

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

LUÍS FELIPE SCHNEIDER

**Aplicação do Processo de Descoberta de Conhecimento em Dados do
Poder Judiciário do Estado do Rio Grande do Sul**

Trabalho de Conclusão de Mestrado
apresentado como requisito parcial para a
obtenção do grau de Mestre em Informática

Prof. Luís Otávio Álvares
Orientador

Porto Alegre, dezembro de 2003

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Schneider, Luís Felipe

Aplicação do Processo de Descoberta de Conhecimento em Dados do Poder Judiciário do Estado do Rio Grande do Sul / Luís Felipe Schneider. - Porto Alegre: Programa de Pós-Graduação em Computação, 2003.

103p.: il.

Trabalho de Conclusão (Mestrado) - Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2003. Orientador: Prof. Luis Otávio Alvares

1.Descoberta de Conhecimento (DCBD) 2.Mineração de Dados 3.Agrupamento 4.Associação 5.*Apriori* 6.*K-médias* 7.*Weka* 8.Poder Judiciário 9.Classificação Processual. I.Álvares, Luis Otavio. II.Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitora: Profa. Wrana Panizzi

Pró-Reitor de Ensino: Prof. José Carlos Ferraz Hennemann

Pró-Reitora Adjunta de Pós-Graduação: Profa. Jocélia Grazia

Diretor do Instituto de Informática: Prof. Phillippe Olivier Alexandre Navaux

Coordenador do PPGC: Prof. Carlos Alberto Heuser

Bibliotecária – Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Gostaria de fazer um agradecimento a todos aqueles que de alguma forma contribuíram para esta realização.

Ao professor Luis Otávio Alvares, que sempre esteve empenhado em me orientar, procurando com que sua experiência e conhecimento pudessem auxiliar na condução deste trabalho.

À minha esposa e família, que sempre estiveram comigo me apoiando em todos os momentos, obrigado pelo incentivo e carinho.

A todos os amigos e colegas, pelo apoio, descontração e amizade.

Meus agradecimentos à administração do Tribunal de Justiça do Estado do Rio Grande do Sul que permitiu que eu utilizasse suas bases de dados para tal objetivo.

SUMÁRIO

LISTA DE FIGURAS.....	6
LISTA DE TABELAS.....	7
RESUMO.....	8
ABSTRACT.....	9
1 INTRODUÇÃO.....	10
1.1 Motivação.....	10
1.2 Apresentação.....	11
2 DESCOBERTA DO CONHECIMENTO.....	12
2.1 Conceito.....	12
2.2 Processo de DCBD.....	12
2.2.1 Etapas do Processo de DCBD.....	13
2.3 Metodologia CRISP – DM.....	14
2.3.1 Entendimento do Negócio.....	15
2.3.2 Entendimento dos Dados.....	16
2.3.3 Preparação de Dados.....	17
2.3.4 Mineração.....	18
2.3.5 Avaliação.....	19
2.3.6 Disponibilização.....	19
3 MINERAÇÃO DE DADOS.....	21
3.1 Conceito.....	21
3.2 Classes de Tipo de Problema e Técnicas de Mineração de Dados.....	21
3.2.1 Classe de Tipo de Problema – Agrupamento.....	22
3.2.2 Classe de Tipo de Problema – Associação.....	23
4 APLICAÇÃO - PODER JUDICIÁRIO DO RS.....	28
4.1 Entendimento do Negócio.....	28
4.1.1 Determinação dos objetivos do negócio.....	28
4.1.2 Avaliação da Situação.....	30
4.1.3 Definição dos Objetivos para Mineração de Dados.....	31

4.1.4	Escolha da Classe de Tipo de Problema e Técnica de Mineração de Dados.....	31
4.1.5	Ferramenta Escolhida.....	32
4.2	Entendimento dos Dados.....	32
4.2.1	Coleção Inicial de Dados.....	32
4.2.2	Modelo Lógico de Dados.....	33
4.2.3	Exploração dos Dados.....	34
4.3	Preparação de dados.....	34
4.3.1	Seleção e Mescla de Dados.....	34
4.3.2	Limpeza de Dados.....	36
4.3.3	Construção de Dados.....	36
4.3.4	Formatação de Dados.....	37
4.4	Mineração.....	38
4.4.1	Mineração de dados.....	38
4.4.2	Resultados da Mineração de Dados.....	39
4.5	Avaliação.....	40
4.5.1	Avaliação dos Resultados.....	40
4.5.2	Revisão do processo.....	54
4.6	Disponibilização.....	56
4.6.1	Revisão do projeto.....	56
5	CONCLUSÃO.....	58
	REFERÊNCIAS.....	60
	ANEXO A DESCRIÇÃO DOS DADOS.....	62
	ANEXO B VALORES DOS ATRIBUTOS.....	66
	ANEXO C ESTATÍSTICAS PRINCIPAIS.....	74
	ANEXO D RESULTADOS DA MINERAÇÃO - TÉCNICA DE SEGMENTAÇÃO.....	86
	ANEXO E RESULTADOS DA MINERAÇÃO – TÉCNICA REGRAS DE ASSOCIAÇÃO.....	97

LISTA DE FIGURAS

Figura 2.1: Processo de DCBD.	13
Figura 2.2: Metodologia CRISP-DM	15
Figura 4.1: Resumo da Estrutura Corporativa do Poder Judiciário.....	29
Figura 4.2: Modelo Lógico de Dados.....	33
Figura 4.3: Seção x Classe x Especialização.....	42
Figura 4.4: Comarca x Seção x Classe x Especialização	43
Figura 4.5: Resultados - Classificação x Tempo - Justiça Comum.....	45
Figura 4.6: Classificação x Comarca x Tempo – Justiça Comum.....	47
Figura 4.7: Classificação x Tempo - Juizados Especiais.....	48
Figura 4.8: Resultados Classificação x Comarca x Tempo - Juizados Especiais.....	49
Figura 4.9: Classificação Processual x Sentença Proferida.....	50
Figura 4.10: Resultados Classificação Processual x Audiência	51
Figura 4.11: Perfil do Réu em Processos Criminais.....	52
Figura 4.12: Perfil do Réu x Natureza.....	54

LISTA DE TABELAS

Tabela 3.1: Espaço de Busca	26
Tabela 3.2: Conjunto de Transações.....	26
Tabela 3.3: Ocorrências de cada item.....	26
Tabela 3.4: Espaço de Busca Filtrado	27
Tabela 3.5: Ocorrências do Conjunto de Itens	27
Tabela 4.1: Pessoas e áreas envolvidas na aplicação de DCBD.....	31
Tabela 4.2: Relação de comarcas escolhidas.....	33
Tabela 4.3: Estrutura da Consulta - Objetivo 1a	35
Tabela 4.4: Estrutura da Consulta - Objetivo 1b	35
Tabela 4.5: Estrutura da Consulta - Objetivo 1c	35
Tabela 4.6: Estrutura da Consulta - Objetivo 1e	35
Tabela 4.7: Estrutura da Consulta - Objetivo 2	36
Tabela 4.8: Construção - Tempos de Tramitação.....	36
Tabela 4.9: Construção - Idade.....	36
Tabela 4.10: Consulta Access.....	37
Tabela 4.11: Formato <i>.cvs</i>	37
Tabela 4.12: Formato <i>.arff</i>	38
Tabela 4.13: Testes com 8 Clusters variando o <i>Seed</i>	40
Tabela 4.14: Intervalos de Tempo - Justiça Comum.....	45
Tabela 4.15: Intervalos de Tempos - Juizados Especiais	48
Tabela 4.16: Intervalos - Idade	51
Tabela 4.17: Resumo dos resultados dos testes 1,2,4 e 5 do Anexo E.....	55
Tabela 4.18: Resumo dos resultados dos testes 3 do Anexo E	56

RESUMO

Para explorar as relações existentes entre os dados abriu-se espaço para a procura de conhecimento e informações úteis não conhecidas, a partir de grandes conjuntos de dados armazenados. A este campo deu-se o nome de Descoberta de Conhecimento em Base de Dados (DCBD), o qual foi formalizado em 1989. O DCBD é composto por um processo de etapas ou fases, de natureza iterativa e interativa. Este trabalho baseou-se na metodologia CRISP-DM .

Independente da metodologia empregada, este processo tem uma fase que pode ser considerada o núcleo da DCBD, a “mineração de dados” (ou modelagem conforme CRISP-DM), a qual está associado o conceito “classe de tipo de problema”, bem como as técnicas e algoritmos que podem ser empregados em uma aplicação de DCBD. Destacaremos as classes associação e agrupamento, as técnicas associadas a estas classes, e os algoritmos *Apriori* e *K-médias*. Toda esta contextualização estará compreendida na ferramenta de mineração de dados escolhida, *Weka (Waikato Environment for Knowledge Analysis)*.

O plano de pesquisa está centrado em aplicar o processo de DCBD no Poder Judiciário no que se refere a sua atividade fim, julgamentos de processos, procurando por descobertas a partir da influência da classificação processual em relação à incidência de processos, ao tempo de tramitação, aos tipos de sentenças proferidas e a presença da audiência. Também, será explorada a procura por perfis de réus, nos processos criminais, segundo características como sexo, estado civil, grau de instrução, profissão e raça. O trabalho apresenta nos capítulos 2 e 3 o embasamento teórico de DCBD, detalhando a metodologia CRISP-DM. No capítulo 4 explora-se toda a aplicação realizada nos dados do Poder Judiciário e por fim, no capítulo 5, são apresentadas as conclusões.

Palavras-chave: Descoberta de Conhecimento (DCBD), Mineração de Dados, Agrupamento, Associação, *Apriori*, *K-médias*, *Weka*, Poder Judiciário, Classificação Processual.

Applying the Knowledge Discovery in Database (KDD) Process to Data of the Judiciary Power of Rio Grande do Sul

ABSTRACT

With the purpose of exploring existing connections among data, a space has been created for the search of Knowledge an useful unknown information based on large sets of stored data. This field was dubbed Knowledge Discovery in Databases (KDD) and it was formalized in 1989. The KDD consists of a process made up of iterative and interactive stages or phases. This work was based on the CRISP-DM methodology.

Regardless of the methodology used, this process features a phase that may be considered as the nucleus of KDD, the “data mining” (or modeling according to CRISP-DM) which is associated with the task, as well as the techniques and algorithms that may be employed in an application of KDD. What will be highlighted in this study is affinity grouping and clustering, techniques associated with these tasks and *Apriori* and *K-means* algorithms. All this contextualization will be embodied in the selected data mining tool, *Weka (Waikato Environment for Knowledge Analysis)*.

The research plan focuses on the application of the KDD process in the Judiciary Power regarding its related activity, court proceedings, seeking findings based on the influence of the procedural classification concerning the incidence of proceedings, the proceduring time, the kind of sentences pronounced and hearing attendance. Also, the search for defendants’ profiles in criminal proceedings such as sex, marital status, education background, professional and race. In chapters 2 and 3, the study presents the theoretical grounds of KDD, explaining the CRISP-DM methodology. Chapter 4 explores all the application preformed in the data of the Judiciary Power, and lastly, in Chapter conclusions are drawn.

Keywords: Knowledge Discovery in Databases (KDD), Data Mining, Clustering, Affinity Grouping, *Apriori*, *K-means*, *Weka*, Judiciary Power, Procedural Classification.

1 INTRODUÇÃO

Com a evolução da Tecnologia da Informação e o aumento do uso de computadores conectados em rede, as empresas vêm buscando soluções baseadas em Sistemas de Informação automatizados para controlar seus processos mais importantes. A base desses sistemas é, em geral um banco de dados, que com o passar do tempo acaba armazenando uma grande quantidade de dados relacionados aos negócios, mas não relacionados entre si. [INM 99]

Para explorar as relações existentes entre os dados abriu-se espaço para a procura de conhecimento e informações úteis não conhecidas, a partir de grandes conjuntos de dados armazenados. A este campo deu-se o nome de Descoberta de Conhecimento em Base de Dados (DCBD), o qual foi formalizado em 1989 [PAR89]. O DCBD é composto por um processo de etapas ou fases, de natureza iterativa e interativa. Este trabalho baseou-se na metodologia CRISP-DM [CHA2000].

Independente da metodologia empregada, este processo tem uma fase que pode ser considerada o núcleo da DCBD, a “mineração de dados” (ou modelagem conforme CRISP-DM). O nome “mineração de dados” (*data mining*) surgiu devido às semelhanças entre a procura de informação importante para o mundo dos negócios (em uma base de dados) e o minerar de uma montanha para encontrar um veio de ouro. Ambos processos requerem uma seleção em um imenso amontoado de material ou um sondar inteligente deste material, para encontrar o sítio onde está o valor desejado [SAN2000].

A maior parte da literatura [BER97] associa, ao termo “mineração de dados” as classes de tipo de problema (também conhecidas como tarefas), bem como as técnicas e algoritmos que podem ser empregados em uma aplicação de DCBD. Neste trabalho, destacaremos: as classes associação e agrupamento; as técnicas associadas a estas classes; e os algoritmos *Apriori* e *K-médias*. Toda esta contextualização estará compreendida na ferramenta de mineração de dados escolhida, Weka (*Waikato Environment for Knowledge Analysis*).

1.1 Motivação

A mineração de dados tem sido aplicada em diversas empresas com a intenção de se obter ganhos e vantagens competitivas nas mais diversas áreas de atuação, como vendas, marketing, finanças, seguros e telecomunicações. E também, em algumas instituições não comerciais, como na saúde e na área jurídica.

Ser mais competitiva possível é o objetivo de qualquer empresa, mas neste trabalho, não estaremos abordando uma organização comercial, e sim, um órgão público, o Poder Judiciário, que não sofre concorrências, ou simplesmente não precisaria conquistar mercado, mas necessita “avançar” na sua metodologia de produção (julgamento de processos), na sua organização e no seu comprometimento com redução de gastos públicos.

O plano de pesquisa deste trabalho está centrado em aplicar o processo de DCBD no Poder Judiciário no que se refere a sua atividade fim, julgamentos de processos. Busca-se descobrir relacionamentos entre os dados dos mais diferentes tipos de processos que possam auxiliar a alta administração no planejamento de todo o Judiciário, reorganizando estruturas que estejam deficientes e antecipando-se a picos de demanda de processos.

Como fator de motivação para este trabalho levou-se em consideração a pequena exploração deste assunto na área jurídica no Brasil, visto que, depois de diversas pesquisas na procura por aplicações e estudo de casos nesta área, encontrou-se pouco material consistente (passos para mineração de dados, volume de dados utilizado, tarefa e técnica escolhidas, ferramenta empregada) que pudesse servir de embasamento para este trabalho. Mesmo identificando que, em outros países, a mineração de dados já está mais disseminada, o material disponível para pesquisa está, ainda, incipiente nesta área. Os casos encontrados estavam associados à área de detecção de crimes [SCH2001], e mesmo assim, o material de pesquisa foi obtido através de propagandas de ferramentas comerciais, ou seja, com poucas informações científicas.

Este trabalho vai procurar encontrar relacionamentos entre os dados de processos disponíveis na base de dados do Poder Judiciário do Estado do Rio Grande do Sul, levando em consideração os processos de 1ª Instância, a fim de procurar conhecimento a partir da influência da classificação processual em relação à incidência de processos, ao tempo de tramitação, aos tipos de sentenças proferidas e a presença da audiência em alguns destes processos. Também, procurar-se-á por perfis de réus, em processos criminais, no que tange características, como sexo, estado civil, grau de instrução, profissão e cor.

1.2 Apresentação

O trabalho está organizado da seguinte maneira: o capítulo 2 descreve o embasamento teórico de descoberta do conhecimento em base de dados, bem como todo o resumo do processo de prospecção de conhecimento, defendido por [FAY96] e o detalhamento da metodologia abordada neste trabalho, CRISP-DM; o capítulo 3 explora uma das fases do processo de DCBD, a “mineração de dados”, apresentando os conceitos associados a este termo, como as classes de tipo de problema, as técnicas e algoritmos a serem utilizados; o capítulo 4 descreve toda aplicação realizada com os dados do Poder Judiciário do R.S, baseado na metodologia escolhida; as conclusões são discutidas no capítulo 5; e por fim, seguem os anexos, sendo os anexos A, B e C a descrição, a distribuição e a exploração dos dados trabalhados e os anexos D e E, os resultados da mineração de dados segundo as técnicas aplicadas.

2 DESCOBERTA DO CONHECIMENTO

Neste capítulo, discutiremos o que chamamos de Descoberta do Conhecimento em Base de Dados, bem como o todo o processo, apresentando a metodologia a ser seguida, CRISP-DM. Dentre as fases do processo, será enfocada a de mineração de dados, na qual exploraremos as classes de tipos de problemas, as técnicas, bem como os algoritmos que serão utilizados ao longo deste trabalho.

2.1 Conceito

O termo Descoberta do Conhecimento em Banco de Dados – DCBD (*Knowledge Discovery in Databases* – KDD) foi formalizado em 1989 [PAR89], em referência ao processo de procurar conhecimento em dados. Este processo tem como objetivo extrair as relações implícitas e previamente desconhecidas entre os dados, na busca de informações potencialmente úteis.

[BRU99] aponta que a DCBD pode ser vista como um processo de descoberta de novas correlações, padrões, e tendências significativas por meio da análise minuciosa de grandes conjuntos de dados estocados. Já, [AGR95] define DCBD como um processo não trivial de identificação de conhecimento válido, novo e compreensível a partir dos dados armazenados. Em resumo, descobrir conhecimento significa extrair, de grandes bases de dados, sem nenhuma formulação prévia de hipóteses, informações genéricas, relevantes e previamente desconhecidas, que podem ser utilizadas para a tomada de decisões [CAB97].

Segundo [BRU99], o DCBD pode ser visto como uma atividade multidisciplinar, pois explora a pesquisa de diversos ramos da Inteligência Artificial, como o Aprendizado de Máquina, Reconhecimento de Padrões, Sistemas Inteligentes, Aquisição de Conhecimento, bem como disciplinas matemáticas, como a Estatística, Teoria da Informação e Processamento de Incertezas.

2.2 Processo de DCBD

Associado a Descoberta de Conhecimento, deve existir um “Processo de Descoberta de Conhecimento”. Inicialmente, [FAY96] sugeriu uma proposição para este processo: “Este processo é composto por uma série de etapas, possuindo uma natureza iterativa e interativa”. Iterativa, pois apesar do processo ser composto por etapas sequenciais, pode haver retorno a etapas anteriores, e interativo, pois neste caso, o usuário poderá optar pela retomada em qualquer uma das etapas deste processo. Sendo assim, o processo de DCBD é cooperativo entre humanos e computadores. Humanos

projetam a arquitetura do sistema, descrevem os problemas e identificam os objetivos a serem atingidos, enquanto, os computadores, processam os dados a procura de padrões, a fim de satisfazer os objetivos anteriormente definidos.

2.2.1 Etapas do Processo de DCBD

As etapas sugeridas por [FAY96] para o processo de DCBD foram as seguintes, como mostra a figura 2.1:

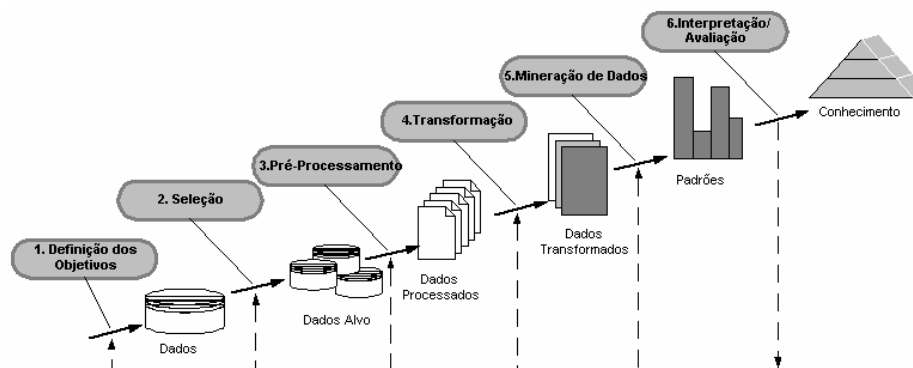


Figura 2.1: Processo de DCBD.

2.2.1.1 Definição dos Objetivos

O processo de DCBD começa com o entendimento do domínio da aplicação e dos objetivos finais a serem atingidos, isto é, nesta etapa o usuário define o que ele deseja conhecer e quais são seus objetivos com o conhecimento a ser adquirido.

2.2.1.2 Seleção

Nesta etapa é realizada uma seleção dos dados alvos de prospecção, podendo ser determinados subconjuntos destes dados.

2.2.1.3 Pré-Processamento

Normalmente, as bases de dados não estão adaptadas para uma “garimpagem”. Esta etapa consiste na limpeza dos dados, através de um pré-processamento, visando adequá-los para a etapa de mineração de dados, que veremos adiante.

2.2.1.4 Transformação

Os dados pré-processados ainda devem passar por uma transformação que os armazene adequadamente, visando reduzi-los, a fim de diminuir o número de variáveis envolvidas no processo. Esta transformação busca atributos que representem todas as características destes dados.

2.2.1.5 Mineração de Dados

Esta etapa pode ser considerada o núcleo do processo de DCBD. Ela consiste na própria aplicação da mineração de dados, ou seja, a busca por padrões nos dados.

2.2.1.6 Interpretação e Avaliação

Ao final da aplicação da mineração de dados, os padrões enumerados devem ser interpretados e avaliados, ou seja, deve ser realizado um trabalho de pós-processamento, a fim de verificar o que constitui as novas descobertas.

2.3 Metodologia CRISP – DM

A partir da proposição defendida por [FAY96] relativa ao processo de DCBD e suas etapas, algumas organizações: DaimlerChrysler; SPSS; e NCR, sentiram a necessidade de detalhar com mais profundidade todo este processo, pois a medida que procuravam aplicá-lo, se deparavam com algumas situações que provocava cada uma a formular o seu próprio processo de DCBD. Baseado nisto, e nas incertezas do que estava correto e incorreto nestes processos particulares, estas três organizações juntaram-se e conceberam uma nova metodologia para este processo, a *CRISP-DM methodology*.

Conforme [CHA2000], CRISP-DM (*Cross-Industry Standard Process for Data Mining*) é uma metodologia para aplicação do processo de DCBD. Ela foi concebida no final do ano de 1996 por três organizações do jovem e incipiente mercado de mineração de dados, ou seja, pela empresa DaimlerChrysler, uma das maiores organizações comerciais e industriais que tem realizado experiências em aplicações de mineração de dados em suas operações de negócio, pela SPSS, que tem prestado serviços baseados em mineração de dados desde 1990, e pela NCR, que tem equipes de consultores de mineração de dados, bem como especialistas na tecnologia, para orientar clientes que se interessem em aplicar a técnica a fim obterem ganhos de diversas espécies.

O processo proposto de DCBD invoca para uma rápida visualização no ciclo de vida de um projeto desta magnitude, isto é, ele é composto de fases, suas respectivas tarefas e dos relacionamentos entre estas fases.

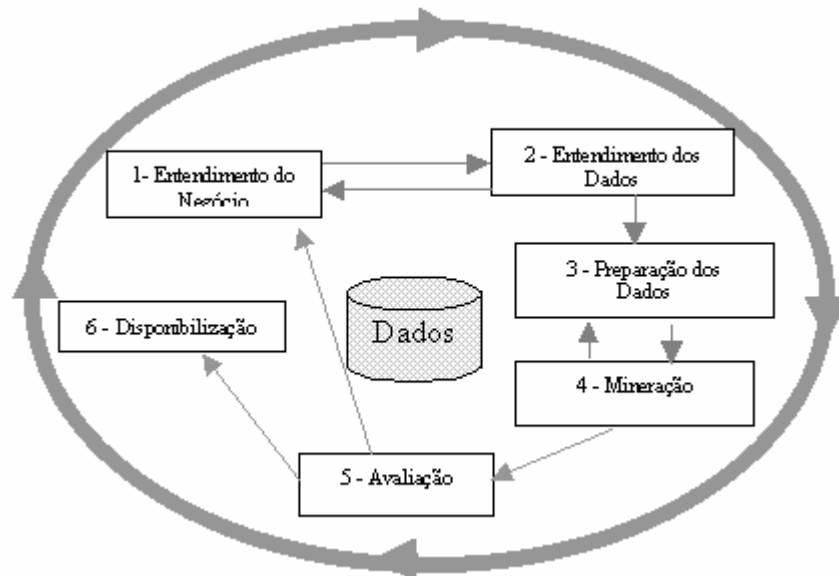


Figura 2.2: Metodologia CRISP-DM

O ciclo de vida proposto por esta metodologia para um projeto de DCBD é composto de 6 fases, sendo a seqüência das mesmas não rígida. Pode-se percorrer de traz para frente e de frente para traz entre as diferentes fases do processo. Este caminho será dependente dos resultados obtidos a cada fase ou de cada tarefa particular da fase. A figura 2.2 mostra as mais importantes e freqüentes dependências entre as fases, tendo o “circulo externo” da mesma representando o ciclo natural do processo de DCBD, no qual as descobertas são fruto de sucessivas iterações e não de uma só iteração.

A metodologia CRISP-DM aponta como integrante do processo as seguintes fases: entendimento do negócio; entendimento dos dados; preparação dos dados; mineração de dados; avaliação; e disponibilização. A seguir, serão exploradas cada uma destas fases.

2.3.1 Entendimento do Negócio

Esta fase inicial focaliza no entendimento dos objetivos do projeto associado às exigências do negócio, com o intuito de definir o problema de descoberta de conhecimento e traçar um plano preliminar para encontrar os tais objetivos. Pode-se detalhar esta fase nas seguintes tarefas:

2.3.1.1 Determinação dos objetivos do negócio

O analista deve adquirir um entendimento total do negócio, buscando identificar o que o cliente quer acompanhar, e procurando por fatores que possam influenciar os resultados do projeto. Efetivamente, o analista deve:

- Registrar, desde o começo do projeto, as informações conhecidas sobre os negócios da organização;
- Descrever os objetivos primeiros do cliente, acrescentando a eles os objetivos adicionais de interesse do negócio;
- Descrever os critérios para êxito nos resultados do projeto em relação ao objetivo do negócio.

2.3.1.2 Avaliação da situação

Esta tarefa sugere a obtenção de maiores detalhes dos fatores envolvidos no negócio, como: recursos; regras; e suposições, a fim de determinar um plano preliminar para execução do projeto, ou seja:

- Listar os recursos envolvidos no projeto, incluindo: pessoas (analista de negócios, o analista de dados, os técnicos de suporte, as pessoas de mineração de dados), dados (acesso aos dados operacionais e *warehouses*), recursos computacionais (plataforma de hardware) e softwares (ferramenta de mineração de dados e outras);
- Listar todas as exigências, incluindo cronograma, suposições que devam ser testadas no processo, e regras envolvidas;
- Listar riscos ou eventos que possam provocar atraso no projeto, bem como planos de contingência;
- Descrever glossário de terminologias relevantes para o projeto, dividido em um glossário com os termos do negócio e outro com a terminologia de mineração de dados;
- Construir análise de custos e benefícios para o projeto.

2.3.1.3 Determinação dos objetivos do processo de descoberta de conhecimento

Os objetivos do processo de DCBD compreendem:

- Descrever os objetivos, em termos técnicos, enfocando os resultados pretendidos e procurando relacioná-lo com os objetivos definidos em nível de negócio;
- Definir os critérios para atingir estes resultados pretendidos.

2.3.1.4 Elaboração do plano do projeto

Esta tarefa engloba a descrição do plano de projeto para encontrar os objetivos tanto em nível de negócio quanto em nível técnico. O plano deve especificar, previamente os passos a serem executados durante o restante do projeto, principalmente no que se refere a seleção inicial das técnicas e ferramentas a serem utilizadas, baseada na classe de tipo de problema de mineração de dados escolhida.

2.3.2 Entendimento dos Dados

A fase de entendimento dos dados inicia com uma coleção inicial de dados e prossegue com atividades relacionadas a obtenção de familiaridade destes dados, com o objetivo de detectar problemas de qualidade, bem como descobrir os primeiros subconjuntos que possam ser interessantes. Esta fase compreende as seguintes tarefas:

2.3.2.1 Coleção inicial de dados

Esta tarefa envolve tudo que está relacionado a obtenção dos dados:

- Pode ser necessária a utilização de alguma ferramenta para carga destes dados;
- Caso os dados estejam distribuídos, será necessário um esforço adicional para integração destes dados.

2.3.2.2 *Descrição dos dados*

Esta tarefa compreende um exame nos dados, e conseqüentemente a geração de relatórios que descrevam as propriedades dos dados, como:

- Entidades e atributos envolvidos;
- Formato dos dados;
- Distribuição dos dados (por entidade, por atributo, etc.).

2.3.2.3 *Exploração dos dados*

Esta tarefa procura atacar questões relativas ao comportamento dos dados em face aos objetivos de mineração de dados preestabelecidos, as quais podem ser exploradas utilizando-se ferramentas de consulta e visualização. Isto inclui:

- Quantidades de dados;
- Resultado de simples agregações;
- Propriedades de sub-populações significantes;
- Análises estatísticas de pequena complexidade.

2.3.2.4 *Verificação da qualidade destes dados*

Aqui, pretende-se analisar a qualidade dos dados envolvidos, e caso os dados apresentem problemas (falta de consistência, valores nulos ou errados, etc.) é importante listar as possíveis soluções que contornem este empecilho.

2.3.3 Preparação de Dados

A fase de preparação de dados cobre todas as atividades para construção do conjunto final de dados, que será modelado a fim de servir de entrada para a ferramenta de mineração de dados escolhida. A preparação de dados compreende as seguintes tarefas:

2.3.3.1 *Seleção de dados*

Baseado nos objetivos definidos para a mineração de dados, bem como na qualidade e características dos dados, e respeitando os limites de volume e tipificação dos mesmos, esta tarefa procura selecionar os dados para a aplicação.

2.3.3.2 *Limpeza de dados*

Esta tarefa sugere que, a partir dos problemas encontrados na tarefa de verificação da qualidade dos dados (item 2.3.2.4), sejam descritas as ações que serão tomadas para resolução de tais problemas nos dados, como por exemplo, a inserção de valores padrão (*defaults*) para dados errôneos ou sem valor.

2.3.3.3 *Construção de dados*

Nesta tarefa procura-se construir, caso necessário, operações que produzam atributos derivados. Exemplo: $\text{área} = \text{comprimento} * \text{largura}$. Ou ainda, a construção de novos conjuntos de dados que respeitem determinados critérios.

2.3.3.4 *Mescla de dados*

Este método sugere combinar a informação de diversas tabelas ou registros a fim de construir um novo conjunto de registros ou valores. A mescla de dados, também, cobre agregações, ou seja, operações onde novos valores são calculados para condensação de informações de múltiplas tabelas e registros.

2.3.3.5 *Formatação de dados*

De vez em quando é necessário alterar a ordem dos registros no conjunto de dados, pois, muitas vezes, a ferramenta de mineração exige que os registros sejam ordenados por um determinado atributo. Também, podem ser necessárias alterações sintáticas, como remover vírgulas dentro de campos texto.

2.3.4 **Mineração¹**

Nesta fase é realizada a mineração de dados propriamente dita, baseada na técnica selecionada e no conjunto de dados pré-processados. Normalmente, pode-se aplicar diversas técnicas para uma mesma classe de tipo de problema de mineração de dados.

2.3.4.1 *Seleção da técnica de mineração*

O primeiro passo da mineração é a seleção da técnica que será utilizada. Esta escolha pode já ter sido realizada na fase de entendimento do negócio. É nesta tarefa que especifica-se, detalhadamente, a técnica escolhida. Caso mais de uma técnica seja aplicada, deve-se realizar esta tarefa para cada técnica, separadamente.

2.3.4.2 *Geração de conjuntos de treinamento*

Antes de executar a mineração, é preciso gerar procedimentos e mecanismos para testar a validade e qualidade do mesmo. Desta forma, dependendo da técnica escolhida, tem que se definir um conjunto de dados para servir de treinamento e, um outro, para validar do modelo gerado.

¹ Mineração: a metodologia CRISP-DM chama esta fase de modelagem, mas para homogeneizar este trabalho, estamos chamando-a de mineração, visto que a maior parte da literatura [BER97, FAY96] utiliza este termo.

2.3.4.3 *Mineração de Dados*

Aqui, é a execução da mineração de dados propriamente dita. É importante salientar que cada ferramenta possui um grande número de parâmetros que devem ser ajustados e testados para execução de uma ou mais minerações.

Como resultado desta tarefa é importante descrevermos um modelo resultante, explicitando a interpretação da mineração, bem como relatando as dificuldades encontradas na execução.

2.3.4.4 *Resultados da Mineração*

Como desfecho desta fase, deve ser apresentado um resumo dos resultados obtidos, descrevendo as “qualidades”, em termos de acurácia, da mineração, bem como o *ranking* desta acurácia. Também, deve ser revisado o conjunto de parâmetros definidos, e caso necessário, executar mais vezes a mineração, ajustando estes parâmetros, até encontrar-se um “melhor modelo”.

2.3.5 **Avaliação**

Antes de disponibilizar os resultados finais do processo de DCBD, deverá ser realizada a avaliação da mineração produzida, bem como revisar as fases anteriores para certificar-se de que os resultados obtidos estão de acordo com os objetivos pré-definidos.

2.3.5.1 *Avaliação dos resultados*

O resultado desta tarefa procura obter a interpretação dos resultados obtidos na fase de mineração de dados, procurando consonância com os objetivos inicialmente estabelecidos.

2.3.5.2 *Revisão do processo*

Esta tarefa recomenda que seja realizada uma revisão no processo a fim de salientar fases ou etapas que não foram exploradas adequadamente, apontando o que deve ser repetido no processo de DCBD.

2.3.5.3 *Determinação dos próximos passos*

Esta tarefa sugere realizar, baseado nos resultados obtidos, uma avaliação para definir se o processo está finalizado ou se será necessário realizar novas iterações em determinadas fases, ou ainda iniciar novos projetos de mineração de dados, sempre buscando atingir os objetivos inicialmente estabelecidos.

2.3.6 **Disponibilização**

Nesta fase são apresentados os resultados finais do processo de DCBD, buscando apontar de que forma o conhecimento obtido pode auxiliar na tomada de decisão.

2.3.6.1 Plano de disponibilização

Elaborar o plano de apresentação de resultados é o enfoque desta tarefa.

2.3.6.2 Plano de monitoração e manutenção

Caso a aplicação de mineração de dados passe a fazer parte da rotina diária da organização, deve ser elaborado plano de monitoração, bem como manutenção do processo de DCBD.

2.3.6.3 Produção do relatório final

Esta tarefa objetiva produzir o relatório final da mineração de dados, com os resultados devidamente resumidos e organizados e de forma que o cliente tenha plenas condições de realizar sua própria avaliação.

2.3.6.4 Revisão do projeto

Como desfecho, esta tarefa propõe um relato de toda a experiência obtida com a aplicação do processo de DCBD, apontando os aspectos positivos, bem como as dificuldades encontradas.

3 MINERAÇÃO DE DADOS

Devido à importância da etapa de mineração de dados, e por ser considerada o núcleo do processo de DCBD, enfocada tanto pela metodologia sugerida por [FAY96], quanto pela metodologia CRISP-DM [CHA2000], muitas vezes, o termo “mineração de dados” é utilizado como sinônimo de todo o processo de DCBD. É necessário explorarmos mais detalhadamente esta fase, pois a ela, estão associados alguns conceitos vitais para o entendimento deste trabalho, como classes de tipo de problemas, técnicas e algoritmos.

3.1 Conceito

Como descrito anteriormente, a chamada mineração de dados (*Data Mining*, DM) é “um passo dentro o processo de DCBD, que consiste na aplicação de algoritmos de descoberta de dados que, sob certas limitações de eficiência computacional aceitáveis, produzem uma enumeração particular de padrões sobre estes dados” [FAY96].

Em resumo, mineração de dados é um termo genérico utilizado para todas as novas técnicas computacionais para a extração de informações úteis a partir de grandes conjuntos de dados armazenados.

3.2 Classes de Tipo de Problema¹ e Técnicas de Mineração de Dados

Para empregar o processo de DCBD e, em consequência, a mineração de dados em alguma aplicação, devemos levar em conta que existem diversas classes de tipo de problema que podem ser utilizadas e, associadas a elas, outras tantas técnicas de mineração de dados. O emprego de uma classe de tipo de problema e de uma técnica irá depender da natureza da aplicação em questão, ou melhor, podemos dizer que obtemos melhores resultados se aplicarmos a tarefa e técnica mais adequada para as condições, necessidades e objetivos que dispomos.

As classes de tipo de problema podem ser entendidas como os tipos de relacionamentos entre os dados que vamos estabelecer para obtermos conhecimento passível de interpretação humana. As técnicas de mineração de dados podem ser consideradas como os fundamentos computacionais, ligados intimamente a área de

¹ Classes de Tipo de Problema: Também conhecida como tarefas, conforme [BER97]. Não foi denominada desta maneira para não gerar confusão com o termo “tarefas” da metodologia CRISP-DM.

inteligência artificial, que propiciarão a construção de algoritmos que possibilitem a busca por padrões escondidos nos dados [SCH2001]. Destacaremos, a seguir, as classes de tipos de problemas descritivas² (agrupamento e associação), bem como as técnicas associadas a cada uma destas classes, juntamente com os algoritmos escolhidos para mineração de dados.

3.2.1 Classe de Tipo de Problema – Agrupamento

Agrupamento é uma classe de tipo de problema que procura segmentar populações heterogêneas em subgrupos homogêneos. Segundo [BER97], nesta classe os registros são agrupados conforme alguma similaridade em si mesmos.

O agrupamento é normalmente realizado como um prelúdio de alguma outra forma de mineração de dados ou modelagem. Por exemplo, agrupamento pode ser o primeiro passo em uma segmentação de mercado. Em vez de tentar propor uma regra que atenda a todos, podemos primeiro dividir os clientes baseados em segmentos ou em pessoas com hábitos de aquisição semelhantes.

Uma das técnicas de maior difusão para aplicar o agrupamento é a segmentação (*Clustering*).

3.2.1.1 Técnica – Segmentação

A principal característica da técnica de segmentação é descobrir grupos homogêneos em uma base de dados, esta necessidade decorre do crescimento destas bases, revelando-se cada vez mais heterogêneas. Esta técnica procura separar a base de dados em segmentos (*clusters*) de registros que compartilhem de mesmas propriedades, e assim sejam considerados homogêneos, ou seja, com características similares entre si (intragrupo), mas distintas entre os grupos (intergrupos).

Para identificar essa homogeneidade entre os registros, já que ela é bastante subjetiva, são utilizados diferentes critérios, sendo um dos mais populares a utilização de uma função chamada “medida de similaridade”. Como podemos aplicar diferentes medidas de similaridade, esta técnica pode ser aplicada para a maioria dos tipos de dados [BER 97].

A vantagem desta técnica se concentra na possibilidade de aplicarmos-la mesmo não conhecendo a estrutura interna do banco de dados, e assim, propiciar a descoberta de estruturas escondidas. Ela pode ser utilizada como preparo para a aplicação de outras técnicas. As desvantagens se resumem na escolha desta medida de similaridade e na dificuldade de interpretar seus resultados.

A técnica de segmentação pode ser classificada conforme a forma que os grupos (*clusters*) são constituídos [WIV99]. As formas básicas são baseadas nos métodos de partição e hierárquico. Dentre os algoritmos mais conhecidos, segundo o método de partição, destacam-se o *K-means* e *Expectation and Maximization (EM)*. Já, entre os hierárquicos, destaca-se o algoritmo *CobWeb*. Em [TOR2001], [GAM2001] são encontradas abordagens didáticas sobre o funcionamento dos algoritmos citados. A

² Descritivas: Existem dois tipos de classes de tipo de problema, as descritivas e as preditivas. As descritivas objetivam encontrar padrões sem um conceito pré-estabelecido. Já as preditivas consistem em usar algumas variáveis para prever um valor desconhecido ou futuro.

seguir será explorado o algoritmo *K-means*, já que o mesmo será utilizado na aplicação prática deste trabalho.

3.2.1.2 Algoritmo – *K-Médias* (K-means)

Conforme [BER97], o algoritmo K-médias é um dos mais utilizados para realizar a técnica de segmentação. Embora ele tenha muitas variações, sua forma original foi primeiro publicado por J. B. MacQueen em 1967.

Informalmente, o agrupamento por K-médias irá encontrar “k” grupos, sendo “k” escolhido a priori, onde a variância dos objetos dentro de cada grupo é mínima. Cada grupo é identificado por um valor chamado de centróide do grupo. Apresentamos, abaixo, os passos envolvidos no algoritmo K-médias:

1. Determinar os “k” centróides iniciais dos grupos (número de *clusters* em que vamos agrupar), segundo algum critério, como por exemplo, aleatoriamente dentro do espaço dos objetos ou de acordo com alguma heurística;
2. Calcular a distância de cada objeto em relação ao centróide de cada grupo já definido;
3. Alocar o objeto ao grupo cuja distância a seu centróide seja menor;
4. Recalcular os centróides dos grupos a partir dos objetos realocados;
5. Repetir os passos 2 a 4 até que algum critério de convergência seja atingido.

Conforme [ENG2002], o cálculo da distância é baseado em alguma medida de similaridade, sendo uma das mais comuns e apropriadas, para atributos reais, a distância *Euclidiana*, e para atributos nominais, a distância de *Hamming*. Quem irá determinar a pertinência do registro dentro de um determinado grupo é a medida de similaridade escolhida. Em [BER97] são apresentados maiores detalhes do algoritmo K-médias.

3.2.2 Classe de Tipo de Problema – Associação

A classe de tipo de problema “associação” tem o intuito de identificar associações entre os dados que de alguma maneira estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam a presença de outros em uma mesma transação [BER97]. Esta classe tem numerosas aplicações e pode ser enquadrada em dois grupos distintos:

- Associação: Um exemplo clássico para entender este grupo está relacionado a compras que fazemos em um supermercado. A especialização desta classe de tipo de problema procura por produtos que são frequentemente comprados em uma mesma transação por cada cliente, revelando afinidades escondidas entre esses produtos;
- Padrões Sequenciais: Este grupo procura por associações entre eventos que ocorram em uma sequência através do tempo, descobrindo padrões sequenciais de aquisição de produtos ou serviços.

A classe de tipo de problema associação é uma abordagem baseada na geração de regras a partir dos dados. Estas regras serão discutidas a seguir na técnica de regras de associação.

3.2.2.1 Técnica – Regras de Associação

Uma regra é composta de duas partes, uma condição e um resultado. Conforme [BER97], o formato geral da regra de associação possui a forma SE “A.”. ENTÃO “B”..., Onde “A” é chamado de condição e “B” é chamado de resultado.

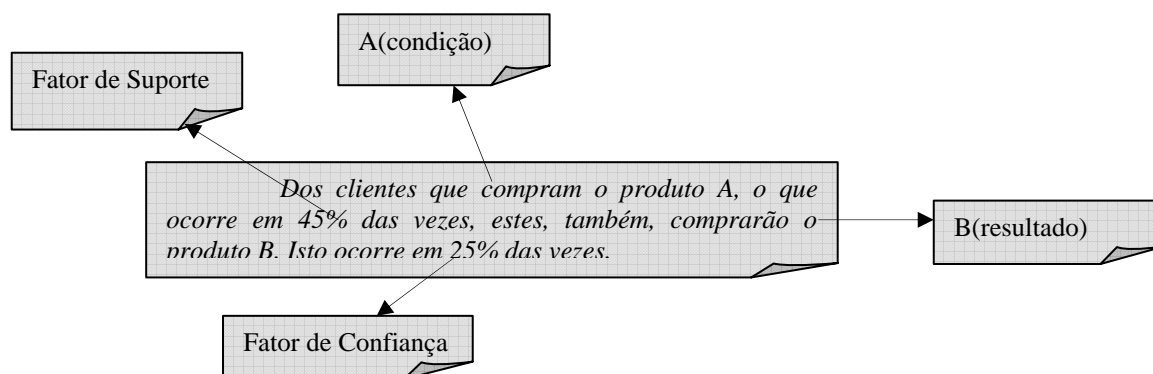


Figura 3.1 - Regra de Associação e seus componentes

Além da condição e do resultado, outros componentes complementam esta regra: fator de suporte; fator de confiança; e *lift*.

- Fator de suporte: é o número de vezes que determinado item ou conjunto de itens aparecem em transações distintas relacionadas com o número total de transações operadas. Fórmula:

$$\text{Fator Suporte} = (\text{ocorrências do (item ou do conjunto de itens)} / \text{número de transações operadas})$$

- Fator de confiança: é o número de transações que suportam o resultado (ENTÃO B...) em relação ao número de transações que suportam a condição (SE A...). Fórmula:

$$\text{Fat. Confiança} = (\text{Fator Suporte de A} / \text{Fator Suporte de B})$$

- Lift: é uma estatística adicional para verificação de dependência entre os atributos envolvidos em determinada regra de associação. Quanto maior for o *lift*, maior será a dependência entre os atributos, e provavelmente, maior será o valor da regra. Fórmula:

$$\text{Lift} = (\text{Fator Suporte de A,B} / (\text{Fator Suporte de A} * \text{Fator de suporte de B}))$$

Esta técnica se caracteriza por contabilizar as ocorrências associadas às combinações possíveis entre itens. Identificadas as transações, procura-se quantificar as ocorrências de um determinado item, devido a presença de um ou de outros na transação. Pode-se considerar que ela trabalha em um primeiro momento buscando

encontrar o fator de suporte, para posteriormente, testar todas as regras verificando o fator de confiança.

A complexidade do emprego da regra de associação dependerá do número de variáveis (itens) envolvidas. Se procurarmos por um item no resultado e por outro na condição teremos uma melhor performance do que se procurarmos por mais itens tanto no resultado quanto na condição.

As regras para procura de padrões seqüenciais divergem das regras de associação pelo fato de introduzir fatores como identificação de “cliente” e período de tempo. Como valores de entrada, sempre teremos informações relativas ao tempo da transação (data, hora), bem como a identificação do “cliente” associado às transações efetuadas por ele.

Embora as regras para padrões seqüenciais diverjam das regras de associação, o fator de suporte tem grande importância para ambas. Este fator determina o número de vezes que o padrão seqüencial ocorreu em relação ao número total de transações. Sendo assim, estabelecido um fator de suporte mínimo a ser avaliado, descobrimos quais “clientes” suportam o padrão seqüencial.

Esta regra reforça a idéia de contarmos com um grande número de registros para encontrarmos os padrões seqüenciais, visto que precisamos de várias transações realizadas por um único “cliente”.

Dentre alguns algoritmos (*Apriori* e *Apriori*Tid [AGR95A], e *Partition* [SAV95] que foram desenvolvidos com o objetivo de gerar regras de associação entre itens de dados, o mais utilizado é o *Apriori*, sendo que os demais ou são extensões deste ou o utilizam. A seguir, será explanado o algoritmo *Apriori*.

3.2.2.2 Algoritmo Apriori

O algoritmo *Apriori* tem como objetivo encontrar todos os conjuntos de itens freqüentes, denominados *large itemsets* (L_k). Ele decompõe a resolução do problema em duas etapas:

- A primeira que se preocupa em encontrar todos os conjuntos de itens freqüentes com fator de suporte acima do suporte mínimo estabelecido;
- E a segunda etapa que tem a função de gerar as regras de associação.

Dado um conjunto de “n” transações T, sendo T um conjunto de itens, e estabelecendo um fator de suporte mínimo (*MinSup*) e um fator de confiança mínimo (*MinConf*), o objetivo do algoritmo concentra-se em obter:

As regras de associação $A \rightarrow B$ tais que:

O Fator de Suporte $\geq \text{MinSup}$ e o Fator de Confiança $\geq \text{MinConf}$

Para atingir tal objetivo, primeiramente, este algoritmo, realiza a contagem das ocorrências dos itens a fim de determinar os *large itemsets*(L_k) de tamanho unitário. Os “k” passos seguintes consistem de duas sub-etapas. Primeiro, os

large itemsets (L_{k-1}), encontrados na passagem anterior (k-1) são utilizados para gerar os conjuntos de itens potencialmente freqüentes, ou seja, os *candidate itemsets* (C_k). Segundo, a partir destes conjuntos potenciais, é realizada uma nova contagem, levando em conta o suporte de cada candidato (C_k).

A geração destes conjuntos potenciais (C_k), a princípio, assume como argumento L_{k-1} (o conjunto de todos k-1 *large itemsets*). Para isso, é utilizada a função denominada *Apriori_gen*, que retorna um superconjunto de todos os conjuntos freqüentes.

Para compreensão desta etapa, vamos a um exemplo, no qual o espaço de busca é composto por conjuntos de transações com 3 itens (tabela 3.1).

Tabela 3.1: Espaço de Busca

{1, 2, 3}		
{1,2}	{1,3}	{2,3}
{1}	{2}	{3}

Conforme já mencionado, a primeira etapa começa com a identificação dos conjuntos de itens freqüentes de tamanho 1. Supondo que o suporte mínimo deva ser de 50% (2 elementos), temos:

Tabela 3.2: Conjunto de Transações

T (ID)	ITEM-SET
100	1, 2, 3
200	1,3
300	1
400	3



Tabela 3.3: Ocorrências de cada item

ITEM-SET	Suporte
{1}	3
{2}	1
{3}	3

É importante destacarmos que se um conjunto está abaixo do suporte mínimo, podemos descartar seus superconjuntos. A partir da exemplificação acima, obtemos o espaço de busca filtrado (tabela 3.4):

Tabela 3.4: Espaço de Busca Filtrado

{1, 2, 3}		
{1,2}	{1,3}	{2,3}
{1}	{2}	{3}

Conforme tabela 3.3, os itens freqüentes de tamanho 1 para o exemplo são: $L_{k1} = \{\{1\}, \{3\}\}$

Os conjuntos de tamanho 2 são obtidos a partir dos conjuntos freqüentes de tamanho 1 (L_{k1}), ou seja, candidatos $C_{k2} = \{\{a,b\} / \{a\}e \{b\} \in F_1\}$. Neste caso temos $C_{k2} = \{1,3\}$

Tabela 3.5: Ocorrências do Conjunto de Itens

ITEM-SET	Suporte
{1,3}	2

Desta maneira o conjunto de itens freqüentes de tamanho 2 (Tabela 3.5), resume-se ao conjunto: $L_{k2} = \{\{1,3\}\}$

Não existem conjuntos freqüentes nem candidatos de tamanho 3, pois os superconjuntos candidatos já haviam sido descartados por não atingirem o suporte mínimo estabelecido.

A segunda etapa do algoritmo consiste em produzir as regras de associação, através da função *Genrules*. A geração de regras, para qualquer conjunto de itens freqüentes (L_k), significa encontrar todos os subconjuntos não vazios l . Assim para todo e qualquer subconjunto A , produz-se uma regra $A \rightarrow (l-A)$ somente se a razão ($\text{suporte}(l)/\text{suporte}(A)$) é ao menos igual a confiança mínima estabelecida.

Para gerar regras com múltiplos consequentes. São considerados todos os subconjuntos. Por exemplo, dados um conjunto de itens ABCD, considera-se primeiro o subconjunto ABC, seguido de AB, etc. Se $ABC \rightarrow D$ não atinge uma confiança suficiente, não é necessário verificar $AB \rightarrow CD$

Em [AGR94] são apresentados detalhes do algoritmo *Apriori*, bem como a codificação do mesmo. Também, em [CAM2002] é explorado, com profundidade, este algoritmo. Cabe salientar que uma abordagem bastante didática, tanto do funcionamento, quanto das variações deste algoritmo é encontrada em [JOR2000].

4 APLICAÇÃO - PODER JUDICIÁRIO DO RS

Neste capítulo apresenta-se o estudo de caso realizado com dados do Poder Judiciário do Estado do Rio Grande do Sul. A organização das seções segue as fases definidas na metodologia CRISP-DM.

4.1 Entendimento do Negócio

Para compreendermos o trabalho proposto, destacaremos no entendimento do negócio as seguintes tarefas: determinação dos objetivos do negócio; avaliação da situação; definição dos objetivos da mineração de dados; escolha da classe de tipo de problema e da técnica de mineração de dados; e escolha da ferramenta de mineração a ser utilizada.

4.1.1 Determinação dos objetivos do negócio

Antes de definirmos os objetivos do negócio, resumiremos a estrutura do Poder Judiciário, bem como a tramitação de um processo a fim de oportunizar uma visão global da área de negócio em destaque neste trabalho.

4.1.1.1 Estrutura Corporativa

O Poder Judiciário é composto de duas instâncias: a primeira que compreende todas as comarcas espalhadas por diversas cidades do Estado; e a segunda que se resume ao Tribunal de Justiça do Estado.

A primeira instância é composta de duas ramificações principais: os Juizados Especiais, que objetivam a solução de problemas de menor complexidade e de valores envolvidos menos significativos (até 40 salários mínimos), os quais sempre buscam o entendimento das partes; e a Justiça Comum. Ambas ramificações incorporam a área cível e criminal. Esta instância é composta por 161 comarcas¹, que se subdividem em foros². Estes, por suas vez, dividem-se em varas³, e em cada uma delas é alocado um Juiz.

¹ comarcas: é a nomenclatura jurídica para indicar as cidades que possuem representação do Poder Judiciário.

² foros: é a nomenclatura jurídica para indicar as instalações físicas de cada comarca. Somente na Comarca de Porto Alegre é que temos mais de um foro (Foro Central e Foros Regionais), nas demais Comarcas temos apenas um.

³ varas: são as estruturas de julgamento de cada foro, compostas por um cartório e pelo gabinete do Juiz.

Já a segunda instância opera no Tribunal de Justiça. Ela divide-se em grupos⁴, que por sua vez, subdividem-se em câmaras⁵. Estas câmaras são normalmente compostas por 4 Desembargadores. Também no Tribunal de Justiça, opera a área administrativa do Poder Judiciário. A figura 4.1 ilustra um resumo da estrutura corporativa.

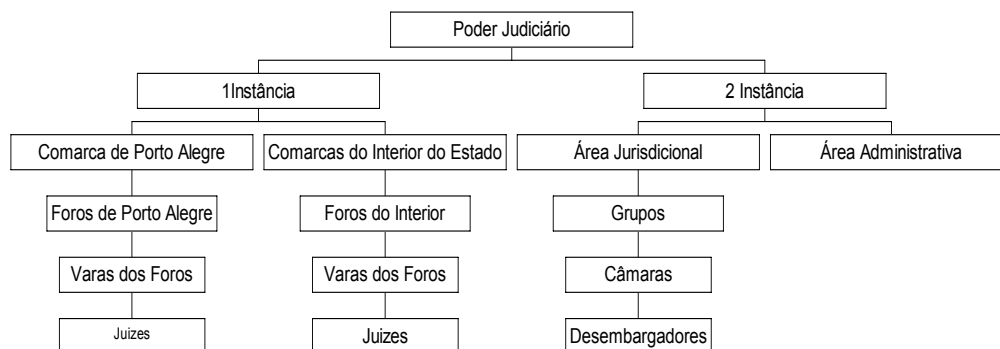


Figura 4.1: Resumo da Estrutura Corporativa do Poder Judiciário

4.1.1.2 Trâmite Processual

Para regular as relações interpessoais de uma sociedade, os indivíduos, que se sentem lesados de alguma forma, podem recorrer ao Poder Judiciário, com o intuito de reaver os seus supostos direitos alegados. Esta recorrência se efetiva através da entrada de um processo, impetrado pela parte ou seu representante. O ingresso do processo, normalmente ocorre em 1ª Instância (existem casos em que a competência inicial é da 2ª Instância), na comarca onde ocorreu o fato reclamado ou discordante.

Todo o processo possui uma classificação, onde a seção⁶, a especialização⁷, a classe⁸ e a natureza⁹ são identificadas. A partir do cadastramento do processo, este será encaminhado a um juiz que julgará a ação interposta. Mas, até o efetivo julgamento do pleito, ocorre uma verdadeira tramitação interna, e por vezes até externa, que deve ser acompanhada através do registro de seu andamento, que chamamos de “movimentação do processo”. Depois de realizada todas as preliminares, é sentenciada uma decisão do processo e a mesma é publicada a fim de comunicar a decisão a sociedade.

Caso a parte não fique satisfeita com a sentença proferida, ela pode recorrer a 2ª Instância, impetrando um recurso¹⁰. A tramitação é muito semelhante a de 1ª instância, o que se modifica é foro de julgamento, que neste caso transcorre no

⁴ grupos: são os conjuntos de câmaras que julgam o mesmo tipo de matéria.

⁵ câmaras: é a denominação dada na 2ª Instância ao que chamamos de “varas” na 1ª Instância

⁶ seção: é a ramificação principal da estrutura de julgamentos do Poder Judiciário.

⁷ especialização: cada seção é subdividida em determinadas áreas para fins de organização judiciária.

⁸ classe: é o instrumento jurídico pelo qual uma “parte” pode ingressar na justiça.

⁹ natureza: é a matéria em si pela qual a “parte” pode pleitear.

¹⁰ recurso: é o nome técnico dado a um processo, quando este representa uma recorrência a decisão de 1ª Instância.

Tribunal de Justiça, e que cada processo é decidido por uma composição de julgadores.¹¹

4.1.1.3 *Objetivos Gerais – Conhecimento a ser adquirido*

Propõem-se, como objetivo deste trabalho, demonstrar uma aplicação do processo de DCBD, utilizando, como estudo de caso, bases de dados do Poder Judiciário do Estado do Rio Grande do Sul, as quais enfocarão os processos jurisdicionais de 1ª Instância.

Inicialmente, devido a abrangência e a distribuição dos dados, bem como, a heterogeneidade sistêmica, ficou complexo de definir os objetivos gerais para a mineração. No entanto, baseado na carência de informações estatísticas da Corregedoria, órgão regulador e controlador das Comarcas do Estado, de processos de 1ª Instância, optou-se em realizar uma aplicação de DCBD que pudesse encontrar padrões de comportamento relacionando dados de classificação processual, com o padrão de sentenças proferidas, bem como com o tempo de tramitação do processo envolvido. Adicionalmente, procurar-se-á pelo perfil dos réus em processos criminais, no que tange as características pessoais do mesmo.

Para atingir tais objetivos, constatou-se que o sistema que possibilitará a aplicação da mineração de dados será o “Themis”, a ser explicitado no item 4.1.2.1, devido a sua característica de armazenar dados com maior consistência e riqueza para geração de informações gerenciais.

4.1.2 **Avaliação da Situação**

Dentre as informações essenciais para aplicarmos o processo de DCDB, destaca-se a estrutura sistêmica e a identificação de pessoas envolvidas.

4.1.2.1 *Estrutura Sistêmica*

Em termos de sistemas e bases de dados que atendem a área jurisdicional, podemos dividi-la em três grandes grupos:

- Comarcas de grande porte e Tribunal de Justiça: sistema “APJ” (Procergs) e base de dados ADABAS (*mainframe*), modo caractere. Este sistema está interligado entre as 9 maiores comarcas do Estado. No Tribunal de Justiça, o sistema atual é o “TJP”(Procergs), que possui a mesma base de dados do “APJ”, embora sua interface seja gráfica;
- Comarcas de médio porte: sistema instalado é o “JUS”(Procergs), sistema *standalone* (interligado em rede no âmbito da Comarca) e base de dados ZIM;
- Comarcas de pequeno porte: estas comarcas até o final de 2001 não eram informatizadas. Hoje, estão implantadas com sistema “Themis”, desenvolvido pelo próprio Departamento de Informática do Poder Judiciário, na arquitetura cliente-servidor, linguagem

¹¹ julgadores: os julgadores na 2ª Instância são chamados de “Desembargadores”. Os de 1ª Instância são chamados de “Juizes”.

Delphi, com base de dados Oracle e arquitetura de banco de dados descentralizada.

4.1.2.2 Identificação de Pessoas

Podemos considerar que as áreas envolvidas neste processo podem ser resumidas pela tabela 4.1:

Tabela 4.1: Pessoas e áreas envolvidas na aplicação de DCBD

Área	Envolvimento	Pessoas
Corregedoria	Responsável pela correição (controle e regulamentação) das comarcas de 1ª Instância	Juizes Corregedores Assistentes de Correição
Departamento de Informática	Responsável pelo fornecimento de ferramentas que auxiliem o trabalho da Corregedoria Responsável pelo sistema Themis implantado em diversas comarcas	Analista de Sistemas Administrador de Banco de Dados Administrador de Dados Assessora Jurídica

4.1.3 Definição dos Objetivos para Mineração de Dados

Especificamente, os objetivos foram, inicialmente, definidos em dois grupos:

1. A incidência de processos considerando a Classificação Processual (seção, classe e especialização):
 - a. Em si e com a comarca de origem;
 - b. Com o tempo de tramitação do processo;
 - c. Com o tipo de sentença proferida;
 - d. Com a influência de ocorrência de audiência.
2. O perfil dos réus em processos criminais, considerando:
 - a. Características pessoais, como sexo, estado civil, grau de instrução, profissão e cor;
 - b. Características pessoais com a natureza do processo;

4.1.4 Escolha da Classe de Tipo de Problema e Técnica de Mineração de Dados

Como os objetivos definidos são bastante abrangentes, a característica desta aplicação não procura definir um alvo para prever um determinado comportamento, fato que nos afasta de realizar predição, e nos aproxima de classes de tipo de problema descritivo. Sendo assim, optou-se por, inicialmente, trabalhar com a classe de agrupamento, a fim de encontrarmos grupos homogêneos de dados e, caso necessário, posteriormente, trabalhar com a classe associação, que irá procurar por relacionamentos de dados de maior intensidade. Em pesquisas anteriores [SCH2001], não foi encontrada nenhuma aplicação de DCBD na área jurídica, de relevância, que se pudesse fundamentar a escolha da classe do tipo de problema. Então, procurou-se

escolher a classe de agrupamento que, muitas vezes, serve como pré-análise para aplicação de outras classes.

Como existe uma estreita relação entre a classe escolhida e a técnica a ser aplicada, nos restringimos as técnicas de segmentação (*clustering*) e de regras de associação.

4.1.5 Ferramenta Escolhida

A ferramenta escolhida para realização do processo de DCBD foi a Weka (Waikato Environment for Knowledge Analysis). Ela é formada por um conjunto de implementações de algoritmos de diversas técnicas de mineração de dados. Foi implementada na linguagem Java, e tem como principal característica a portabilidade, podendo ser executada em diversas plataformas. É um software de domínio público e está disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>.

A escolha desta ferramenta deveu-se ao fato dela reunir características favoráveis a aplicação em questão, pois além de implementar tanto algoritmos de regra de associação, quanto de segmentação, possui funções de pré-processamento que se tornam indispensáveis para a aplicação do processo de DCBD.

4.1.5.1 Algoritmos Utilizados

Para a técnica de segmentação, a ferramenta Weka implementa os algoritmos *Expectation and Maximization (EM)*, K-médias (*Kmeans*) e *Cobweb*, já citados e comentados no item 3.2.1.1. Dentre estes, escolheu-se o algoritmo K-médias, basicamente pela sua maior difusão técnico-científica. Já para a técnica de regras de associação, o único algoritmo que a ferramenta implementa é o *Apriori*.

4.2 Entendimento dos Dados

A partir do entendimento do negócio, o próximo passo é conhecer os dados que serão trabalhados. Para isto, destacamos as tarefas a seguir:

4.2.1 Coleção Inicial de Dados

As bases selecionadas para a aplicação do processo de DCBD serão as que fornecem dados de 1ª Instância, das comarcas de menor expressão, ou seja, as bases do sistema Themis, visto que, devido à consistência de seus dados, são, aparentemente, as mais apropriadas para o encontro de possíveis descobertas de conhecimento.

Como as bases de dados escolhidas estão descentralizadas, ou seja, cada comarca possui o seu servidor de banco de dados, Oracle, e não dispomos de um repositório único, será necessário um esforço adicional para integração destes dados. Já que o número de comarcas é bastante extenso, cerca de 80 nestas características, decidiu-se trabalhar com uma amostra de bases de dados.

A administração do Tribunal de Justiça divide o Estado em 14 regionais. Decidiu-se escolher uma comarca de cada regional, que, entre si, possuam

características semelhantes, como: entrância = inicial; número de varas = 1; comarca interligada com o TJ; e data de implantação do sistema no ano de 2001 ou 2002.

Baseado nestes critérios, a tabela 4.2 mostra a relação de comarcas escolhidas:

Tabela 4.2: Relação de comarcas escolhidas

Regional	Comarca
Alegrete	Cacequi
Bagé	Lavras do Sul
Caxias do Sul	Gramado
Erechim	Getúlio Vargas
Ijuí	Augusto Pestana
Lajeado	Arvorezinha
Osório	Palamres do Sul
Palmeira das Missões	Rodeio Bonito
Passo Fundo	Não-Me-Toque
Pelotas	São José do Norte
Porto Alegre	Triunfo
Santa Cruz do Sul	Candelária
Santa Maria	Faxinal do Soturno
Santa Rosa	Horizontina

Os dados destas comarcas foram replicados *on line*, a partir de conexão remota, e posteriormente, serão centralizados em um banco de dados Microsoft Access.

4.2.2 Modelo Lógico de Dados

Abaixo, segue modelo lógico de dados para compreensão do domínio de dados a ser explorado (figura 4.2).

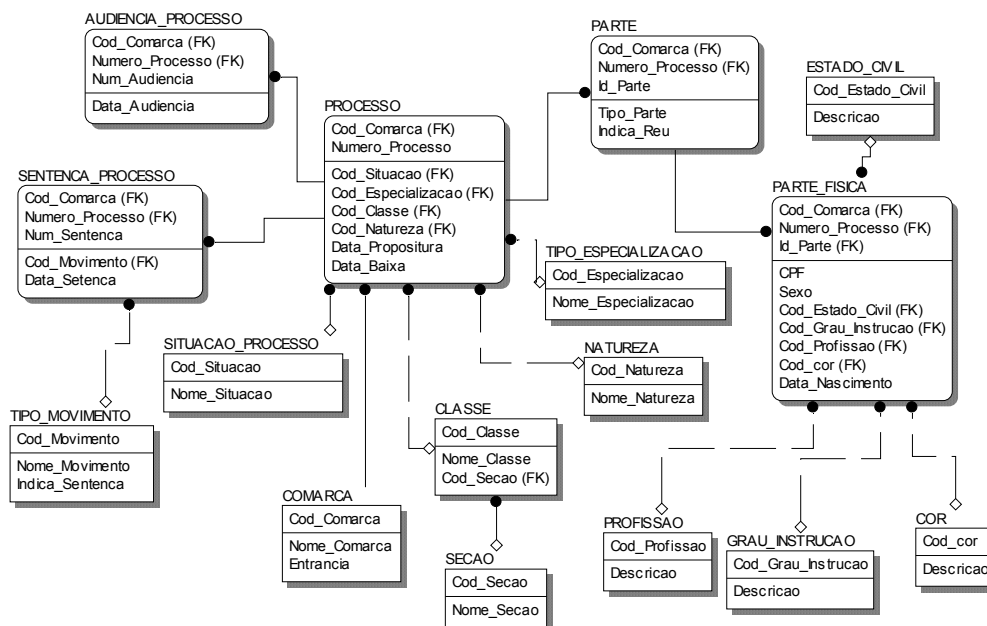


Figura 4.2: Modelo Lógico de Dados

4.2.2.1 *Descrição dos dados*

As entidades e atributos, compreendidos no modelo lógico de dados na figura 4-2, bem como suas descrições e tipos de dados estão descritos no Anexo A.

4.2.2.2 *Valores dos Atributos*

Conforme as entidades descritas no Anexo A, os valores dos atributos são apresentados no Anexo B.

4.2.3 **Exploração dos Dados**

As quantidades e estatísticas principais das entidades e dados envolvidos são apresentadas no Anexo C.

4.3 **Preparação de dados**

Baseado nos objetivos preestabelecidos, no entendimento dos dados, bem como nas exigências da ferramenta de mineração escolhida (Weka), as seções a seguir irão detalhar o procedimento de preparação de dados.

4.3.1 **Seleção e Mescla de Dados**

Como passo inicial da preparação de dados foram realizados os seguintes procedimentos:

- Conexão remota com os servidores de banco de dados (Oracle) das 14 localidades escolhidas, conforme já comentado no item 4.2.1;
- Seleção e importação das tabelas da aplicação transacional (Sistema Themis), necessárias para os objetivos propostos, em cada um destes servidores;
- Integração dos dados destes 14 servidores em um repositório único de dados, migrando para a plataforma de banco de dados Microsoft Access;
- Implementação de consultas, baseado nos objetivos propostos:
 1. Objetivo 1:
 - a. Item “1a”:
 - Descrição: *A incidência de processos considerando a Comarca de origem e a classificação processual, onde a classificação é composta pela seção, especialização e classe;*
 - Estrutura da Consulta:

Tabela 4.3: Estrutura da Consulta - Objetivo 1a

Comarca	Processo	Seção	Especialização	Classe
126	10100000607	1	CV	36
149	10100000607	1	FM	45

b. Item “1b”:

- Descrição: *A incidência de processos considerando a classificação processual com o tempo de tramitação do processo;*
- Estrutura da Consulta:

Tabela 4.4: Estrutura da Consulta - Objetivo 1b

Comarca	Processo	Seção	Especialização	Classe	Situação	Propositura	Baixa	Tempo
126	10100000623	1	CV	31	R	27-out-99	11-jun-02	958
126	10100000631	1	CV	31	R	05-jul-00	01-out-02	818

- Informações adicionais: Este arquivo foi segmentado em dois: Justiça Comum (seções 1 e 2); e Juizados Especiais (seções 3 e 4), pois o tempo médio de tramitação, em princípio, é bastante diferente nestas duas ramificações.

c. Item “1c”:

- Descrição: *A incidência de processos considerando a classificação processual com o tipo de sentença proferida;*
- Estrutura da Consulta:

Tabela 4.5: Estrutura da Consulta - Objetivo 1c

Comarca	Processo	Seção	Especialização	Classe	Tipo_Sentença
126	10100000674	1	CV	31	12
126	10100000569	1	CV	31	12

d. Item “1d”

- Descrição: *A influência da ocorrência de audiência em relação ao tempo de tramitação do processo, baseado na classificação processual;*
- Estrutura da Consulta:

Tabela 4.6: Estrutura da Consulta - Objetivo 1e

Comarca	Processo	Seção	Classe	Especialização	Situação	Propositura	Baixa	Tempo	Audiência
126	10100000399	1	34	CV	R	12-abr-01	26-abr-02	379	10-abr-02
126	10100000860	1	34	CV	R	03-set-01	06-fev-02	156	

2. Objetivo 2:

a. Itens “2a” e “2b”:

- Descrição: *Definir o perfil dos réus em processos criminais, considerando as características*

pessoais e, em um segundo momento segmentando por natureza processual;

- Estrutura da Consulta:

Tabela 4.7: Estrutura da Consulta - Objetivo 2

Comarca	Processo	Especialização	Classe	Natureza	sexo	Estado Civil	Grau Instrução	Profissao	Cor	Idade
126	20100000078	CR	102	244	M	1	1	123	2	40
126	20100000078	CR	102	244	M	2	2	123	1	69

4.3.2 Limpeza de Dados

No momento de definir os objetivos, e conseqüentemente, o conjunto de dados selecionado, foi realizada análise com o intuito de trabalhar, o quanto possível, com dados consistentes, procedimento que facilitou a tarefa de limpeza de dados, pois a quantidade de dados nulos e errôneos, para o 1º objetivo, era insignificante.

Como exemplificação, encontraram-se alguns processos baixados, sem data de baixa, ou com tempo de tramitação negativo (data de baixa errônea). Para contornar tais ocorrências, entrou-se em contato com os responsáveis pelos processos para que os mesmos efetuassem as correções nas bases de origem, e posteriormente atualizou-se o repositório de nossa aplicação.

Já para o 2º objetivo tivemos que nos deparar com uma quantidade de dados nulos (profissão, estado civil, grau de instrução). Não realizamos alterações, como supressão de registros, apenas adequamos os campos nulos para o formato de entrada requerido pela ferramenta de mineração.

4.3.3 Construção de Dados

Dentre as construções de pré-processamento que foram necessárias realizar destacam-se:

- Cálculo do tempo de tramitação de cada processo, conforme tabela 4.9, abaixo:

Tabela 4.8: Construção - Tempos de Tramitação

Tempos de Tramitação
Tramitação até a baixa: [Data da Baixa] – [Data da Propositura], com situacao_processo = (baixado ou arquivado).
Tramitação até a audiência = [Data da Audiência] – [Data da Propositura], com situacao_processo = (ativo, baixado ou arquivado).

- Cálculo da idade do réu a partir da sua data de nascimento:

Tabela 4.9: Construção - Idade

Idade
Idade= [Data da Hoje] – [Data de Nascimento]

- Utilização da função de pré-processamento *Filter Discretize*, disponível na ferramenta Weka, com o objetivo de transformar em intervalos contínuos:

1. Os tempos de tramitação de processos;
2. As idades dos réus.

4.3.4 Formatação de Dados

A ferramenta de mineração de dados escolhida estabelece um padrão para os dados de entrada. Sendo assim, foram executados os seguintes passos para cada uma das consultas implementadas:

- As consultas, descritas no item 4.3.1, foram convertidas em arquivos planos (texto), no formato *.csv* (separação dos dados por vírgulas);
- Estes arquivos texto foram adaptados para o formato *.arff* (*Attribute Relation Format File*), formato que a ferramenta Weka admite a entrada de dados. A explicação completa sobre o significado de cada um dos elementos, bem como as representações possíveis deste formato são encontradas em <http://www.cs.waikato.ac.nz/~ml/weka/arff.html>.

Como exemplo dos procedimentos executados para cada uma das consultas a fim de transformá-las em formato *.arff*, segue roteiro:

1. Consulta no formato *.mdb* (Microsoft Access):

Tabela 4.10: Consulta Access

Comarca	Processo	Seção	Especialização	Classe
126	101000006071		CV	36
149	101000006071		FM	45

2. Arquivo no formato *.csv*:

Tabela 4.11: Formato *.csv*

126,10100000607,1,36,CV
 126,10100000615,1,34,CV
 126,20100000205,2,102,CR
 126,10100000623,1,31,CV

3. Arquivo no formato *.arff*

Tabela 4.12: Formato *.arff*

```

@relation Objetivo1a
@attribute COD_COMARCA {126,96,85,50,89,149,139,82,158,151}
@attribute PROCESSO real
@attribute COD_SECAO {1,2,3,4}
@attribute COD_CLASSE {109,104,33,35,105,38,41,37,43,108 }
@attribute COD_ESPECIALIZACAO {AC,AJ,AT,CI,CR,CS,CV,DC }

@Data
126,10100000607,1,36,CV
126,10100000615,1,34,CV
126,20100000205,2,102,CR
126,10100000623,1,31,CV
....

```

Observação: Os campos nulos devem ser inseridos com o caractere “?” no formato *.arff*.

4.4 Mineração

Como esta fase é extremamente prática, descreveremos aqui um resumo das tarefas realizadas na ferramenta Weka, bem como os resultados obtidos a partir dos objetivos preestabelecidos e dos arquivos de entrada preparados. A seguir, seguem os procedimentos realizados:

4.4.1 Mineração de dados

Como ponto de partida, definiu-se que dentre as técnicas escolhidas (segmentação e regras de associação) para a aplicação, começar-se-ia pela segmentação (*clustering*). Para tanto, escolheu-se o algoritmo K-médias, realizando as seguintes variações, para atender o objetivo 1, item “a”:

- Os filtros aplicados em relação aos seus atributos foram:
 1. Seção x Classe x Especialização;
 2. Comarca x Seção x Classe x Especialização.
- O algoritmo K-médias (implementado na ferramenta Weka) permite a configuração de três parâmetros:
 1. Número de Clusters: para cada um dos filtros do arquivo de entrada executou-se a mineração para três conjuntos de clusters: 4; 8; e 12;
 2. Número randômico para escolha (alocação) dos centróides iniciais (*seed*): para os testes realizados com 8 clusters, utilizou-se os seguintes números randômicos: 1; 2; 5; 10 (padrão da ferramenta); 25; 50; e 100;

3. Modo de cluster: para todos os testes foi utilizado o modo padrão “use training set”, ou seja, depois da geração dos clusters, o algoritmo classifica as instâncias de treinamento dentro dos clusters já representados, e calcula a porcentagem de instâncias em cada cluster.

Na definição das configurações do algoritmo para aplicar o objetivo 1, itens de “b” à “d”, foram considerados os resultados obtidos com o item “a”, as quais permaneceram inalteradas no que se refere aos conjuntos de clusters testados. Apenas fixou-se o *seed* no padrão da ferramenta, igual a 10. Já quanto aos filtros, como comentado anteriormente, para o arquivo que envolvia tempo de tramitação do processo (objetivo 1, item “b”) aplicou-se uma função de pré-processamento da ferramenta, *Discretize Filter*¹², para realizar a “discretização” do tempo, ou seja, a função distribuiu os tempos dos processos em intervalos contínuos.

Já para o objetivo 2 realizou-se testes com as seguintes variações:

- Os filtros aplicados foram:
 1. Sexo x Estado Civil x Idade;
 2. Sexo x Estado Civil x Grau de Instrução x Idade;
 3. Sexo x Estado Civil x Grau de Instrução x Profissão x Idade;
 4. Sexo x Estado Civil x Grau de Instrução x Profissão x Cor x Idade;
 5. Natureza x Sexo x Estado Civil x Grau de Instrução x Profissão x Cor x Idade;
- As configurações do algoritmo k-médias foram fixadas em:
 1. Número de Clusters: 8
 2. Número randômico para escolha dos centróides iniciais (*seed*): 10
 3. Modo de cluster: foi utilizado o modo padrão “use training set”.

4.4.2 Resultados da Mineração de Dados

Os resultados da mineração de dados para todos os objetivos propostos estão disponíveis no Anexo D.

¹² Discretize Filter: esta função de pré-processamento permite a configuração de alguns parâmetros. Os mais significativos são dois: o primeiro no qual determina-se o número de intervalos; e o segundo que define se o modo de estabelecer estes intervalos é dependente distribuição dos dados ou é, simplesmente, fixo. Para o primeiro parâmetro foi selecionado 20 intervalos, tanto para o tempo como para a idade, e para o segundo escolheu-se o modo dependente da distribuição dos dados. Embora tenha-se definido um parâmetro inicial de 20 intervalos, observou-se que a função ajustou este número e os intervalos em si, conforme a distribuição dos dados.

4.5 Avaliação

Nesta fase, serão avaliados os resultados apresentados na fase anterior, organizada conforme os objetivos preestabelecidos.

4.5.1 Avaliação dos Resultados

Para a avaliação dos resultados, concentraram-se os esforços nos testes realizados com 8 *clusters*, pois os testes realizados com 12 *clusters* geravam agrupamentos com quantidades de instâncias muito pequenas, que acabavam tornando-se insignificantes. Em contra-partida os testes realizados com 4 *clusters* ocultavam agrupamentos em potencial, entre 5 e 10% de instâncias.

4.5.1.1 Objetivo 1 – Classificação Processual

A decisão de concentrar-se nos testes de 8 *clusters* resumiu a alternância dos resultados, para atender este objetivo, a variação do parâmetro *seed*, já explicitado no item 4.4.1. Apenas variando-se este parâmetro, verificou-se distinções tanto em nível dos agrupamentos formados, quanto nas quantidades de instâncias atribuídas a cada agrupamento (tabela 4.14). Acredita-se que estas distinções são atribuídas a porção de aleatoriedade do algoritmo K-médias (implementado na ferramenta Weka) no momento de estabelecer os agrupamentos iniciais. Como não se chegou a uma conclusão sobre este comportamento, resolveu-se trabalhar todos os demais testes com o *seed* padrão.

Além disso, cabe destacar que alguns agrupamentos, como por exemplo, na tabela 4-14: teste C, *clusters* 5 e 6; teste D, *cluster* 7; teste F, *cluster* 4 e 7; e teste G, *cluster* 5, embora listados, não apresentam quantitativos de instâncias. Provavelmente, isto acontece pelo fato do agrupamento, inicialmente atribuído pelo algoritmo, não possuir registros em quantidades mínimas de instâncias. Na ocorrência de tal fato, desconsiderou-se os *clusters* com estas características.

Tabela 4.13: Testes com 8 Clusters variando o *Seed*

Teste A -Número de Clusters = 8 seed=1					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	102	CR	27900	36
Cluster 1	4	108	JC	8931	12
Cluster 2	1	45	FM	15350	20
Cluster 3	1	31	CV	10928	14
Cluster 4	1	36	CV	5729	7
Cluster 5	2	107	CS	5950	8
Cluster 6	1	41	FC	618	1
Cluster 7	1	31	FP	2065	3
Teste B -Número de Clusters = 8 seed=2					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	4	108	JC	9248	12
Cluster 1	1	45	FA	16139	21
Cluster 2	1	42	PV	5540	7
Cluster 3	1	45	FM	11185	14
Cluster 4	1	36	CV	15159	20
Cluster 5	3	43	JP	3808	5
Cluster 6	2	102	CR	13515	17

Cluster 7	3	43	JE	2877	4
Teste C - Número de Clusters = 8 seed=5					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	4	108	JC	8931	12
Cluster 1	3	43	JP	6687	9
Cluster 2	1	42	FM	32773	42
Cluster 3	2	107	CS	5950	8
Cluster 4	1	36	CV	15248	20
Cluster 5	1	109	AC		
Cluster 6	1	109	AC		
Cluster 7	2	102	CR	7882	10
Teste D - Número de Clusters = 8 – Seed = 10					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	4	108	JC	8931	12
Cluster 1	3	43	JP	6687	9
Cluster 2	1	31	FA	21071	27
Cluster 3	1	45	FM	15322	20
Cluster 4	1	36	CV	11628	15
Cluster 5	2	107	PC	5950	8
Cluster 6	2	102	CR	7882	10
Cluster 7	1	109	AC		
Teste E - Número de Clusters = 8 - Seed=25					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	108	JC	26505	34
Cluster 1	1	42	PV	5496	7
Cluster 2	1	31	FA	6952	9
Cluster 3	1	44	FA	1836	2
Cluster 4	1	45	FP	4975	6
Cluster 5	1	45	FM	11157	14
Cluster 6	2	102	CR	13865	18
Cluster 7	3	43	JP	6685	9
Teste F - Número de Clusters = 8 - Seed = 50					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	45	FM	30802	40
Cluster 1	2	102	CR	13865	18
Cluster 2	1	33	FA	6993	9
Cluster 3	4	108	JC	8898	11
Cluster 4	1	109	AC		
Cluster 5	1	37	VS	1710	2
Cluster 6	1	36	CV	15203	20
Cluster 7	1	109	AC		
Teste G - Número de Clusters = 8 - Seed=100					
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	45	FM	25827	33
Cluster 1	2	102	CR	11449	15
Cluster 2	4	108	JC	8898	11
Cluster 3	2	103	CS	2416	3
Cluster 4	1	36	CV	15248	20
Cluster 5	1	109	AC		
Cluster 6	3	43	JP	6685	9
Cluster 7	1	33	FA	6948	9

Esclarecidos os detalhes técnicos, apresenta-se a avaliação dos resultados considerando a classificação processual (Figura 4.3), desconsiderando, num primeiro momento, o atributo comarca de origem:

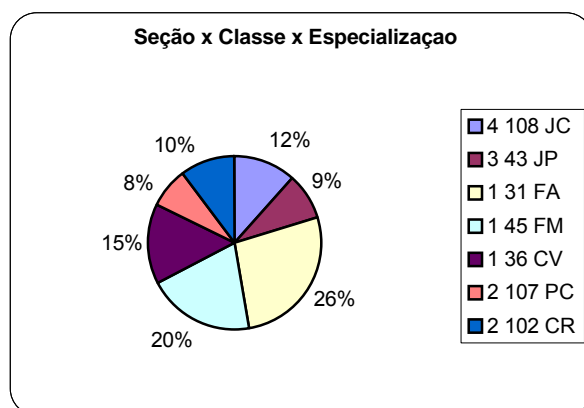


Figura 4.3: Seção x Classe x Especialização

1. 26% dos casos estão agrupados na seção: *cível (1)*, na classe: *processo de conhecimento (31)* e na especialização: *família (FA)*;
2. 20% dos casos estão agrupados na seção: *cível (1)* na classe: *processo de execução fiscal (45)* e na especialização: *fazenda pública (FM)*;
3. 15% dos casos estão agrupados em seção: *cível (1)*, na classe: *processos de execução (36)* e na especialização: *cível (CV)*;
4. 12% dos casos estão agrupados na seção: *juizado especial crime(4)*, *classe associadas e especialização* (obs: nos juizados especiais a especialização e classe tem a mesma nomenclatura, não existindo diferenciação).
5. 10% dos casos estão agrupados na seção: *crime (2)*, na classe: *procedimento ordinário-reclusão (36)* e na especialização: *crime (CR)*;
6. 9% dos casos estão agrupados na seção: *juizado especial cível (3)*, na classe e especialização associadas.
7. 8% dos casos estão agrupados na seção: *crime (2)*, na classe: *precatórias criminais (36)* e na especialização: *precatória crime (PC)*;

Comentário:

A alta incidência de processos cíveis, que tramitam na Justiça Comum, de especialização família (26%) e fazenda pública (20%), pode indicar que a administração deva promover a especialização de varas, ou seja, alocação de juizes

específicos para julgamento destes tipos de processo, ou então, passar estas especializações para os Juizados Especiais, que tem trâmite abreviado, em relação à Justiça Comum.

Destaca-se, também, a supremacia do Juizados Especiais Criminais (12%) ao dos Juizados Especiais Cíveis (9%), pois se pensava que a maior contingência de processos, nos Juizados Especiais, fossem cíveis. Antigamente, pequenos delitos (ofensas, ameaças e etc.) eram absorvidos pelas delegacias e não chegavam até o Judiciário, mas com a criação da lei dos Juizados Especiais Criminais, que tinha a intenção de recobrar a impunidade, as delegacias são obrigadas a encaminhar qualquer ocorrência ao Judiciário, o que provoca um grande volume deste tipo de processo, pois são muitos menos formais que os processos do Juizado Especial Cível. Talvez, essa lei deva ser revista para tentar aliviar esta demanda.

Ainda, cabe destacar que o índice de 8% em precatórias criminais é significativo, visto que este tipo de processo ocorre, normalmente, quando é necessário intimar uma pessoa, mas ela não se encontra na jurisdição da Comarca onde ocorreu o fato. Isto acontece com frequência entre municípios que tenham pequena distância entre si e não tenham mesma jurisdição, provocando trabalho adicional a Comarca vizinha, já que a precatória crime tramita na Comarca onde a pessoa intimada reside no momento e não onde ocorreu o fato.

Acrescentando a comarca de origem, os clusters que mais se destacam são:

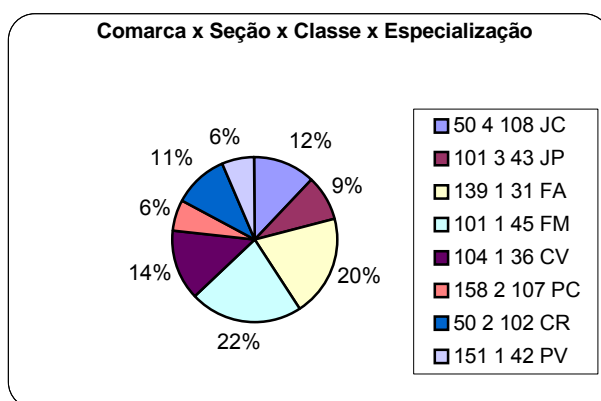


Figura 4.4: Comarca x Seção x Classe x Especialização

1. 22% dos casos estão agrupados na seção: *cível (1)*, na classe: *processo de execução fiscal (45)*, na especialização: *fazenda pública (FM)* e a comarca que se destaca é a de *Gramado (101)*;
2. 20% dos casos estão agrupados na seção: *cível (1)*, na classe: *processo de conhecimento (31)*, na especialização: *família (FA)* e a comarca que se destaca é a de *Triunfo (139)*;
3. 14% dos casos estão agrupados em seção: *cível (1)*, na classe: *processos de execução (36)*, na especialização: *cível (CV)* e a comarca que se destaca é a de *Horizontina (104)*;

4. 12% dos casos estão agrupados na seção: *juizado especial crime(4)*, na classe e especialização associadas, e a comarca que se destaca é a de *Getúlio Vargas (50)*;
5. 11% dos casos estão agrupados na seção: *crime (2)*, na classe: *procedimento ordinário- reclusão (102)*, na especialização: *crime (CR)* e a comarca que se destaca é a de *Getulio Vargas (50)*;
6. 9% dos casos estão agrupados na seção: *juizado especial cível (3)*, na classe e especialização associadas, e a comarca que se destaca é a de *Gramado (101)*;
7. 6% dos casos estão agrupados na seção: *crime (2)*, na classe: *precatórias criminais (107)*, na especialização: *precatória crime (PC)* e a comarca que se destaca é a de *Rodeio Bonito (158)*;
8. 6% dos casos estão agrupados na seção: *cível (1)*, na classe: *precatórias cíveis (42)*, na especialização: *precatória cível (PV)* e a comarca que se destaca é a de *Palmares do Sul (151)*;

Comentário:

Tanto na Justiça Comum, quanto nos Juizados Especiais, a comarca que se destaca em processos criminais é a de Getúlio Vargas, fato que deve ser noticiado ao Juiz e Prefeito da cidade para que os mesmos possam realizar suas próprias avaliações. Já, a Comarca de Gramado aparece em dois agrupamentos relacionados, destaque na incidência de processos cíveis nos Juizados Especiais e, também, no grande volume de processos de execução fiscal (fazenda pública). Isto pode estar evidenciando que as relações comerciais, nesta cidade, estão provocando desavenças, bem como gerando uma grande inadimplência no pagamento de impostos.

4.5.1.2 Objetivo 1 – Classificação Processual x Tempo

A avaliação deste item do 1º objetivo está segmentada em dois grupos: o primeiro que elucidará a influência da classificação de processos no tempo de tramitação do mesmo na Justiça Comum; e o segundo que mostrará a tendência da mesma proposição nos Juizados Especiais. Para ambos os grupos, também, somente foram considerados os testes de 8 *clusters*.

Conforme já informado na preparação de dados, no item 4.3.1, este objetivo envolve o atributo tempo, para o qual foi aplicado a função de pré-processamento *Discretize Filter*, mencionada nos itens 4.3.3 e 4.4.1.

4.5.1.2.1 Justiça Comum

Para os processos da Justiça Comum, os intervalos obtidos a partir da aplicação da função e dos resultados apresentados foram:

Tabela 4.14: Intervalos de Tempo - Justiça Comum

Intervalos
(-inf – 75,1)
[75,1 – 150,2)
[150,2 – 225,3)
[225,3 – 300,4)
[300,4 – 375, 5)
.... ¹³ (não informado)
[1051, 4 – inf)

A seguir, apresenta-se a avaliação dos resultados deste teste para processos da Justiça Comum, sem considerar a comarca de origem:

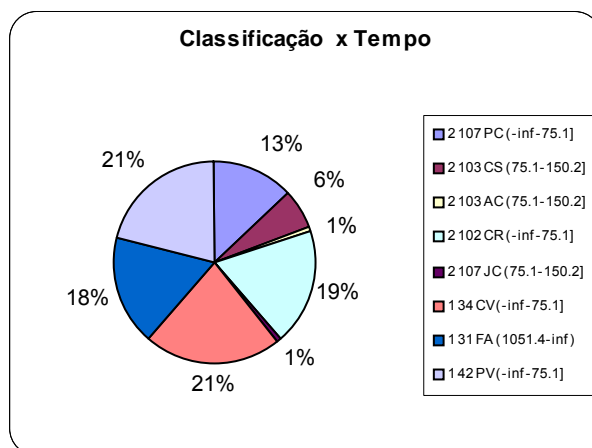


Figura 4.5: Resultados - Classificação x Tempo - Justiça Comum

1. Com tempos inferiores a 75,1 dias de tramitação encontram-se:
 - a. Os processos da seção: *cível* (1), da classe: *precatória cível* (42) e especialização: *precatória cível* (PV), com 21% dos casos;
 - b. Os processos da seção: *cível* (1), da classe: *procedimento especial de jurisdição contenciosa*(34) e especialização: *cível* (CV), com 21% dos casos;
 - c. Os processos da seção: *crime* (2), da classe: *procedimento ordinário-reclusão* (102) e especialização: *crime*(CR), com 19% dos casos;

¹³ os intervalos constantes entre 375,5 e 1051,4 não foram apresentados em nenhum dos testes realizados.

- d. Os processos da seção: *crime* (2), da classe: *precatória crime* (107) e especialização: *precatória crime* (PC), com 13% dos casos;
2. Com tempos entre 75,1 e 150,2 dias de tramitação, encontram-se os processos da seção: *crime* (2), da classe: *procedimento sumário-detenção* (103) e especialização: *crime* (CR), com 6% do total de casos.
3. Com tempos superiores a 1051,4 dias encontram-se os processos da seção: *cível*, da classe: *processo de conhecimento* (31) e especialização: *família* (FA), totalizando 18% dos casos.

Comentário:

Embora não se possa comparar, efetivamente, os tempos apresentados pelos agrupamentos, com os tempos médios agrupados por classe e especialização, pois estes tempos médios não levam em consideração a distribuição dos dados, realizou-se esta consulta para confirmar determinados tempos de tramitação, que, aparentemente, estavam muito enxutos. E as seguintes constatações foram extraídas:

- No geral, os tempos médios são superiores, duas a três vezes, aos apresentados pelos agrupamentos, sendo a maior discrepância a diferença apresentada pelo agrupamento de processos cíveis de classe *procedimento especial de jurisdição contenciosa* que tem tempos médios em torno de 596,7 dias e estão identificados pelo agrupamento como inferiores a 75,1 dias.
- O único agrupamento deste teste que correspondeu ao tempo médio foi o de processos criminais de *classe precatória crime*, que tem tramitação em torno de 57,9 dias;
- Os tempos apresentados pelo agrupamento dos processos cíveis, da classe *processo de conhecimento* e especialização *família* foi superior ao tempo médio desta tipificação, que é de 556,4 dias, praticamente a metade identificada no agrupamento.

A lista completa dos tempos médios, encontra-se no Anexo C, no item tempos médios de processos na Justiça Comum.

Com relação a análise dos resultados deste teste, no sentido de descobertas de conhecimento, confirma-se a preocupação em procurar-se mecanismos para otimizar os processos cíveis de especialização *família*, que além de numerosos, têm tempos de tramitação excessivos.

Incorporando a comarca de origem nesta avaliação, resumimos:

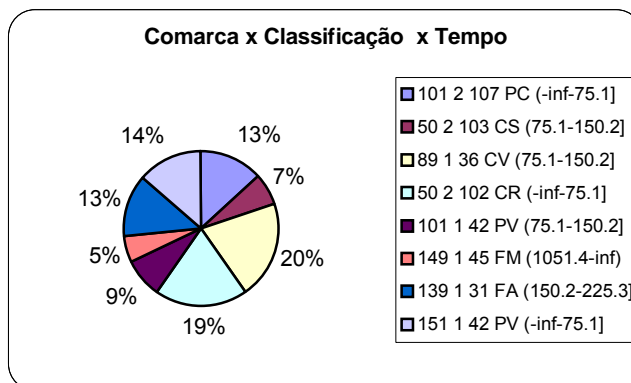


Figura 4.6: Classificação x Comarca x Tempo – Justiça Comum

1. Com tempos inferiores a 75,1 dias, concentram-se:
 - a. Os processos da comarca de Palmares do Sul (151), da seção: *cível (1)*, da classe e especialização: *precatória cível*;
 - b. Os processos da comarca de Gramado (101), da seção *crime (2)*, da classe e especialização: *precatória crime*;
 - c. Os processos da comarca de Getúlio Vargas (50) da seção: *crime (2)*, da classe: *procedimento ordinário-reclusão (103)*.
2. Com tempos entre 75,1 e 150, 2 dias, estão:
 - a. Os processos da comarca de Getúlio Vargas (50) da seção: *crime(2)* e da classe: *procedimento sumário-detenção (103)*;
 - b. Os processos da comarca de Candelária (149) da seção: *cível* e da classe: *processos de execução*;
 - c. Os processos da comarca de Gramado (101), da seção: *cível (1)*, da classe e especialização: *precatória cível*;
3. Com tempos entre 150, 2 e 225,3 dias, estão os processos da comarca de Triunfo (139), da seção: *cível(1)*, da classe: *processo de conhecimento(31)* e especialização: *família (FA)*;
4. Com tempos superiores a 1051,4 dias, estão os processos da comarca de Augusto Pestana (149), da seção: *cível (1)*, da classe: *processo de execução fiscal (45)* e especialização: *fazenda pública (FM)*;

Comentário:

O tempo de tramitação dos processos cíveis de especialização *família*, na comarca de Triunfo é bem inferior ao da média das comarcas (1051,4 dias, conforme

resultado da técnica de segmentação e de 554 dias, de acordo com a média geral), ficando entre 150,2 e 225,3 dias.

Também, cabe destacar o excessivo tempo de tramitação dos processos cíveis de classe *execução fiscal* e especialização *fazenda pública*, na comarca de Augusto Pestana, com mais de 1051,4 dias. Neste caso, salienta-se que algumas comarcas realizam convênios com as prefeituras locais, para abreviar o trâmite jurídico destes tipos de processos. Normalmente, a prefeitura cede alguns funcionários para a comarca, para que os mesmos entrem em contato com os devedores a fim de negociar as dívidas. Este procedimento reduz a quantidade de processos que necessitam passar por toda tramitação convencional. A comarca de Augusto Pestana não estabeleceu este convênio, ainda.

4.5.1.2.2 Juizados Especiais

Os intervalos relativos ao tempos processuais nos Juizados Especiais, estabelecidos após a aplicação da função de pré-processamento, foram:

Tabela 4.15: Intervalos de Tempos - Juizados Especiais

Intervalos
(-inf – 41,79)
[41,79 – 83,58)
[83,58 – 125,37)
[125,37 – inf)

Com base nos experimentos realizados, segue a avaliação relativa aos tempos processuais nos Juizados Especiais, inicialmente, não incluindo o atributo comarca:

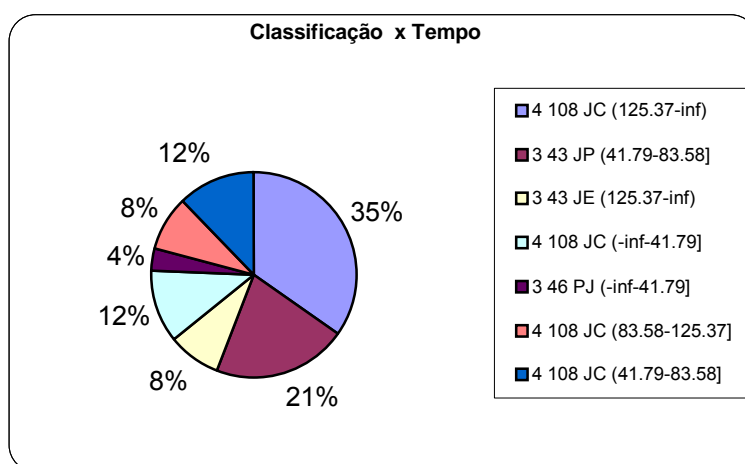


Figura 4.7: Classificação x Tempo - Juizados Especiais

- Os processos da seção, classe e especialização *juizados especiais criminais (4)*, estão enquadrados em todos os intervalos de tempo estabelecidos pela ferramenta;

2. Os processos da seção, classe e especialização: *juizados especiais cíveis*(3), têm tempos de tramitação entre 41,79 e 83,59 dias;
3. Os processos da seção e classe: *juizados especiais cíveis* (3), e especialização: *execução* (JE) têm tempos de tramitação superiores a 125,37 dias;
4. Os processos da seção: *juizados especiais cíveis* (3), da classe e especialização: *precatórias*(PJ) têm tempos de tramitação inferiores a 41,79 dias;

Comentário:

Conforme os tempos médios calculados (Anexo C, item: Tempos Médios – Juizados Especiais), o tempo de tramitação dos processos nos *Juizados Especiais Criminais* é maior que 125,37 dias, cerca de 200 dias, o que vai ao encontro dos resultados obtidos neste experimento, onde a maioria dos processos desta tipificação encontram-se no agrupamento que exprime esta tendência.

Também, para os processos dos *Juizados Especiais Cíveis*, os tempos médios acompanham os agrupamentos destacados na figura 4.7, dando ênfase aos processos de especialização de *execução*, que possuem tempo de tramitação superior as demais especializações deste Juizado.

Incluindo o atributo comarca, a figura 4.8 mostra os agrupamentos destacáveis:

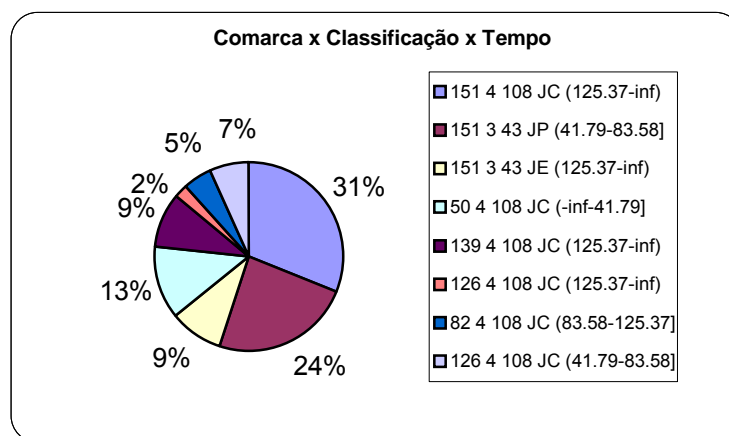


Figura 4.8: Resultados Classificação x Comarca x Tempo - Juizados Especiais

Os processos dos Juizados Especiais Criminais têm comportamento semelhante nas comarcas de Rodeio Bonito (151), Triunfo (139), com tempos de tramitação superiores a 125,37 dias. Nas demais comarcas o tempo destes tipos de processo apresentam variação de uma para outra, destacando a comarca de Getulio Vargas que, embora seja a comarca que se destaca pela quantidade desta tipificação, tem tempo de tramitação inferior a 41,79 dias.

Demais resultados acompanham a tendência dos apresentados na figura 4.7, nos quais não está inserido o atributo comarca.

4.5.1.3 Objetivo 1 - Classificação Processual x Sentença

Procurou-se analisar, neste item, a influência da classificação processual no padrão de sentenças proferidas. A figura 4.9 mostra os resultados obtidos:

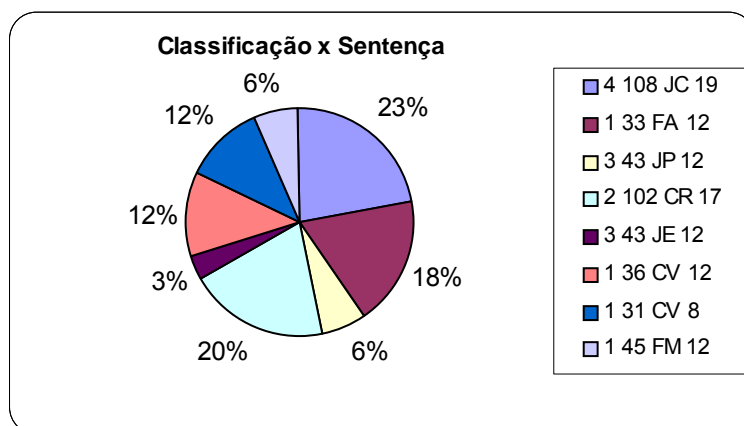


Figura 4.9: Classificação Processual x Sentença Proferida

1. Para os processos das seções *cível (1)* e *juizado especial cível (3)*, não foi encontrada informação relevante, pois os tipos de sentença constantes nos agrupamentos, foram, na maioria, tipos genéricos, ou seja, *Demais sentenças e decisões (12)* e *Outras decisões terminativas (17)*. Para extrair-se algum conhecimento a partir da tipificação de sentenças, a aplicação transaccional deve ser revista a fim de qualificar estas categorias;
2. Para os processos da seção: *juizados especiais criminais (4)*, classe e especialização associadas, chama a atenção o alto índice (23% dos casos) do tipo de sentença: *Extinção de punibilidade (19)*, fato que revela que muitas das ocorrências do grande volume de processos dos juizados especiais criminais não são levadas adiante ou não ocorre a “representação” dentro do prazo legal.

Observação: não foi considerada a influência da comarca neste teste.

4.5.1.4 Objetivo 1 – Classificação Processual x Audiência

Analisando este teste, concluiu-se que as descobertas relevantes não foram baseadas nas classificações processuais que tinham maior número de ocorrências, caso dos juizados especiais criminais (4), com 45%, dos processos *crimes ordinários de reclusão (102)*, com 21%, ou dos processos *cíveis de especialização família (FA)*, com 16%, pois as audiências nestas categorias são uma constante. Os agrupamentos interessantes estavam concentrados nos de menores ocorrências. Conforme resultados apresentados na figura 4.10, identificou-se como agrupamento incomum os processos da seção: *cível (1)*, classe e especialização *precatórias cíveis (PV)*, terem incidência em audiências. Esta tendência se justifica pois em muitos processos desta classificação existe a necessidade de ouvir testemunhas.

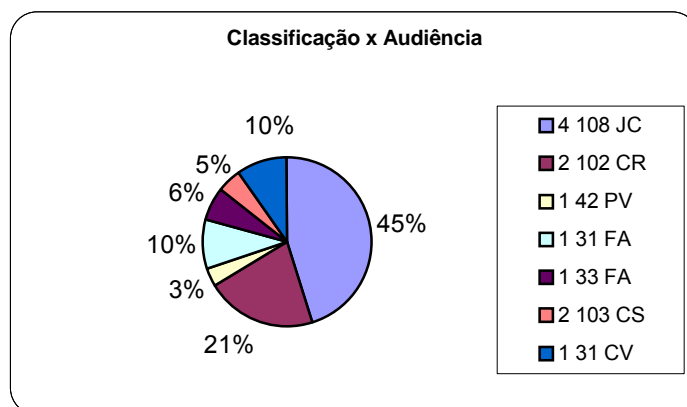


Figura 4.10: Resultados Classificação Processual x Audiência

4.5.1.5 Objetivo 2 – Perfil dos Réus em Processos Crimes

Baseado nos experimentos, a seguir descritos (figura 4.11 e 4.12), procurou-se verificar a existência de padrões de sexo, estado civil, escolaridade e raça dos réus em processos criminais e a dependência destes padrões em relação a matéria (natureza) criminal. Estes experimentos envolveram o atributo idade que foi “discretizado” nos seguintes intervalos:

Tabela 4.16: Intervalos - Idade

Intervalos - Idade
(-inf – 17,4)
[17,4 – 22,5)
[22,5 – 27,6)
[27,6 – 32,7)
[32,7 – 38,8)
[38,8 – 42,9)
[42,9 – 48)
[48 – 53,1)
[53,1 – 58,2)
[58,2 – inf)

Na figura abaixo, mostra-se os resultados obtidos, primeiramente, a partir do perfil dos réus nos processos criminais:

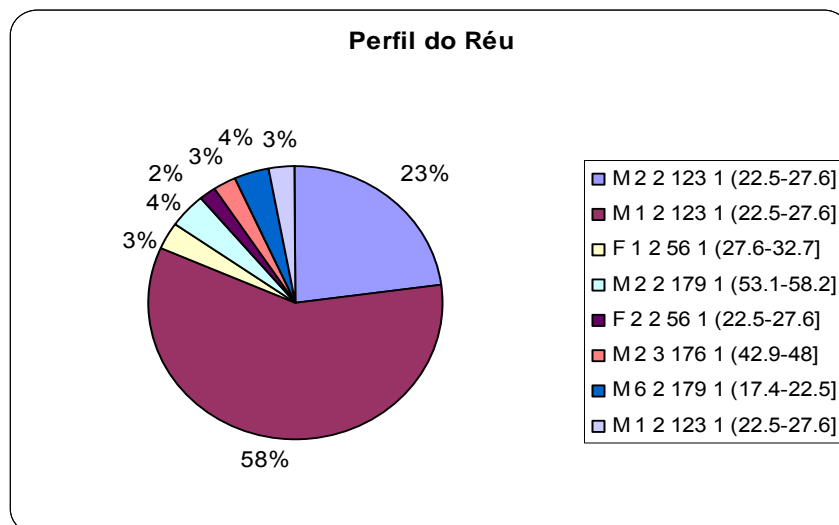


Figura 4.11: Perfil do Réu em Processos Criminais

1. Com 58% dos casos, o agrupamento predominante tem as seguintes características:
 - a. Sexo: *masculino (M)*
 - b. Estado Civil: *solteiro (1)*
 - c. Grau de Instrução: *1º Grau (2)*
 - d. Profissão: *Outras (123)*
 - e. Cor : *branca (1)*
 - f. Idade: entre 22,5 e 27,6 anos
2. Com 23% dos casos, encontra-se um outro agrupamento muito semelhante ao predominante, diferenciando-se apenas no estado civil, que neste caso prevalece o *casado (2)*;
3. Com 8% dos casos estão dois agrupamentos em que o sexo é *masculino (M)*, a profissão é *agricultor (179)*, o grau de instrução é 1º Grau, diferenciando-se:
 - a. Em um deles o estado civil é *amigado (6)* e a idade está entre 17,4 e 22,5 anos;
 - b. No outro o estado civil é *casado (2)* e a idade está entre 53,1 e 58,2 anos.
4. Um outro agrupamento, com 3% dos casos, tem como diferenciais:
 - a. A profissão que é *comerciante (176)*;
 - b. O grau de instrução que é *2º Grau (3)*;
 - c. A idade que está entre 42,9 e 48 anos.
5. E por último, destaca-se dois agrupamentos, onde:
 - a. O sexo é *feminino (F)*;

- b. A profissão é *dona de casa* (56);
- c. O grau de instrução é *1º Grau* (2);
- d. Diferenciando-se no estado civil, onde:
 - i. Em um prevalece *o solteiro*, sendo a idade entre 22,5 e 27,6 anos;
 - ii. E no outro *o casado*, sendo a idade entre 27,6 e 32,7 anos.

Comentário:

Os agrupamentos gerados neste teste parecem bastante homogêneos, identificando um réu jovem, predominantemente do sexo masculino, sem formação e sem profissão fixa.

Em todos os agrupamentos destacados, os réus tem raça branca, fato que, a princípio, principalmente pelo preconceito existente, pode-se estranhar, mas se explica, pois a população dos municípios considerados é, provavelmente, predominante nesta raça.

Outro ponto que merece destaque, são os agrupamentos onde o sexo feminino prevalece, que evidencia a profissão, ou falta de, destas mulheres que cometem delitos como *dona de casa*.

4.5.1.6 Objetivo 2 – Perfil dos Réus x Natureza Criminal

Os testes realizados, incluindo-se o atributo matéria (natureza) criminal ao perfil dos réus, evidenciaram que os tipos de crimes cometidos, a partir dos três agrupamentos predominantes, onde o sexo foi *masculino* (M), o grau de instrução foi *1º Grau* (2) e a idade estava entre 22,5 e 27,6 anos, foram:

1. *Crimes de Furto* (244);
2. *Crimes de Roubo e Extorsão* (245);
3. *Estelionato e Fraudes* (247).

Para os agrupamentos onde ressalta-se o sexo *feminino* (F), além dos crimes de furto (244) destaca-se os *crimes contra a honra* (298) , ou seja, ofensas, calúnias e difamações.

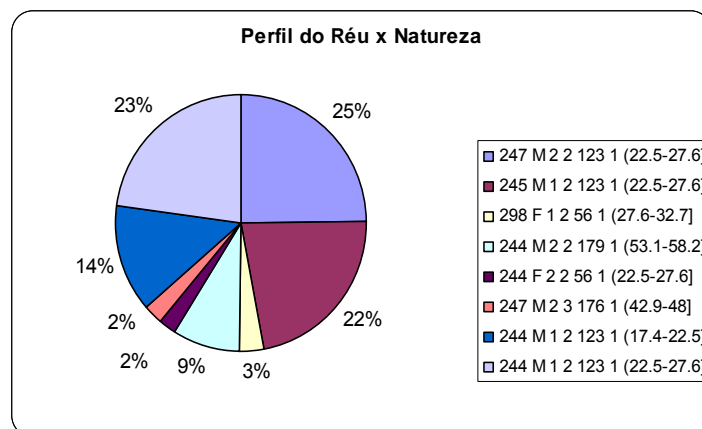


Figura 4.12: Perfil do Réu x Natureza

4.5.2 Revisão do processo

Com base na avaliação dos resultados, resolveu-se aplicar a técnica de regras de associação no 2º objetivo proposto, a fim de estabelecer a análise de dependências entre os atributos envolvidos neste processo, e de certa maneira, confirmar os resultados apresentados pela técnica de segmentação. Como já mencionado, o algoritmo utilizado para aplicar esta técnica de regras de associação foi o *Apriori*, e igualmente como explicitamos sobre os testes realizados para a técnica de segmentação, detalharemos os filtros aplicados e parâmetros estabelecidos:

- A partir do arquivo de entrada do 2º objetivo, os filtros aplicados foram:
 1. Natureza x Estado Civil x Grau de Instrução x Profissão x Cor x Idade;
 2. Comarca x Natureza x Profissão x Idade;
 3. Comarca x Natureza x Sexo (Feminino) x Profissão x Idade.
- Parâmetros do algoritmo *Apriori*
 1. Suporte Mínimo: foram testados com 10%, 5% e 1%;
 2. Confiança Mínima: foi fixada em 25%;
 3. Número de regras: foi estabelecido um número de 50 regras.

4.5.2.1 Resultados da Mineração de Dados – Técnica Regras de Associação

Os resultados da mineração de dados da técnica de regras de associação encontram-se no Anexo E.

4.5.2.2 Avaliação dos Resultados – Técnica Regras de Associação

Para os testes realizados optou-se por definir um suporte baixo, devido ao grande número de atributos não obrigatórios (profissão, estado civil, grau de instrução e raça) terem valor efetivo por volta de 50% do conjunto total de dados. E, também, para tentar identificar determinadas regras que, embora não ocorram frequentemente, possam ter fator de confiança alto. Cabe salientar, que pelos mesmos objetivos definiu-se um

grande número de regras como padrão. Além disso, para os primeiros testes, não foi considerado o sexo, visto que, mais de 90 % dos infratores são do sexo masculino. No último teste, apenas foram consideradas ocorrências do sexo feminino.

Para realizar a análise de dependência das regras obtidas a partir dos testes realizados, incluiu-se uma estatística adicional, o *lift*. A partir disso, foram selecionadas algumas das regras geradas que, certamente, irão complementar os resultados da técnica já explorada. As tabelas 4.17 e 4.18 mostram estes novos resultados:

Tabela 4.17: Resumo dos resultados dos testes 1,2,4 e 5 do Anexo E

Regras obtidas independente do sexo		
Atributo = Natureza		
Suporte	Quando a natureza for Tráfico de Drogas, o que ocorre em 5,5 % das vezes	
1 Confiança	A Comarca é Palmares do Sul, em 42 % das vezes.	Lift = 4,2
2 Confiança	A idade é entre 22,5 e 27 anos, em 31 % das vezes.	Lift = 1,5
Suporte	Quando a natureza for Crimes de Roubo e Extorsão, o que ocorre em 7 % das vezes,	
1 Confiança	A profissão é Outras ¹⁴ , em 33% das vezes	Lift = 1,69
Atributo = Idade		
Suporte	Quando a idade for entre 17,4 e 22,5 anos, o que ocorre em 9,9 % das vezes,	
1 Confiança	O estado civil é solteiro, em 81 % das vezes.	Lift = 2,14
Suporte	Quando a idade for entre 22,5 e 27,6 anos, o que ocorre em 15,1 % das vezes,	
1 Confiança	O estado civil é solteiro, em 74 % das vezes.	Lift = 1,97
4 Confiança	O grau de instrução é 1º Grau Incompleto, em 43% das vezes	Lift = 1,51
Suporte	Quando a idade for entre 27,6 e 32,7 anos, o que ocorre em 11,6 % das vezes,	
1 Confiança	O estado civil é solteiro, em 54 % das vezes.	Lift = 1,42
Atributo = Comarca		
Suporte	Quando a Comarca for Gramado, o que ocorre em 7,7% das vezes	
1 Confiança	A natureza é Estelionato e Fraudes, em 27% das vezes.	Lift = 2,32
Suporte	Quando a Comarca for Getúlio Vargas, o que ocorre em 11,1% das vezes	
1 Confiança	A profissão é Outras, em 36% das vezes.	Lift = 1,86
Suporte	Quando a Comarca for Candelária, o que ocorre em 15,7% das vezes	
1 Confiança	A natureza é Crimes de Furto, em 61% das vezes.	Lift = 1,44

Comentários:

A Comarca de Palmares do Sul tem alta incidência de processos de natureza *tráfico de drogas* em relação as demais comarcas escolhidas. Isto devido ao fato que é uma comarca localizada próxima do litoral, onde o tráfico de drogas, durante o veraneio, é intenso.

¹⁴ Outras: Entenda-se por outras como sem profissão definida

As regras obtidas, onde a cabeça da regra é o estado civil, não acrescentam grande conhecimento, pois os dados relativos a este atributo são predominantemente associados, em 38% dos casos, ao estado civil *solteiro*. Enquanto que em 26% dos casos não são informados.

Ainda, pode-se comentar sobre a incidência da natureza *estelionato e fraudes* na Comarca de Gramado, regra que só vem a confirmar os resultados obtidos nos testes da aplicação da técnica de segmentação e já comentados no item 4.5.1.1.

Tabela 4.18: Resumo dos resultados do testes 3 do Anexo E

Regras obtidas especificamente do sexo feminino	
Atributo = Comarca	
Suporte	Quando a Comarca for Não Me Toque o que ocorre em 4,8% das vezes,
1 Confiança	A profissão é dona de casa, em 55% das vezes.
Suporte	Quando a Comarca for Faxinal do Soturno e a profissão é dona de casa, o que ocorre em 4 % das vezes,
1 Confiança	A natureza é Crimes Contra Honra , em 51% das vezes.
Suporte	Quando a Comarca for Gramado, o que ocorre em 12,6 % das vezes,
1 Confiança	A natureza Estelionato e Fraudes, em 42% das vezes.
2 Confiança	A natureza Crimes de Furto, em 36% das vezes.
Atributo = Natureza	
Suporte	Quando a Natureza for Crimes contra a Administração da Justiça (261) o que ocorre em 4 % das vezes
1 Confiança	A profissão é dona de casa, em 50% das vezes.

Comentário:

Para os testes realizados, considerando somente o sexo feminino, não foi avaliado o *lift* da regra, pois esta estatística apresentava valores inferiores a 1, ou seja, uma dependência fraca entre os atributos das regras geradas. No caso, somente foi levado em consideração a confiança. Devido a isto, não se pode tirar conclusões efetivas a partir deste teste. Apenas, pode-se confirmar os resultados apresentados pela técnica de segmentação no que se refere aos agrupamentos em que o sexo feminino foi predominante.

4.6 Disponibilização

Nesta fase abordaremos as disposições finais sob o enfoque técnico-científico, relatando, no tópico revisão do projeto, a experiência adquirida com a aplicação do processo de DCBD.

4.6.1 Revisão do projeto

Com o objetivo de relatar a experiência vivida na aplicação do processo de DCBD, seguem comentários sobre os pontos mais relevantes de todo este processo:

- A primeira e talvez maior dificuldade encontrada nesta aplicação foi estabelecer os objetivos para a aplicação do processo. Torna-se complexo objetivar algo que não se sabe. No caso do Poder Judiciário, as maiores necessidades estão associadas a projetos de “*datawarehouse*” e um projeto sistemático de DCBD tem de amadurecer, para que se consiga mostrar aos especialistas o alcance de um projeto desta natureza;
- Outro ponto não trivial é estabelecer qual classe de tipo de problema e técnica de mineração de dados é mais apropriada para os objetivos propostos. Acredita-se que uma das escolhas corretas é começar com a técnica de segmentação, principalmente para quem não tem familiaridade com os dados;
- Outro fator importantíssimo é analisar a qualidade dos dados a serem minerados, para que depois de inúmeros testes não se conclua que os dados em questão em nada contribuirão para novas descobertas;
- Pela diversidade de ferramentas comerciais, talvez seja de grande valia realizar um estudo prévio de algumas ferramentas para selecionar a que seja mais proveitosa em relação aos objetivos definidos;
- Cabe salientar, que é preponderante analisar, quando possível, o funcionamento dos algoritmos a serem utilizados, caracterizando a influência dos parâmetros suportados e, principalmente, identificando o grau de aleatoriedade destes algoritmos. Esta providência facilitará bastante no momento da análise dos resultados;
- Mesmo para aplicações que pretenda-se, somente, utilizar a técnica de regras de associação, provavelmente, seja interessante aplicar a técnica de segmentação, pois ela pode evidenciar determinado conjuntos de dados que as regras de associação não identificariam, mesmo em níveis baixos de suporte;
- A definição mais apropriada para tal processo passa pelas palavras “iteratividade” e “interatividade”, pois o número de vezes que se retoma a fases anteriores a fim de provocar mais “descobertas” são incontáveis.

5 CONCLUSÃO

Os aspectos importantes extraídos deste trabalho são de grande valia tanto em nível técnico-científico, quanto em nível organizacional. Em nível científico, baseado nos poucos trabalhos práticos realizados até o momento na área de DCBD e talvez no inusitado conjunto de dados escolhido, acredita-se que a experiência relatada com a aplicação do processo possa orientar futuros trabalhos, abreviando dificuldades que possam vir a surgir. Até mesmo em níveis conceituais, procurou-se abranger as diferentes proposições de alguns autores [BER97, CHA2000, FAY96] que denominam de maneira diversa elementos da mesma área científica, da qual não entram em consenso nem no nome desta área [BRU99].

Podemos comentar, também, sobre a metodologia aplicada (CRISP-DM). Ela mostrou-se completa, permitindo um fluxo consistente, coerente e detalhado da aplicação de DCBD realizada. A única dificuldade encontrada foi em documentar determinadas tarefas que eram essencialmente práticas. E também, o local adequado de documentar “revisões” de determinada fase ou tarefa, que, porventura, seja necessária retomar para realizar novos experimentos.

Ainda em nível científico, cabe salientar aspectos relativos a ferramenta Weka. Como já mencionado, as suas características principais estão na sua portabilidade e quantidade de algoritmos que implementa. Seus fatores negativos estão associados ao pós-processamento para visualização de dados, que parece bastante incipiente e, principalmente, a falta de documentação para a utilização da ferramenta no que se refere a informação sobre a influência das parametrizações dos algoritmos, bem como das funções de pré-processamento.

Já em nível organizacional, podemos dizer que, atualmente, a administração do Tribunal de Justiça está necessitando de um projeto de *Datawarehouse* para consolidar os dados dos diversos sistemas e bases transacionais em produção, e assim passe a ter informações estatísticas corretas que auxiliem nas ações futuras. Este passo será necessário para que projetos de DCDB tornem-se viáveis e sensíveis ao Poder Judiciário.

Considera-se que este trabalho serviu como uma semente para que trabalhos futuros possam ser desenvolvidos. Ele teve limitações, principalmente, relacionadas as bases de dados disponíveis, visto que optou-se por um conjunto de dados amostral, já que não tínhamos interligação plena entre as comarcas e os dados estavam totalmente descentralizados. Com a interligação de todas as comarcas, prevista para o final deste ano, e com o intuito de centralizar parte destes dados, poderemos ampliar o conjunto dos mesmos a fim de avançar nos resultados encontrados e procurar por novas descobertas. Também, poderemos partir para novos objetivos devido a qualidade e diversidade dos dados da aplicação transacional.

É importante ressaltar que a proposição deste trabalho alerta para um projeto de DCBD, que não tenha apenas dados do Judiciário envolvidos, mas também dados do IBGE (Instituto Brasileiro de Geografia e Estatística), bem como dados da Segurança Pública, ou até mesmo dados dos Municípios, visto que um conjunto de dados pode comprovar ou esclarecer descobertas de um conjunto afim. Como exemplo, podemos citar os resultados encontrados relativos a raça dos réus que foi predominou a cor branca, caso tivéssemos distribuição racial das regiões escolhidas, poderíamos avaliar o grau de importância desta descoberta.

REFERÊNCIAS

- [AGR94] AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, VLDB, 20., 1994, Santiago, Chile. **Proceedings...**Hove: Morgan Kaufmann, 1994.
- [AGR95] AGRAWAL, R.; PSAILA, G.; Active data mining. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 1995, Montreal. **Proceedings...**[S.l.: s.n.], 1995.
- [AGR95A] AGRAWAL, R et al. **Fast Discovery of Association Rules: Advances in Knowledge Discovery and Datamining.** [S.l.]: AAAI/MIT Press, 1995.
- [BER97] BERRY, M.J.A.; LINOFF, G. **Data Mining Techniques: for Marketing, Sales and Customer Support.** New York: John Wiley & Sons, 1997. 454p.
- [BRU99] BRUHA, I. **Data Mining, KDD, and Knowledge Integration: Methodology and A Case Study.** Canadá, 1999. Disponível em: <<http://www.ssgrr.it/en/ssgrr2000/papers/288.pdf>>. Acesso em: 10 out. 2002.
- [CAB97] CABENA, P. et al. **Discovering Data Mining from Concept to Implementation.** New Jersey, USA: Prentice Hall PTR, 1997.
- [CAM2002] CAMARGO, S. da S. **Mineração de regras de associação no problema da cesta de compras aplicada ao comércio varejista de confecção.** 2002. 103p. Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre.
- [CHA2000] CHAPMAN, P. et al. **CRISP-DM 1.0, Step-by-step Datamining guide.** Disponível em: <<http://www.crisp-dm.org/CRISPWP-0800.pdf>>. Acesso em: 10 jan. 2003.
- [ENG2002] ENGEL, P. M. **Tópicos em Mineração de Dados: Técnica de Agrupamento - Algoritmo K-Médias.** Porto Alegre: PPGC da UFRGS, 2002. Notas de aula. Disponível em:<www.inf.ufrgs.br/~engel/tmd5.pdf>. Acesso em: 10 jan. 2003.
- [FAY96] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **Advances in Knowledge Discovery and Data Mining.** Menlo Park: AAAI, 1996. 611p.

- [GAM2001] GAMA, J. **Métodos de Agrupamento (Clustering)**. LIACC, FEP. Disponível em: <<http://www.niaad.liacc.up.pt/~pbrazdil/Ensino/ecdas/cluster.pdf>>. Acesso em: 20 abr. 2003.
- [JOR2000] JORGE, A. **Descoberta de Regras de Associação, Algoritmo Apriori**. Disponível em: <<http://www.niaad.liacc.up.pt/~amjorge/Aulas/madsad/ecd2/>>. Acesso em: 12 abr. 2003.
- [INM99] INMON, W.H. **Gerenciando Data Warehouse**. São Paulo: Makron Books, 1999.
- [NOG2000] NOGUEZ, J. H. **Técnicas de Mineração de Dados no Processo de Descoberta de Conhecimento**. 2000. 47f. Trabalho Individual (Mestrado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre.
- [PAR89] PARSAYE, K. **Intelligent Data Bases: object-oriented, deductive and hypermedia technologies**. New York: John Wiley, 1989.
- [SAN2000] SANTOS, J.; HENRIQUES, N.A.C.; REIS, V. **Data Mining / Data Warehousing - FCT**. Disponível em: <<http://students.fct.unl.pt/users/nach/dmdw/prologo>>. Acesso em: 10 jul. 2000.
- [SAV95] SAVASERE, A. et al. An Efficient Algorithm for Mining Association Rules in Large Databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, VLDB, 21., 1995, Zurich. **Proceedings...**São Francisco: Morgan Kaufmann, 1995. p.432-444.
- [SCH2001] SCHNEIDER, L. F. **Mineração de Dados: conceitos, tarefas e técnicas- aplicações na área jurídica**. 2001. 42f. Trabalho Individual (Mestrado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre.
- [TOR2001] TORGO, L. **Agrupamento – Clustering: extração de conhecimento de dados**. Disponível em: <<http://www.liacc.up.pt/~ltorgo/Ensino/madsad/ecd1>>. Acesso em: 20 abr. 2003.
- [WIV99] WIVES, L.K. Um estudo sobre agrupamentos de documentos textuais em processamento de informações não estruturadas usando técnicas de Clustering. 1999. 102f. Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre.

ANEXO A DESCRIÇÃO DOS DADOS

Entidade: Processo

Tabela A.1: Entidade Processo

Entidade: Processo		
Atributo	Descrição	Tipo de Dados
Cod_Comarca	Código da Comarca	Numérico
Número_Processo	Número do Processo	Numérico
Cod_Situacao	Situação do processo	Caractere
Cod_Especializacao	Código da Especialização	Numérico
Cod_Classe	Código da Classe	Numérico
Cod_Natureza	Código da Natureza	Numérico
Data_Propositura	Data do ingresso da Ação (Processo)	Data
Data_Baixa	Data no qual o processo foi encerrado	Data

Entidade: Comarca

Tabela A.2: Entidade Comarca

Entidade: Comarca		
Atributo	Descrição	Tipo de Dados
Cod_Comarca	Código da Comarca	Numérico
Nome_Comarca	Nome da Comarca	Caractere
Entrancia	Se a comarca é de Entrância: (I = Inicial; M = Intermediária; e F = Final)	Caractere

Entidade: Classe

Tabela A.3: Entidade Classe

Entidade: Classe		
Atributo	Descrição	Tipo de Dados
Cod_Classe	Código da Classe	Numérico
Nome_Classe	Nome da Classe	Caractere
Cod_Secao	Código da Seção	Numérico

Entidade: Situacao_Processo

Tabela A.4: Entidade Situacao_Processo

Entidade: Situacao_Processo		
Atributo	Descrição	Tipo de Dados
Cod_Situacao	Código da Situação do Processo	Caractere
Nome_Situacao	Situação do Processo	Caractere

Entidade: Seção

Tabela A.5: Entidade Seção

Entidade: Secao		
Atributo	Descrição	Tipo de Dados
Cod_Secao	Código da Seção	Numérico
Nome_Secao	Nome da Seção	Caractere

Entidade: Tipo_Especializacao

Tabela A.6: Entidade Tipo_Especializacao

Entidade: Tipo_Especializacao		
Atributo	Descrição	Tipo de Dados
Cod_Especializacao	Código da Especialização	Numérico
Nome_Especializacao	Nome da Especialização	Caractere

Entidade: Sentença_Processo

Tabela A.7: Entidade Processo_Sentenca

Entidade: Sentença_Processo		
Atributo	Descrição	Tipo de Dados
Cod_Comarca	Código da Comarca	Numérico
Número_Processo	Número do Processo	Numérico
Num_Sentença	Número Sequencial da Sentença no Processo	Numérico
Cod_Movimento	Código do Movimento de Sentença	Numérico
Data_Sentença	Data da Sentença	Data

Entidade: Tipo_Movimento

Tabela A.8: Entidade Tipo_Movimento

Entidade: Tipo_Movimento		
Atributo	Descrição	Tipo de Dados
Cod_Movimento	Código do Movimento	Numérico
Nome_Movimento	Nome do Movimento	Caractere

Indica_Sentença	Indica se é movimento de sentença. 'S' se sim	Caractere
-----------------	---	-----------

Entidade: Audiencia_Processo

Tabela A.9: Entidade Audiencia_Processo

Entidade: Audiencia_Processo		
Atributo	Descrição	Tipo de Dados
Cod_Comarca	Código da Comarca	Numérico
Número_Processo	Número do Processo	Numérico
Num_Audiência	Número seqüencial da audiência no processo	Numérico
Data_Audiência	Data da Audiência	Numérico

Entidade: Natureza

Tabela A.10: Entidade Natureza

Entidade: Natureza		
Atributo	Descrição	Tipo de Dados
Cod_Classe	Código da Natureza	Numérico
Nome_Natureza	Nome da Natureza	Caractere

Entidade: Parte

Tabela A.11: Entidade Parte

Entidade: Parte		
Atributo	Descrição	Tipo de Dados
Cod_Comarca	Código da Comarca	Numérico
Número_Processo	Número do Processo	Numérico
Id_Parte	Identificador de parte no processo	Numérico
Tipo_Parte	Se a parte é Física (F) ou Jurídica (J)	Numérico
Indica_Reu	Indica se a parte é ré no processo	Caractere

Entidade: Parte_Fisica

Tabela A.12: Entidade Parte_Fisica

Entidade: Parte_Fisica		
Atributo	Descrição	Tipo de Dados
Cod_Comarca	Código da Comarca	Numérico
Número_Processo	Número do Processo	Numérico
Id_Parte	Identificador de parte no processo	Numérico
CPF	Cadastro de Pessoa Física	Numérico
Sexo	Se a parte é do sexo masculino (M) ou feminino (F)	Caractere
Cod_Estado_Civil	Código do Estado Civil	Numérico
Cod_Grau_Instrucao	Código do Grau de Instrução	Numérico

Cod_Profissao	Código da Profissão	Numérico
Cod_Cor	Código da Cor da parte	Numérico
Data_Nascimento	Data de nascimento do réu	Data

Entidade: Estado_Civil

Tabela A.13: Entidade Estado_Civil

Entidade: Estado_Civil		
Atributo	Descrição	Tipo de Dados
Cod_Estado_Civil	Código do Estado Civil	Numérico
Descricao	Descrição do Estado Civil	Caractere

Entidade: Grau_Instrucao

Tabela A.14: Entidade Grau_Instrucao

Entidade: Grau_Instrucao		
Atributo	Descrição	Tipo de Dados
Cod_Grau_Instrucao	Código do Grau de Instrução	Numérico
Descricao	Descrição do Grau de Instrução	Caractere

Entidade: Profissão

Tabela A.15: Entidade Profissão

Entidade: Profissao		
Atributo	Descrição	Tipo de Dados
Cod_Profissao	Código da Profissão	Numérico
Descricao	Descrição da Profissao	Caractere

Entidade: Cor

Tabela A.16: Entidade Cor

Entidade: Cor		
Atributo	Descrição	Tipo de Dados
Cod_Cor	Código da Cor	Numérico
Descricao	Descrição da Cor	Caractere

ANEXO B VALORES DOS ATRIBUTOS

Comarca

Tabela B.17: Identificação das Comarcas

Código	Comarca
50	Comarca de Getúlio Vargas
82	Comarca de Arvorezinha
85	Comarca de Cacequi
89	Comarca de Candelária
96	Comarca de Faxinal do Soturno
101	Comarca de Gramado
104	Comarca de Horizontina
108	Comarca de Lavras do Sul
112	Comarca de Não-Me-Toque
126	Comarca de São José do Norte
139	Comarca de Triunfo
149	Comarca de Augusto Pestana
151	Comarca de Palmares do Sul
158	Comarca de Rodeio Bonito
Quantidade de Comarcas	14

Seção

Tabela B.18: Identificação das Seções

Código	Seção
1	Cível
2	Crime
3	Juizado Especial Cível
4	Juizado Especial Crime
Quantidade de Seções	4

Tipo Especialização

Tabela B.19: Identificação das Especializações

Código	Tipo Especialização
AC	Vara de Trânsito

AJ	Arquivo Judicial
AT	Acidente de Trabalho
CI	Crime de Imprensa
CR	Crime
CS	Crime Sumário
CV	Cível
DC	Direção do Foro – Crime
DF	Direção do Foro – Administrativo
DV	Direção do Foro – Cível
FA	Família
FC	Falência e Concordatas
FE	Fazenda Pública Estadual
FM	Fazenda Pública Municipal
FP	Fazenda Pública
JC	Juizado Especial Crime
JE	Execução do Juizado Especial Cível
JP	Juizado Especial Cível
JT	Juizado Especial Crime – Trânsito
JU	Júri
PC	Precatória Crime
PF	Precatória Família
PJ	Precatória Juizado Especial Cível
PL	Cautelares Plantão Judicial
PP	Posto do Juizado Especial Cível
PV	Precatória Cível
RP	Registros Públicos
VE	Varas de Execuções Criminais
VS	Vara de Sucessões
Quantidade de Especializações	
29	

Classe

Tabela B.20: Identificação de Classes

Código	Classe
31	Processo de Conhecimento
32	Processo Cautelar
33	Alimentos, Separação e Divórcios Litigiosos
34	Procedimento Especial de Jurisdição Contenciosa
35	Embargos do Devedor
36	Processos de Execução
37	Inventários e Arrolamentos
38	Falências e Concordatas
39	Separação e Divórcio Consensuais
40	Procedimento Especial de Jurisdição Voluntária
41	Incidentes Processuais
42	Precatórias Cíveis
43	Juizado Especial Cível (JEC)
44	Processo de Execução de Alimentos

45	Processo de Execução Fiscal	
46	Precatórias - Juizado Especial Cível (JEC)	
101	Procedimento do Júri	
102	Procedimento Ordinário - Reclusão	
103	Procedimento Sumário - Detenção	
104	Ações Constitucionais	
105	Especiais	
106	Secundários	
107	Precatórias Criminais	
108	Juizado Especial Crime (JECR)	
109	Acautelatórias	
Quantidade de Classes		25

Tipo Movimento

Tabela B.21: Identificação do Tipo de Movimento

Código	Tipo Movimento	
4	Sentença Homologatória	
8	Sentença de Mérito	
11	Acordos obtidos em audiência	
12	Demais sentenças e decisões	
14	Sentença Homologatória LF 9099/95	
15	Absolutórias	
16	Impronúncia e absolvição sumária	
17	Demais decisões terminativas	
18	Condenatórias - com pena	
19	Extinção da punibilidade	
20	Sentença de Pronúncia	
21	Desclassificação do Tribunal do Júri	
27	Condenatória - sem pena	
Quantidade de Tipos de Sentenças		13

Situação Processo

Tabela B.22: Identificação de Situações do Processo

Código	Situação Processo	
A	Ativo	
B	Baixado	
C	Em Cadastramento	
E	Excluído	
I	Inquérito	
R	Arquivado	
X	Cancelado	
Quantidade de Situações		7

Naturezas da Especialização Criminal

Tabela B.23: Identificação das Naturezas

Código	Naturezas da Especialização Criminal
105	Interpelação
146	Habeas Corpus
153	Mandado de Segurança
189	Outros
240	Lesões Corporais Graves
241	Lesões Corporais seguidas de Morte
242	Periclitacão da Vida e da Saúde
243	Crimes contra a Liberdade Pessoal
244	Crimes de Furto
245	Crimes de Roubo e Extorsão
246	Crimes de Apropriação Indébita
247	Estelionato e Fraudes
248	Receptação Dolosa e Especial
249	Crimes contra a Organização do Trabalho
250	Crimes contra o Respeito aos Mortos
251	Crimes contra a Liberdade Sexual
252	Sedução e Corrupção de Menores
253	Rapto Violento ou Mediante Fraude
254	Lenocínio e Tráfico de Mulheres
255	Crime contra a Família
256	Crimes contra a Incolumidade Pública
257	Crimes contra a Paz Pública
258	Crimes contra a Fé Pública
259	Crimes praticados por Particular contra a Administração
260	Crimes contra a Administração em Geral
261	Crimes contra a Administração da Justiça
262	Crime de Corrupção de Menores (Lei 2252/54)
263	Crime de Parcelamento do Solo Urbano (Lei 6766/79)
265	Preconceito de Raça ou Cor (Lei 6766/79)
266	Crimes contra Crianças e Adolescentes (Lei 8069/90)
267	Crimes contra a Ordem Pública (Lei 8137/90)
268	Ordinário
269	Porte de Arma
275	Crime contra Inviolabilidade de Correspondência
297	Crime de Responsabilidade dos Funcionários Públicos
298	Crimes contra a Honra
299	Crimes contra a Propriedade Imaterial - DL 7903/45 - Lei 7646/87
300	Abuso de Autoridade - Lei 4898/65
301	Tráfico de Drogas – Lei 6368/76
303	Contravenções Florestais - Lei 4771/65
304	Jogo do Bicho - Lei 1508/51
305	Crimes Falimentares
306	Crimes contra a Economia Popular
308	Exceções
310	Restituição de Coisas Apreendidas
311	Seqüestro de Bens - DL 3240/41, Lei 3164/67 - 3502/58.
314	Incidente de Insanidade Mental
316	Reabilitação
326	Precatória de Acompanhamento de Suspensão Condicional
327	Precatória de Acomp./Fiscalização/Proposta Susp. Cond. do Processo
329	Explicação em Juízo
330	Pedido de Explicações

Código	Naturezas da Especialização Criminal
337	Estupro
Quantidade de Naturezas criminais	54

Profissão

Tabela B.24: Identificação das Profissões

Código	Profissão
1	Administrador
2	Advogado
3	Agenciador de propaganda
6	Agente de viagem e guia de turismo
7	Agrônomo
10	Aposentado
12	Arquiteto
13	Arquivista
20	Autônomo
21	Auxiliar de escritório e assemelhados
22	Auxiliar de laboratório
23	Avaliador
24	Bailarino
25	Bancário
26	Barbeiro
30	Bolsista
32	Cabeleireiro
34	Camareiro
41	Contador
44	Corretor de imóveis, seguros, títulos e valores
45	Cozinheiro
47	Decorador
48	Delegado de polícia
51	Desenhista
52	Despachante
56	Dona de casa
58	Economista
59	Eletricista e assemelhados
61	Empregado doméstico
62	Empresário
63	Enfermeiro
64	Engenheiro
66	Estagiário
70	Estivador
71	Estudante
72	Farmacêutico
73	Faxineiro
79	Funcionário público
81	Garçom

82	Garimpeiro
87	Industriário
88	Instalador de gás, água, esgoto
89	Jardineiro
92	Jornalista
97	Manicura
100	Marceneiro
101	Marinheiro
102	Massagista
104	Mecânico
105	Médico
107	Mestre e contramestre
108	Metalúrgico
110	Militar
115	Motorista
117	Músico
119	Odontólogo
120	Oficiais das forças armadas e forças auxiliares
121	Operador
122	Ourives
123	Outros
124	Pensionista
125	Piloto
126	Pintor
127	Porteiro
128	Prefeito
131	Professor
135	Publicitário
136	Químico
137	Radialista
138	Recepcionista
140	Representante
141	Sacerdote
142	Secretário
143	Securitário
145	Serventuário de justiça
146	Servidor público estadual
147	Servidor público federal
148	Servidor público municipal
149	Siderúrgico
153	Técnico de contabilidade e de estatística
155	Técnico de eletrônica
157	Técnico de mecânica
158	Técnico de química
159	Técnico de telecomunicações
160	Técnico em agronomia e agrimensura
165	Trabalhador agrícola
166	Trabalhador da pecuária

167	Trabalhador da pesca	
168	Trabalhador de artes gráficas	
169	Trabalhador de construção civil	
170	Vendedor de comércio varejista e atacadista	
171	Vendedor praticista	
172	Vereador	
173	Veterinário	
174	Zelador	
176	Comerciante	
177	Comerciário	
178	Servidor Público	
179	Agricultor	
Quantidade de Profissões		99

Estado Civil

Tabela B.25: Identificação dos Estados Civis

Código	Estado Civil	
1	Solteiro	
2	Casado	
3	Viúvo	
4	Separado	
5	Desquitado	
6	Amigado	
8	Divorciado	
Quantidade de Estados Civis		7

Grau de Instrução

Tabela B.26: Identificação dos Graus de Instrução

Código	Grau de Instrução	
1	Analfabeto	
2	Primeiro Grau	
3	Segundo Grau	
4	Superior Incompleto	
5	Não observado	
6	Superior Completo	
Quantidade de Graus de Instrução		6

Raça

Tabela B.27: Identificação das raças

Código	Raça
1	Branca
2	Negra
3	Mulata

4	Amarela
5	Indiatica
Quantidade	
	5

ANEXO C ESTATÍSTICAS PRINCIPAIS

Estatísticas gerais de processos

Tabela C.28: Estatísticas de Processos

Descrição	Quantidade (unidades)
Processos	77471
Processos Ativos (Ativos + Em Cadastramento + Inquérito)	50545
Processos Encerrados (Baixados + Arquivados + Excluídos + Cancelados)	27196
Processos com Sentença Proferida	29290
Processos com Audiência	16180

Estatísticas de processos distribuídos por seção

Tabela C.29: Distribuição por Seção

Código	Seção	Qtde	%
1	Cível	48023	61,99%
2	Crime	13865	17,90%
3	Juizado Especial Cível	6685	8,63%
4	Juizado Especial Crime	8898	11,49%
	Total	77471	100,00%

Estatísticas de processos distribuídos por especialização

Tabela C.30: Distribuição por especialização

Código	Especialização	Qtde	%
AC	Vara de Trânsito	210	0,27%
AT	Acidente de Trabalho	66	0,09%
CI	Crime de Imprensa	7	0,01%
CR	Crime	8107	10,46%
CS	Crime Sumário	2225	2,87%
CV	Cível	14730	19,01%
DC	Direção do Foro - Crime	126	0,16%

DF	Direção do Foro - Administrativo	3	0,00%
DV	Direção do Foro - Cível	236	0,30%
FA	Família	6993	9,03%
FC	Falência e Concordatas	494	0,64%
FE	Fazenda Pública Estadual	1535	1,98%
FM	Fazenda Pública Municipal	11258	14,53%
FP	Fazenda Pública	5442	7,02%
JC	Juizado Especial Crime	8851	11,42%
JE	Execução do Juizado Especial Cível	2935	3,79%
JP	Juizado Especial Cível	3312	4,28%
JT	Juizado Especial Crime - Trânsito	454	0,59%
JU	Júri	668	0,86%
PC	Precatória Crime	2115	2,73%
PF	Precatória Família	1354	1,75%
PJ	Precatória Juizado Especial Cível	534	0,69%
PV	Precatória Cível	3619	4,67%
RP	Registros Públicos	487	0,63%
VS	Vara de Sucessões	1710	2,21%
	Total	77471	100,00%

Estadísticas de processos distribuídos por classe

Tabela C.31: Distribuição por Classe

Classe	Qtde	%
Acautelatórias	6	0,01%
Ações Constitucionais	6	0,01%
Alimentos, Separação e Divórcios Litigiosos	1871	2,42%
Embargos do Devedor	1755	2,27%
Especiais	848	1,09%
Falências e Concordatas	199	0,26%
Incidentes Processuais	785	1,01%
Inventários e Arrolamentos	1609	2,08%
Juizado Especial Cível (JEC)	6187	7,99%
Juizado Especial Crime (JECR)	8897	11,48%
Precatórias - Juizado Especial Cível (JEC)	498	0,64%
Precatórias Cíveis	5495	7,09%
Precatórias Criminais	2967	3,83%
Procedimento do Júri	558	0,72%
Procedimento Especial de Jurisdição Contenciosa	2956	3,82%
Procedimento Especial de Jurisdição Voluntária	1873	2,42%
Procedimento Ordinário – Reclusão	6960	8,98%
Procedimento Sumário – Detenção	2387	3,08%
Processo Cautelar	1456	1,88%
Processo de Conhecimento	6670	8,61%
Processo de Execução de Alimentos	1836	2,37%
Processo de Execução Fiscal	15082	19,47%
Processos de Execução	5731	7,40%
Secundários	134	0,17%

Separação e Divórcio Consensuais	705	0,91%
Total	77471	100,00%

Estatísticas de processos distribuídos por comarca

Tabela C.32: Distribuição por Comarca

Comarca	Qtde	%
Comarca de Arvorezinha	2780	3,59%
Comarca de Augusto Pestana	3152	4,07%
Comarca de Cacequi	4086	5,27%
Comarca de Candelária	5589	7,21%
Comarca de Faxinal do Soturno	4629	5,98%
Comarca de Getúlio Vargas	8642	11,16%
Comarca de Gramado	13155	16,98%
Comarca de Horizontina	7261	9,37%
Comarca de Lavras do Sul	1765	2,28%
Comarca de Não-Me-Toque	4448	5,74%
Comarca de Palmares do Sul	8088	10,44%
Comarca de Rodeio Bonito	4388	5,66%
Comarca de São José do Norte	3290	4,25%
Comarca de Triunfo	6198	8,00%
Total	77471	100,00%

Estatísticas de processos distribuídos por situação

Tabela C.33: Distribuição por Situação

Código	Situação Processo	Qtde	%
A	Ativo	47946	61,89%
B	Baixado	7772	10,03%
C	Em Cadastramento	420	0,54%
E	Excluído	106	0,14%
I	Inquérito	2179	2,81%
R	Arquivado	18699	24,14%
X	Cancelado	349	0,45%
	total	77471	100,00%

Estatísticas de processos distribuídos por comarca x seção

Tabela C.34: Distribuição por Comarca x Seção

Comarca	Seção	Qtde	%
Comarca de Getúlio Vargas	Cível	5037	58,29%
Comarca de Getúlio Vargas	Crime	2146	24,83%
Comarca de Getúlio Vargas	Juizado Especial Cível	410	4,74%
Comarca de Getúlio Vargas	Juizado Especial Crime	1049	12,14%
	Sub Total	8642	100,00%
Comarca de Arvorezinha	Cível	1393	50,11%
Comarca de Arvorezinha	Crime	625	22,48%

Comarca de Arvorezinha	Juizado Especial Cível	392	14,10%
Comarca de Arvorezinha	Juizado Especial Crime	370	13,31%
Sub Total		2780	100,00%
Comarca de Cacequi	Cível	2434	59,57%
Comarca de Cacequi	Crime	818	20,02%
Comarca de Cacequi	Juizado Especial Cível	451	11,04%
Comarca de Cacequi	Juizado Especial Crime	383	9,37%
Sub Total		4086	100,00%
Comarca de Candelária	Cível	3678	65,81%
Comarca de Candelária	Crime	1035	18,52%
Comarca de Candelária	Juizado Especial Cível	388	6,94%
Comarca de Candelária	Juizado Especial Crime	488	8,73%
Sub Total		5589	100,00%
Comarca de Faxinal do Soturno	Cível	2130	46,01%
Comarca de Faxinal do Soturno	Crime	1155	24,95%
Comarca de Faxinal do Soturno	Juizado Especial Cível	592	12,79%
Comarca de Faxinal do Soturno	Juizado Especial Crime	752	16,25%
Sub Total		4629	100,00%
Comarca de Gramado	Cível	9377	71,28%
Comarca de Gramado	Crime	1267	9,63%
Comarca de Gramado	Juizado Especial Cível	1404	10,67%
Comarca de Gramado	Juizado Especial Crime	1107	8,42%
Sub Total		13155	100,00%
Comarca de Horizontina	Cível	4989	68,71%
Comarca de Horizontina	Crime	1120	15,42%
Comarca de Horizontina	Juizado Especial Cível	647	8,91%
Comarca de Horizontina	Juizado Especial Crime	505	6,95%
Sub Total		7261	100,00%
Comarca de Lavras do Sul	Cível	1267	71,78%
Comarca de Lavras do Sul	Crime	190	10,76%
Comarca de Lavras do Sul	Juizado Especial Cível	121	6,86%
Comarca de Lavras do Sul	Juizado Especial Crime	187	10,59%
Sub Total		1765	100,00%
Comarca de Não-Me-Toque	Cível	2653	59,64%
Comarca de Não-Me-Toque	Crime	1171	26,33%
Comarca de Não-Me-Toque	Juizado Especial Cível	165	3,71%
Comarca de Não-Me-Toque	Juizado Especial Crime	459	10,32%
Sub Total		4448	100,00%
Comarca de São José do Norte	Cível	1976	60,06%
Comarca de São José do Norte	Crime	577	17,54%
Comarca de São José do Norte	Juizado Especial Cível	117	3,56%
Comarca de São José do Norte	Juizado Especial Crime	620	18,84%
Sub Total		3290	100,00%
Comarca de Triunfo	Cível	3777	60,94%
Comarca de Triunfo	Crime	1123	18,12%
Comarca de Triunfo	Juizado Especial Cível	308	4,97%
Comarca de Triunfo	Juizado Especial Crime	990	15,97%
Sub Total		6198	100,00%

Comarca de Augusto Pestana	Cível	2239	71,03%
Comarca de Augusto Pestana	Crime	399	12,66%
Comarca de Augusto Pestana	Juizado Especial Cível	193	6,12%
Comarca de Augusto Pestana	Juizado Especial Crime	321	10,18%
Sub Total		3152	100,00%
Comarca de Palmares do Sul	Cível	5065	62,62%
Comarca de Palmares do Sul	Crime	1181	14,60%
Comarca de Palmares do Sul	Juizado Especial Cível	1065	13,17%
Comarca de Palmares do Sul	Juizado Especial Crime	777	9,61%
Sub Total		8088	100,00%
Comarca de Rodeio Bonito	Cível	2008	45,76%
Comarca de Rodeio Bonito	Crime	1058	24,11%
Comarca de Rodeio Bonito	Juizado Especial Cível	432	9,85%
Comarca de Rodeio Bonito	Juizado Especial Crime	890	20,28%
Sub Total		4388	100,00%
Total		77471	

Estatísticas de Tempos Médios por Classe x Especialização Justiça Comum

Tabela C.35: Tempos Médios – Justiça Comum

Classe	Especialização	Tempos Médios	Qtde. de Processos	%
31	AT	875,67	3	0%
31	CV	871,54	570	3%
31	FA	556,36	346	2%
31	FM	508,00	1	0%
31	FP	547,91	334	2%
31	RP	3506,00	1	0%
31	VS	2147,50	2	0%
32	CV	410,24	319	2%
32	FA	285,91	160	1%
32	RP	238,00	1	0%
33	FA	351,80	883	5%
34	AT	148,00	1	0%
34	CV	596,68	589	3%
34	FA	423,29	28	0%
34	FC	1076,14	7	0%
34	FM	422,67	3	0%
34	FP	1114,15	13	0%
34	RP	296,66	107	1%
35	CV	351,27	196	1%
35	FA	305,71	7	0%
35	FC	248,00	1	0%
35	FE	571,88	8	0%
35	FM	370,38	16	0%
35	FP	380,83	24	0%
35	JE	69,89	9	0%

35	PV	75,33	9	0%
36	CV	1292,38	691	4%
36	FA	207,13	8	0%
36	FE	322,00	1	0%
36	FP	924,05	22	0%
37	VS	637,55	491	3%
38	FC	911,19	27	0%
39	FA	355,01	363	2%
40	AT	38,00	1	0%
40	CV	235,24	433	2%
40	DV	372,50	2	0%
40	FA	307,07	296	2%
40	FP	83,00	1	0%
40	RP	219,38	128	1%
40	VS	159,20	15	0%
41	CV	335,42	77	0%
41	FA	168,80	10	0%
41	FC	262,53	34	0%
41	FP	186,67	3	0%
41	PV	76,00	1	0%
41	VS	138,33	9	0%
42	CV	432,40	20	0%
42	DF	8,00	3	0%
42	DV	22,48	192	1%
42	FE	208,30	115	1%
42	FM	173,00	53	0%
42	PF	45,02	1178	6%
42	PJ	195,32	19	0%
42	PV	154,00	2415	13%
44	FA	343,36	744	4%
45	FE	1222,14	28	0%
45	FM	681,49	549	3%
45	FP	1042,32	114	1%
101	JU	610,34	174	1%
102	CR	260,55	3185	17%
103	AC	383,55	104	1%
103	CR	68,00	3	0%
103	CS	509,45	932	5%
104	CR	234,00	1	0%
105	CI	179,00	5	0%
105	CR	440,26	384	2%
106	CR	293,80	15	0%
106	CS	403,50	6	0%
106	JC	188,08	12	0%
106	JU	251,00	5	0%
107	AC	64,50	14	0%
107	CR	493,25	130	1%
107	CS	369,17	6	0%

107	DC	19,75	118	1%
107	JC	63,46	228	1%
107	JT	77,38	45	0%
107	JU	41,85	85	0%
107	PC	57,93	1773	9%
109	CR	249,00	2	0%
Total			18908	100%

Estatísticas de Tempos Médios por Classe x Especialização Juizados Especiais

Tabela C.36: Tempos Médios – Juizados Especiais

Classe	Especialização	Tempos Médios	Qtde. de Processos	%
43	JE	224,03	789	10%
43	JP	111,13	1439	19%
46	PJ	163,74	272	4%
108	JC	200,23	4883	65%
108	JT	209,17	180	2%
Total			7563	100%

Estatísticas de réus distribuídos por sexo

Tabela C.37: Distribuição por sexo

Código	Sexo	Qtde.	%
F	Feminino	867	10%
M	Masculino	7811	90%
Total		8678	100%

Estatísticas de réus distribuídos por estado civil

Tabela C.38: Distribuição por Estado Civil

Código	Estado Civil	Qtde.	%
?	Não Informado	2254	26%
1	Solteiro	3287	38%
2	Casado	1901	22%
3	Viúvo	53	1%
4	Separado	364	4%
5	Desquitado	36	0%
6	Amigado	709	8%
8	Divorciado	74	1%
Total		8678	100%

Estatísticas de réus distribuídos por grau de instrução

Tabela C.39: Distribuição por Graus de Instrução

Código	Grau de Instrução	Qtde.	%
--------	-------------------	-------	---

?	Não Informado	4921	57%
1	Analfabeto	168	2%
2	Primeiro Grau	2504	29%
3	Segundo Grau	455	5%
4	Superior Incompleto	63	1%
5	Não observado	447	5%
6	Superior Completo	120	1%
Total		8678	100%

Estadísticas de réus distribuídos por profissão

Tabela C.40: Distribuição por Profissão

Código	Profissão	Qtde.	%
1	Administrador	5	0%
3	Agenciador de propaganda	1	0%
6	Agente de viagem e guia de turismo	3	0%
7	Agrônomo	4	0%
12	Arquiteto	3	0%
13	Arquivista	1	0%
21	Auxiliar de escritório e assemelhados	16	0%
22	Auxiliar de laboratório	1	0%
23	Avaliador	1	0%
24	Bailarino	1	0%
25	Bancário	21	0%
26	Barbeiro	1	0%
30	Bolsista	12	0%
32	Cabeleireiro	7	0%
34	Camareiro	2	0%
41	Contador	18	0%
44	Corretor de imóveis, seguros, títulos e valores	12	0%
45	Cozinheiro	7	0%
47	Decorador	1	0%
48	Delegado de polícia	1	0%
51	Desenhista	1	0%
52	Despachante	16	0%
58	Economista	1	0%
61	Empregado doméstico	36	0%
62	Empresário	38	0%
63	Enfermeiro	1	0%
64	Engenheiro	28	0%
66	Estagiário	7	0%
70	Estivador	2	0%

72	Farmacêutico	1	0%
73	Faxineiro	3	0%
88	Instalador de gás, água, esgoto	1	0%
89	Jardineiro	16	0%
92	Jornalista	3	0%
97	Manicura	3	0%
102	Massagista	2	0%
105	Médico	10	0%
107	Mestre e contramestre	1	0%
108	Metalúrgico	18	0%
110	Militar	41	0%
117	Músico	6	0%
119	Odontólogo	1	0%
120	Oficiais das forças armadas e forças auxiliares	1	0%
121	Operador	24	0%
122	Ourives	1	0%
124	Pensionista	8	0%
125	Piloto	2	0%
127	Porteiro	1	0%
128	Prefeito	3	0%
131	Professor	26	0%
135	Publicitário	5	0%
136	Químico	1	0%
137	Radialista	4	0%
138	Recepcionista	1	0%
140	Representante	14	0%
141	Sacerdote	2	0%
142	Secretário	7	0%
143	Securitário	2	0%
145	Serventuário de justiça	2	0%
146	Servidor público estadual	7	0%
147	Servidor público federal	1	0%
148	Servidor público municipal	9	0%
149	Siderúrgico	5	0%
153	Técnico de contabilidade e de estatística	9	0%
155	Técnico de eletrônica	5	0%
157	Técnico de mecânica	1	0%
158	Técnico de química	1	0%
159	Técnico de telecomunicações	1	0%
160	Técnico em agronomia e agrimensura	1	0%
166	Trabalhador da pecuária	7	0%

167	Trabalhador da pesca	23	0%
168	Trabalhador de artes gráficas	1	0%
171	Vendedor praticista	13	0%
172	Vereador	15	0%
173	Veterinário	2	0%
174	Zelador	1	0%
177	Comerciário	35	0%
178	Servidor Público	8	0%
2	Advogado	45	1%
10	Aposentado	80	1%
20	Autônomo	73	1%
59	Eletricista e assemelhados	44	1%
71	Estudante	97	1%
79	Funcionário público	127	1%
81	Garçom	46	1%
82	Garimpeiro	54	1%
87	Industriário	49	1%
100	Marceneiro	52	1%
101	Marinheiro	57	1%
104	Mecânico	121	1%
126	Pintor	121	1%
165	Trabalhador agrícola	51	1%
169	Trabalhador de construção civil	111	1%
170	Vendedor de comércio varejista e atacadista	102	1%
179	Agricultor	831	10%
123	Outros	1673	19%
56	Dona de casa	197	2%
115	Motorista	168	2%
176	Comerciante	407	5%
?	Não Informado	3569	41%
Total		8678	1

Estadísticas de réus distribuídos por cor

Tabela C.41: Distribuição por Cor

Código	Cor	Qtde	%
?	Não Informado	4264	49%
1	Branco	3599	41%
2	Negro	349	4%
3	Mista	457	5%
4	Amarela	6	0%
5	Indiático	3	0%

	Total	8678	100%
--	-------	------	------

Estatísticas de réus por comarca

Tabela C.42: Distribuição de Réus por Comarca

Código	Comarca	Qtde.	%
50	Comarca de Getúlio Vargas	1148	13%
82	Comarca de Arvorezinha	301	3%
85	Comarca de Cacequi	596	7%
89	Comarca de Candelária	645	7%
96	Comarca de Faxinal do Soturno	665	8%
101	Comarca de Gramado	676	8%
104	Comarca de Horizontina	1360	16%
108	Comarca de Lavras do Sul	120	1%
112	Comarca de Não-Me-Toque	527	6%
126	Comarca de São José do Norte	396	5%
139	Comarca de Triunfo	863	10%
149	Comarca de Augusto Pestana	167	2%
151	Comarca de Palmares do Sul	864	10%
158	Comarca de Rodeio Bonito	350	4%
	total	8678	100%

Estatísticas de réus distribuídos por natureza criminal

Tabela C.43: Distribuição de Réus por Natureza

Código	Natureza	Qtde.	%
105	Interpelação	4	0%
146	Habeas Corpus	4	0%
240	Lesões Corporais Graves	227	3%
241	Lesões Corporais seguidas de Morte	10	0%
242	Periclitação da Vida e da Saúde	51	1%
243	Crimes contra a Liberdade Pessoal	27	0%
244	Crimes de Furto	3673	42%
245	Crimes de Roubo e Extorsão	641	7%
246	Crimes de Apropriação Indébita	217	3%
247	Estelionato e Fraudes	995	11%
248	Receptação Dolosa e Especial	317	4%
249	Crimes contra a Organização do Trabalho	1	0%
250	Crimes contra o Respeito aos Mortos	10	0%
251	Crimes contra a Liberdade Sexual	228	3%
252	Sedução e Corrupção de Menores	21	0%
253	Rapto Violento ou Mediante Fraude	2	0%
254	Lenocínio e Tráfico de Mulheres	79	1%
255	Crime contra a Família	38	0%
256	Crimes contra a Incolumidade Pública	52	1%
257	Crimes contra a Paz Pública	19	0%
258	Crimes contra a Fé Pública	271	3%

259	Crimes praticados por Particular contra a Administração	35	0%
260	Crimes contra a Administração em Geral	100	1%
261	Crimes contra a Administração da Justiça	173	2%
262	Crime de Corrupção de Menores (Lei 2252/54)	12	0%
263	Crime de Parcelamento do Solo Urbano (Lei 6766/79)	9	0%
265	Preconceito de Raça ou Cor (Lei 6766/79)	5	0%
266	Crimes contra Crianças e Adolescentes (Lei 8069/90)	24	0%
267	Crimes contra a Ordem Pública (Lei 8137/90)	21	0%
268	Ordinário	101	1%
269	Porte de Arma	167	2%
275	Crime contra Inviolabilidade de Correspondência	4	0%
297	Crime de Responsabilidade dos Funcionários Públicos	48	1%
298	Crimes contra a Honra	418	5%
299	Crimes contra a Propriedade Imaterial - DL 7903/45 - Lei 7646/87	3	0%
300	Abuso de Autoridade - Lei 4898/65	89	1%
301	Tráfico de Drogas - Lei 6368/76	482	6%
303	Contravenções Florestais - Lei 4771/65	45	1%
304	Jogo do Bicho - Lei 1508/51	4	0%
305	Crimes Falimentares	2	0%
306	Crimes contra a Economia Popular	45	1%
329	Explicação em Juízo	1	0%
330	Pedido de Explicações	1	0%
337	Estupro	2	0%
	Total	8678	100%

ANEXO D RESULTADOS DA MINERAÇÃO - TÉCNICA DE SEGMENTAÇÃO

Objetivo 1a – Classificação Processual

Testes com variação de Clusters

- Algoritmo: K-médias

Teste 1 - Seção x Classe x Especialização

Tabela D.44: Teste 1 - Seção x Classe x Especialização

Filtro: Seção x Classe x Especialização					
Número de Clusters = 4			Seed = 10		
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	2	108	JC	22763	29
Cluster 1	3	43	JP	6687	9
Cluster 2	1	31	CV	32699	42
Cluster 3	1	45	FM	15322	20
Número de Clusters = 8			Seed = 10		
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	4	108	JC	8931	12
Cluster 1	3	43	JP	6687	9
Cluster 2	1	31	FA	21071	27
Cluster 3	1	45	FM	15322	20
Cluster 4	1	36	CV	11628	15
Cluster 5	2	107	PC	5950	8
Cluster 6	2	102	CR	7882	10
Cluster 7	1	109	AC		
Número de Clusters = 12			Seed = 12		
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	2	103	CS	3037	4
Cluster 1	3	43	JP	6687	9
Cluster 2	1	31	FP	18737	24
Cluster 3	1	45	FM	12421	16
Cluster 4	1	36	CV	11215	14
Cluster 5	2	107	PC	2952	4
Cluster 6	2	102	CR	7877	10

Cluster 7	1	109	AC		
Cluster 8	1	109	AC		
Cluster 9	1	33	FA	5648	7
Cluster 10	1	109	AC		
Cluster 11	4	108	JC	8897	11

Teste 2 - Comarca x Seção x Classe x Especialização

Tabela D.45: Teste 2 - Comarca x Seção x Classe x Especialização

Filtro: Comarca x Seção x Classe x Especialização							
Número de Clusters = 4				Seed = 10			
Clusters	Comarca	Seção	Classe	Especialização	Qtde.	%	
Cluster 0	50	2	108	CR	24302	31	
Cluster 1	101	3	43	JP	6632	9	
Cluster 2	139	1	31	CV	29452	38	
Cluster 3	101	1	41	FM	17085	22	
Número de Clusters = 8				Seed = 10			
Clusters	Comarca	Seção	Classe	Especialização	Qtde.	%	
Cluster 0	50	4	108	JC	9424	12	
Cluster 1	101	3	43	JP	6866	9	
Cluster 2	139	1	31	FA	15440	20	
Cluster 3	101	1	45	FM	16914	22	
Cluster 4	104	1	36	CV	10740	14	
Cluster 5	158	2	107	PC	4733	6	
Cluster 6	50	2	102	CR	8389	11	
Cluster 7	151	1	42	PV	4965	6	
Número de Clusters = 12				Seed = 10			
Clusters	Comarca	Seção	Classe	Especialização	Qtde.	%	
Cluster 0	50	2	103	CS	3162	4	
Cluster 1	101	3	43	JP	6630	9	
Cluster 2	139	1	31	FP	14996	19	
Cluster 3	101	1	45	FM	14891	19	
Cluster 4	104	1	36	CV	10698	14	
Cluster 5	101	2	107	PC	2876	4	
Cluster 6	50	2	102	CR	7765	10	
Cluster 7	82	1	44	FA	3278	4	
Cluster 8	85	1	45	FM	1154	1	
Cluster 9	50	1	33	FA	3005	4	
Cluster 10	108	2	102	CR	118	0	
Cluster 11	101	4	108	JC	8898	11	

Testes com variação de Seed

Teste 3 - Seção x Classe x Especialização

Tabela D.46: Teste 3 - Seção x Classe x Especialização

Filtro: Seção x Classe x Especialização					
Número de Clusters = 8			Seed = 1		
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	102	CR	27900	36

Cluster 1	4	108	JC	8931	12
Cluster 2	1	45	FM	15350	20
Cluster 3	1	31	CV	10928	14
Cluster 4	1	36	CV	5729	7
Cluster 5	2	107	CS	5950	8
Cluster 6	1	41	FC	618	1
Cluster 7	1	31	FP	2065	3
Número de Clusters = 8		Seed = 2			
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	4	108	JC	9248	12
Cluster 1	1	45	FA	16139	21
Cluster 2	1	42	PV	5540	7
Cluster 3	1	45	FM	11185	14
Cluster 4	1	36	CV	15159	20
Cluster 5	3	43	JP	3808	5
Cluster 6	2	102	CR	13515	17
Cluster 7	3	43	JE	2877	4
Número de Clusters = 8		Seed = 5			
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	4	108	JC	8931	12
Cluster 1	3	43	JP	6687	9
Cluster 2	1	42	FM	32773	42
Cluster 3	2	107	CS	5950	8
Cluster 4	1	36	CV	15248	20
Cluster 5	1	109	AC		
Cluster 6	1	109	AC		
Cluster 7	2	102	CR	7882	10
Número de Clusters = 8		Seed = 25			
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	108	JC	26505	34
Cluster 1	1	42	PV	5496	7
Cluster 2	1	31	FA	6952	9
Cluster 3	1	44	FA	1836	2
Cluster 4	1	45	FP	4975	6
Cluster 5	1	45	FM	11157	14
Cluster 6	2	102	CR	13865	18
Cluster 7	3	43	JP	6685	9
Número de Clusters = 8		Seed = 50			
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	45	FM	30802	40
Cluster 1	2	102	CR	13865	18
Cluster 2	1	33	FA	6993	9
Cluster 3	4	108	JC	8898	11
Cluster 4	1	109	AC		
Cluster 5	1	37	VS	1710	2
Cluster 6	1	36	CV	15203	20
Cluster 7	1	109	AC		
Número de Clusters = 8		Seed = 100			
Clusters	Seção	Classe	Especialização	Qtde.	%
Cluster 0	1	45	FM	25827	33

Cluster 1	2	102	CR	11449	15
Cluster 2	4	108	JC	8898	11
Cluster 3	2	103	CS	2416	3
Cluster 4	1	36	CV	15248	20
Cluster 5	1	109	AC		
Cluster 6	3	43	JP	6685	9
Cluster 7	1	33	FA	6948	9

Teste 4 - Comarca x Seção x Classe x Especialização

Tabela D.47: Teste 4 - Comarca x Seção x Classe x Especialização

Filtro: Comarca x Seção x Classe x Especialização						
Número de Clusters = 8				Seed = 25		
Clusters	Comarca	Seção	Classe	Especialização	Qtde.	%
Cluster 0	139	1	42	CV	23233	30
Cluster 1	101	4	108	JC	9771	13
Cluster 2	112	1	31	FA	6962	9
Cluster 3	82	1	44	FA	2231	3
Cluster 4	151	1	45	FP	7339	9
Cluster 5	101	1	45	FM	9948	13
Cluster 6	50	2	102	CR	11472	15
Cluster 7	101	3	43	JP	6515	8
Número de Clusters = 8				Seed = 50		
Clusters	Comarca	Seção	Classe	Especialização	Qtde.	%
Cluster 0	101	1	45	FM	20798	27%
Cluster 1	50	2	102	CR	13590	18
Cluster 2	96	1	33	FA	7859	20
Cluster 3	101	4	108	JC	9016	12
Cluster 4	151	1	45	FM	4732	6
Cluster 5	158	1	37	VS	2892	4
Cluster 6	101	1	36	CV	13643	18
Cluster 7	112	1	42	PV	4941	6
Número de Clusters = 8				Seed = 100		
Clusters	Comarca	Seção	Classe	Especialização	Qtde.	%
Cluster 0	101	1	45	FM	21879	28
Cluster 1	50	2	102	CR	11040	14
Cluster 2	101	4	108	JC	6962	9
Cluster 3	101	2	103	CS	2231	3
Cluster 4	104	1	36	CV	7339	9
Cluster 5	139	1	42	FA	9948	13
Cluster 6	101	3	43	JP	11472	15
Cluster 7	89	1	44	FA	6515	8

Objetivo 1b – Classificação Processual x Tempo Justiça Comum

Teste 1 - Seção x Classe x Especialização x Tempo

Tabela D.48: Teste 1 - Seção x Classe x Especialização x Tempo

Filtro: Seção x Classe x Especialização x Tempo						
Número de Clusters = 4			Seed = 10			
Clusters	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	1	42	CV	(-inf-75.1]	11681	62
Cluster 1	2	103	CS	(75.1-150.2]	1326	7
Cluster 2	2	107	PC	(-inf-75.1]	2365	13
Cluster 3	2	102	CR	(-inf-75.1]	3536	19
Número de Clusters = 8			Seed = 10			
Clusters	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	2	107	PC	(-inf-75.1]	2444	13
Cluster 1	2	103	CS	(75.1-150.2]	1212	6
Cluster 2	2	103	AC	(75.1-150.2]	108	1
Cluster 3	2	102	CR	(-inf-75.1]	3536	19
Cluster 4	2	107	JC	(75.1-150.2]	97	1
Cluster 5	1	34	CV	(-inf-75.1]	4181	22
Cluster 6	1	31	FA	(1051.4-inf)	3350	18
Cluster 7	1	42	PV	(-inf-75.1]	3980	21
Número de Clusters = 12			Seed = 10			
Clusters	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	2	107	PC	(-inf-75.1]	2443	13
Cluster 1	2	103	CS	(-inf-75.1]	953	5
Cluster 2	2	103	AC	(75.1-150.2]	198	1
Cluster 3	2	102	CR	(-inf-75.1]	3535	19
Cluster 4	2	107	JC	(75.1-150.2]	98	1
Cluster 5	1	45	FM	(-inf-75.1]	1723	9
Cluster 6	1	33	FA	(150.2-225.3]	2970	16
Cluster 7	1	42	PV	(-inf-75.1]	3893	21
Cluster 8	1	31	CV	(1051.4-inf)	2455	13
Cluster 9	1	109	AC	(-inf-75.1]		
Cluster 10	1	35	CV	(75.1-150.2]	640	3
Cluster 11	1	109	AC	(-inf-75.1]		

Teste 2 - Comarca x Seção x Classe x Especialização x Tempo

Tabela D.49: Teste 2 - Comarca x Seção x Classe x Especialização x Tempo

Filtro: Comarca x Seção x Classe x Especialização x Tempo						
Número de Clusters = 4			Seed = 10			
Clusters	Comarca	Seção	Classe	Espec.	Tempo	Qtde. %
Cluster 0	50	1	42	PV	(-inf-75.1]	7674 41
Cluster 1	85	2	103	CS	(75.1-150.2]	2018 11
Cluster 2	89	1	31	CV	(75.1-150.2]	3929 21
Cluster 3	50	2	102	CR	(-inf-75.1]	5287 28
Número de Clusters = 8			Seed = 10			
Clusters	Comarca	Seção	Classe	Espec.	Tempo	Qtde. %

Cluster 0	101	2	107	PC	(-inf-75.1]	2512	13
Cluster 1	50	2	103	CS	(75.1-150.2]	1263	7
Cluster 2	89	1	36	CV	(75.1-150.2]	3848	20
Cluster 3	50	2	102	CR	(-inf-75.1]	3598	19
Cluster 4	101	1	42	PV	(75.1-150.2]	1612	9
Cluster 5	149	1	45	FM	(1051.4-inf)	1007	5
Cluster 6	139	1	31	FA	(150.2-225.3]	2502	13
Cluster 7	151	1	42	PV	(-inf-75.1]	2566	14
Número de Clusters = 12				Seed = 10			
Clusters	Comarca	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	96	2	107	PC	(-inf-75.1]	2272	12
Cluster 1	50	2	103	CS	(75.1-150.2]	1167	5
Cluster 2	89	1	40	FA	(75.1-150.2]	276	15
Cluster 3	50	2	102	CR	(-inf-75.1]	3044	16
Cluster 4	101	2	107	JC	(75.1-150.2]	220	1
Cluster 5	149	1	45	FM	(1051.4-inf)	1326	7
Cluster 6	139	1	33	FA	(150.2-225.3]	1487	8
Cluster 7	151	1	42	PV	(-inf-75.1]	3871	20
Cluster 8	151	1	31	CV	(300.4-375.5]	1599	8
Cluster 9	139	2	107	PC	(-inf-75.1]	184	1
Cluster 10	112	1	35	CV	(75.1-150.2]	493	3
Cluster 11	112	2	102	CR	(-inf-75.1]	489	3

Objetivo 1b – Classificação Processual x Tempo Juizados Especiais

Teste 1 - Seção x Classe x Especialização x Tempo

Tabela D.50: Teste 1 - Seção x Classe x Especialização x Tempo

Filtro: Seção x Classe x Especialização x Tempo							
Número de Clusters = 4				Seed = 10			
Clusters	Seção	Classe	Espec.	Tempo	Qtde.	%	
Cluster 0	4	108	JC	(125.37-inf)	4184	55	
Cluster 1	3	43	JP	(41.79-83.58]	1810	24	
Cluster 2	3	43	JE	(125.37-inf)	690	9	
Cluster 3	4	108	JC	(-inf-41.79]	879	12	
Número de Clusters = 8				Seed = 10			
Clusters	Seção	Classe	Espec.	Tempo	Qtde.	%	
Cluster 0	4	108	JC	(125.37-inf)	2616	35	
Cluster 1	3	43	JP	(41.79-83.58]	1608	21	
Cluster 2	3	43	JE	(125.37-inf)	620	8	
Cluster 3	4	108	JC	(-inf-41.79]	879	12	
Cluster 4	3	46	PJ	(-inf-41.79]	272	4	
Cluster 5	3	109	AC	(-inf-41.79]			
Cluster 6	4	108	JC	(83.58-125.37]	636	8	
Cluster 7	4	108	JC	(41.79-83.58]	932	12	
Número de Clusters = 12				Seed = 10			
Clusters	Seção	Classe	Espec.	Tempo	Qtde.	%	
Cluster 0	4	108	JC	(125.37-inf)	2616	35	
Cluster 1	3	43	JP	(41.79-83.58]	1186	16	

Cluster 2	3	43	JE	(125.37-inf)	743	10
Cluster 3	4	108	JC	(-inf-41.79]	879	12
Cluster 4	3	109	AC	(-inf-41.79]		
Cluster 5	3	109	AC	(-inf-41.79]		
Cluster 6	4	108	JC	(83.58-125.37]	636	8
Cluster 7	4	108	JC	(41.79-83.58]	932	12
Cluster 8	4	109	AC	(-inf-41.79]		
Cluster 9	4	109	AC	(-inf-41.79]		
Cluster 10	4	109	AC	(-inf-41.79]		
Cluster 11	3	43	JP	(41.79-83.58]	571	8

Teste 2 - Comarca x Seção x Classe x Especialização x Tempo

Tabela D.51: Teste 2 - Comarca x Seção x Classe x Especialização x Tempo

Filtro: Comarca x Seção x Classe x Especialização x Tempo							
Número de Clusters = 4				Seed = 10			
Clusters	Comarca	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	139	4	108	JC	(125.37-inf)	4131	55
Cluster 1	151	3	43	JP	(41.79-83.58]	1798	24
Cluster 2	151	3	43	JE	(125.37-inf)	690	9
Cluster 3	50	4	108	JC	(-inf-41.79]	944	12
Número de Clusters = 8				Seed = 10			
Clusters	Comarca	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	151	4	108	JC	(125.37-inf)	2358	31
Cluster 1	151	3	43	JP	(41.79-83.58]	1798	24
Cluster 2	151	3	43	JE	(125.37-inf)	690	9
Cluster 3	50	4	108	JC	(-inf-41.79]	957	13
Cluster 4	139	4	108	JC	(125.37-inf)	700	9
Cluster 5	126	4	108	JC	(125.37-inf)	176	2
Cluster 6	82	4	108	JC	(83.58-125.37]	381	5
Cluster 7	126	4	108	JC	(41.79-83.58]	503	7
Número de Clusters = 12				Seed = 10			
Clusters	Comarca	Seção	Classe	Espec.	Tempo	Qtde.	%
Cluster 0	151	4	108	JC	(125.37-inf)	1763	23
Cluster 1	151	3	43	JP	(-inf-41.79]	1186	16
Cluster 2	151	3	43	JE	(125.37-inf)	743	10
Cluster 3	50	4	108	JC	(-inf-41.79]	945	12
Cluster 4	139	4	108	JC	(125.37-inf)	700	9
Cluster 5	126	4	108	JC	(125.37-inf)	176	2
Cluster 6	82	4	108	JC	(83.58-125.37]	381	5
Cluster 7	126	4	108	JC	(41.79-83.58]	466	6
Cluster 8	96	4	108	JC	(125.37-inf)	185	2
Cluster 9	50	4	108	JC	(125.37-inf)	410	5
Cluster 10	89	4	108	JC	(41.79-83.58]	40	1
Cluster 11	151	3	43	JP	(41.79-83.58]	568	8

Objetivo 1c – Classificação Processual x Sentença

Teste 1 - Seção x Classe x Especialização x Sentença

Tabela D.52: Teste 1 - Seção x Classe x Especialização x Sentença

Filtro: Seção x Classe x Especialização x Sentença						
Número de Clusters = 4				Seed = 10		
Clusters	Seção	Classe	Especialização	Sentença	Qtde.	%
Cluster 0	4	108	JC	19	6825	23
Cluster 1	1	31	CV	12	14533	48
Cluster 2	3	43	JP	12	2888	10
Cluster 3	2	102	CR	17	5992	20
Número de Clusters = 8				Seed = 10		
Clusters	Seção	Classe	Especialização	Sentença	Qtde.	%
Cluster 0	4	108	JC	19	6825	23
Cluster 1	1	33	FA	12	5376	18
Cluster 2	3	43	JP	12	1918	6
Cluster 3	2	102	CR	17	5992	20
Cluster 4	3	43	JE	12	970	3
Cluster 5	1	36	CV	12	3680	12
Cluster 6	1	31	CV	8	3576	12
Cluster 7	1	45	FM	12	1901	6
Número de Clusters = 12				Seed = 10		
Clusters	Seção	Classe	Especialização	Sentença	Qtde.	%
Cluster 0	4	108	JC	19	6825	23
Cluster 1	1	33	FA	12	4288	14
Cluster 2	3	43	JP	12	1918	6
Cluster 3	2	102	CR	17	5992	20
Cluster 4	3	43	JE	12	970	3
Cluster 5	1	36	CV	8	3120	10
Cluster 6	1	31	CV	8	3550	12
Cluster 7	1	45	FM	12	1901	6
Cluster 8	1	37	VS	4	746	2
Cluster 9	1	40	CV	12	738	2
Cluster 10	1	41	FC	12	190	1
Cluster 11	1	109	AC	4		

Teste 2 - Comarca x Seção x Classe x Especialização x Sentença

Tabela D.53: Teste 2 - Comarca x Seção x Classe x Especialização x Sentença

Filtro: Comarca x Seção x Classe x Especialização x Sentença						
Número de Clusters = 4				Seed = 10		
Clusters	Comarca	Seção	Classe	Especialização	Sentença	Qtde. %
Cluster 0	139	4	108	JC	19	7130 24
Cluster 1	101	1	31	CV	12	14276 47
Cluster 2	151	3	43	JP	12	2926 10
Cluster 3	50	2	102	CR	17	5906 20
Número de Clusters = 8				Seed = 10		
Clusters	Comarca	Seção	Classe	Especialização	Sentença	Qtde. %
Cluster 0	139	4	108	JC	19	6877 23

Cluster 1	50	1	33	FA	12	5344	18
Cluster 2	151	3	43	JP	12	2201	7
Cluster 3	50	2	102	CR	17	5925	20
Cluster 4	101	3	43	JE	12	728	2
Cluster 5	85	1	36	CV	12	3574	12
Cluster 6	89	1	31	CV	8	3561	12
Cluster 7	101	1	45	FM	12	2028	7
Número de Clusters = 12				Seed = 10			
Clusters	Comarca	Seção	Classe	Especialização	Sentença	Qtde.	%
Cluster 0	96	1	37	VS	4	1608	5
Cluster 1	139	1	33	FA	12	3802	13
Cluster 2	151	3	43	JP	12	2273	8
Cluster 3	50	2	102	CR	17	6184	20
Cluster 4	101	3	43	JE	12	727	2
Cluster 5	85	1	36	CV	8	2780	9
Cluster 6	89	1	31	CV	8	3426	11
Cluster 7	101	1	45	FM	12	1928	6
Cluster 8	50	4	108	JC	19	5421	18
Cluster 9	50	1	40	CV	12	951	3
Cluster 10	101	1	41	FC	12	210	1
Cluster 11	139	4	108	JC	19	928	3

Objetivo 1d – Classificação Processual x Audiência

Teste 1 - Seção x Classe x Especialização x Audiência

Tabela D.54: Teste 2 - Seção x Classe x Especialização x Audiência

Filtro: Seção x Classe x Especialização x Audiência						
Número de Clusters = 4				Seed = 10		
Clusters	Seção	Classe	Especialização	Qtde.	%	
Cluster 0	4	108	JC	11451	45	
Cluster 1	2	102	CR	6518	26	
Cluster 2	2	102	CR	551	2	
Cluster 3	1	31	FA	6948	27	
Número de Clusters = 8				Seed = 10		
Clusters	Seção	Classe	Especialização	Qtde.	%	
Cluster 0	4	108	JC	10034	39	
Cluster 1	2	102	CR	5237	21	
Cluster 2	2	102	CR	440	2	
Cluster 3	1	31	FA	3549	14	
Cluster 4	1	33	FA	2014	14	
Cluster 5	2	103	CS	1633	6	
Cluster 6	1	31	CV	2325	9	
Cluster 7	4	108	JC	236	1	
Número de Clusters = 12				Seed = 10		
Clusters	Seção	Classe	Especialização	Qtde.	%	
Cluster 0	4	108	JC	8635	34	
Cluster 1	2	102	CR	5225	21	

Cluster 2	2	102		CR	413	2
Cluster 3	1	31		FA	3363	13
Cluster 4	1	33		FA	2009	8
Cluster 5	2	103		CS	1633	6
Cluster 6	1	31		CV	2320	9
Cluster 7	4	108		JC	236	1
Cluster 8	4	108		JC	699	3
Cluster 9	4	108		JC	698	3
Cluster 10	1	40		RP	237	1
Cluster 11	1	109		AC		

Objetivo 2a - Perfil de Réus

Teste 1 – Perfil Réu

Tabela D.55: Teste 1 – Perfil do Réu

Filtro: Comarca x Natureza x Sexo x Estado Civil x Grau de Instrução x Profissão x Cor x Idade										
Número de Clusters = 8					Seed = 10					
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0	50	247	M	2	2	123	1	(22.5-27.6]	2314	27
Cluster 1	96	245	M	1	2	123	1	(22.5-27.6]	1820	21
Cluster 2	139	244	F	1	3	56	1	(27.6-32.7]	209	2
Cluster 3	82	247	M	2	2	179	1	(53.1-58.2]	359	4
Cluster 4	104	244	F	2	2	56	1	(22.5-27.6]	269	3
Cluster 5	101	247	M	2	6	176	1	(42.9-48]	184	2
Cluster 6	96	244	M	2	2	179	1	(17.4-22.5]	864	10
Cluster 7	104	244	M	2	2	123	1	(22.5-27.6]	2659	31
Número de Clusters = 8					Seed = 10					
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0		247	M	2	2	123	1	(22.5-27.6]	2145	25
Cluster 1		245	M	1	2	123	1	(22.5-27.6]	1933	22
Cluster 2		298	F	1	2	56	1	(27.6-32.7]	284	3
Cluster 3		244	M	2	2	179	1	(53.1-58.2]	756	9
Cluster 4		244	F	2	2	56	1	(22.5-27.6]	186	2
Cluster 5		247	M	2	3	176	1	(42.9-48]	216	2
Cluster 6		244	M	1	2	123	1	(17.4-22.5]	1178	14
Cluster 7		244	M	1	2	123	1	(22.5-27.6]	1980	23
Número de Clusters = 8					Seed = 10					
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0		247	M	2	2	123		(22.5-27.6]	2145	25
Cluster 1		245	M	1	2	123		(22.5-27.6]	1933	22
Cluster 2		298	F	1	2	56		(27.6-32.7]	284	3
Cluster 3		244	M	2	2	179		(27.6-32.7]	756	9
Cluster 4		244	F	2	2	56		(22.5-27.6]	186	2
Cluster 5		247	M	2	3	176		(42.9-48]	216	2
Cluster 6		244	M	1	2	123		(17.4-22.5]	1178	14

Cluster 7		244	M	1	2	123		(22.5-27.6]	1980	23
Número de Clusters = 8				Seed = 10						
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0		247	M	2	2			(22.5-27.6]	2424	28
Cluster 1		245	M	1	2			(22.5-27.6]	1713	20
Cluster 2		298	F	1	2			(27.6-32.7]	410	5
Cluster 3		244	M	1	2			(27.6-32.7]	1091	13
Cluster 4		244	F	2	2			(22.5-27.6]	262	3
Cluster 5		244	M	2	6			(63.3-inf)	111	1
Cluster 6		244	M	1	2			(17.4-22.5]	706	8
Cluster 7		244	M	1	2			(22.5-27.6]	1961	23
Número de Clusters = 8				Seed = 10						
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0		247	M	6				(22.5-27.6]	1696	20
Cluster 1		245	M	1				(22.5-27.6]	1904	22
Cluster 2		298	F	1				(27.6-32.7]	460	5
Cluster 3		244	M	4				(53.1-58.2]	338	4
Cluster 4		244	F	6				(22.5-27.6]	175	2
Cluster 5		244	M	2				(42.9-48]	1124	13
Cluster 6		244	M	1				(17.4-22.5]	1117	13
Cluster 7		244	M	1				(22.5-27.6]	1864	21
Número de Clusters = 8				Seed = 10						
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0		247	M					(22.5-27.6]	2970	34
Cluster 1		245	M					(22.5-27.6]	639	7
Cluster 2		298	F					(27.6-32.7]	539	6
Cluster 3		246	M					(53.1-58.2]	229	3
Cluster 4		247	F					(22.5-27.6]	303	3
Cluster 5		244	M					(27.6-32.7]	1285	15
Cluster 6		244	M					(17.4-22.5]	687	8
Cluster 7		244	M					(22.5-27.6]	2026	23
Número de Clusters = 8				Seed = 10						
Clusters	Comarca	Natureza	Sexo	Estado civil	Grau de Instrução	Profissão	Cor	Idade	Qtde.	%
Cluster 0			M	2	2	123	1	(22.5-27.6]	1991	23
Cluster 1			M	1	2	123	1	(22.5-27.6]	5070	58
Cluster 2			F	1	2	56	1	(27.6-32.7]	283	3
Cluster 3			M	2	2	179	1	(53.1-58.2]	366	4
Cluster 4			F	2	2	56	1	(22.5-27.6]	166	2
Cluster 5			M	2	3	176	1	(42.9-48]	230	3
Cluster 6			M	6	2	179	1	(17.4-22.5]	323	4
Cluster 7			M	1	2	123	1	(22.5-27.6]	249	3

ANEXO E RESULTADOS DA MINERAÇÃO – TÉCNICA REGRAS DE ASSOCIAÇÃO

Testes levando em consideração a confiança

Teste 1 – Natureza x Estado Civil x Grau de Instrução x Profissão x Cor x Idade

Tabela E.56: Teste 1 - Natureza x Est. Civil x Grau de Instr. x Profissão x Cor x Idade

```

=== Run information ===
Scheme: weka.associations.Apriori -R -N 50 -T 0 -C 0.3 -D 0.01 -U
0.42325420603825764 -M 0.01 -S -1.0
Relation:  Objetivo6-weka.filters.AttributeFilter-V-R5,7-11-
weka.filters.DiscretizeFilter-E-O-B20-Rfirst-last
Instances: 8678
Attributes: 6
      COD_NATUREZA
      COD_ESTADO_CIVIL
      COD_GRAU_INSTRUCAO
      COD_PROFISSAO
      COD_COR
      IDADE
=== Associator model (full training set) ===

Apriori
=====
Minimum support: 0.05
Minimum metric <confidence>: 0.3
Number of cycles performed: 37

Generated sets of large itemsets:

Size of set of large itemsets L(1): 17
Size of set of large itemsets L(2): 23
Size of set of large itemsets L(3): 8

Best rules found:
1. IDADE=(17.4-22.5]' 869 ==> COD_ESTADO_CIVIL=1 704  conf: (0.81)
2. COD_COR=1 IDADE=(22.5-27.6]' 688 ==> COD_ESTADO_CIVIL=1 544  conf: (0.79)
3. IDADE=(22.5-27.6]' 1313 ==> COD_ESTADO_CIVIL=1 978  conf: (0.74)

```

4. COD_NATUREZA=244 COD_PROFISSAO=123 697 ==> COD_ESTADO_CIVIL=1 481 conf: (0.69)
5. COD_GRAU_INSTRUCAO=2 COD_PROFISSAO=123 757 ==> COD_ESTADO_CIVIL=1 505 conf: (0.67)
6. COD_PROFISSAO=123 COD_COR=1 764 ==> COD_ESTADO_CIVIL=1 474 conf: (0.62)
7. COD_PROFISSAO=123 1673 ==> COD_ESTADO_CIVIL=1 1033 conf: (0.62)
8. COD_GRAU_INSTRUCAO=2 2504 ==> COD_COR=1 1511 conf: (0.6)
9. COD_NATUREZA=244 COD_COR=1 1196 ==> COD_ESTADO_CIVIL=1 718 conf:(0.6)
10. COD_NATUREZA=244 COD_GRAU_INSTRUCAO=2 1006 ==> COD_ESTADO_CIVIL=1 592 conf:(0.59)
11. COD_NATUREZA=244 COD_GRAU_INSTRUCAO=2 1006 ==> COD_COR=1 584 conf:(0.58)
12. COD_ESTADO_CIVIL=1 COD_GRAU_INSTRUCAO=2 1313 ==> COD_COR=1 748 conf:(0.57)
13. COD_PROFISSAO=179 831 ==> COD_COR=1 470 conf:(0.57)
14. COD_ESTADO_CIVIL=2 1901 ==> COD_COR=1 1075 conf:(0.57)
15. IDADE='(32.7-37.8]' 819 ==> COD_COR=1 463 conf:(0.57)
16. COD_ESTADO_CIVIL=1 IDADE='(22.5-27.6]' 978 ==> COD_COR=1 544 conf:(0.56)
17. IDADE='(17.4-22.5]' 869 ==> COD_NATUREZA=244 481 conf: (0.55)
18. IDADE='(27.6-32.7]' 1008 ==> COD_ESTADO_CIVIL=1 543 conf: (0.54)
19. COD_ESTADO_CIVIL=1 3287 ==> COD_COR=1 1759 conf:(0.54)
20. IDADE='(27.6-32.7]' 1008 ==> COD_COR=1 536 conf: (0.53)
21. COD_GRAU_INSTRUCAO=2 2504 ==> COD_ESTADO_CIVIL=1 1313 conf: (0.52)
22. IDADE='(22.5-27.6]' 1313 ==> COD_COR=1 688 conf: (0.52)
23. COD_NATUREZA=244 COD_ESTADO_CIVIL=1 1374 ==> COD_COR=1 718 conf:(0.52)
24. COD_GRAU_INSTRUCAO=2 COD_COR=1 1511 ==> COD_ESTADO_CIVIL=1 748 conf:(0.5)
25. COD_NATUREZA=247 995 ==> COD_COR=1 492 conf:(0.49)
26. COD_ESTADO_CIVIL=1 COD_PROFISSAO=123 1033 ==> COD_GRAU_INSTRUCAO=2 505 conf:(0.49)
27. COD_COR=1 3599 ==> COD_ESTADO_CIVIL=1 1759 conf:(0.49)
28. COD_NATUREZA=244 COD_COR=1 1196 ==> COD_GRAU_INSTRUCAO=2 584 conf:(0.49)
29. COD_ESTADO_CIVIL=1 COD_PROFISSAO=123 1033 ==> COD_NATUREZA=244 481 conf:(0.47)
30. COD_ESTADO_CIVIL=1 COD_PROFISSAO=123 1033 ==> COD_COR=1 474 conf:(0.46)
31. COD_PROFISSAO=123 1673 ==> COD_COR=1 764 conf: (0.46)
32. COD_PROFISSAO=123 1673 ==> COD_GRAU_INSTRUCAO=2 757 conf: (0.45)
33. COD_ESTADO_CIVIL=1 COD_GRAU_INSTRUCAO=2 1313 ==> COD_NATUREZA=244 592 conf: (0.45)
34. IDADE='(22.5-27.6]' 1313 ==> COD_NATUREZA=244 574 conf: (0.44)
35. IDADE='(22.5-27.6]' 1313 ==> COD_GRAU_INSTRUCAO=2 571 conf: (0.43)
36. COD_NATUREZA=244 COD_ESTADO_CIVIL=1 1374 ==> COD_GRAU_INSTRUCAO=2 592 conf: (0.43)
37. COD_ESTADO_CIVIL=1 COD_COR=1 1759 ==> COD_GRAU_INSTRUCAO=2 748 conf: (0.43)
38. COD_COR=1 3599 ==> COD_GRAU_INSTRUCAO=2 1511 conf: (0.42)
39. COD_ESTADO_CIVIL=1 3287 ==> COD_NATUREZA=244 1374 conf: (0.42)
40. COD_PROFISSAO=123 1673 ==> COD_NATUREZA=244 697 conf: (0.42)
41. IDADE='(22.5-27.6]' 1313 ==> COD_ESTADO_CIVIL=1 COD_COR=1 544 conf: (0.41)
42. COD_ESTADO_CIVIL=1 COD_COR=1 1759 ==> COD_NATUREZA=244 718 conf:(0.41)
43. COD_GRAU_INSTRUCAO=2 2504 ==> COD_NATUREZA=244 1006 conf: (0.4)
44. COD_ESTADO_CIVIL=1 3287 ==> COD_GRAU_INSTRUCAO=2 1313 conf: (0.4)
45. COD_GRAU_INSTRUCAO=2 COD_COR=1 1511 ==> COD_NATUREZA=244 584 conf: (0.39)
46. COD_ESTADO_CIVIL=1 COD_GRAU_INSTRUCAO=2 1313 ==> COD_PROFISSAO=123 505 conf: (0.38)
47. COD_NATUREZA=244 3673 ==> COD_ESTADO_CIVIL=1 1374 conf: (0.37)
48. COD_NATUREZA=244 COD_ESTADO_CIVIL=1 1374 ==> COD_PROFISSAO=123 481 conf: (0.35)
49. COD_COR=1 3599 ==> COD_NATUREZA=244 1196 conf: (0.33)
50. COD_NATUREZA=244 3673 ==> COD_COR=1 1196 conf: (0.33)

Teste 2 – Comarca x Natureza x Profissão x Idade

Tabela E.57: Teste 2 – Comarca x Natureza x Profissão x Idade

=== Run information ===

Scheme: weka.associations.Apriori -R -N 50 -T 0 -C 0.3 -D 0.01 -U
 0.42325420603825764 -M 0.01 -S -1.0
 Relation: Objetivo6-weka.filters.AttributeFilter-V-R1,5,9,11-
 weka.filters.DiscretizeFilter-E-O-B20-Rfirst-last
 Instances: 8678
 Attributes: 4

COD_COMARCA
 COD_NATUREZA
 COD_PROFISAO
 IDADE

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.01

Minimum metric <confidence>: 0.3

Number of cycles performed: 41

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44

Size of set of large itemsets L(2): 56

Size of set of large itemsets L(3): 6

Best rules found:

1. COD_COMARCA=89 645 ==> COD_NATUREZA=244 393 conf: (0.61)
2. COD_PROFISAO=123 IDADE='(17.4-22.5]' 244 ==> COD_NATUREZA=244 143 conf: (0.59)
3. COD_COMARCA=112 527 ==> COD_NATUREZA=244 295 conf: (0.56)
4. IDADE='(17.4-22.5]' 869 ==> COD_NATUREZA=244 481 conf: (0.55)
5. COD_COMARCA=50 IDADE='(22.5-27.6]' 227 ==> COD_PROFISAO=123 121 conf: (0.53)
6. COD_COMARCA=104 1360 ==> COD_NATUREZA=244 678 conf: (0.5)
7. COD_PROFISAO=123 IDADE='(22.5-27.6]' 413 ==> COD_NATUREZA=244 200 conf: (0.48)
8. COD_COMARCA=85 596 ==> COD_NATUREZA=244 282 conf: (0.47)
9. IDADE='(22.5-27.6]' 1313 ==> COD_NATUREZA=244 574 conf: (0.44)
10. COD_COMARCA=158 350 ==> COD_NATUREZA=244 153 conf: (0.44)
11. COD_COMARCA=126 396 ==> COD_NATUREZA=244 169 conf: (0.43)
12. COD_COMARCA=104 COD_PROFISAO=123 356 ==> COD_NATUREZA=244 150 conf: (0.42)
13. COD_COMARCA=151 864 ==> COD_NATUREZA=244 364 conf: (0.42)
14. COD_NATUREZA=301 482 ==> COD_COMARCA=151 202 conf: (0.42)
15. COD_PROFISAO=123 1673 ==> COD_NATUREZA=244 697 conf: (0.42)
16. COD_COMARCA=50 COD_PROFISAO=123 411 ==> COD_NATUREZA=244 171 conf: (0.42)
17. COD_PROFISAO=123 IDADE='(27.6-32.7]' 306 ==> COD_NATUREZA=244 125 conf: (0.41)
18. COD_COMARCA=50 COD_NATUREZA=244 428 ==> COD_PROFISAO=123 171 conf: (0.4)
19. COD_COMARCA=101 676 ==> COD_NATUREZA=244 255 conf: (0.38)
20. IDADE='(27.6-32.7]' 1008 ==> COD_NATUREZA=244 377 conf: (0.37)
21. COD_COMARCA=50 1148 ==> COD_NATUREZA=244 428 conf: (0.37)
22. COD_PROFISAO=179 831 ==> COD_NATUREZA=244 298 conf: (0.36)
23. COD_COMARCA=50 1148 ==> COD_PROFISAO=123 411 conf: (0.36)
24. COD_NATUREZA=244 IDADE='(22.5-27.6]' 574 ==> COD_PROFISAO=123 200 conf: (0.35)
25. COD_NATUREZA=244 IDADE='(27.6-32.7]' 377 ==> COD_PROFISAO=123 125 conf: (0.33)
26. COD_NATUREZA=245 641 ==> COD_PROFISAO=123 209 conf: (0.33)
27. IDADE='(22.5-27.6]' 1313 ==> COD_PROFISAO=123 413 conf: (0.31)
28. COD_COMARCA=139 863 ==> COD_NATUREZA=244 266 conf: (0.31)
29. COD_NATUREZA=301 482 ==> IDADE='(22.5-27.6]' 148 conf: (0.31)
30. IDADE='(27.6-32.7]' 1008 ==> COD_PROFISAO=123 306 conf: (0.3)

Teste 3 – Sexo Feminino - Comarca x Natureza x Profissão x Idade

Tabela E.58: Teste 3 - Sexo Feminino

=== Run information ===

Scheme: weka.associations.Apriori -N 50 -T 0 -C 0.3 -D 0.01 -U 1.0 -M 0.01 -S -1.0

Relation: Objetivo61-weka.filters.AttributeFilter-V-R1-2,6,8-weka.filters.DiscretizeFilter-B20-Rfirst-last

Instances: 867

Attributes: 4

COD_COMARCA
COD_NATUREZA
COD_PROFISAO
IDADE

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.01

Minimum metric <confidence>: 0.3

Number of cycles performed: 98

Generated sets of large itemsets:

Size of set of large itemsets L(1): 41

Size of set of large itemsets L(2): 37

Size of set of large itemsets L(3): 1

Best rules found:

1. COD_COMARCA=112 42 ==> COD_PROFISAO=56 23 conf: (0.55)
2. COD_COMARCA=96 COD_PROFISAO=56 35 ==> COD_NATUREZA=298 18 conf: (0.51)
3. COD_NATUREZA=298 COD_PROFISAO=56 36 ==> COD_COMARCA=96 18 conf: (0.5)
4. COD_NATUREZA=261 36 ==> COD_PROFISAO=56 18 conf: (0.5)
5. IDADE='(18.25-22.5]' 51 ==> COD_NATUREZA=244 24 conf: (0.47)
6. COD_COMARCA=96 123 ==> COD_NATUREZA=298 55 conf: (0.45)
7. COD_COMARCA=101 109 ==> COD_NATUREZA=247 46 conf: (0.42)
8. COD_NATUREZA=298 148 ==> COD_COMARCA=96 55 conf: (0.37)
9. COD_COMARCA=101 109 ==> COD_NATUREZA=244 39 conf: (0.36)
10. COD_COMARCA=85 62 ==> COD_NATUREZA=298 22 conf: (0.35)
11. COD_COMARCA=96 COD_NATUREZA=298 55 ==> COD_PROFISAO=56 18 conf: (0.33)
12. IDADE='(35.25-39.5]' 70 ==> COD_PROFISAO=56 22 conf: (0.31)
13. IDADE='(31-35.25]' 106 ==> COD_COMARCA=104 32 conf: (0.3)

Testes levando em consideração o *lift*

Teste 4 –Natureza x Estado Civil x Grau de Instrução x Profissão x Cor x Idade

Tabela E.59: Teste 1 - Natureza x Est. Civil x Grau de Instr. x Profissão x Cor x Idade

=== Run information ===

Scheme: weka.associations.Apriori -N 50 -T 1 -C 0.3 -D 0.01 -U 1.0 -M 0.01 -S -1.0
Relation: Objetivo6-weka.filters.AttributeFilter-V-R5,7-11-weka.filters.DiscretizeFilter-E-O-B20-Rfirst-last
Instances: 8678

```

Attributes: 6
            COD_NATUREZA
            COD_ESTADO_CIVIL
            COD_GRAU_INSTRUCAO
            COD_PROFISSAO
            COD_COR
            IDADE
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.06
Minimum metric <lift>: 0.3
Number of cycles performed: 94

Generated sets of large itemsets:

Size of set of large itemsets L(1): 16
Size of set of large itemsets L(2): 19
Size of set of large itemsets L(3): 5

Best rules found:

  1. IDADE='(17.4-22.5]' 869 ==> COD_ESTADO_CIVIL=1 704   conf:(0.81) < lift:(2.14)>
  2. COD_ESTADO_CIVIL=1 3287 ==> IDADE='(17.4-22.5]' 704   conf:(0.21) < lift:(2.14)>
  3. COD_COR=1 IDADE='(22.5-27.6]' 688 ==> COD_ESTADO_CIVIL=1 544   conf:(0.79) <
lift:(2.09)>
  4. COD_ESTADO_CIVIL=1 3287 ==> COD_COR=1 IDADE='(22.5-27.6]' 544   conf:(0.17) <
lift:(2.09)>
  5. IDADE='(22.5-27.6]' 1313 ==> COD_ESTADO_CIVIL=1 COD_COR=1 544   conf:(0.41) <
lift:(2.04)>
  6. COD_ESTADO_CIVIL=1 COD_COR=1 1759 ==> IDADE='(22.5-27.6]' 544   conf:(0.31) <
lift:(2.04)>
  7. IDADE='(22.5-27.6]' 1313 ==> COD_ESTADO_CIVIL=1 978   conf:(0.74) < lift:(1.97)>
  8. COD_ESTADO_CIVIL=1 3287 ==> IDADE='(22.5-27.6]' 978   conf:(0.3) < lift:(1.97)>
  9. COD_NATUREZA=244 COD_COR=1 1196 ==> COD_GRAU_INSTRUCAO=2 584   conf:(0.49) <
lift:(1.69)>
 10. COD_GRAU_INSTRUCAO=2 2504 ==> COD_NATUREZA=244 COD_COR=1 584   conf:(0.23) <
lift:(1.69)>
 11. COD_PROFISSAO=123 1673 ==> COD_ESTADO_CIVIL=1 1033   conf:(0.62) < lift:(1.63)>
 12. COD_ESTADO_CIVIL=1 3287 ==> COD_PROFISSAO=123 1033   conf:(0.31) < lift:(1.63)>
 13. COD_NATUREZA=244 COD_COR=1 1196 ==> COD_ESTADO_CIVIL=1 718   conf:(0.6) <
lift:(1.58)>
 14. COD_ESTADO_CIVIL=1 3287 ==> COD_NATUREZA=244 COD_COR=1 718   conf:(0.22) <
lift:(1.58)>
 15. COD_PROFISSAO=123 1673 ==> COD_GRAU_INSTRUCAO=2 757   conf:(0.45) < lift:(1.57)>
 16. COD_GRAU_INSTRUCAO=2 2504 ==> COD_PROFISSAO=123 757   conf:(0.3) < lift:(1.57)>
 17. COD_NATUREZA=244 COD_GRAU_INSTRUCAO=2 1006 ==> COD_ESTADO_CIVIL=1 592
conf:(0.59) < lift:(1.55)>
 18. COD_ESTADO_CIVIL=1 3287 ==> COD_NATUREZA=244 COD_GRAU_INSTRUCAO=2 592
conf:(0.18) < lift:(1.55)>
 19. IDADE='(22.5-27.6]' 1313 ==> COD_GRAU_INSTRUCAO=2 571   conf:(0.43) < lift:(1.51)>
 20. COD_GRAU_INSTRUCAO=2 2504 ==> IDADE='(22.5-27.6]' 571   conf:(0.23) < lift:(1.51)>
 21. COD_GRAU_INSTRUCAO=2 2504 ==> COD_NATUREZA=244 COD_ESTADO_CIVIL=1 592
conf:(0.24) < lift:(1.49)>
 22. COD_NATUREZA=244 COD_ESTADO_CIVIL=1 1374 ==> COD_GRAU_INSTRUCAO=2 592
conf:(0.43) < lift:(1.49)>
 23. COD_ESTADO_CIVIL=1 COD_COR=1 1759 ==> COD_GRAU_INSTRUCAO=2 748   conf:(0.43) <
lift:(1.47)>
 24. COD_GRAU_INSTRUCAO=2 2504 ==> COD_ESTADO_CIVIL=1 COD_COR=1 748   conf:(0.3) <
lift:(1.47)>
 25. COD_COR=1 3599 ==> COD_GRAU_INSTRUCAO=2 1511   conf:(0.42) < lift:(1.46)>
 26. COD_GRAU_INSTRUCAO=2 2504 ==> COD_COR=1 1511   conf:(0.6) < lift:(1.46)>
 27. IDADE='(27.6-32.7]' 1008 ==> COD_ESTADO_CIVIL=1 543   conf:(0.54) < lift:(1.42)>
 28. COD_ESTADO_CIVIL=1 3287 ==> IDADE='(27.6-32.7]' 543   conf:(0.17) < lift:(1.42)>
 29. COD_COR=1 3599 ==> COD_NATUREZA=244 COD_GRAU_INSTRUCAO=2 584   conf:(0.16) <
lift:(1.4)>
 30. COD_NATUREZA=244 COD_GRAU_INSTRUCAO=2 1006 ==> COD_COR=1 584   conf:(0.58) <
lift:(1.4)>
 31. COD_GRAU_INSTRUCAO=2 2504 ==> COD_ESTADO_CIVIL=1 1313   conf:(0.52) < lift:(1.38)>
 32. COD_ESTADO_CIVIL=1 3287 ==> COD_GRAU_INSTRUCAO=2 1313   conf:(0.4) < lift:(1.38)>

```

```

lev:(0.04) [364] conv:(89170022.26)
33. COD_COR=1 3599 ==> COD_ESTADO_CIVIL=1 COD_GRAU_INSTRUCAO=2 748   conf:(0.21) < lift:(1.37)>
34. COD_ESTADO_CIVIL=1 COD_GRAU_INSTRUCAO=2 1313 ==> COD_COR=1 748   conf:(0.57) < lift:(1.37)>
35. COD_COR=1 3599 ==> COD_ESTADO_CIVIL=2 1075   conf:(0.3) < lift:(1.36)>
36. COD_ESTADO_CIVIL=2 1901 ==> COD_COR=1 1075   conf:(0.57) < lift:(1.36)> 1
37. COD_ESTADO_CIVIL=1 IDADE='(22.5-27.6] 978 ==> COD_COR=1 544   conf:(0.56) < lift:(1.34)>
38. COD_COR=1 3599 ==> COD_ESTADO_CIVIL=1 IDADE='(22.5-27.6] 544   conf:(0.15) < lift:(1.34)>
39. COD_GRAU_INSTRUCAO=2 COD_COR=1 1511 ==> COD_ESTADO_CIVIL=1 748   conf:(0.5) < lift:(1.31)>
40. COD_ESTADO_CIVIL=1 3287 ==> COD_GRAU_INSTRUCAO=2 COD_COR=1 748   conf:(0.23) < lift:(1.31)>

```

Teste 5 – Comarca x Natureza x Profissão x Idade

Tabela E.60: Teste 2 – Comarca x Natureza x Profissão x Idade

```

=== Run information ===

Scheme:          weka.associations.Apriori -N 50 -T 1 -C 1.0 -D 0.01 -U 1.0 -M 0.01 -S -1.0
Relation:        Objetivo6-weka.filters.AttributeFilter-V-R1,5,9,11-
weka.filters.DiscretizeFilter-E-O-B20-Rfirst-last
Instances:       8678
Attributes:      4
                 COD_COMARCA
                 COD_NATUREZA
                 COD_PROFISSAO
                 IDADE
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.01
Minimum metric <lift>: 1
Number of cycles performed: 98

Generated sets of large itemsets:

Size of set of large itemsets L(1): 34

Size of set of large itemsets L(2): 26

Size of set of large itemsets L(3): 1

Best rules found:

 1. COD_NATUREZA=301 482 ==> COD_COMARCA=151 202   conf:(0.42) < lift:(4.21)>
 2. COD_COMARCA=151 864 ==> COD_NATUREZA=301 202   conf:(0.23) < lift:(4.21)>
 3. COD_NATUREZA=247 995 ==> COD_COMARCA=101 180   conf:(0.18) < lift:(2.32)>
 4. COD_COMARCA=101 676 ==> COD_NATUREZA=247 180   conf:(0.27) < lift:(2.32)>
 5. IDADE='(22.5-27.6] 1313 ==> COD_NATUREZA=244 COD_PROFISSAO=123 200 conf:(0.15) < lift:(1.9)>
 6. COD_NATUREZA=244 COD_PROFISSAO=123 697 ==> IDADE='(22.5-27.6] 200 conf:(0.29) < lift:(1.9)>
 7. COD_PROFISSAO=123 1673 ==> COD_COMARCA=50 411   conf:(0.25) < lift:(1.86)>
 8. COD_COMARCA=50 1148 ==> COD_PROFISSAO=123 411   conf:(0.36) < lift:(1.86)>
 9. COD_NATUREZA=244 IDADE='(22.5-27.6] 574 ==> COD_PROFISSAO=123 200 conf:(0.35) < lift:(1.81)>
10. COD_PROFISSAO=123 1673 ==> COD_NATUREZA=244 IDADE='(22.5-27.6] 200 conf:(0.12) < lift:(1.81)>
11. COD_PROFISSAO=123 1673 ==> COD_NATUREZA=245 209   conf:(0.12) < lift:(1.69)>
12. COD_NATUREZA=245 641 ==> COD_PROFISSAO=123 209   conf:(0.33) < lift:(1.69)>
13. IDADE='(22.5-27.6] 1313 ==> COD_PROFISSAO=123 413   conf:(0.31) < lift:(1.63)>
14. COD_PROFISSAO=123 1673 ==> IDADE='(22.5-27.6] 413   conf:(0.25) < lift:(1.63)>

```

```
15. IDADE='(27.6-32.7]' 1008 ==> COD_PROFISAO=123 306 conf:(0.3) < lift:(1.57)>
16. COD_PROFISAO=123 1673 ==> IDADE='(27.6-32.7]' 306 conf:(0.18) < lift:(1.57)>
17. IDADE='(17.4-22.5]' 869 ==> COD_PROFISAO=123 244 conf:(0.28) < lift:(1.46)>
18. COD_PROFISAO=123 1673 ==> IDADE='(17.4-22.5]' 244 conf:(0.15) < lift:(1.46)>
19. COD_NATUREZA=244 3673 ==> COD_COMARCA=89 393 conf:(0.11) < lift:(1.44)>
20. COD_COMARCA=89 645 ==> COD_NATUREZA=244 393 conf:(0.61) < lift:(1.44)>
21. COD_PROFISAO=123 1673 ==> COD_COMARCA=104 356 conf:(0.21) < lift:(1.36)>
22. COD_COMARCA=104 1360 ==> COD_PROFISAO=123 356 conf:(0.26) < lift:(1.36)>
23. COD_NATUREZA=244 3673 ==> COD_COMARCA=112 295 conf:(0.08) < lift:(1.32)>
24. COD_COMARCA=112 527 ==> COD_NATUREZA=244 295 conf:(0.56) < lift:(1.32)>
25. IDADE='(17.4-22.5]' 869 ==> COD_NATUREZA=244 481 conf:(0.55) < lift:(1.31)>
26. COD_NATUREZA=244 3673 ==> IDADE='(17.4-22.5]' 481 conf:(0.13) < lift:(1.31)>
27. IDADE='(22.5-27.6]' 1313 ==> COD_COMARCA=50 227 conf:(0.17) < lift:(1.31)>
28. COD_COMARCA=50 1148 ==> IDADE='(22.5-27.6]' 227 conf:(0.2) < lift:(1.31)>
29. IDADE='(32.7-37.8]' 819 ==> COD_PROFISAO=123 196 conf:(0.24) < lift:(1.24)>
30. COD_PROFISAO=123 1673 ==> IDADE='(32.7-37.8]' 196 conf:(0.12) < lift:(1.24)>
31. COD_NATUREZA=244 3673 ==> COD_COMARCA=104 678 conf:(0.18) < lift:(1.18)>
32. COD_COMARCA=104 1360 ==> COD_NATUREZA=244 678 conf:(0.5) < lift:(1.18)>
33. COD_PROFISAO=123 IDADE='(22.5-27.6]' 413 ==> COD_NATUREZA=244 200 conf:(0.48) <
lift:(1.14)>
34. COD_NATUREZA=244 3673 ==> COD_PROFISAO=123 IDADE='(22.5-27.6]' 200 conf:(0.05)
< lift:(1.14)>
35. COD_NATUREZA=244 3673 ==> COD_COMARCA=85 282 conf:(0.08) < lift:(1.12)> [29]
36. COD_COMARCA=85 596 ==> COD_NATUREZA=244 282 conf:(0.47) < lift:(1.12)>
37. IDADE='(22.5-27.6]' 1313 ==> COD_NATUREZA=244 574 conf:(0.44) < lift:(1.03)>
38. COD_NATUREZA=244 3673 ==> IDADE='(22.5-27.6]' 574 conf:(0.16) < lift:(1.03)>
```