|   |   |   |   |   |   |   |   |   |   DEVOL = N
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 12: N (288.0/94.0)
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 12
|   |   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 13: S (128.0/55.0)
|   |   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 13: N (166.0/54.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 14
|   |   |   |   |   |   |   |   DEVOL = S
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 30: S (92.0/37.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 30: N (6.0)
|   |   |   |   |   |   |   |   DEVOL = N
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 44
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 3211: N (7964.0/1689.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 3211
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 15: S (55.0/21.0)
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 15: N (1470.0/377.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 44
|   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 48
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 2008: N (5.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 2008
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 2563: S (25.0/3.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 2563: N (4.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 48: N (12.0)
|   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   VALOR_NF > 4198
|   |   |   |   |   QTDE_ITENS_NF <= 15
|   |   |   |   |   |   QTDE_ITENS_NF <= 11: S (1248.0/221.0)
|   |   |   |   |   |   QTDE_ITENS_NF > 11
|   |   |   |   |   |   |   VALOR_NF <= 6278
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: S (10.0/1.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: S (0.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (2.0)
```

```
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (0.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   |   |   DEVOL = S: S (10.0/1.0)
|   |   |   |   |   |   |   |   |   DEVOL = N
|   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 12: N (124.0/53.0)
|   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 12
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 14: S (90.0/34.0)
|   |   |   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 14: N (80.0/37.0)
|   |   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: S (0.0)
|   |   |   |   |   |   |   VALOR_NF > 6278: S (299.0/56.0)
|   |   |   |   |   QTDE_ITENS_NF > 15
|   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: S (63.0/6.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (9.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   VALOR_NF <= 6839
|   |   |   |   |   |   |   |   DEVOL = S
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 21: S (24.0/3.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 21: N (16.0/2.0)
|   |   |   |   |   |   |   |   DEVOL = N: N (1873.0/681.0)
|   |   |   |   |   |   |   VALOR_NF > 6839
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 27: S (763.0/274.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 27: N (298.0/81.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   NF_EDI = N
|   |   |   |   VALOR_NF <= 781
|   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: S (108.0/19.0)
|   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: S (0.0)
|   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   |   |   |   CLASSIFICA = NOVO
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 8
|   |   |   |   |   |   |   |   VALOR_NF <= 518: S (5.0/1.0)
|   |   |   |   |   |   |   |   VALOR_NF > 518: N (5.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 8: S (6.0/1.0)
|   |   |   |   |   |   CLASSIFICA = ANTIGO: S (326.0/103.0)
|   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (5.0)
|   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   EDI = S
|   |   |   |   |   |   |   VALOR_NF <= 210
|   |   |   |   |   |   |   |   DEVOL = S
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 11: N (3.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 11: S (5.0/1.0)
|   |   |   |   |   |   |   |   DEVOL = N: N (88.0/30.0)
|   |   |   |   |   |   |   VALOR_NF > 210: S (175.0/59.0)
|   |   |   |   |   |   EDI = N: S (58.0/12.0)
|   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: S (0.0)
|   |   |   |   VALOR_NF > 781
|   |   |   |   |   QTDE_ITENS_NF <= 22: S (1738.0/281.0)
|   |   |   |   |   QTDE_ITENS_NF > 22
|   |   |   |   |   |   VALOR_NF <= 26061
```

```
|   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS
|   |   |   |   |   |   |   |   EDI = S: S (10.0/1.0)
|   |   |   |   |   |   |   |   EDI = N: N (31.0/9.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: S (0.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (13.0/1.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: S (0.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 27: S (36.0/15.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 27
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 41: N (12.0/1.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 41: S (3.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: S (0.0)
|   |   |   |   |   |   VALOR_NF > 26061: S (60.0/7.0)
|   |   REGIAO = SUDESTE: S (1512.0/130.0)
|   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   REGIAO = NORTE: S (36.0/2.0)
|   |   REGIAO = NORDESTE
|   |   |   VALOR_NF <= 3851
|   |   |   |   NF_EDI = S
|   |   |   |   |   CLASSIFICA = NOVO
|   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS: N (48.0/12.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: N (0.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA: S (10.0)
|   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   |   CLASSIFICA = ANTIGO
|   |   |   |   |   |   VALOR_NF <= 261
|   |   |   |   |   |   |   TIPO_FORNECEDOR = DIVERSOS
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 6: S (6.0/1.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 6
|   |   |   |   |   |   |   |   |   VALOR_NF <= 147: S (5.0/2.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 147: N (7.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO: S (1.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS: N (0.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   |   |   |   |   |   VALOR_NF <= 215: N (14.0/1.0)
|   |   |   |   |   |   |   |   VALOR_NF > 215: S (2.0)
|   |   |   |   |   |   |   TIPO_FORNECEDOR = TRANSPORTE: N (0.0)
|   |   |   |   |   |   VALOR_NF > 261: S (815.0/269.0)
|   |   |   |   NF_EDI = N: S (745.0/195.0)
|   |   |   VALOR_NF > 3851
|   |   |   |   QTDE_ITENS_NF <= 21: S (2577.0/300.0)
|   |   |   |   QTDE_ITENS_NF > 21
|   |   |   |   |   NF_EDI = S
|   |   |   |   |   |   QTDE_ITENS_NF <= 31
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 25: S (22.0/7.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 25: N (9.0/1.0)
|   |   |   |   |   |   QTDE_ITENS_NF > 31: S (5.0)
|   |   |   |   |   NF_EDI = N: S (41.0/2.0)
```

```
|   |   REGIAO = EXTERIOR: S (136.0/4.0)
NF_PE = N
|   EDI = S
|   |   TIPO_FORNECEDOR = DIVERSOS
|   |   |   DEVOL = S: S (15.0/4.0)
|   |   |   DEVOL = N
|   |   |   |   QTDE_ITENS_NF <= 11
|   |   |   |   |   VALOR_NF <= 545
|   |   |   |   |   |   NF_EDI = S
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   |   |   |   |   REGIAO = SUL: S (51.0/20.0)
|   |   |   |   |   |   |   |   REGIAO = SUDESTE: N (11.0/4.0)
|   |   |   |   |   |   |   |   REGIAO = CENTRO_OESTE: S (0.0)
|   |   |   |   |   |   |   |   REGIAO = NORTE: S (0.0)
|   |   |   |   |   |   |   |   REGIAO = NORDESTE: N (2.0)
|   |   |   |   |   |   |   |   REGIAO = EXTERIOR: S (0.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 1
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 5: N (130.0/19.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 5
|   |   |   |   |   |   |   |   |   VALOR_NF <= 439: S (33.0/8.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 439: N (8.0)
|   |   |   |   |   |   NF_EDI = N: N (506.0/58.0)
|   |   |   |   |   VALOR_NF > 545: N (855.0/71.0)
|   |   |   |   QTDE_ITENS_NF > 11
|   |   |   |   |   VALOR_NF <= 3200: S (67.0/14.0)
|   |   |   |   |   VALOR_NF > 3200: N (94.0/2.0)
|   |   TIPO_FORNECEDOR = IMPORTACAO: N (0.0)
|   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   VALOR_NF <= 15: N (25.0/6.0)
|   |   |   |   VALOR_NF > 15
|   |   |   |   |   REGIAO = SUL: S (716.0/92.0)
|   |   |   |   |   REGIAO = SUDESTE: S (11.0/2.0)
|   |   |   |   |   REGIAO = CENTRO_OESTE: S (0.0)
|   |   |   |   |   REGIAO = NORTE: S (0.0)
|   |   |   |   |   REGIAO = NORDESTE: N (7.0/1.0)
|   |   |   |   |   REGIAO = EXTERIOR: S (0.0)
|   |   |   QTDE_ITENS_NF > 1
|   |   |   |   VALOR_NF <= 244
|   |   |   |   |   VALOR_NF <= 72: N (113.0/14.0)
|   |   |   |   |   VALOR_NF > 72
|   |   |   |   |   |   NF_EDI = S
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 2: N (47.0/11.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 2
|   |   |   |   |   |   |   |   VALOR_NF <= 119: S (11.0/2.0)
|   |   |   |   |   |   |   |   VALOR_NF > 119: N (26.0/5.0)
|   |   |   |   |   |   NF_EDI = N
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 6
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 2: N (92.0/42.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 2: S (34.0/8.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 6
```

```
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 9: N (6.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 9
|   |   |   |   |   |   |   |   |   VALOR_NF <= 218: S (4.0/1.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 218: N (2.0)
|   |   |   |   VALOR_NF > 244
|   |   |   |   |   QTDE_ITENS_NF <= 5
|   |   |   |   |   |   QTDE_ITENS_NF <= 2
|   |   |   |   |   |   |   REGIAO = SUL: S (579.0/217.0)
|   |   |   |   |   |   |   REGIAO = SUDESTE
|   |   |   |   |   |   |   |   VALOR_NF <= 1677: N (6.0/2.0)
|   |   |   |   |   |   |   |   VALOR_NF > 1677: S (2.0)
|   |   |   |   |   |   |   REGIAO = CENTRO_OESTE: S (0.0)
|   |   |   |   |   |   |   REGIAO = NORTE: S (0.0)
|   |   |   |   |   |   |   REGIAO = NORDESTE: N (2.0)
|   |   |   |   |   |   |   REGIAO = EXTERIOR: S (0.0)
|   |   |   |   |   |   QTDE_ITENS_NF > 2
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 4
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 3: S (208.0/45.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 3
|   |   |   |   |   |   |   |   |   VALOR_NF <= 670: N (55.0/20.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 670: S (82.0/10.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 4: S (42.0/3.0)
|   |   |   |   |   QTDE_ITENS_NF > 5
|   |   |   |   |   |   VALOR_NF <= 1296: N (82.0/24.0)
|   |   |   |   |   |   VALOR_NF > 1296: S (126.0/36.0)
|   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS
|   |   |   QTDE_ITENS_NF <= 11: N (56.0/4.0)
|   |   |   QTDE_ITENS_NF > 11: S (2.0)
|   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   NF_EDI = S: S (1755.0/452.0)
|   |   |   NF_EDI = N
|   |   |   |   REGIAO = SUL
|   |   |   |   |   DEVOL = S: S (16.0/1.0)
|   |   |   |   |   DEVOL = N
|   |   |   |   |   |   CLASSIFICA = NOVO: S (14.0/1.0)
|   |   |   |   |   |   CLASSIFICA = ANTIGO
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 2: N (1236.0/316.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 2
|   |   |   |   |   |   |   |   VALOR_NF <= 114: N (59.0/13.0)
|   |   |   |   |   |   |   |   VALOR_NF > 114
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 8: N (314.0/141.0)
|   |   |   |   |   |   |   |   |   QTDE_ITENS_NF > 8
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 1429
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 568: S (16.0/4.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 568: N (18.0/4.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 1429: S (37.0/7.0)
|   |   |   |   REGIAO = SUDESTE
|   |   |   |   |   QTDE_ITENS_NF <= 1: S (132.0/22.0)
|   |   |   |   |   QTDE_ITENS_NF > 1
|   |   |   |   |   |   QTDE_ITENS_NF <= 5: S (30.0/9.0)
|   |   |   |   |   |   QTDE_ITENS_NF > 5: N (5.0)
```

```
|   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   REGIAO = NORTE: N (0.0)
|   |   |   |   REGIAO = NORDESTE
|   |   |   |   |   QTDE_ITENS_NF <= 1: S (162.0/12.0)
|   |   |   |   |   QTDE_ITENS_NF > 1
|   |   |   |   |   |   QTDE_ITENS_NF <= 2: N (58.0/23.0)
|   |   |   |   |   |   QTDE_ITENS_NF > 2: S (33.0/6.0)
|   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   TIPO_FORNECEDOR = TRANSPORTE: N (50.0/2.0)
|   EDI = N
|   |   TIPO_FORNECEDOR = DIVERSOS
|   |   |   VALOR_NF <= 665: N (6564.0/303.0)
|   |   |   VALOR_NF > 665
|   |   |   |   VALOR_NF <= 1080
|   |   |   |   |   REGIAO = SUL
|   |   |   |   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   |   |   |   VALOR_NF <= 761
|   |   |   |   |   |   |   |   VALOR_NF <= 755
|   |   |   |   |   |   |   |   |   VALOR_NF <= 723
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 719
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 685
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 683: N (33.0/5.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 683: S (19.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 685: N (47.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 719: S (76.0/6.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 723: N (56.0/3.0)
|   |   |   |   |   |   |   |   VALOR_NF > 755: S (144.0/11.0)
|   |   |   |   |   |   |   VALOR_NF > 761
|   |   |   |   |   |   |   |   VALOR_NF <= 834
|   |   |   |   |   |   |   |   |   VALOR_NF <= 791: N (33.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 791
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 798: S (16.0/4.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 798: N (57.0/4.0)
|   |   |   |   |   |   |   |   VALOR_NF > 834
|   |   |   |   |   |   |   |   |   VALOR_NF <= 988
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 987
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 972
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 862
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 836: S (13.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 836: N (37.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 862
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 864: S (55.0/2.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 864
|   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 898: N (33.0/2.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 898
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 971
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    VALOR_NF <= 953: S
(105.0/39.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 953: N (13.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 971: S (12.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 972: N (34.0)
```

```
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 987: S (38.0/2.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 988
|   |   |   |   |   |   |   |   |   |   VALOR_NF <= 1007: N (78.0/1.0)
|   |   |   |   |   |   |   |   |   |   VALOR_NF > 1007
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 1008: S (14.0)
|   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 1008
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF <= 1079: N (76.0/24.0)
|   |   |   |   |   |   |   |   |   |   |   |   VALOR_NF > 1079: S (9.0/1.0)
|   |   |   |   |   |   QTDE_ITENS_NF > 1: N (209.0/20.0)
|   |   |   |   |   REGIAO = SUDESTE: N (271.0/21.0)
|   |   |   |   |   REGIAO = CENTRO_OESTE: N (10.0)
|   |   |   |   |   REGIAO = NORTE: N (3.0/1.0)
|   |   |   |   |   REGIAO = NORDESTE: N (568.0/20.0)
|   |   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   |   |   VALOR_NF > 1080: N (6840.0/285.0)
|   |   TIPO_FORNECEDOR = IMPORTACAO
|   |   |   QTDE_ITENS_NF <= 2: N (469.0/28.0)
|   |   |   QTDE_ITENS_NF > 2
|   |   |   |   VALOR_NF <= 155807: N (36.0/9.0)
|   |   |   |   VALOR_NF > 155807: S (7.0)
|   |   TIPO_FORNECEDOR = MANUTENCAO_E_CONSTRUCAO
|   |   |   DEVOL = S: S (21.0/2.0)
|   |   |   DEVOL = N
|   |   |   |   REGIAO = SUL: N (1585.0/322.0)
|   |   |   |   REGIAO = SUDESTE
|   |   |   |   |   CLASSIFICA = NOVO
|   |   |   |   |   |   VALOR_NF <= 5073
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 3
|   |   |   |   |   |   |   |   QTDE_ITENS_NF <= 1
|   |   |   |   |   |   |   |   |   VALOR_NF <= 799: N (4.0)
|   |   |   |   |   |   |   |   |   VALOR_NF > 799: S (4.0)
|   |   |   |   |   |   |   |   QTDE_ITENS_NF > 1: N (4.0/1.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 3: S (2.0)
|   |   |   |   |   |   VALOR_NF > 5073: S (11.0)
|   |   |   |   |   CLASSIFICA = ANTIGO: N (286.0/42.0)
|   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   REGIAO = NORTE: S (6.0/1.0)
|   |   |   |   REGIAO = NORDESTE: N (2484.0/186.0)
|   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   TIPO_FORNECEDOR = MAQUINAS_E_EQUIPAMENTOS
|   |   |   QTDE_ITENS_NF <= 2: N (192.0/20.0)
|   |   |   QTDE_ITENS_NF > 2
|   |   |   |   REGIAO = SUL: S (151.0/14.0)
|   |   |   |   REGIAO = SUDESTE: N (1.0)
|   |   |   |   REGIAO = CENTRO_OESTE: S (0.0)
|   |   |   |   REGIAO = NORTE: S (0.0)
|   |   |   |   REGIAO = NORDESTE: N (37.0/2.0)
|   |   |   |   REGIAO = EXTERIOR: N (3.0/1.0)
|   |   TIPO_FORNECEDOR = MATERIA_PRIMA
|   |   |   DEVOL = S: S (24.0/1.0)
|   |   |   DEVOL = N
```

```
|   |   |   |   REGIAO = SUL
|   |   |   |   |   NF_EDI = S: S (3.0)
|   |   |   |   |   NF_EDI = N: N (841.0/304.0)
|   |   |   |   REGIAO = SUDESTE
|   |   |   |   |   VALOR_NF <= 309
|   |   |   |   |   |   VALOR_NF <= 58
|   |   |   |   |   |   |   QTDE_ITENS_NF <= 1: N (3.0/1.0)
|   |   |   |   |   |   |   QTDE_ITENS_NF > 1: S (4.0/1.0)
|   |   |   |   |   |   VALOR_NF > 58: N (18.0/1.0)
|   |   |   |   |   VALOR_NF > 309: S (90.0/32.0)
|   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   REGIAO = NORTE: N (0.0)
|   |   |   |   REGIAO = NORDESTE
|   |   |   |   |   VALOR_NF <= 1315: N (47.0/1.0)
|   |   |   |   |   VALOR_NF > 1315: S (8.0/1.0)
|   |   |   |   REGIAO = EXTERIOR: N (0.0)
|   |   TIPO_FORNECEDOR = TRANSPORTE
|   |   |   QTDE_ITENS_NF <= 7: N (2680.0/9.0)
|   |   |   QTDE_ITENS_NF > 7
|   |   |   |   REGIAO = SUL: S (2.0)
|   |   |   |   REGIAO = SUDESTE: S (1.0)
|   |   |   |   REGIAO = CENTRO_OESTE: N (0.0)
|   |   |   |   REGIAO = NORTE: N (0.0)
|   |   |   |   REGIAO = NORDESTE: N (30.0)
|   |   |   |   REGIAO = EXTERIOR: N (0.0)

Number of Leaves  :       344

Size of the tree :    567
```

ANEXO II

NOTAS_FISCAIS_RECEBIMENTOS

Name	Null?	Type
COD_EMPRESA	NOT NULL	NUMBER(3)
NRO_NOTA_FISCAL	NOT NULL	NUMBER(6)
SERIE_NOTA_FISCAL	NOT NULL	VARCHAR2(4)
COD_SEQ_PESSOA_FISICA	NOT NULL	NUMBER(6)
IDENTIF_NOTA_FISCAL	NOT NULL	NUMBER(2)
ESPECIE_DOCMTO	NOT NULL	NUMBER(3)
STATUS_NF	NOT NULL	NUMBER(1)
COD_FORNECEDOR		NUMBER(5)
COD_CLIENTE		NUMBER(5)
DATA_EMISSAO	NOT NULL	DATE
DATA_ENTRADA	NOT NULL	DATE
COD_TRANSPORTADORA		NUMBER(5)
FRETE		NUMBER(1)
QTD_VOLUMES		NUMBER(5)
PESO_BRUTO		NUMBER(10,3)
PESO_LIQUIDO		NUMBER(10,3)
COD_UM_PESO		NUMBER(2)
PERC_FRETE		NUMBER(10,6)
BASE_ICMS		NUMBER(18,6)
VALOR_ICMS		NUMBER(18,6)
BASE_ICMS_SUBST		NUMBER(18,6)
VALOR_ICMS_SUBST		NUMBER(18,6)
PERC_DESCONTO		NUMBER(9,6)
VALOR_PRODUTOS		NUMBER(18,6)
VALOR_FRETE		NUMBER(18,6)
VALOR_FRETE_CF		NUMBER(18,6)
VALOR_ICMS_FRETE_CF		NUMBER(18,6)
VALOR_SEGURO		NUMBER(18,6)
BASE_IPI		NUMBER(18,6)
VALOR_IPI		NUMBER(18,6)
VALOR_DESCONTO		NUMBER(18,6)
VALOR_DESP_ACESSORIAS		NUMBER(18,6)
VALOR_DESP_FINANCEIRAS		NUMBER(18,6)
VALOR_DESP_ADUANEIRAS		NUMBER(18,6)
VALOR_TOTAL		NUMBER(18,6)

VALOR_INSUMOS		NUMBER(18,6)
COD_DESC_FINANCEIRO		NUMBER(1)
VALOR_DESC_FINANCEIRO		NUMBER(18,6)
OBSERVACAO_FINANCEIRO		VARCHAR2(100)
COD_CONDICAO_PAGTO		NUMBER(2)
ANO_MES_RECEBIMENTO	NOT NULL	NUMBER(6)
OBSERVACAO_DEVOLUCAO		VARCHAR2(255)
CONTABILIDADE_ATUALIZADA		VARCHAR2(1)
DATA_ATUALIZACAO		DATE
RECIBO_ENTREGA		VARCHAR2(1)
TIPO_BOLETIM		NUMBER(1)
COD_DEPOSITO		NUMBER(3)
COD_FABRICA		NUMBER(3)
COD_SETOR		NUMBER(2)
TURNO_CENTRO_CUSTO		NUMBER(1)
COD_LOCALIZACAO		NUMBER(6)
ATUALIZA_SQLFISCAL		VARCHAR2(1)
VALOR_TOTAL_SEGURIDADE_SOCIAL		NUMBER(18,6)
GERACAO_EDI	NOT NULL	VARCHAR2(1)
DATA_RECEBIMENTO		DATE
COD_FUNCIONARIO_RECEBIMENTO		NUMBER(6)
COD_TRANSPORTADORA_RECEBIMENTO		NUMBER(3)
HORA_ENTREGA		DATE
VALOR_TOTAL_IR_FONTE		NUMBER(18,6)
USUARIO_ENTRADA		VARCHAR2(8)
HORA_ENTRADA		DATE
INTEGRACAO_FATURA		VARCHAR2(1)
DEVOLUCAO		NUMBER(1)
DATA_DEVOLUCAO		DATE
VALOR_ICMS_PAGO_ANTECIPADO		NUMBER(18,6)
VALOR_TOTAL_ISSQN		NUMBER(18,6)
LIBERADO_EDI		VARCHAR2(1)
NRO_DI		NUMBER(10)
NRO_AIDF		VARCHAR2(10)
DATA_AIDF		DATE
BASE_CALCULO_INSS		NUMBER(18,6)
LIBERADO_ARMAZENAMENTO		VARCHAR2(1)
VALOR_RETENCAO_CONSOLIDADO		NUMBER(18,6)
NRO_PO		NUMBER(6)
DIAS_PRORROG_MULTA		NUMBER(3)
COD_MOTIVO_DEVOLUCAO		NUMBER(3)
CHAVE_ACESSO_NF_E		VARCHAR2(44)
NRO_PROTOCOLO_NF_E		NUMBER(15)
NRO_RECIBO_NF_E		NUMBER(15)

ITENS_NOTAS_RECEBIMENTOS

Name	Null?	Type
COD_EMPRESA	NOT NULL	NUMBER(3)
NRO_NOTA_FISCAL	NOT NULL	NUMBER(6)
SERIE_NOTA_FISCAL	NOT NULL	VARCHAR2(4)
COD_SEQ_PESSOA_FISICA	NOT NULL	NUMBER(6)
NRO_ITEM_NOTA	NOT NULL	NUMBER(3)

STATUS_ITEM	NOT NULL	NUMBER(1)
COD_CFO	NOT NULL	NUMBER(4)
COD_MATERIAL_ALTERN		NUMBER(9)
COD_MATERIAL		NUMBER(9)
COD_CONTA_CONTABIL		NUMBER(6)
COD_CLASS_FISCAL	NOT NULL	NUMBER(4)
COD_ORIGEM_MERCADORIA	NOT NULL	NUMBER(1)
COD_TRIBUTACAO_ICMS	NOT NULL	NUMBER(2)
COD_TRIBUTACAO_IPI	NOT NULL	NUMBER(2)
COD_SIT_TRIBUT_FEDERAL	NOT NULL	NUMBER(5)
COD_UM_ENTRADA	NOT NULL	NUMBER(2)
FATOR_CONVERSAO		NUMBER(10,3)
CREDITA_ICMS	NOT NULL	VARCHAR2(1)
CREDITA_IPI	NOT NULL	VARCHAR2(1)
COD_FABRICA		NUMBER(3)
COD_SETOR		NUMBER(2)
TURNO_CENTRO_CUSTO		NUMBER(1)
COD_CONTA_ORCAMENTO		NUMBER(3)
TIPO_PRODUTO	NOT NULL	NUMBER(2)
OBJETIVO_MERCADORIA	NOT NULL	NUMBER(3)
TIPO_REMESSA	NOT NULL	NUMBER(3)
QTD_RECEBIDA_ENTR		NUMBER(14,3)
QTD_REAL_RECEBIDA_ENTR		NUMBER(14,3)
QTD_REAL_RECEBIDA_ESTQ		NUMBER(14,3)
QTD_RECEBIDA_LOTE		NUMBER(14,3)
PRECO_UNITARIO		NUMBER(18,6)
PERC_DESCONTO		NUMBER(9,6)
VALOR_DESCONTO		NUMBER(18,6)
VALOR_ITEM		NUMBER(18,6)
VALOR_DESCONTO_RATEADO		NUMBER(18,6)
VALOR_IR_FONTE		NUMBER(18,6)
PERC_DESP_FINANCEIRAS		NUMBER(9,6)
VALOR_DESP_FINANCEIRAS		NUMBER(18,6)
PERC_DESP_ACESSORIAS		NUMBER(9,6)
VALOR_DESP_ACESSORIAS		NUMBER(18,6)
VALOR_DESP_ADUANEIRAS		NUMBER(18,6)
BASE_IPI		NUMBER(18,6)
PERC_IPI		NUMBER(9,6)
VALOR_IPI		NUMBER(18,6)
VALOR_TOTAL_ITEM		NUMBER(18,6)
VALOR_FRETE_RATEADO		NUMBER(18,6)
VALOR_ICMS_FRETE_RATEADO		NUMBER(18,6)
VALOR_OUTRAS_DESPESAS		NUMBER(18,6)
PERC_DIFERENCIAL_ICMS		NUMBER(9,6)
VALOR_DIFERENCIAL_ICMS		NUMBER(18,6)
PERC_REDUCAO_ICMS		NUMBER(9,6)
BASE_ICMS		NUMBER(18,6)
BASE_ICMS_SUBST		NUMBER(18,6)
PERC_ICMS		NUMBER(9,6)
VALOR_ICMS		NUMBER(18,6)
VALOR_ICMS_SUBST		NUMBER(18,6)
COD_EMPRESA_NF_REMESSA		NUMBER(3)
NRO_NOTA_FISCAL_REMESSA		NUMBER(6)
SERIE_NOTA_FISCAL_REMESSA		VARCHAR2(4)

PRECO_UNITARIO_INSUMOS	NUMBER(18,6)
VALOR_INSUMOS	NUMBER(18,6)
NRO_PROTOCOLO	NUMBER(7)
CONSUMO_DIRETO	VARCHAR2(1)
PERC_SEGURIDADE_SOCIAL	NUMBER(9,6)
VALOR_SEGURIDADE_SOCIAL	NUMBER(18,6)
COD_FORN_PROCEDENCIA	NUMBER(5)
DATA_ATUALIZACAO	DATE
VALOR_ICMS_PAGO_ANTECIPADO	NUMBER(18,6)
VALOR_ISSQN	NUMBER(18,6)
COD_TRANSACAO_CONSUMO	NUMBER(2)
COD_CONTA_CONTABIL_CONSUMO	NUMBER(6)
COD_CFO_SAIDA	NUMBER(4)
VALOR_PIS	NUMBER(18,6)
COD_PROMOCAO	NUMBER(2)
COD_LINHA	NUMBER(3)
COD_REFERENCIA	NUMBER(3)
COD_COR	NUMBER(6)
COD_GRADE	NUMBER(2)
NRO_CALCADO	NUMBER(3,1)
BASE_CALCULO_INSS	NUMBER(18,6)
VALOR_COFINS	NUMBER(18,6)
VALOR_RETENCAO_CONTRIB_SOCIAL	NUMBER(18,6)
VALOR_RETENCAO_CONSOLIDADO	NUMBER(18,6)
VALOR_RETENCAO_COFINS	NUMBER(18,6)
VALOR_RETENCAO_PIS	NUMBER(18,6)
VALOR_FRETE	NUMBER(18,6)
VALOR_FUNRURAL	NUMBER(18,6)
COD_EMPRESA_NF_SAIDA	NUMBER(3)
NRO_NOTA_SAIDA	NUMBER(6)
NRO_SERIE_NF_SAIDA	VARCHAR2(4)
NRO_AUTORIZACAO_IMOB	NUMBER(5)