

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

RSAPP, UM ALGORITMO BASEADO EM
ROUGH SETS PARA AUXÍLIO AO
PROCESSO DE DESCOBERTA DE
CONHECIMENTO EM BANCO DE
DADOS

JULIANO GOMES DA SILVEIRA

DISSERTAÇÃO DE MESTRADO

ORIENTADOR: PROF. DR. DUNCAN DUBUGRAS ALCOBA RUIZ

PORTO ALEGRE
2013

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**RSAPP, UM ALGORITMO BASEADO EM
ROUGH SETS PARA AUXÍLIO AO
PROCESSO DE DESCOBERTA DE
CONHECIMENTO EM BANCO DE
DADOS**

Dissertação apresentada como requisito à
obtenção parcial ao título de Mestre em
Ciência da Computação da Pontifícia
Universidade Católica do Rio Grande do
Sul.

ORIENTADOR: PROF. DR. DUNCAN DUBUGRAS ALCOBA RUIZ

PORTO ALEGRE
2013

S587R Silveira, Juliano Gomes da
 RSAPP, um algoritmo baseado em rough sets para auxílio ao
 processo de descoberta de conhecimento em banco de dados / Juliano
 Gomes da Silveira. – Porto Alegre, 2013.
 161 f.

Diss. (Mestrado) – Fac. de Informática, PUCRS.
Orientador: Prof. Dr. Duncan Dubugras Alcoba Ruiz.

1. Informática. 2. Banco de Dados. 3. Mineração de Dados. I. Ruiz,
Duncan Dubugras Alcoba. II. Título.

CDD 005.74

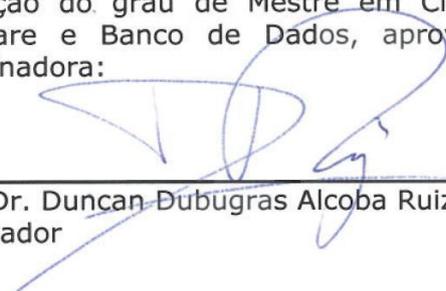
**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Rsapp, Um Algoritmo Baseado em *Rough Sets* para Auxílio ao Processo de Descoberta de Conhecimento em Banco de Dados" apresentada por Juliano Gomes da Silveira como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Engenharia de Software e Banco de Dados, aprovada em 14/03/2013 pela Comissão Examinadora:



Prof. Dr. Duncan Dubugras Alcoba Ruiz -
Orientador

PPGCC/PUCRS



Prof. Dr. Rafael Prikladnicki -

PPGCC/PUCRS



Profa. Dra. Ana Trindade Winck -

UFSM

Homologada em 08/08/2013, conforme Ata No. 013 pela Comissão Coordenadora.



p/ Prof. Dr. Paulo Henrique Lemelle Fernandes
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 - P32- sala 507 - CEP: 90619-900

Fone: (51) 3320-3611 - Fax (51) 3320-3621

E-mail: ppgcc@pucrs.br

www.pucrs.br/facin/pos

“A ciência se torna fascinante quando você não fica só na teoria”

Marcelo Gleiser

AGRADECIMENTOS

Agradeço à minha família e minha namorada Karine, por me apoiarem nesta etapa e por compreenderem quando não pude estar por perto.

Ao meu orientador, professor Duncan, pela oportunidade deste aprendizado, bem como por ter compreendido as minhas ausências durante o curso. Aos colegas do GPIN pela constante troca de conhecimentos e experiências.

Agradeço aos colegas de trabalho, por me apoiarem nesta qualificação, especialmente ao Giovanni e Débora por investirem no meu desenvolvimento profissional.

RSAPP, UM ALGORITMO BASEADO EM ROUGH SETS PARA AUXÍLIO AO PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS

RESUMO

As técnicas de *Business Intelligence* (BI) firmaram-se como grandes aliadas das organizações nas tarefas de transformar dados em conhecimento, apoiando a média e alta gestão na tomada de decisões. As ferramentas de BI em sua composição são fundadas em técnicas de gestão do conhecimento, tais como *Data Warehouse* (DW), OLAP (*Online Analytical Processing*), Mineração de Dados (MD), entre outras. Neste contexto, observa-se que em muitos casos, projetos de MD acabam sendo inviabilizados por alguns fatores, tais como, custo do projeto, duração e principalmente, a incerteza na obtenção de resultados que retornem o investimento despendido no projeto. O presente trabalho busca minimizar os fatores acima por meio um diagnóstico sobre os dados, através de um algoritmo baseado em *Rough Sets Theory* (Teoria dos Conjuntos Aproximados (TCA)). O algoritmo desenvolvido, nomeado *Rough Set App* (RSAPP) objetiva criar um diagnóstico sobre os dados persistidos no DW, a fim de mapear quais atributos possuem maior potencial de gerar modelos de mineração mais precisos e resultados mais interessantes. Desta forma, entende-se que o diagnóstico gerado por RSAPP pode complementar o processo de KDD (*Knowledge Discovery in Database*), reduzindo o tempo gasto nas atividades de entendimento e redução da dimensionalidade dos dados. No trabalho se faz uma descrição detalhada acerca do algoritmo implementado, bem como o relato dos testes

que foram executados. Ao final faz-se uma análise empírica sobre os resultados a fim de estimar a eficácia do algoritmo quanto a sua proposta.

Palavras Chave: Mineração de dados, Teoria dos Conjuntos Aproximados, *Business Intelligence*, *Data Warehouse*, *Knowledge Discovery in Database* – *KDD*.

RSAPP, A ROUGH SET BASED ALGORITHM FOR THE SUPPORT OF KNOWLEDGE DISCOVERY IN DATABASE PROCESS

ABSTRACT

Techniques of Business Intelligence (BI) became one of the main allies of organizations in tasks of transforming data into knowledge, supporting the middle and upper management levels in decision making. BI tools in their composition are based on techniques of knowledge management, such as Data Warehouse (DW), OLAP (Online Analytical Processing), Data Mining (DM), among others. In this context, it is observed that in many cases, DM projects become unfeasible by some factors, such as project costs, duration and specially the uncertainty in obtaining results that return the investment spent on the project. This work seeks to minimize these factors through a diagnosis on data, by an algorithm based on Rough Sets Theory. The algorithm, named Rough Set App (RSAPP) aims to create a diagnosis on data persisted in DW, in order to map which attributes have the greatest potential for generating more accurate mining models and more interesting results. Thus, it is expected that the diagnosis generated by RSAPP can complement the KDD (Knowledge Discovery in Database) process, reducing the time spent on activities of understanding and reducing data dimensionality. This work presents a detailed description about the implemented algorithm, as well as the report of the executed tests. At the end there is an empirical analysis of the results, in order to estimate the effectiveness of the proposed algorithm.

Keywords: Data Mining, Business Intelligence, Rough Sets Theory Data Warehouse, Knowledge Discovery in Database – KDD.

LISTA DE FIGURAS

Figura 1 – Comparação OLTP e DW (OLAP).....	25
Figura 2 – Tabela Fato e Dimensão	26
Figura 3 – Modelo Estrela	27
Figura 4 – Modelo Floco de Neve.....	28
Figura 5 – Granularidade	30
Figura 6 – Multidimensionalidade OLAP.....	33
Figura 7 – Processo de KDD	34
Figura 8 – Técnicas de Mineração	43
Figura 9 – Processo de Geração de Modelo Preditivo.....	44
Figura 10 – Árvore de Decisão	47
Figura 11 – Aproximação superior/inferior de TCA.....	53
Figura 12 – Fluxo do diagnóstico através de RSAPP.....	60
Figura 13 – Processo KDD usando RSAPP.....	62
Figura 14 – Modelo conceitual.....	63
Figura 15 – Diagrama de componentes de RSAPP.....	68
Figura 16 – Gráfico comparativo para redutos de tamanho 1	75
Figura 17 – Gráfico comparativo para redutos de tamanho 2	76
Figura 18 – Gráfico comparativo para redutos de tamanho 3	77
Figura 19 – Gráfico comparativo para redutos de tamanho 4	77
Figura 20 – Gráfico comparativo para redutos de tamanho 5	78
Figura 21 – Gráfico comparativo para redutos de tamanho 6	78
Figura 22 – Gráfico comparativo para redutos de tamanho 7	79
Figura 23 – Gráfico comparativo para redutos de tamanho 8	79

Figura 24 – Gráfico comparativo para redutos de tamanho 9	80
Figura 25 – Gráfico comparativo para redutos de tamanho 10.....	80
Figura 26 – Gráfico comparativo para redutos de tamanho 11.....	81
Figura 27 – Gráfico comparativo para redutos de tamanho 12.....	81
Figura 28 – Gráfico comparativo para redutos de tamanho 13.....	82
Figura 29 – Gráfico comparativo para redutos de tamanho 14.....	82
Figura 30 – Gráfico comparativo para redutos de tamanho 15.....	83
Figura 31 – Gráfico comparativo para redutos de tamanho 16.....	83
Figura 32 – Gráfico comparativo para redutos de tamanho 17.....	84
Figura 33 – Gráfico comparativo para redutos de tamanho 18.....	84

LISTA DE TABELAS

Tabela 1 – Binarização dos Dados. Adaptado de [TAN06].....	40
Tabela 2 – Matriz de Confusão. Adaptado de [WIT05].....	47
Tabela 3 – Medidas do Modelo. Adaptado de [WIT05].....	48
Tabela 4 – Exemplo TCA de [PAW02]	50
Tabela 5 – Quantidade de atributos.....	66
Tabela 6 – Estrutura hierárquica.....	67
Tabela 7 – Tabela <i>STEP_A</i>	70
Tabela 8 – Tabela de diagnóstico gerado por RSAPP (<i>STEP_B</i>)	71
Tabela 9 – Variação de acurácia entre os melhores e piores redutos	86

LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligense</i>
DM	<i>Data Mining</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extract Transform Load</i>
FN	Falso Negativo
FP	Falso Positivo
GPIN	Grupo de Pesquisa em Inteligência de Negócios
GUI	<i>Graphical User Interface</i>
KDD	<i>Knowledge Discovery in Database</i>
MD	Mineração de Dados
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online Transaction Processing</i>
PCA	Análise de Componentes Principais
PL/SQL	<i>Procedural Language/Structured Query Language</i>
PPGCC	Programa de Pós-Graduação em Ciência da Computação
RSAPP	<i>Rough Set App</i>
SGDB	Sistema de Gerenciamento de Banco de Dados
SQL	<i>Structured Query Language</i>
TCA	Teoria dos Conjuntos Aproximados
UML	<i>Unified Modeling Language</i>

VN Verdadeiro Negativo

VP Verdadeiro Positivo

SUMÁRIO

1-	Introdução.....	20
2-	Referencial Teórico.....	23
2.1	<i>Data Warehouse</i>	23
2.1.1	ETL (<i>Extract Transform Load</i>).....	28
2.1.2	Granularidade.....	29
2.1.3	Aplicações OLAP.....	30
2.2	<i>Knowledge Discovery in Database – KDD</i>	33
2.2.1	Seleção dos Dados.....	35
2.2.2	Pré-processamento e Transformação.....	35
2.2.3	Mineração de dados.....	41
2.2.4	Avaliação e Interpretação de Resultados.....	46
2.3	Teoria dos Conjuntos Aproximados.....	49
2.3.1	Exemplo de Aplicação TCA.....	50
2.3.2	Sistema de Informação e Aproximações da TCA.....	51
2.3.3	Precisão e Qualidade de Aproximação.....	53
2.4	Considerações do Capítulo.....	54
3-	Descrição do Cenário.....	56
3.1	Caracterização do Problema.....	56
3.2	Caracterização da Contribuição.....	57

3.3	Contexto de Implementação da Pesquisa.....	59
3.4	Considerações do Capítulo.....	61
4-	Implementação RSAPP.....	62
4.1	Cenário Real de Aplicação.....	63
4.2	Algoritmo RSAPP.....	67
4.2.1	<i>Lower App</i>	68
4.2.2	<i>Quality App</i>	70
4.3	Softwares Utilizados.....	72
4.4	Considerações do Capítulo.....	73
5-	Descrição dos Resultados.....	74
5.1	Diagnóstico RSAPP e a Acurácia de Mineração.....	74
5.1.1	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 2).....	76
5.1.2	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 3).....	77
5.1.3	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 4).....	77
5.1.4	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 5).....	78
5.1.5	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 6).....	78

5.1.6	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 7).....	79
5.1.7	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 8).....	79
5.1.8	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 9).....	80
5.1.9	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 10).....	80
5.1.10	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 11).....	81
5.1.11	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 12).....	81
5.1.12	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 13).....	82
5.1.13	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 14).....	82
5.1.14	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 15).....	83
5.1.15	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 16).....	83
5.1.16	Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 17).....	84

5.1.17 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 18).....	84
5.2 Diagnóstico RSAPP – Melhores e Piores Redutos e a Variação da Acurácia	85
5.3 Considerações do Capítulo.....	87
6- Considerações Finais e Trabalhos Futuros.....	88
Referências.....	90
Anexo A	95
Anexo B	120
Anexo C	157

1- INTRODUÇÃO

O crescimento do volume de dados aportados por sistemas de informação estabelece um cenário crítico para a captura e análise de informações estratégicas. Uma vez que o ambiente corporativo mostra-se cada vez mais competitivo, a gestão tática e estratégica necessita de subsídios de informação que os dê suporte às decisões executivas. Em muitos casos, os dados estão persistidos em estruturas de armazenamento heterogêneas, tais como bancos de dados relacionais, arquivos, documentos, etc. Este contexto levou a concepção de um tipo de banco de dados diferenciado, chamado *Data Warehouse* (DW). Esse banco de dados propõe a integração das diversas estruturas de dados da corporação em um ambiente de persistência único, orientado à recuperação da informação. Por deter essas características, possui alto performance no apoio à análise de grandes volumes de informações, de forma consolidada e integrada [INM05].

Sob o mesmo argumento da concepção dos DW, as aplicações OLAP (*Online Analytical Processing*) são ferramentas preparadas para a análise dos dados. Geralmente, esse tipo de aplicação conecta-se ao DW e suporta diversas funções de análise. Desta forma, o software OLAP fornece uma interface amigável, onde o usuário especialista do domínio pode manipular os seus recursos, assistindo a informação consolidada sob diversas perspectivas. Gráficos, tabelas pivô e relatórios analíticos são interfaces comuns desse tipo de aplicação. As tecnologias de DW e OLAP são amplamente difundidas, através de soluções de software e serviços. Normalmente, projetos de DW e OLAP são

complementares, pois aplicações OLAP preferencialmente usam DW como origem dos dados, aproveitando-se de sua organização orientada à recuperação de informação. Essas tecnologias fazem parte do processo conhecido como *Business Intelligence* (BI), que essencialmente refere-se à prática de capturar, organizar e analisar de informações para suporte à decisão [TUR08].

De fato, aplicações OLAP sobre estruturas de DW têm apoiado as organizações, subsidiando informação estratégica. No entanto, em aplicações OLAP a percepção do especialista de domínio prende-se às informações resultantes da manipulação dos dados, nos seus diferentes níveis de agregação. Neste cenário, o processo KDD (*Knowledge Discovery in Database*) vai um passo adiante, transcendendo essas possibilidades [HAN11]. Tipicamente, a descoberta de conhecimento sobre bancos de dados é feita por Mineração de Dados (*Data Mining*). A Mineração de Dados (MD) é um processo exploratório interdisciplinar, que usa técnicas de aprendizagem de máquina, banco de dados e modelos estatísticos. Basicamente, MD propõe descobrir conhecimento que está implícito nos dados, evidenciando padrões e associações.

Organizações de diferentes ramos de atividade têm buscado em KDD informações que agreguem valor competitivo ao negócio. No entanto, a longa duração dos projetos, alto investimento e principalmente a incerteza na obtenção dos resultados são fatores que em muitas vezes inviabilizam projetos desse tipo [COL11]. Conforme [HAN11], os sistemas de mineração de dados tem potencial de gerar milhares ou até milhões de padrões ou regras. No entanto, nem todos os padrões são considerados interessantes. O autor define algumas características para que os padrões e regras sejam considerados interessantes. São elas: (1) facilmente compreensível por humanos – (2) válido para novos

dados com algum grau de certeza – (3) potencialmente útil – e (4) novo. Em suma, um padrão interessante representa conhecimento novo para os especialistas do domínio estudado.

Inserido neste contexto, este trabalho concentra-se na elaboração de um algoritmo para auxílio à MD sobre dados de DW, dando suporte no momento inicial de pré-processamento dos dados, quando o conhecimento de domínio ainda é superficial para seleção eficaz dos dados. O algoritmo aqui proposto, chamado *Rough Set App* (RSAPP) faz uso da chamada Teoria dos Conjuntos Aproximados – TCA (*Rough Sets Theory*) para criar um mecanismo semiautomático de diagnóstico dos dados. Com base nesse diagnóstico, tem-se a intenção de focar as tarefas de mineração sobre os dados com maiores chances de produzir padrões mais precisos e interessantes para a área de domínio, logo nas etapas iniciais do projeto de KDD. Tal técnica pode ser utilizada em uma prova de conceito ou anteprojeto, na tentativa de reduzir a dúvida de obtenção dos resultados significativos, podendo-se assim inferir com mais propriedade o retorno sobre o investimento do projeto de KDD.

Esta pesquisa dá continuidade ao trabalho desenvolvido por [COL11], que por sua vez propôs uma técnica de indução e ranqueamento de árvores de decisão sobre modelos OLAP. O trabalho aqui desenvolvido e a pesquisa de [COL11] são frutos dos estudos do GPIN-PPGCC (Grupo de Pesquisa em Inteligência de Negócios do Programa de Pós-Graduação em Ciência da Computação da PUCRS) e convergem na intenção de criar mecanismos de avaliação prévia dos dados, que possibilitem a indicação de tarefas de mineração mais assertivas em projetos de KDD sobre DW.

2- REFERENCIAL TEÓRICO

O presente capítulo faz uma breve revisão da bibliografia acerca dos principais temas relacionados à pesquisa. Primeiramente discorre-se sobre os bancos de dados *Data Warehouse* e sua modelagem dimensional. Na segunda seção faz-se um apanhado sobre os conceitos que envolvem KDD, detalhando cada etapa deste processo. Por fim, examina-se a Teoria dos Conjuntos Aproximados (*Rough Sets Theory*) relacionando as suas definições e expondo um exemplo de seu cálculo sobre um conjunto de dados sucinto, encontrado na bibliografia.

2.1 *Data Warehouse*

Em meados da década de oitenta surgem os primeiros DW para subsidiar a tomada de decisão sobre grandes volumes de dados. Subsídios esses, que os bancos de dados existentes na época não atendiam com eficiência. Desde sua concepção, o modelo usado em DW, chamado de Modelo Dimensional [KIM02] prioriza a recuperação de informação, diferentemente do modelo de padrão relacional (normalizado), que é orientado ao processamento de transações (*Online Transaction Processing – OLTP*). Neste cenário, os DW consolidaram-se como técnica de armazenamento de informações analíticas e atualmente integra o processo de *Business Intelligence* (BI).

Ralph Kimball [KIM02] e William Inmon [INM05] são importantes pesquisadores sobre o tema, suas pesquisas e métodos são usados como base da

arquitetura de DW. A abordagem de [INM05] baseia-se essencialmente na criação de um banco de dados corporativo, usando um modelo dimensional integrado e singular. Desta forma, montando uma fonte de informações apropriada com granularidade satisfatória, para suporte as decisões estratégicas. Em linhas gerais, [INM05] caracteriza DW como “uma coleção de dados orientada por assunto, integrada, não volátil, variante em relação ao tempo, que dá apoio às decisões gerenciais”. É importante fragmentar este conceito e aprofundá-lo para compreender melhor o seu significado:

- Orientado por assuntos: estruturado conforme a perspectiva que se quer analisar, ou seja, é organizado em assuntos de negócio. Por exemplo, assuntos como estoque e vendas são comuns na concepção de DW.
- Integrado: com frequência os dados estão distribuídos em diversas plataformas, tais como arquivos, bancos de dados relacionais, entre outros. Segundo o autor, faz parte das atribuições do DW organizar e integrar esses dados, independentemente de sua origem, visando obter uma visão única das informações. Para tanto, faz-se necessário o uso do processo chamado ETL (*Extract Transform Load*). Na subseção 2.1.1 faz-se uma breve descrição sobre o processo de ETL.
- Não volátil: diferente dos sistemas de bancos de dados orientados às aplicações OLTP, em DW os dados não sofrem atualização ou exclusão pelas aplicações. Neste caso, as atualizações dos sistemas origem geram novas entradas de dados e as informações desatualizadas são mantidas em histórico. A Figura 1, baseada

em [INM05] ilustra a diferença entre DW, que é dirigido às aplicações OLAP e bancos de dados orientados às aplicações OLTP.

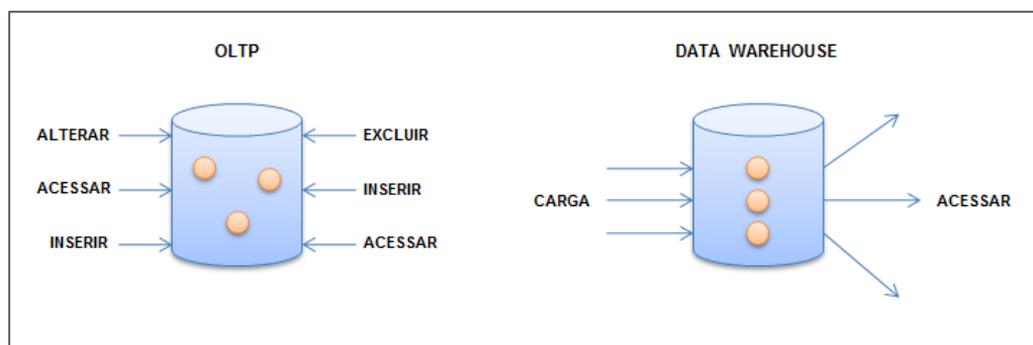


Figura 1 – Comparação OLTP e DW (OLAP)

Adptado de [INM05]

- Variante no tempo: os dados persistidos em DW devem possuir um horizonte temporal significativamente maior que os dados armazenados em bancos relacionais (OLTP). Assim, por padrão o DW armazena o histórico temporal dos dados. Essa característica torna possível que as informações sejam analisadas sob o prisma temporal, por exemplo, hora, dia, mês, ano, etc.

[KIM02] define uma abordagem ligeiramente diferente de [INM05]. Entretanto, ambos convergem na maioria das suas definições. Para [KIM02], o DW deve ser concebido de forma fragmentada sobre estruturas conceituais chamadas *Data Marts*. Os *Data Marts* essencialmente remetem ao conceito de assuntos, sendo cada *Data Mart* a transcrição de um assunto de negócio específico. Em suma, na abordagem de [KIM02] o DW é formado pela união de projetos de *Data Marts*. Neste contexto o autor argumenta que projetos de DW

são evolutivos, ou seja, novos assuntos podem ser acrescentados ao DW ao longo do tempo, desde que este esteja preparado para recebê-los.

Na modelagem dimensional proposta por [KIM02], o banco de dados é composto por dois tipos de entidades (tabelas), são elas: tabelas *Fato* e tabelas *Dimensões*. As tabelas dimensionais comportam as informações qualitativas dos dados, já as tabelas de fatos armazenam as informações quantitativas. Por exemplo, no caso de venda de produtos, a quantidade e valor vendidos permanecem computados na tabela fato e as informações descritivas do produto (nome, categoria...) persistem na tabela dimensional. Na Figura 2, adaptada de [KIM02] estão ilustradas típicas tabelas *Fato* (esquerda) e *Dimensão* (direita).

F FATO_VENDAS		
	QUANTIDADE	NUMBER
	VALOR	NUMBER (10,2)
F *	DIM_PRODUTO_PRODUTO	NUMBER
F *	DIM_TEMPO_TEMPO	NUMBER
F *	DIM_LOJA_LOJA	NUMBER

D DIM_PRODUTO		
P *	PRODUTO	NUMBER
	DESCRICAO	VARCHAR2 (50)
	MARCA	VARCHAR2 (50)
	CATEGORIA	VARCHAR2 (50)
	DEPARTAMENTO	VARCHAR2 (50)
	TAMANHO	VARCHAR2 (50)
	EMBALAGEM	VARCHAR2 (50)
*	PESO	NUMBER (10,4)
 PRODUTO_PK (PRODUTO)		

Figura 2 – Tabela Fato e Dimensão

Adaptado de [KIM02]

Para [KIM02] a modelagem dimensional permite duas abordagens distintas, são elas: *Star Schema* (modelo estrela) e *Snowflake* (modelo floco de neve). Tecnicamente essas abordagens definem como as tabelas se relacionam, ou seja, fornecem um desenho bem definido para as entidades de cada *Data Mart*. O Modelo Estrela, conforme pode ser visto na Figura 3 é simplificado, e ao contrário da abordagem relacional (OLTP), não exige normalização.

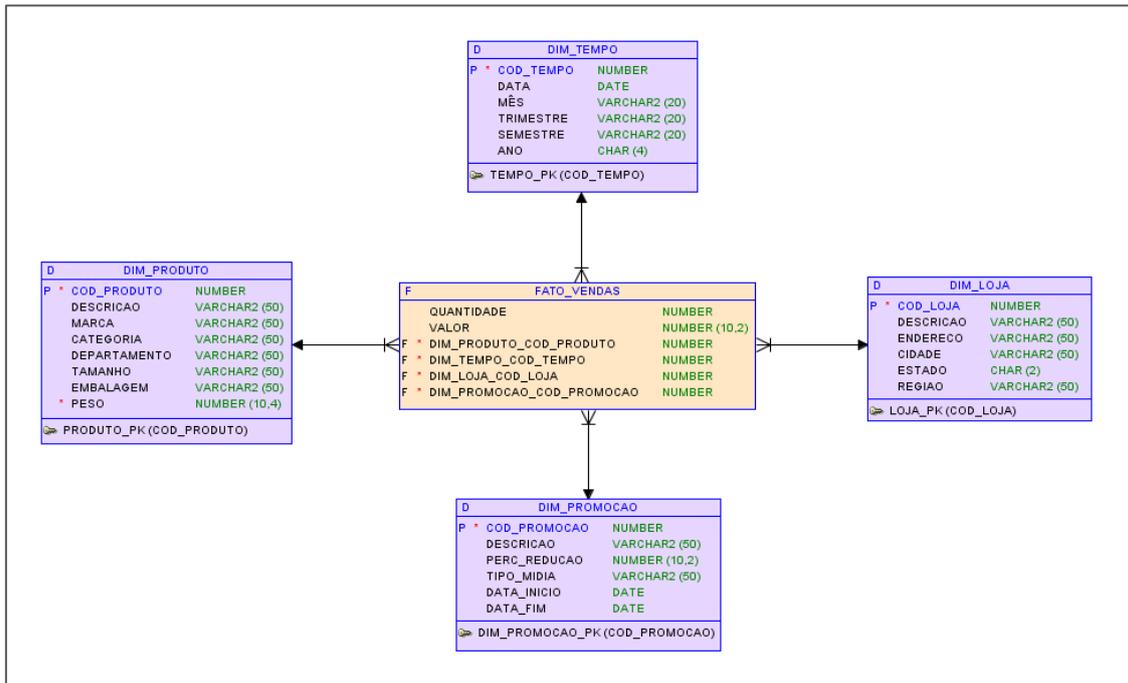


Figura 3 – Modelo Estrela

Adaptado de [KIM02]

A abordagem do Modelo Floco de Neve é um pouco mais complexa, permitindo alguma normalização das dimensões e relacionamentos entre elas, conforme pode ser visualizado na Figura 4.

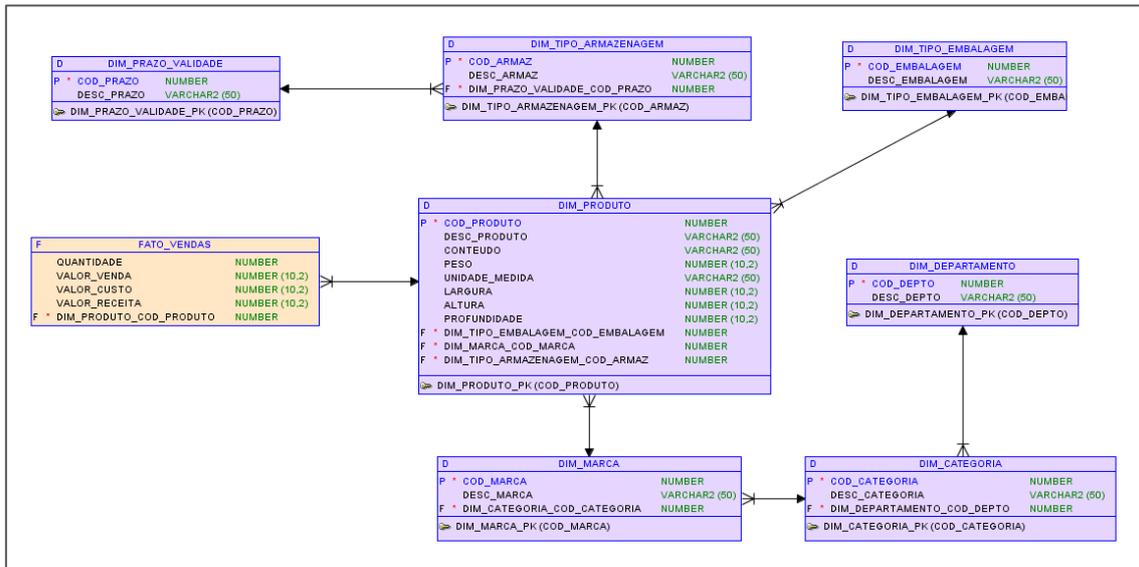


Figura 4 – Modelo Floco de Neve

Adaptado de [KIM02]

Nas subseções a seguir faz-se um apanhado de conteúdos relacionados ao tema dessa seção. Discorre-se sobre o processo de ETL e suas motivações, sobre granularidade dos dados e por fim, uma sucinta descrição das aplicações OLAP e a relação sua estreita com DW.

2.1.1 ETL (*Extract Transform Load*)

Segundo [TUR08] o processo de ETL basicamente consiste em (1) extração (leitura das informações dos bancos de dados de origem) – (2) transformação (eventuais tratamentos, junções e conversões necessárias) e (3) carga (persistência da informação no DW). Desta forma, a principal finalidade de ETL é garantir a integridade e limpeza dos dados do DW, que podem provir de fontes heterogêneas. Para o autor, “Processos de ETL mal projetados apresentam manutenção, alteração e atualização dispendiosas”.

Em sua obra, [BAR01] destaca que as ferramentas de ETL são fundamentais no projeto de DW. Para esse autor, tais ferramentas são encarregadas de transformar os dados do formato operacional (OLTP) em dados “informativos”, ou seja, organizados dimensionalmente.

2.1.2 Granularidade

A granularidade dos dados é o atributo que caracteriza o nível de detalhe ou resumo das informações de DW. Nesse contexto, quanto mais detalhada estiver persistida a informação diz-se que a granulação é menor, do contrário, quando mais resumida, maior a granulação. Segundo [INM05] a granularidade é aspecto muito importante na concepção de projetos de DW, pois tem um reflexo direto no custo do processamento das consultas, bem como o volume de dados armazenados. Sendo assim, a escolha da granularidade deve levar em conta o custo e também a necessidade do negócio envolvido, ou seja, qual o nível de detalhe que se deseja observar a informação. Desta forma, evitando desperdícios de recursos computacionais e o descontentamento do cliente.

Para [KIM02] a escolha da granulação dos dados é um ponto determinante para o sucesso do projeto de DW. O autor afirma que embora primordial, essa escolha não é uma definição trivial para equipe técnica do projeto. No texto de [INM05] destaca-se o exemplo ilustrado na Figura 5, abaixo.

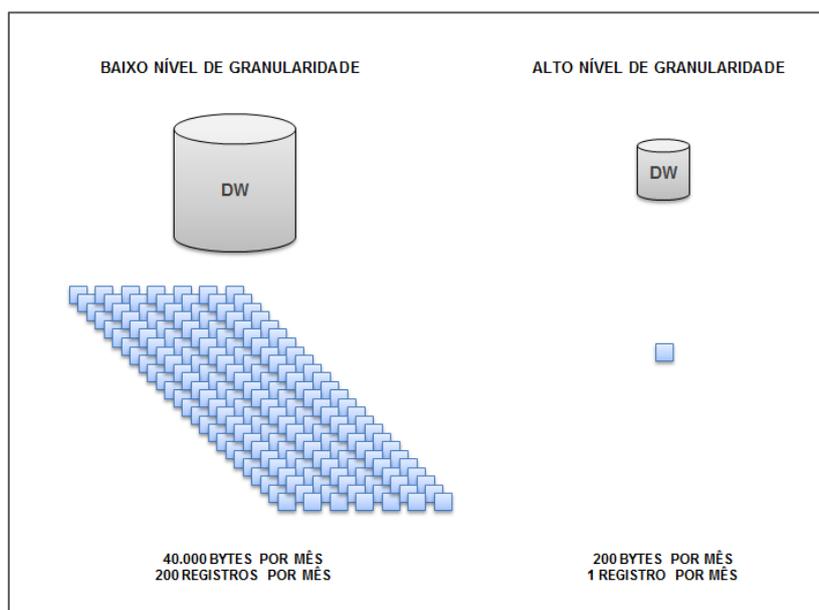


Figura 5 – Granularidade

Adaptado de [INM05]

No exemplo acima se tem a comparação de registros de ligações telefônicas de um cliente em diferentes níveis e granulação. À esquerda da figura os dados estão detalhados por registros individuais de ligações no período de um mês (200 ligações), somando 40.000 bytes, representando baixo nível de granularidade. A direita da figura tem-se os mesmos registros (200), porém de forma sumarizada, abstraindo os detalhes individuais de cada ligação.

2.1.3 Aplicações OLAP

As habituais operações de sistemas de informação, tais como, criar, modificar e recuperar informações no detalhe elementar são baseadas em OLTP. Por exemplo, uma operação de saque em caixa eletrônico, ou ainda, uma transação individual de venda são típicas funções deste tipo de aplicação. Tais

operações são eficientes no contexto operacional, porém deixam a desejar quando a necessidade é recuperar e apresentar informações de apoio à decisão estratégica sobre grandes volumes de dados. Neste cenário que as aplicações OLAP protagonizam [TUR08].

Uma consulta de OLAP pode acessar um banco de dados com gigabytes ou terabytes de informações sobre vários anos de vendas a fim de encontrar todas as vendas de produtos em cada região para cada tipo de produto. Após revisar os resultados, um analista pode refinar ainda mais a pesquisa com o objetivo de encontrar o volume para cada canal de vendas dentro de uma região, ou dentro de determinadas classificações do produto. Uma última etapa pode envolver o analista executando comparações entre anos ou trimestres para cada canal de vendas. O processo todo pode ser executado online, com um tempo rápido de resposta para que o processo de análise ocorra sem interferências. [TUR08], página 109.

Para [DAT99] OLAP caracteriza-se pelo processo dinâmico de criar, gerenciar e produzir relatórios analíticos sobre os dados da organização. Dados estes que podem ser oriundos de DW ou cubos multidimensionais. Segundo este autor, uma vez que tais ferramentas permitem que o próprio usuário possa customizar de forma ágil as suas consultas, o subsídio de informação se torna dirigido. Desta forma é comum a busca por respostas de questões de negócio, padrões e tendências indicadas por medidas de desempenho.

As ferramentas OLAP são baseadas na arquitetura de dados multidimensional. Comumente essas aplicações são nomeadas como “Cubos de Dados” (uma referência à dimensionalidade de um cubo). Os cubos multidimensionais são assim chamados por permitir a análise de fatos nas suas múltiplas perspectivas (dimensões). A análise das informações sobre cubos são conduzidas pelas funções chamadas *roll-up*, *drill-down*, *slicing* e *dicing*. As

funções *roll-up* e *drill-down* remetem ao conceito de “hierarquias” dimensionais. Essas hierarquias são mapeadas no cubo para que a informação possa ser explorada dinamicamente, navegando entre os diferentes níveis de detalhe. Neste caso, a função *drill-down* aprofunda a análise de uma determinada medida (fato) para um nível mais detalhado e a função *roll-up* retorna ao nível resumido da dimensão analisada. Em ambos os casos, quando essas funções são aplicadas é papel do cubo agregar ou detalhar a informação nas categorias da dimensão. Por exemplo, é possível usar uma hierarquia da dimensão de tempo para observar a informação de vendas por trimestre e dinamicamente, através da função *drill-down* chegar ao detalhe de mês. Neste caso, obviamente a aplicação distribui o valor de venda de cada trimestre nos seus respectivos meses de venda. De forma análoga, um novo estímulo de detalhamento sobre o mês poderia detalhar as vendas por dia. Tecnicamente essas funções quando aplicadas sobre o DW são representadas através de consultas que utilizam a cláusula *GROUP BY*. As operações *slicing* e *dicing* são usadas para selecionar uma parte específica do cubo, filtrando assim os dados de uma ou mais dimensões em algum nível hierárquico específico, neste caso, obtem-se um subcubo de dados parciais [HAN11]. A Figura 6 abaixo, baseada em [HAN11] ilustra a multidimensionalidade conceitual de OLAP bem como a sua representação hierárquica.

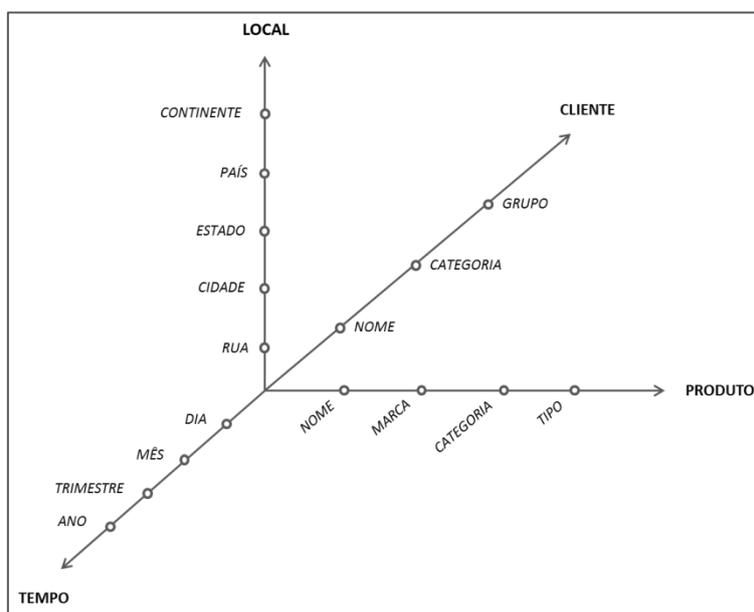


Figura 6 – Multidimensionalidade OLAP

Adaptado de [HAN11]

Na representação acima o cubo OLAP é composto por quatro dimensões, sendo (1) tempo – (2) localidade – (3) cliente e (4) produto. Desta forma, o fato (vendas) pode ser visualizado através do prisma dessas quatro dimensões e detalhado em seus níveis de hierarquia, usando as funções *roll-up/drill-down*.

2.2 Knowledge Discovery in Database – KDD

Organizações dos mais diversos nichos procuram manter suas informações armazenadas em DW a fim de posteriormente utilizar tais recursos como subsídio para decisões estratégicas. Neste sentido, o processo de KDD é utilizado na descoberta automática de conhecimento sobre grandes volumes de dados, uma vez que capacidade de análise manual se torna inviável. Por exemplo, utilizando um banco de dados de uma empresa de bens de consumo,

através de KDD é possível traçar correlações de vendas de certos grupos de clientes e então usar esse conhecimento para introduzir propagandas e novas ofertas direcionadas [FAY96].

O valor do volume de armazenamento de dados depende da nossa capacidade de extrair relatórios úteis, detectar eventos interessantes e tendências, decisões estratégicas baseadas na análise estatística e inferência, e explorar os dados como diferencial competitivo de negócios, operacionais ou objetivos científicos. [FAY96], página 28.

Basicamente KDD é composto por uma sequência bem definida de etapas que ao final buscam encontrar, automaticamente, informações úteis em grandes bases de dados, estabelecendo padrões, associações e previsões, que por sua vez passariam despercebidas sob a análise manual. As etapas que compõem o processo do KDD são: (1) seleção dos dados – (2) pré-processamento – (3) transformação – (4) mineração de dados e (5) avaliação e interpretação dos resultados. A Figura 7 de [FAY96] visualizada logo abaixo demonstra a visão geral do processo de KDD.

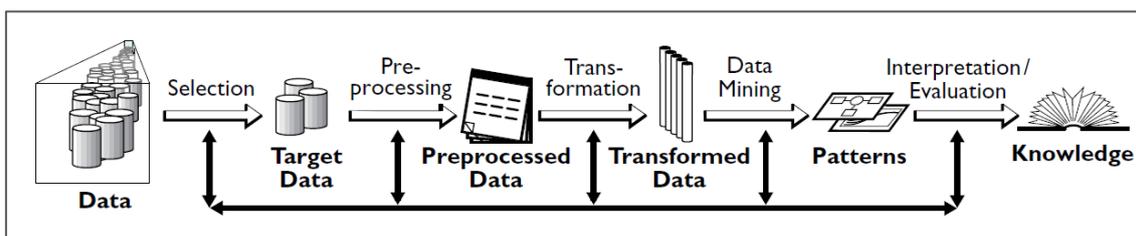


Figura 7 – Processo de KDD

Fonte [FAY96]

Cabe estudar cada etapa do processo separadamente, em especial as referências de pré-processamento e mineração. Portanto, ao longo das próximas

subseções (2.2.1, 2.2.2, 2.2.3 e 2.2.4) faz-se um apanhado acerca dos estágios do processo de KDD.

2.2.1 Seleção dos Dados

Nesta etapa faz-se o levantamento de quais dados serão usados no processo de KDD. Para tanto é necessário estudar as necessidades de negócio e quais dados podem estar relacionados com essas questões. Nesse contexto, para [FAY96] é necessário que o domínio seja estudado, a fim de otimizar a seleção das informações.

Segundo [HAN11] as informações podem ser oriundas de plataformas diferentes, tais como bancos de dados relacionais, arquivos de dados, entre outras fontes. Para este autor os DW estão intimamente relacionados com o processo de KDD. Segundo ele o uso de DW é conveniente, pois se faz uso da sua prévia preparação dos dados, bem como a sua orientação para consultas e modelagem dimensional. Neste contexto, os dados de KDD são mantidos no DW, selecionados através do processo de ETL (subseção 2.1.1).

2.2.2 Pré-processamento e Transformação

Segundo [HAN11] bancos de dados do “mundo real” são altamente suscetíveis a ruídos, valores ausentes e inconsistências de dados. Para o autor, a baixa qualidade dos dados implica diretamente nos resultados de mineração. Embora extraídos pelo processo de ETL e persistidos em DW (abordagem de [HAN11]), comumente os dados precisam ser pré-processados, eventualmente transformados e reduzidos para a execução das tarefas de mineração. Esta etapa

responde por grande parte do esforço e tempo gasto em KDD. Mesmo que o investimento na preparação dos dados seja significativo, esse trabalho tem reflexo direto sobre a confiabilidade dos resultados obtidos. [TAN06] sugere algumas etapas de preparação para a mineração, que podem ser empregadas ou não, dependendo do problema. A seguir são detalhadas essas etapas.

2.2.2.1 Agregação

Frequentemente os dados estão disponíveis em um nível granular muito baixo e dependendo do caso é conveniente aumentar esta granularidade, processo semelhante ao empregado em ferramentas OLAP.

Quando as informações estão dispostas no nível elementar, tem-se uma visão mais operacional desses dados. Portanto, a fim de se obter resultados de menor nível de detalhe é possível usar a agregação. Ainda, em condições normais o nível granular é proporcional ao volume de objetos a ser processado. Portanto, em casos quando a tarefa de mineração demanda alto custo de processamento, o emprego de agregação dos dados pode reduzir esse custo computacional.

Tecnicamente atributos quantitativos são agregados aplicando-se funções de soma ou média. Os atributos qualitativos podem ser simplesmente suprimidos ou podem ser substituídos por informações de maior nível granular na hierarquia física. A agregação deve ser aplicada com cautela, visto que podem ser perdidos detalhes interessantes dos dados. Por exemplo, no caso de uma agregação de vendas por mês abstaindo os detalhes de dia pode-se ignorar algum padrão relacionado às sazonalidades dos dias.

2.2.2.2 Amostragem

A amostragem é uma técnica estatística para seleção de um subconjunto dos dados a serem analisados. No caso da disciplina de estatística, freqüentemente não é possível trabalhar com a totalidade dos dados por não ser possível coletar as informações de toda a população. Neste caso então, busca-se um conjunto parcial dos dados chamado de “amostra representativa” tanto para os experimentos preliminares quanto para as análises finais. A amostragem também se mostra útil na mineração, no entanto, diferente da motivação estatística, em mineração de dados o estímulo é reduzir a complexidade computacional, visto que, em alguns casos o processamento de determinado algoritmo de mineração sobre um grande conjunto de dados pode ser impraticável.

2.2.2.3 Redução de Dimensionalidade

A dimensionalidade representa as diferentes informações (características) que se tem sobre um mesmo objeto. Conforme a dimensionalidade do conjunto cresce, normalmente os dados ficam mais dispersos e conseqüentemente é mais difícil de se obter padrões precisos. A redução de dimensionalidade busca eliminar as características irrelevantes ou combinar duas ou mais dimensões para execução da tarefa de mineração, visto que tal redução pode levar a resultados mais compreensíveis. Além disso, quanto maior a quantidade de dimensões do conjunto de dados, maior o esforço computacional do algoritmo minerador. Desta forma, reduzem-se as dimensões com pouca ou nenhuma significância para a mineração.

Diferenciar quais dimensões tem maior potencial de gerar padrões classificáveis com maior precisão é uma tarefa que exige análise sobre os dados. Em muitos casos, o técnico que está executando as tarefas de mineração não possui pleno conhecimento sobre o assunto dos dados. Por esse motivo, com frequência utilizam-se técnicas de redução. Algumas técnicas conhecidas utilizam álgebra linear como base para suprimir dimensões, como por exemplo, a Análise de Componentes Principais (PCA).

2.2.2.4 Seleção de Subconjuntos

Intuitivamente entende-se que aplicar mineração apenas sobre um subconjunto dos dados perde-se informações e conseqüentemente os modelos gerados podem não ser válidos. No entanto, quando os dados do conjunto possuem características redundantes ou irrelevantes é possível selecionar apenas um subconjunto, sem perda de informação relevante. Dados redundantes são aqueles que possuem pouca ou nenhuma variação, como por exemplo, uma taxa de cálculo de imposto, que em alguns casos não varia. As informações irrelevantes são aquelas que não produzem conhecimento ao modelo, como por exemplo, uma chave primária sequencial de um conjunto de dados. As redundâncias e irrelevâncias além de não agregarem conhecimento, podem abreviar a qualidade dos grupos gerados.

Quando há conhecimento de domínio torna-se mais fácil suprimir redundâncias e irrelevâncias. No entanto, em alguns casos o conhecimento de domínio não é suficiente para realizar esse trabalho, assim uma abordagem sistemática pode ser usada. Uma abordagem poderia ser executar experimentos de mineração sobre todas as possibilidades de subconjuntos, e então selecionar

os subconjuntos que ofereçam os melhores resultados. Obviamente essa abordagem não é viável, pois sendo n atributos, haverá 2^n possibilidades. Neste caso, existem três técnicas para seleção de subconjuntos, são elas: (1) interna – (2) filtro e (3) envoltório.

- Interna: nessa abordagem atribui-se a seleção de subconjunto ao algoritmo de mineração. Neste caso não há pré-processamento, pois se deixa que o próprio algoritmo decida quais os atributos deve ou não usar para geração do modelo.
- Filtro: neste caso faz-se uma seleção antes do processamento do algoritmo, baseada em um critério que o técnico ou especialista de domínio podem definir, como por exemplo, uma seleção baseada em correlação.
- Envoltório: a seleção feita nessa abordagem baseia-se na aplicação de um algoritmo de mineração exclusivamente para seleção dos atributos a serem trabalhados. Neste caso a execução do algoritmo que irá indicar os melhores atributos para execução dos próximos experimentos de mineração.

2.2.2.5 Criação de Recursos

Em algumas condições é possível construir novos atributos baseados em combinações. Essa técnica busca extrair as informações mais importantes através da junção, além do ganho provido pela redução de dimensionalidade. A criação de recursos baseia-se em três métodos: (1) extração de características – (2) mapeamento dos dados para novo espaço e (3) construção de recursos.

2.2.2.6 Discretização e Binarização

Determinados algoritmos de mineração exigem que as informações sejam discretizadas ou binarizadas. Especialmente em tarefas de classificação o atributo classificador deve estar no formato categórico. De forma similar, para execução de tarefas de padrões de associação as informações devem possuir formato binário.

Frequentemente a categorização não é uma tarefa trivial, pois os atributos classificadores podem estar contidos em uma grande escala de valores. Assim, valores que possuem pouca frequência teriam sua própria descrição. Neste caso, convenientemente a escala de valor pode ser fragmentada em categorias que englobem um intervalo da escala e não todas as variações. Por exemplo, valores contínuos de 0 a 100 podem ser enquadrados em categorias de pequeno, médio e grande. Na binarização a tarefa é transformar os atributos contínuos em atributos booleanos. Neste contexto, verifica-se a escala do atributo que se quer binarizar e depois se cria novos atributos baseado nessa escala, transpondo cada valor no formato de coluna, neste caso, se o objeto se enquadra naquele atributo, então é verdadeiro, senão falso. A tabela abaixo, de [TAN06] exemplifica o processo de binarização.

Tabela 1 – Binarização dos Dados. Adaptado de [TAN06]

Valor Categorizado	Valor Inteiro	x_1	x_2	x_3	x_4	x_5
Terrível	0	1	0	0	0	0
Fraco	1	0	1	0	0	0
Satisfatório	2	0	0	1	0	0
Bom	3	0	0	0	1	0
Excelente	4	0	0	0	0	1

No exemplo acima, de [TAN06] o atribuo de valor contínuo foi convertido em cinco novas propriedades, sendo ($0 \rightarrow x_1, \dots, 4 \rightarrow x_5$), onde cada objeto foi identificado por 0 ou 1 em cada nova propriedade. Existem diversas técnicas disponíveis para auxílio de discretização, pois uma vez que os atributos contínuos são aleatórios as categorizações devem manter a semântica dos dados. Na bibliografia da disciplina de análise multivariada são propostas diversas formas de discretização de dados.

2.2.2.7 Transformação de Variáveis

A transformação de variáveis (atributos) baseia-se na aplicação de um cálculo sobre o valor elementar a fim de obter outro valor de referência. Por exemplo, a conversão de uma unidade de medida, ou ainda, uma função de transformação de número absoluto para um número de referência. Neste caso é possível também combinar diferentes variáveis de um objeto para inferir outra variável. Por exemplo, tendo as informações das coordenadas geográficas de um objeto, pode-se transpor essas informações em categorias de localidade, como cidade, CEP, etc.

2.2.3 Mineração de dados

Embora muitas vezes tratada individualmente, a mineração de dados compõe o processo de KDD [FAY96]. Esse processo combina técnicas de banco de dados, estatística e aprendizagem de máquina [TAN06]. Em [HAN11] o autor define mineração de dados como o tema essencial de KDD, onde algoritmos inteligentes são aplicados sobre os dados a fim de extrair novos padrões. Segundo [TAN06] mineração de dados é um processo de descoberta automática

de conhecimento útil sobre grandes conjuntos de dados. Para o autor, os algoritmos de mineração são construídos com a intenção de encontrar padrões não evidentes nos dados e prever informações futuras, conhecimento esse que poderia não ser detectado sob análise convencional.

Segundo [TUR08], mineração de dados é um domínio de análise de informações, fundamentado em banco de dados que podem prever comportamentos futuros. Por exemplo, organizações de varejo podem usar os algoritmos de mineração para mapear classes de clientes com interesses comuns. Para o autor, a mineração pode viabilizar a descoberta de novas e importantes correlações, padrões e tendências através de métodos estatísticos e de identificação de padrões.

Algoritmos de mineração podem ser classificados em dois tipos: descritivos e preditivos. Neste aspecto, as tarefas de mineração com abordagem **descritiva** propõem explorar propriedades de características que estão implícitas nos dados, para tanto, o modelo é orientado a um **atributo alvo** e explicado através dos demais atributos (**condicionais**). Os algoritmos preditivos concentram-se em prever informações de um dado atributo a partir dos valores dos demais atributos disponíveis, sendo chamada de **atributo alvo** a informação a ser prevista e **variáveis independentes** ou **explicativas** aquelas usadas para operar a previsão. As abordagens de mineração podem ser divididas em subcategorias, sendo **classificação**, **regressão** e **detecção de anomalias** as preditivas e **associação** e **agrupamento** as descritivas. A Figura 8 mostra as diferentes técnicas de mineração. Nas próximas subseções faz-se uma pequena explanação de cada técnica segundo [TAN06].

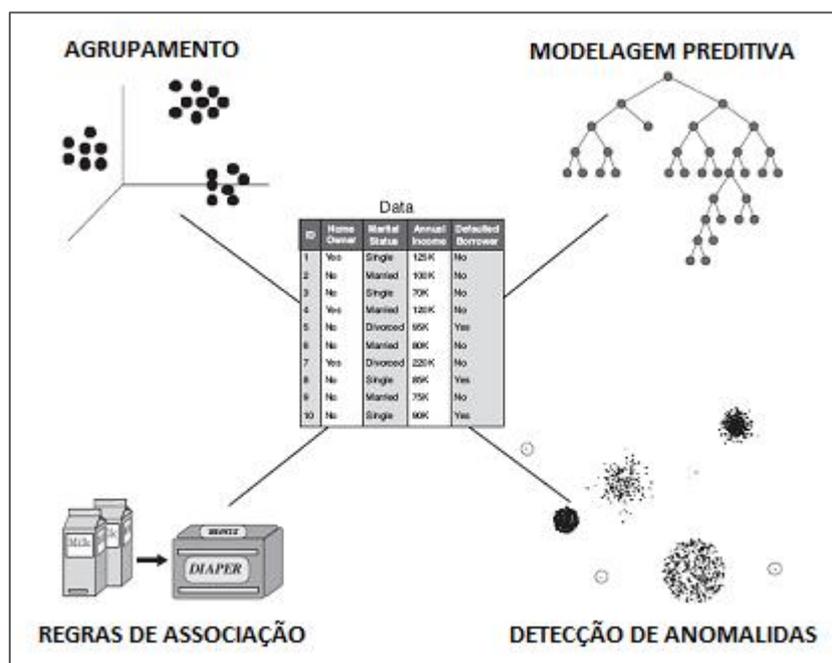


Figura 8 – Técnicas de Mineração

Fonte [TAN06]

2.2.3.1 Classificação e Regressão

Basicamente a tanto a classificação como a regressão buscam criar um modelo a partir do atributo alvo em relação aos demais atributos do conjunto. Nesse caso tem-se um conjunto de dados de amostra ou população e o objetivo é prever valores de alvo (classe) futuros ou simplesmente mapear as características que classificam esse atributo alvo. A classificação é usada para execuções sobre atributos alvo categóricos e a regressão para atributos alvo do tipo contínuo. Comumente os modelos baseados em classificação são representados por grafos chamados árvores de decisão, que mapeiam condições para enquadrar os objetos em uma categoria específica do atributo alvo. Por exemplo, prever se um usuário de uma livraria online fará ou não uma compra. Neste exemplo o atributo classe é a compra, que neste caso possui dois domínios

(sim ou não), as demais informações como assunto pesquisado, tempo de pesquisa, cidade, entre outros, são informações que irão classificá-lo no modelo, identificando condições para enquadrá-lo em um potencial comprador ou não.

Na classificação e regressão, para gerar o modelo o algoritmo primeiramente executa a etapa conhecida como **indução** sobre um conjunto de **treino**, onde os valores do atributo classe são conhecidos. Logo após, esse modelo passa pelo processo de **dedução**, onde é submetido a uma fração do conjunto de dados onde os valores do atributo classe não são conhecidos, conjunto esse chamado **teste**. A etapa de teste verifica o quanto o modelo conseguiu classificar os dados segundo a propriedade alvo. A imagem abaixo (Figura 9) exemplifica o processo de um classificador típico.

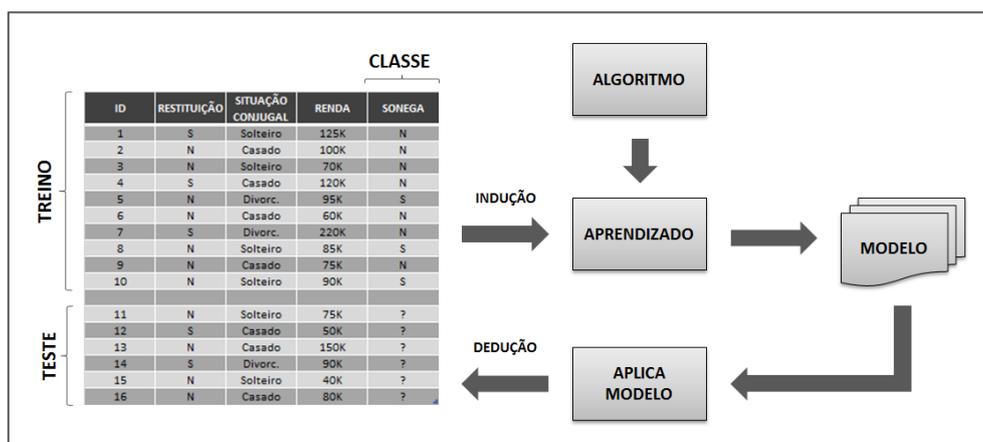


Figura 9 – Processo de Geração de Modelo Preditivo

Adaptado de [TAN06]

2.2.3.2 Detecção de Anomalias

A detecção de anomalias busca objetos que apresentam características de exceção perante o conjunto dos demais objetos. Neste contexto, aponta as informações que representam alguma anormalidade nos dados, ou seja, que

estejam inconsistentes com os padrões dos demais objetos do conjunto. Uma aplicação comum deste tipo de mineração é a detecção de fraudes, por exemplo, a fraude de compra com cartão de crédito. Nesse exemplo, um modelo é criado a partir do padrão de consumo do portador do cartão, assim cada compra é comparada com esse modelo e eventualmente, caso identificar um comportamento fora do padrão tal transação pode não ser autorizada.

2.2.3.3 Associação

Esse tipo de algoritmo examina associações entre objetos, associações essas que não seriam detectadas empiricamente. Os relacionamentos entre esses objetos são representados na forma de **regras de associação**. Neste contexto o algoritmo examina os valores de cada atributo e identifica os padrões mais frequentes. Em muitos casos esses padrões podem ser óbvios, porém o objetivo das regras de associação é detectar dentre esses padrões, os mais interessantes e que intuitivamente não seriam identificados. Um exemplo comum de aplicação desse tipo de algoritmo é a associação de itens de cesta de supermercado. Nesse caso, itens que não possuem relação direta podem estar associados pela frequência que estão presentes nas cestas.

2.2.3.4 Agrupamento

Na análise de agrupamento os dados são distribuídos em grupos que possuem alguma similaridade. Neste caso o algoritmo faz um enquadramento do objeto dentre os grupos disponíveis e posiciona-o no grupo que melhor o descreve. Normalmente os agrupamentos são observados em gráficos de dispersão. Aplicações comuns para esse tipo de algoritmo são desde a descoberta

de relações de padrões genéticos até identificação de comportamento de índices financeiros.

2.2.4 Avaliação e Interpretação de Resultados

Cada técnica de mineração possui uma saída de formato específico. Como é o caso do modelo de agrupamento, que pode ser visualizado através de um gráfico de dispersão, sendo sua precisão medida através do **erro quadrático**.

No caso das técnicas de associação a saída é apresentada pelas categorias mais frequentemente associadas dentro do conjunto, acompanhada de medidas de aderência ao modelo gerado, como **suporte** e **confiança**.

As técnicas preditivas possuem ampla opção de algoritmos, sendo que alguns desses são capazes de gerar modelos através das árvores de decisão. Nesse contexto a árvore classifica o conjunto de dados de forma intuitiva, sendo os nodos folha o resultado da classificação e os nodos acima representam os pontos condicionais de classificação da classe. A Figura 10 demonstra um exemplo da árvore que classifica declarações de imposto de renda, sendo o atributo alvo a informação sonega (SIM/NÃO) e os demais atributos que induzem o resultado.

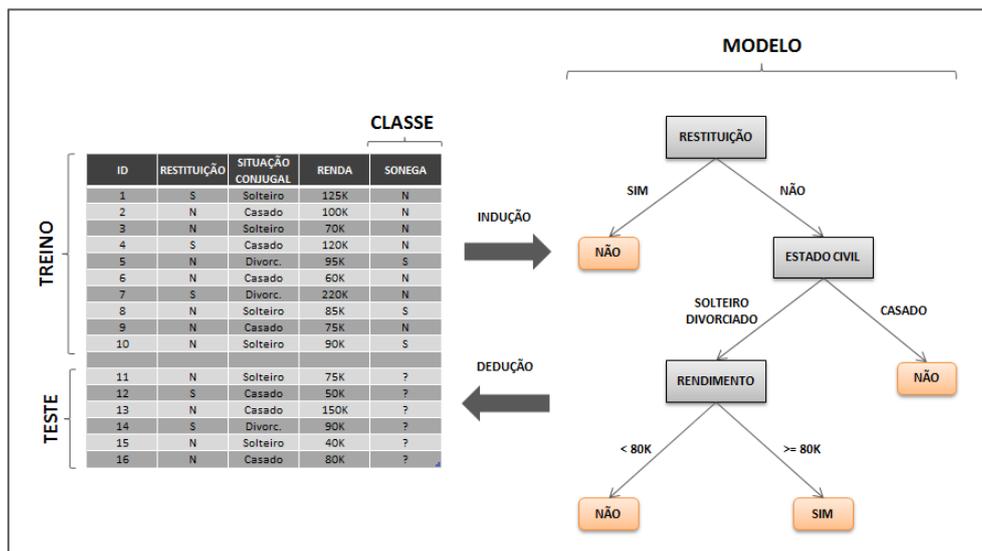


Figura 10 – Árvore de Decisão
Adaptado de [TAN06]

Os algoritmos preditivos também possuem medidas que verificam a aderência do modelo em relação ao conjunto de teste, tais como: acurácia, taxa de erro, confiabilidade positiva, confiabilidade negativa, sensibilidade e especificidade. Tais medidas são calculadas a partir da **matriz de confusão**. Essa matriz verifica os objetos que foram enquadrados (classificados) de forma correta e de forma errada através da dedução do modelo sobre o conjunto de teste. A Tabela 2, baseada em [WIT05] representa uma matriz de confusão. No quadro subsequente (Tabela 3) tem-se a formalização das medidas de aderência obtidas através da matriz de confusão.

Tabela 2 – Matriz de Confusão. Adaptado de [WIT05]

Atributo Preditivo		Classe Preditiva	
		Classe = 1	Classe = 0
Classe Real	Classe = 1	VP	FN
	Classe = 0	FP	VN

Tabela 3 – Medidas do Modelo. Adaptado de [WIT05]

$ACURÁCIA = \frac{VP + VN}{VP + VN + FP + FN}$
$TAXA DE ERRO = \frac{FP + FN}{VP + VN + FP + FN}$
$CONFIABILIDADE POSITIVA = \frac{VP}{VP + FP}$
$CONFIABILIDADE NEGATIVA = \frac{VN}{VN + FN}$
$SENSIBILIDADE = \frac{VP}{VP + FN}$
$ESPECIFICIDADE = \frac{VN}{FP + VN}$

Embora as ferramentas de mineração forneçam subsídios para analisar os resultados, a interação com o especialista e estudo do domínio é elementar para discernir os resultados de forma eficiente. Geralmente o especialista do domínio não é responsável pela execução de KDD, mas sim por dar sentido às informações disponibilizadas. Sendo assim, o conhecimento especialista é importante desde a etapa inicial, onde os dados são escolhidos, até a análise dos resultados de mineração [FAY96].

2.3 Teoria dos Conjuntos Aproximados

A teoria dos conjuntos aproximados – TCA (*Rough Sets Theory*) foi proposta por [PAW82] como uma alternativa às já existentes teorias que trabalham com conhecimento imperfeito, como Conjuntos *Fuzzy*, Redes Bayesianas, entre outras. Segundo [PAW82], TCA pode ser entendida como uma ferramenta de alto nível para pré-processamento de dados a serem utilizados por algoritmos de aprendizagem de máquina, sendo empregada em diversos domínios onde há dificuldade de análise dos dados.

A TCA baseia-se na suposição de que o conhecimento está associado a algumas informações do conjunto universo de dados. Nesse contexto, os atributos caracterizados pela mesma informação são considerados indiscerníveis. Essa relação indiscernibilidade que diferencia os atributos é a base do cálculo matemático da TCA. O conjunto de todos os objetos indiscerníveis é chamado de conjunto elementar. A relação de indiscernibilidade de atributos constrói os conjuntos parciais chamados redutos. Por sua vez os redutos são formados por atributos que mantêm as mesmas propriedades do conjunto universo. Cada reduto é medido quanto a sua discernibilidade através da aproximação inferior e superior. Nesse contexto a aproximação inferior consiste em todos os objetos que certamente pertencem ao reduto e a aproximação superior contém todos os objetos que possivelmente pertencem ao reduto. Usando um exemplo transcrito de [PAW02], nas subseções a seguir parte do formalismo da TCA é explicado.

2.3.1 Exemplo de Aplicação TCA

O exemplo de [PAW02] ilustra a aplicação de TCA sobre um sistema de informação de rotatividade de clientes telecomunicações. A análise baseada em TCA inicia-se a partir de uma tabela de dados chamada “tabela de decisão”, onde as colunas são os atributos as linhas os objetos. Os atributos da tabela de decisão são divididos em dois grupos. São eles: atributos de condição e atributo de decisão¹. Na Tabela 4 demonstrada abaixo se tem os elementos relativos a seis segmentos de cliente.

Tabela 4 – Exemplo TCA de [PAW02]

Segmento	<i>In</i>	<i>Out</i>	<i>Change</i>	<i>Churn</i>	N
1	médio	médio	baixo	não	200
2	alto	alto	baixo	não	100
3	baixo	baixo	baixo	não	300
4	baixo	baixo	alto	sim	150
5	médio	médio	baixo	sim	220
6	médio	baixo	baixo	sim	30

Acima tem-se os atributos: “*In*” que corresponde às chamadas recebidas – “*Out*” que representa as chamadas efetuadas para a mesma operadora – “*Change*” que corresponde às chamadas efetuadas para outras operadoras de telefonia móvel – “*Churn*” que descreve se o cliente foi mantido ou não (se *Churn* = S então cliente não foi, se *Churn* = N então cliente foi mantido) – “N” identifica a frequência de casos semelhantes. Em suma, esse exemplo tenta identificar quais são os fatores mais relevantes para explicar a rotatividade de

¹ Na TCA o atributo de decisão é a informação que se quer classificar. Fazendo uma associação à mineração de dados esta informação é representada pelo atributo alvo.

clientes. Porém de forma intuitiva esse problema não pode ser resolvido em função da inconsistência do conjunto, ou seja, 1 e 5 têm o mesmo perfil, mas diferentes valores para o domínio para o atributo de decisão (*Churn*). Observamos a análise abaixo:

- segmentos de 2 e 3 (4 e 6) podem ser classificados como conjuntos de clientes que certamente $Churn = \text{não}$;
- segmentos 1, 2, 3 e 5 (1, 4, 5 e 6) podem ser classificados como conjuntos de clientes que possivelmente $Churn = \text{não}$;
- segmentos 1 e 5 são conjuntos indecidíveis de clientes.

Dada a análise acima temos a seguintes noção:

- o conjunto $\{2,3\}$ ($\{4,6\}$) é a aproximação inferior do conjunto $\{1,2,3\}$ ($\{4,5,6\}$);
- o conjunto $\{1,2,3,5\}$ ($\{1,4,5,6\}$) é a aproximação superior do conjunto $\{1,2,3\}$ ($\{4,5,6\}$);
- o conjunto $\{1,5\}$ é a região de fronteira do conjunto $\{1,2,3\}$ ($\{4,5,6\}$).

Na próxima subseção a aproximação superior/inferior e região de fronteira são vistos em detalhe.

2.3.2 Sistema de Informação e Aproximações da TCA

A fim de identificar as aproximações primeiramente defini-se o conjunto de dados, nesse contexto titulado como “sistema de informação”. Um sistema de

informação é definido como $S = (U, A)$, onde U e A são finitos e não vazios, sendo o U o conjunto de objetos de S e A o conjunto de atributos de S . Para todo atributo $a \in A$ associa-se um conjunto V_a , de seus valores, chamados de domínio de a . Para qualquer subconjunto de B de A aplica-se a relação de indiscernibilidade $I(B)$ em U , definido da seguinte forma: $(x, y) \in I(B)$ se e somente se $a(x) = a(y)$ para todo $a \in A$, onde $a(x)$ representa o valor do atributo a para o objeto x . Portanto $I(B)$ é uma relação de equivalência. O grupo de todas as classes de equivalência de $I(B)$ é denotado por $U/I(B)$, ou simplesmente por U/B . Uma classe de equivalência de $I(B)$ é denotado por $B(x)$. Se (x, y) pertence a $I(B)$ diz-se que x e y são indiscerníveis em relação a B . Classes de equivalência da relação $I(B)$ são referidos como o conjunto *B-elementares*.

Para um sistema de informação $S = (U, A)$, $X \subseteq U$, e $B \subseteq A$ defini-se duas operações atribuídas a cada $X \subseteq U$. São os conjuntos $B_*(X)$ e $B^*(X)$ denominados de *aproximação B-inferior* e *B-superior*, respectivamente. Abaixo a formalização de ambas as operações:

$$B_*(X) = \bigcup_{x \in U} \{B(x) : B(x) \subseteq X\},$$

$$B^*(X) = \bigcup_{x \in U} \{B(x) : B(x) \cap X \neq \emptyset\}$$

Portanto a aproximação inferior de um dado conjunto é a união de todos de *B-elementares* deste conjunto. Por outro lado a aproximação superior é a união de todos *B-elementares* que possuem intersecção não vazia com o conjunto. A Figura 11, adaptada de [PAW82] ilustra a noção de aproximação

inferior e superior de $S = (U, A)$ em uma representação bidimensional de quadrados elementares.

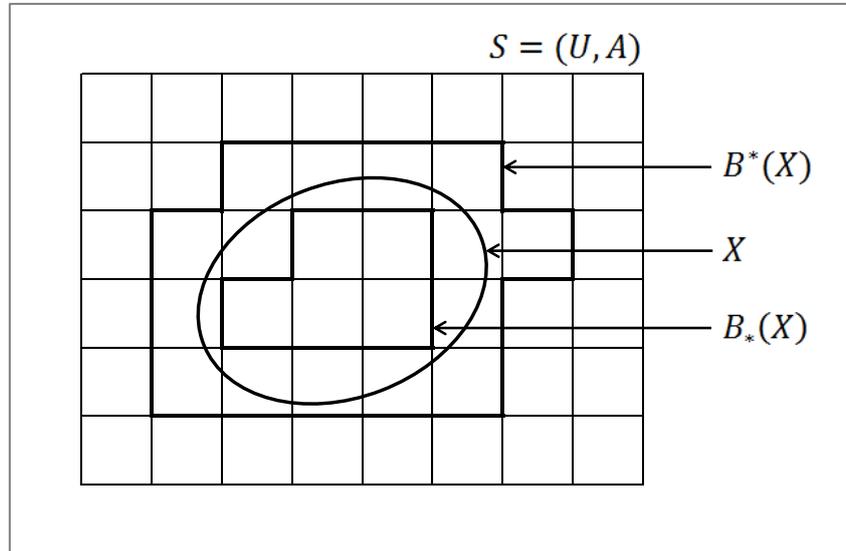


Figura 11 – Aproximação superior/inferior de TCA

Adaptado de [PAW82]

Uma terceira região do conjunto, nomeada conjunto *B-fronteira* inclui os objetos que não podem ser enquadrados em X . Para tanto, aplica-se a diferença entre as aproximações (superior e inferior), conforme formalização a seguir.

$$BN_B(X) = B^*(X) - B_*(X)$$

2.3.3 Precisão e Qualidade de Aproximação

Cada subconjunto pode ser medido quanto a sua acurácia. Nesse contexto tem-se as medidas de *precisão de aproximação* e *qualidade de*

aproximação. Em [GRE01] são demonstradas ambas as medidas. Abaixo a representação das medidas adaptadas para a ilustração de [PAW02].

Para X , $X \neq \emptyset$, por meio dos atributos de B define-se a *precisão de aproximação*:

$$\alpha_P(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

A *qualidade de aproximação* de X para por meio dos atributos de B é definida conforme abaixo:

$$\gamma_P(X) = \frac{|B_*(X)|}{|X|}$$

A qualidade de aproximação representa a frequência relativa dos objetos corretamente classificados por meio dos atributos de B .

2.4 Considerações do Capítulo

Neste capítulo apresentou-se a fundamentação teórica dos principais temas relacionados a este trabalho. Buscou-se referências bibliográficas essenciais, bem como exemplos e ilustrações a fim de evidenciar o conteúdo da pesquisa.

Primeiramente, na seção 2.1 foram percorridos os conceitos de DW bem como as características de sua arquitetura e aplicação. Tal fundamentação ancora-se principalmente nas obras de [KIM02] e [INM05]. Na segunda seção (2.2) as referências de KDD foram esclarecidas detalhando cada uma das etapas do processo. Nesse contexto, utilizou-se principalmente as obras de [FAY96],

[TAN06] e [HAN11]. Na seção 2.3 fez-se uma descrição da Teoria dos Conjuntos Aproximados, usando como referência [PAW82], [PAW02] e [GRE01].

3- DESCRIÇÃO DO CENÁRIO

A pesquisa concentra-se em criar um algoritmo de pré-processamento baseado em TCA onde o resultado da execução forneça um diagnóstico sobre os dados. O diagnóstico resultante provê uma base de conhecimento ao técnico minerador para auxiliá-lo na seleção dos dados, principalmente quando há carência de entendimento do domínio.

Os dados utilizados nesse estudo provêm de um DW de operação real, onde se focou em um de seus *data marts*. Ao longo das próximas seções são detalhadas questões como a caracterização do problema, contribuição e o contexto de implementação da pesquisa.

3.1 Caracterização do Problema

Para [FAY96] a alta dimensionalidade e o entendimento dos dados são fatores críticos em projetos de KDD. A grande quantidade de atributos pode levar a indução de modelos sobre informações que não são relevantes ao negócio. Neste cenário, o autor destaca a importância do conhecimento do domínio em todas as etapas de KDD.

Mesmo dispondo de padrões de projeto bem definidos, como por exemplo, CRISP-DM [CRI03], segundo [GON08] projetos de mineração até então estão sendo desenvolvidos mais como uma arte do que como uma ciência. Para esse autor isso se deve ao fato de que se depende muito da experiência e do

conhecimento que o técnico minerador tem sobre o domínio em questão, ou seja, o conhecimento que tem sobre os dados de forma que possa trabalhá-los e a capacidade técnica para interpretar os modelos gerados.

Segundo [KOP02] o conhecimento de domínio é crucial para o sucesso do projeto, principalmente na fase inicial (seleção e pré-processamento) e na fase final (análise dos resultados). Portanto devem ser desenvolvidas ferramentas que auxiliem o técnico minerador a compreender o domínio estudado.

Os fatores acima tendem a aumentar o tempo e conseqüentemente o custo de projetos de mineração, independente do método utilizado. Soma-se também o fato de que não há como garantir que o valor do conhecimento obtido supere o investimento despendido, uma vez que os dados podem ou não demonstrar novo conhecimento. De forma empírica tem-se observado que a incerteza do retorno sobre o investimento em KDD por muitas vezes inviabiliza projetos desse tipo.

3.2 Caracterização da Contribuição

Encontra-se na literatura diversos estudos sobre o uso da TCA no pré-processamento dos dados, mais especificamente, para auxílio na redução de dimensionalidade. O trabalho de [THA09] titulado “*Dimensionality Reduction Based On Rough Set Theory: A Review*” faz uma revisão das contribuições até então publicadas sobre redução de dimensionalidade usando TCA, descrevendo resumidamente a proposta de cada trabalho. Nesta revisão é possível perceber que grande parte das publicações são propostas de variações do aparato matemático de TCA com vistas à resolução de algum tipo de problema, como

por exemplo, o trabalho de [LEU06], que faz uma análise sobre aquisição de conhecimento sobre conjuntos de dados incompletos.

Dada argumentação da seção 3.1 entende-se que quanto maior a compreensão do domínio, mais eficaz se torna o processo de KDD. Do ponto de vista prático, ao dominar o assunto sabe-se discernir os dados e, portanto a melhor forma de selecionar, pré-processar, reduzir dimensionalidade, analisar resultados, etc. Também é sabido que o estudo do domínio tem uma curva de aprendizagem. Logo, a criação de ferramentas que auxiliem o técnico minerador a compreender os dados tende a tornar o processo mais efetivo, bem como minimizar o tempo/custo de projeto. Nessa direção, o presente trabalho apresenta um algoritmo baseado em TCA, nomeado *Rough Set App* (RSAPP), para geração de um diagnóstico sobre os dados. Tal diagnóstico infere medidas de qualidade de aproximação de diferentes redutos. Após a execução de RSAPP tem-se uma base de conhecimento, onde o técnico pode consultar quais combinações são mais representativas para um determinado atributo alvo, ou ainda, quais atributos condicionais podem ser suprimidos (redução de dimensionalidade) sem prejuízos. O algoritmo aqui proposto pretende criar um mecanismo que providencie melhores resultados de mineração, logo nas etapas iniciais do projeto, uma vez que, desta forma o técnico minerador dispõe de uma ferramenta que o induz a redução do universo de dados para uma dimensionalidade de maior relevância. Cabe salientar que tal algoritmo não isenta o essencial estudo de domínio e a interação com o especialista, mas propõe servir como um artifício nesse processo de aprendizado.

O algoritmo implementado foi executado sobre um banco de dados OLAP a fim de identificar dentre as dimensões e suas hierarquias quais são os

dados mais relevantes para mineração, ou seja, os dados com maior potencial de produzir modelos interessantes com precisão satisfatória. Em suma, o diagnóstico gerado por RSAPP pretende reduzir o tempo de projeto de KDD nas tarefas de entendimento dos dados e redução de dimensionalidade.

Com base na revisão de [THA09] é possível verificar diversas frentes onde TCA é empregada para redução de dimensionalidade. Porém se entende como diferencial dessa pesquisa as seguintes características: (1) uso de um banco de dados OLAP com dados de operação real – (2) apresentação dos resultados de RSAPP de forma intuitiva (diagnóstico) e (3)² desenvolvimento com uso de linguagem nativa de banco de dados (SQL).

3.3 Contexto de Implementação da Pesquisa

Este estudo é fruto do GPIN (PPGCC-PUCRS). Uma das frentes de pesquisa que o grupo tem trabalhado é a criação mecanismos de avaliação prévia dos dados, que possibilite a indicação de tarefas de mineração mais assertivas em projetos de KDD. Por exemplo, no trabalho de [COL11] foi proposta uma técnica de indução e ranqueamento de árvores de decisão sobre modelos OLAP. O presente trabalho se mantém nessa direção, ou seja, concentra-se em criar subsídios a fim elevar a eficácia do projeto de KDD.

A pesquisa foi desenvolvida sobre o DW de uma organização de grande porte que opera no mercado de transações eletrônicas multisserviços. Este DW foi implementado em 2011 e atualmente é utilizado por aplicações OLAP, auxiliando na tomada de decisão de diversas frentes, como por exemplo,

² O diagnóstico permanece no formato de metadado no banco de dados onde RSAPP é executado, ou seja, não é necessário outro tipo de aplicativo além do próprio SGDB.

medição do volume transacional, análise de aderência a novos produtos, controle de ativos, entre outros assuntos. Cabe pontuar que o estudo usou apenas uma fração do DW disponível. A arquitetura do DW segue o padrão de modelagem dimensional, usando o modelo estrela proposta por [KIM02]. A rigor o modelo estrela empregado está organizado em constelação de fatos, desta forma os assuntos são integrados através de pelo menos uma dimensão. Até o momento deste trabalho a organização não utilizava KDD em suas rotinas de análise de informações provindas do DW. Porém, tem feito intenso uso de aplicações OLAP na busca por informações que subsidiem a tomada de decisões estratégicas. Nesse contexto o algoritmo RSAPP foi aplicado e segue o fluxo de implementação ilustrado na Figura 12, abaixo:

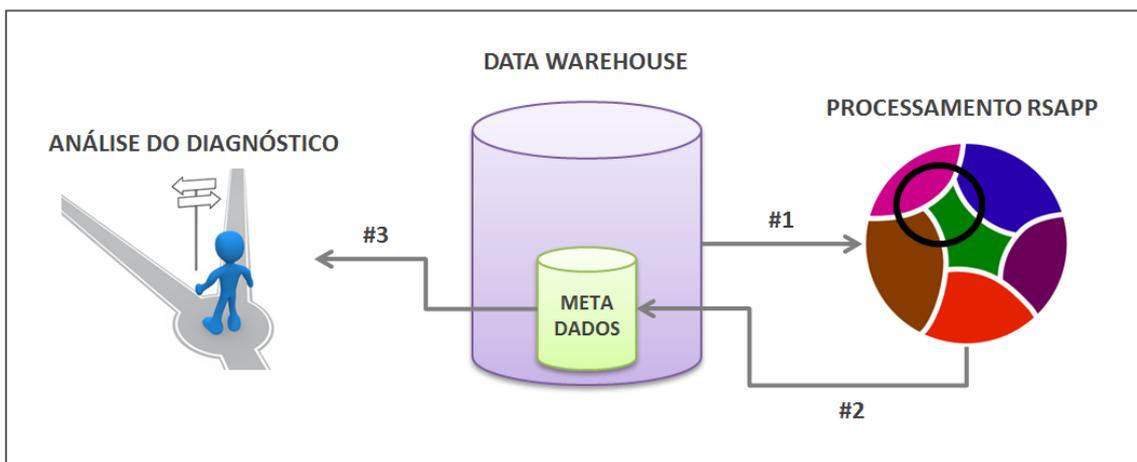


Figura 12 – Fluxo do diagnóstico através de RSAPP

Na primeira etapa (1) os dados são processados pelo algoritmo, logo após o resultado é gravado em tabelas de metadados (2), feito isso o técnico minerador tem em mãos o diagnóstico (3) e pode analisar quais as potencias

combinações que possivelmente classificam melhor os dados, segundo o argumento da TCA.

A fim de testar a efetividade da ferramenta foram propostas algumas heurísticas, onde as combinações de atributos condicionais (reduzidos) processadas por RSAPP e suas respectivas medidas de qualidade de aproximação foram comparadas com a precisão dos modelos de classificação induzidos pela ferramenta de mineração. Mais detalhes acerca dos testes realizados estão documentados no capítulo 5.

3.4 Considerações do Capítulo

Neste capítulo foi apresentado o panorama da pesquisa. Para tanto, na seção 3.4.1 foram referenciados alguns desafios de projetos de KDD. Nesta seção procurou-se evidenciar os problemas enfrentados, onde foi percebida a essencial relação entre compreensão do domínio e eficácia do processo de KDD. Na seção seguinte (3.4.2) argumentou-se sobre a solução proposta a fim de justificá-la em relação a sua contribuição para a área. Por último, na seção 3.4.3 fez-se uma breve descrição do cenário onde esta pesquisa está inserida bem como resumo do seu desenvolvimento.

4- IMPLEMENTAÇÃO RSAPP

O algoritmo RSAPP se deteve a processar os dados do DW e armazenar o diagnóstico no formato de metadados. A Figura 13 representa o ambiente onde RSAPP foi desenvolvido, sendo a sua implementação compreendida nos passos #2 e #3.

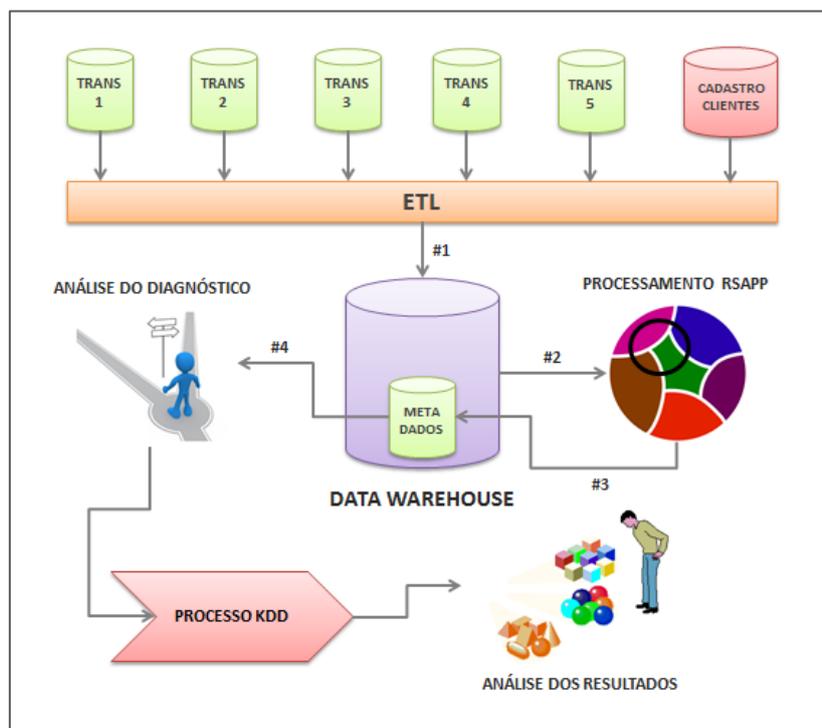


Figura 13 – Processo KDD usando RSAPP

A pesquisa fez uso de um DW de uma empresa de grande porte. Embora o processo de ETL não tenha sido implementado nessa pesquisa (passo #1 da Figura 13), faz-se necessário descrever os dados usados, quanto a sua modelagem e características, portanto na seção 4.1.1 essas informações são

detalhadas. Os passos #2 e #3 da Figura 13 representam o desenvolvimento de RSAPP propriamente dito, portanto na seção 4.1.2 são percorridas as etapas desse desenvolvimento. O diagnóstico gerado por RSAPP (passo #4 da Figura 13) é explicado pormenor na seção 4.1.3. Entende-se que o diagnóstico gerado por RSAPP pode integrar o processo de KDD independente da metodologia utilizada, visto que tal ferramenta posiciona-se como artifício de pré-processamento, por isso na Figura 13 o uso de RSAPP é visto como uma etapa geradora de conhecimento para a sequência de processos de KDD.

4.1 Cenário Real de Aplicação

Os dados utilizados provêm de um modelo dimensional, organizado no formato estrela, onde estão persistidas informações sobre inadimplência. A Figura 14 apresenta o modelo conceitual, composto por uma tabela fato e sete tabelas dimensionais.

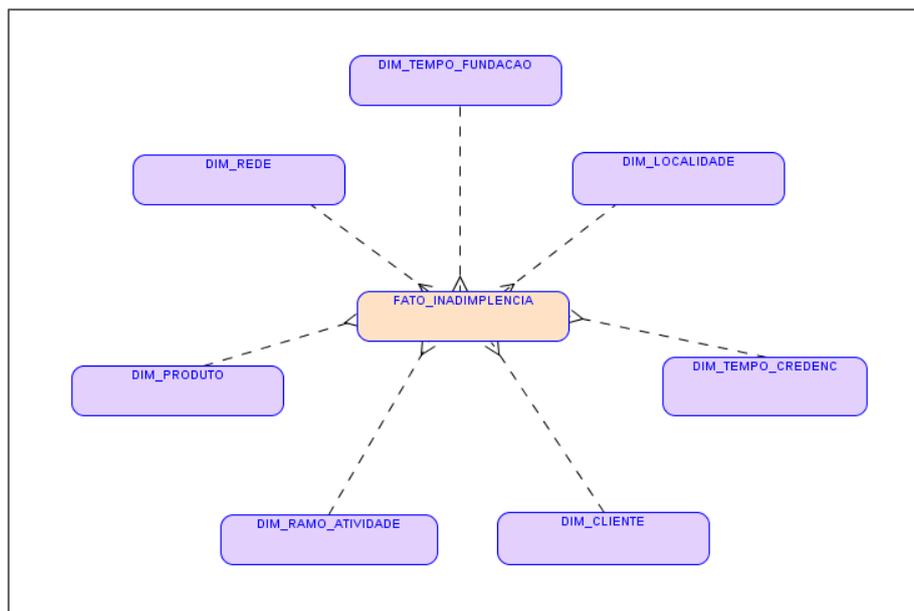


Figura 14 – Modelo conceitual

Abaixo uma descrição breve descrição das propriedades de cada uma das entidades.

- DIM_TEMPO_CREDENC: dimensão de informações de tempo de quando o cliente foi credenciado, ou seja, quando determinado cliente passou a compor a carteira de clientes. Atributos como ano e mês persistem nessa entidade.
- DIM_TEMPO_FUNDACAO: dimensão de informações de tempo de fundação da empresa cliente. Atributos como ano e mês persistem nessa entidade.
- DIM_LOCALIDADE: dimensão de informações de localidade do cliente. Nesta estão contidas informações como região e UF.
- DIM_CLIENTE: dimensão de informações de cadastro de cliente. Nesse contexto o cliente é um ponto de venda. Tal dimensão armazena informações como, natureza jurídica, classificação, entre outras informações cadastrais.
- DIM_PRODUTO: os clientes estão aptos a executar diversos tipos de transações. Esses diferentes tipos de transações eletrônicas são identificadas como produtos. Por exemplo, transações de pagamento com cartão de vale refeição, ou ainda, recarga de celular pré-pago são típicos produtos. Essa dimensão identifica quais os produtos que o cliente está apto a operar.

- DIM_REDE: essa dimensão distingue se aquele registro de fato, provém de um cliente que pertence ou não a uma rede de clientes. No caso de pertencer, esta dimensão vai identificar qual a rede.
- DIM_RAMO_ATIVIDADE: dimensão de informação do ramo de atividade do cliente. Também nesta consta o subramo de atividade. Um ramo de atividade pode ser, por exemplo, alimentação, vestuário, etc. No caso do ramo de alimentação, pode haver subramos, como por exemplo, restaurante, cafeteria, etc.
- FATO_INADIMPLENCIA tabela fato onde estão computados os registros de inadimplência. Tais registros identificam se o cliente encontra-se adimplente ou inadimplente em determinado título.

RSAPP foi aplicado a fim de identificar dentre as informações disponíveis, quais subconjuntos (redutos) podem produzir melhores resultados de mineração. Nesse contexto, as dimensões possuem diferentes tamanhos (quantidade de atributos). Abaixo (Tabela 5) são mostradas as quantidades de atributos de cada uma das dimensões e da tabela fato.

Tabela 5 – Quantidade de atributos

Tabela	Quantidade de Atributos ³
DIM_TEMPO_CREDENC	3
DIM_TEMPO_FUNDACAO	3
DIM_LOCALIDADE	4
DIM_CLIENTE	8
DIM_PRODUTO	2
DIM_REDE	3
DIM_RAMO_ATIVIDADE	3
FATO_TRANSACAO	8

Algumas dimensões do modelo estão organizadas de forma hierárquica, porém há casos híbridos, ou seja, parte dos atributos da dimensão está organizada na forma hierárquica e outra parte são atributos adicionais que descrevem o objeto. Por exemplo, nas duas dimensões de tempo tem-se a hierarquia formada por ano e mês. Há casos em que nenhuma hierarquia está formada na dimensão, ou seja, a dimensão está no seu nível elementar, é o caso da dimensão de produto. Na Tabela 6 são descritas essas propriedades.

¹ A quantidade de atributos informada não inclui chaves primárias e estrangeiras. Cabe salientar essas chaves foram desconsideradas no processamento de RSAPP.

Tabela 6 – Estrutura hierárquica

Dimensão	Organização	Hierarquia
DIM_TEMPO_CREDENC	Hierarquizada	Ano e mês
DIM_TEMPO_FUNDACAO	Hierarquizada	Ano e mês
DIM_LOCALIDADE	Hierarquizada	País, região e UF
DIM_CLIENTE	Parcialmente hierarquizada	Perfil e classificação
DIM_PRODUTO	Não hierarquizada	
DIM_REDE	Hierarquizada	Identificação e rede ou varejo e descrição das redes
DIM_RAMO_ATIVIDADE	Hierarquizada	Ramo e subramo

4.2 Algoritmo RSAPP

O algoritmo RSAPP foi construído através da linguagem procedural PL/SQL (*Procedural Language/Structured Query Language*), sendo esta uma extensão da linguagem SQL para o SGDB *Oracle Database* [ORA12]. Basicamente o algoritmo foi dividido em duas etapas, nomeadas *LOWER_APP* e *QUALITY_APP*. A primeira etapa (*LOWER_APP*) é executada para extrair a **aproximação inferior** de cada um dos redutos. O cálculo da aproximação inferior é explicado em detalhes na subseção 2.3.2. Com base do resultado gerado por *LOWER_APP* o procedimento *QUALITY_APP* é executado e então extrai a **qualidade da aproximação** de cada reduto. O cálculo da aproximação inferior é explicado em detalhes na subseção 2.3.2. Na Figura 15 é representado o diagrama dos componentes e as suas interações.

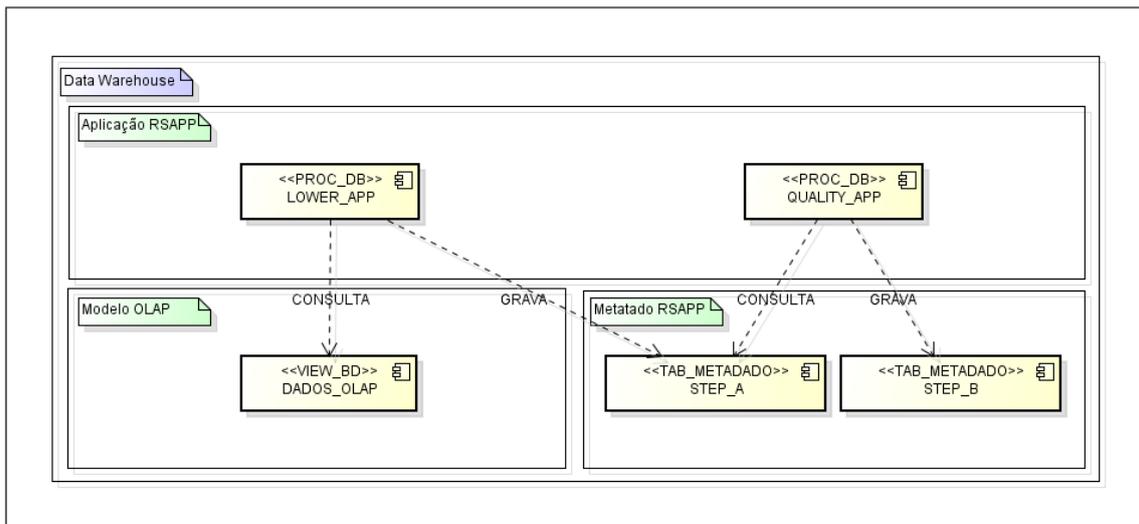


Figura 15 – Diagrama de componentes de RSAPP

Nas próximas subseções faz-se a descrição detalhada do funcionamento de *LOWER_APP* e *QUALITY_APP*:

4.2.1 *Lower App*

Esse procedimento aplica diversas consultas sobre os dados. Esses dados podem estar contidos em uma *view* ou em uma consulta padrão (*SELECT*) inserida diretamente no seu fonte. O procedimento *LOWER_APP* recebe como parâmetros a consulta (ou nome da *view*) e a identificação do atributo alvo. Com base nessas informações o procedimento cria diferentes subconjuntos (redutos) e calcula a aproximação inferior de cada combinação (conforme fórmula detalhada na subseção 2.3.1.3). Ao final o procedimento grava os resultados em uma tabela de metadado intermediária, chamada *STEP_A*. No quadro abaixo o algoritmo do procedimento *LOWER_APP* é descrito e comentado:

```

create or replace procedure sp_lower_app as
--inicialização de variáveis
vquery varchar2(4000);
vwhere varchar2(4000);
vnome varchar2(4000);
vqtd number;

begin
--limpa tabela parcial do diagnóstico (step_a)
delete from step_a where 1 = 1;
commit;
--loop para seleção de cada um dos domínios do atributo alvo
for rec in (select distinct CLASS
            from /*nome da tabela ou view*/ l)
loop
--loop para seleção dos atributos condicionais que compõem o reduto
for coll in (select 0 column_id -- Para cada id de coluna, faça:
             from dual
             union all
             select column_id
             from sys.all_tab_cols c
             where table name = /*nome da view ou consulta*/
               and column_name != 'CLASS' --descarta o atributo classe dos redutos
             order by column_id )

loop
--escreve na variável vwhere o valor do domínio do atributo alvo que será usado neste loop
vwhere := 'CLASS in (select distinct CLASS from /*nome da view ou consulta*/ where CLASS !=
'''||rec.CLASS||''')';
vnome := null;
--incrementa atributos condicionais para estabelecer diferentes redutos
for col in (select column_name
            from sys.all_tab_cols c
            where table name = /*nome da view ou consulta*/
              and column_name != 'CLASS'
              and column_id > coll.column_id
            order by column_id )

loop
--testa se o reduto é vazio, senão incrementa até haver os diferentes redutos formados
if vnome is null then
vnome := col.column_name;
vqtd := 1;
else
vnome := vnome||' '||col.column_name;
vqtd := vqtd + 1;
end if;
vwhere := vwhere || ' and gn.'||col.column_name||' = gs.'||col.column_name;
--escreve na variável vquery a classe testada o conjunto escrito em vnome e a quantidade de objetos
contidos na aproximação inferior
vquery := '
insert into step a
select '''||rec.CLASS||''' as decisao, '''||vnome||''' as reduto,
      '''||vqtd||''' as qtd, count(1) as aprox_inf
from /*nome da view ou consulta*/ gn
where CLASS = '''||rec.CLASS||'''
and not exists (select 1
                from /*nome da view ou consulta*/ gs
                where '||vwhere||')
'; --subconsulta forma a variável para relação de indiscernibilidade de TRS

DBMS_OUTPUT.ENABLE(1000000);
dbms_output.put_line(vquery);
execute immediate vquery;

end loop;

end loop;

end loop;
commit;

end;

```

O procedimento *LOWER_APP* grava os metadados parciais na tabela de *STEP_A*. O layou de *STEP_A* é ilustrada na Tabela 7 conforme abaixo.

Tabela 7 – Tabela *STEP_A*

Alvo	Reduto	Tamanho do Reduto	Aprox. Inferior
NÃO	ATRIB_1	1	10
NÃO	ATRIB_1 + ATRIB_2	2	20
NÃO	ATRIB_2	1	15
NÃO	ATRIB_N	N	N
↓	↓	↓	↓
SIM	ATRIB_1	1	10
SIM	ATRIB_1 + ATRIB_2	2	20
SIM	ATRIB_2	1	15
SIM	ATRIB_N	N	N
↓	↓	↓	↓

A tabela de metadados *STEP_A* (ilustrada acima) serve de origem de dados para execução do procedimento seguinte (*QUALITY_APP*), onde será calculada a qualidade da aproximação do subconjunto.

4.2.2 *Quality App*

A qualidade da aproximação de TCA identifica quanto os redutos se aproximam do conjunto total. O procedimento *QUALITY_APP* implementa tal cálculo (conforme fórmula detalhada na subseção 2.3.2) e após grava os resultados na tabela de metadados *STEP_B*, que por sua vez representa o

diagnóstico propriamente dito. No quadro abaixo o algoritmo do procedimento *QUALITY_APP* é descrito e comentado:

```

create or replace procedure sp_quality_app as
--inicialização da variável
  vquery varchar2(4000);
begin
--limpa tabela do diagnóstico (step_b)
  delete from step_b where l = 1;
  commit;
--aplica consulta sobre o procedimento anterior (LOWER_APP) que está persistido em STEP_A
  vquery := '
    insert into step_b
    select --seleciona o reduto
      reduto,
      --em step a os resutos estão detalhados por domínio do atributo alvo, portanto
      --faz-se a média quantidade de objetos, pois a qualidade de aproximação é medida por reduto e não é
      --detalhado por domínio do alvo
      avg(r.qtd) as qtd,
      --seleciona a cardinalidade total do conjunto
      (select count(1) from /*nome da view ou consulta*/) as card u,
      --soma a aproximação inferior calculada no procedimento LOWER_APP
      sum(aprox inf) as aprox inferior,
      --calcula a qualidade de aproximação do reduto
      sum(r.aprox_inf) / (select count(1) from /*nome da view ou consulta*/) as
qualidade_aprox,
    from step_a r
    group by reduto
  ';

  dbms_output.put_line(vquery);
  execute immediate vquery;

  commit;

end;

```

A Tabela 8 de dados figurados demonstra a disposição do resultado final de RSAPP, no formato de diagnóstico.

Tabela 8 – Tabela de diagnóstico gerado por RSAPP (*STEP_B*)

Reduto	Cardinalidade do Reduto	Aprox. Inferior	Qualidade da Aproximação
ATRIB_1	1	10	0,3
ATRIB_1 + ATRIB_2	2	20	0,35
ATRIB_2	1	15	0,33
ATRIB_N	N	N	N
↓	↓	↓	↓

Com base nas informações disponíveis no diagnóstico o técnico minerador pode verificar quais atributos ou quais as combinações de atributos possuem uma medida de qualidade de aproximação mais conveniente e, então balancear com a quantidade de atributos condicionais. Por exemplo, uma combinação de 5 atributos pode ser mais qualificada que uma combinação de 8 atributos, e as execuções possivelmente mais rápidas e precisas. No próximo capítulo é demonstrado o diagnóstico sobre os dados da pesquisa bem como descritos os testes realizados e resultados obtidos.

4.3 Softwares Utilizados

Na implementação de RSAPP foram usados os seguintes softwares:

- *Astah Community*: ferramenta de diagramação UML, versão 6.6.4. Provida por *Change Vision Software* [AST12];
- *Oracle Database*: sistema de gerenciamento de banco de dados, versão 11g, *release 2*. Provido por *Oracle Corporation* [ORA12];
- *Oracle Data Modeler*: ferramenta de diagramação E/R, versão 3.1.1.703. Provida por *Oracle Corporation* [ORA12];
- *Oracle SQL Developer*: ferramenta de administração de SGDB, versão 2.0.04. Provida por *Oracle Corporation* [ORA12];
- WEKA: ferramenta de mineração de dados, versão 3.6.4. Provida por *Waikato University* [WEK12].

4.4 Considerações do Capítulo

Neste capítulo os dados utilizados na pesquisa foram contextualizados. Também neste fez-se a descrição do algoritmo RSAPP, detalhando a sua arquitetura bem como o seu código fonte e o layout do seu resultado. Por fim foram referenciadas as ferramentas utilizadas na construção de RSAPP.

5- DESCRIÇÃO DOS RESULTADOS

O diagnóstico gerado por RSAPP sobre os dados usados nessa pesquisa pode ser visualizado integralmente no Anexo A.

Para aferir os resultados gerados pelo diagnóstico RSAPP em relação a sua proposta, aplicou-se uma série de tarefas de mineração sobre o mesmo conjunto de dados onde RSAPP foi executado. Foram usadas como medidas comparativas a **acurácia** dos modelos de mineração e a **qualidade de aproximação** do diagnóstico.

Para as execuções de mineração utilizou-se 4 algoritmos (J48, *Random Forest*, *Simple Cart* e *Naive Bayes*), ou seja, os 190 redutos foram submetidos aos quatro algoritmos de mineração a fim de apurar a acurácia de cada um. Adicionalmente foi criada a métrica de média dos quatro algoritmos. O Anexo B demonstra os resultados das tarefas de mineração.

Nas próximas seções são descritos os testes e as respectivas referências aos anexos, onde estão documentados por completo.

5.1 Diagnóstico RSAPP e a Acurácia de Mineração

O conjunto de dados onde RSAPP foi aplicado possuía vinte atributos, sendo um alvo e dezenove atributos condicionais. RSAPP gerou 190 redutos, com tamanhos que variam de 1 até 19 atributos condicionais. Os resultados dos testes de mineração podem ser observados por completo no Anexo B.

Para comparar a qualidade de aproximação de um reduto com a acurácia deste reduto, os resultados foram plotados aqui graficamente. Nesse contexto a qualidade de aproximação gerada por RSAPP foi ordenada decrescentemente. Logo, a análise deve voltar-se a verificar se a curva dos resultados de acurácia acompanha ou não os resultados de qualidade de aproximação inferidos pelo diagnóstico.

Abaixo faz-se uma análise sobre o primeiro gráfico, com redutos de cardinalidade 1.

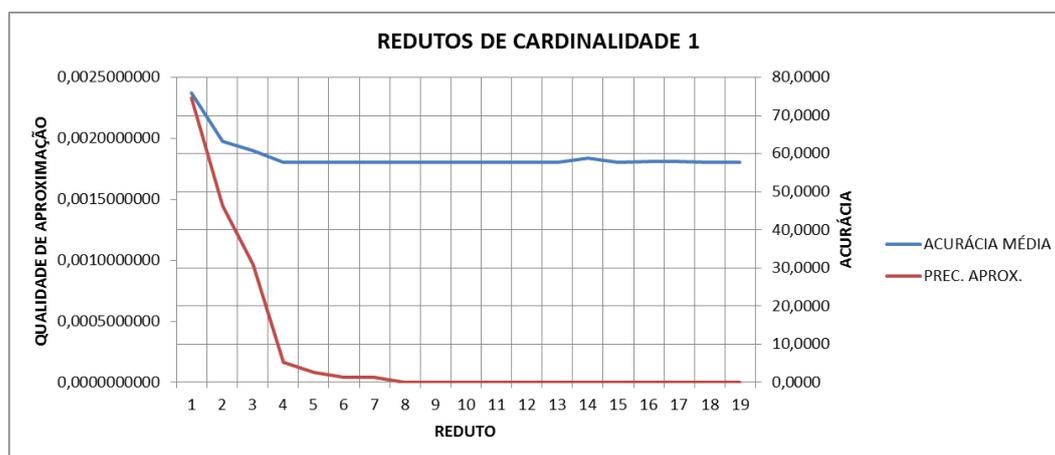


Figura 16 – Gráfico comparativo para redutos de tamanho 1

O eixo à esquerda representa a métrica de qualidade de aproximação apurada pelo diagnóstico. À direita tem-se média de acurácia das execuções de mineração dos quatro diferentes algoritmos.

No caso acima pode ser visto que a acurácia da mineração teve uma curva alinhada com o resultado da métrica de qualidade de aproximação do diagnóstico. No reduto 14 observa-se uma pequena oscilação da acurácia, que não está refletida na qualidade de aproximação. Porém essa oscilação é bastante

pequena. Em suma, no gráfico fica evidente que, para redutos dessa cardinalidade, o algoritmo conseguiu identificar quais atributos gerariam modelos de mineração mais precisos.

Ao longo das próximas subseções é demonstrado cada um dos gráficos comparativos, sendo cada um desses a referência de uma cardinalidade.

5.1.1 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 2)

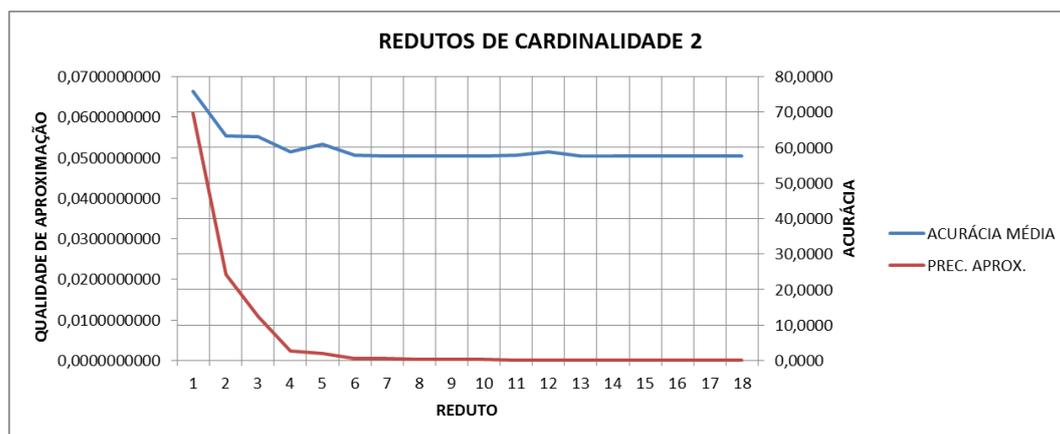


Figura 17 – Gráfico comparativo para redutos de tamanho 2

5.1.2 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 3)

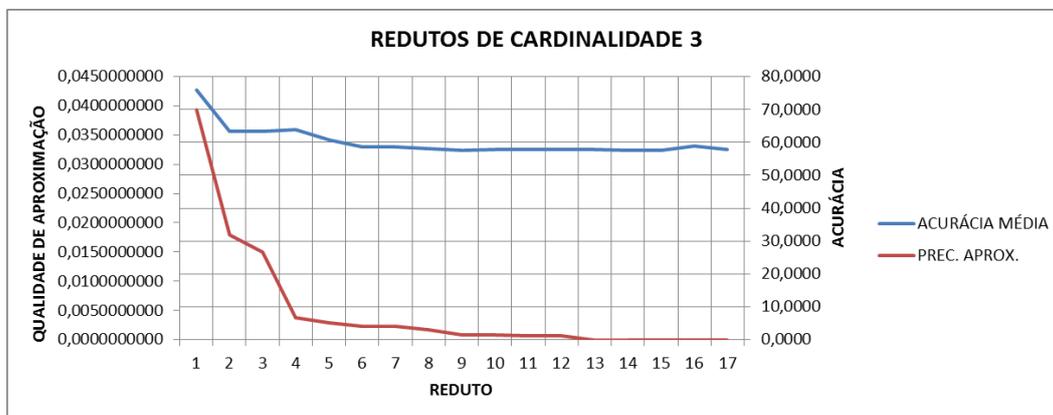


Figura 18 – Gráfico comparativo para redutos de tamanho 3

5.1.3 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 4)

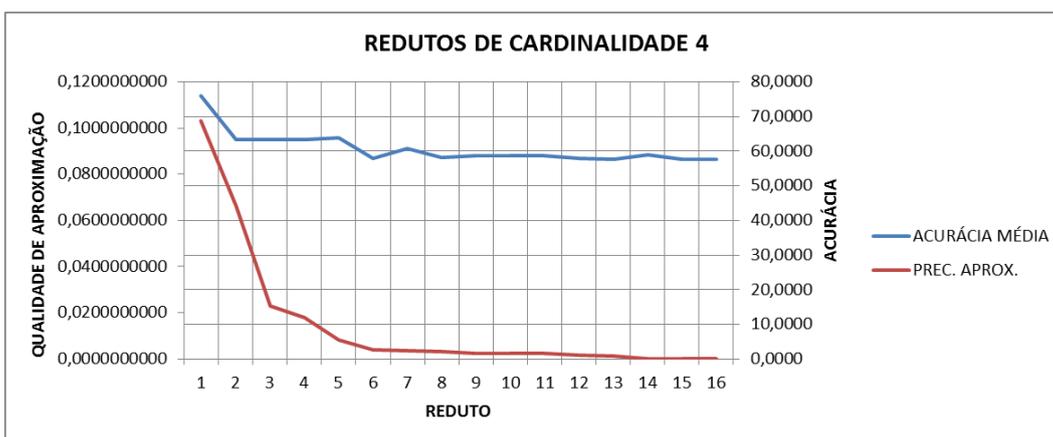


Figura 19 – Gráfico comparativo para redutos de tamanho 4

5.1.4 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 5)

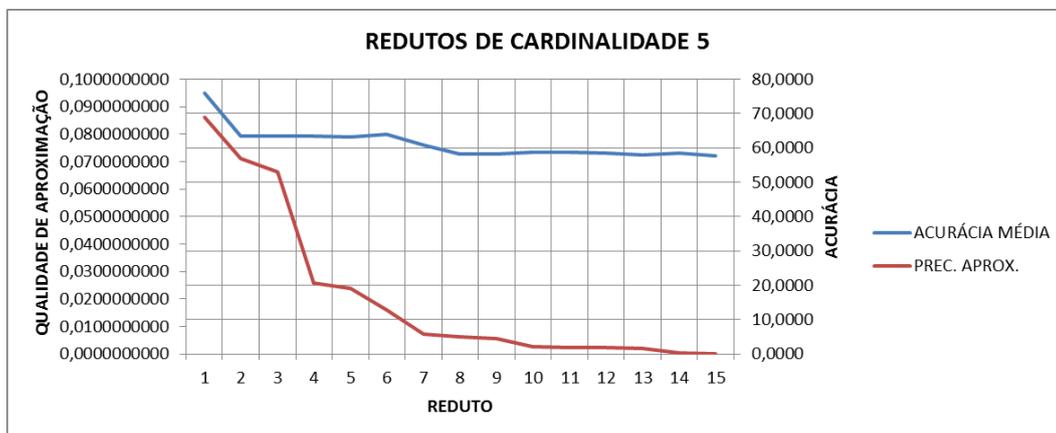


Figura 20 – Gráfico comparativo para redutos de tamanho 5

5.1.5 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 6)

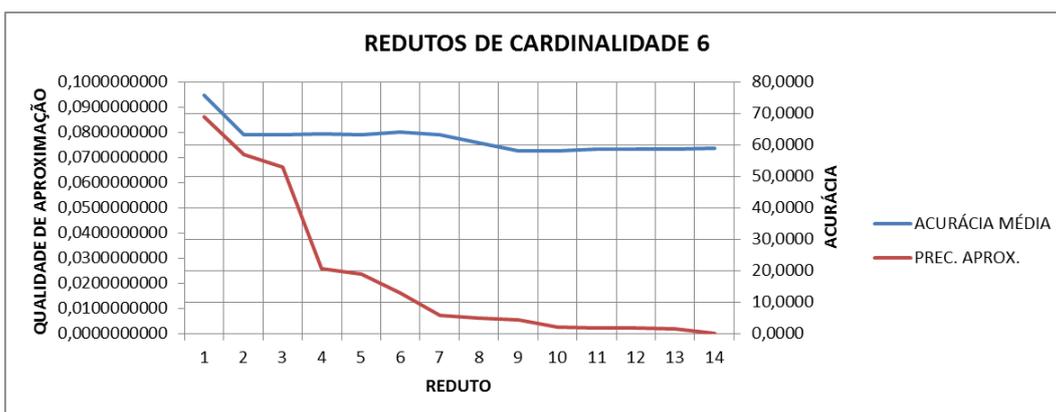


Figura 21 – Gráfico comparativo para redutos de tamanho 6

5.1.6 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 7)

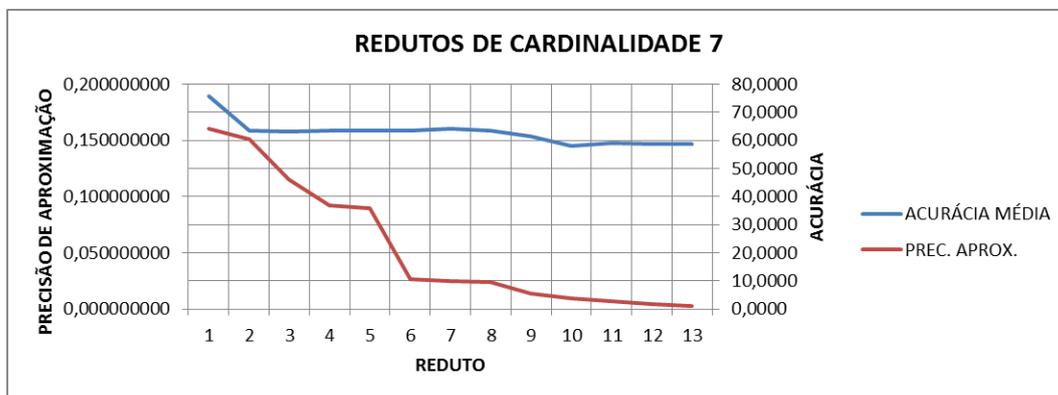


Figura 22 – Gráfico comparativo para redutos de tamanho 7

5.1.7 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 8)

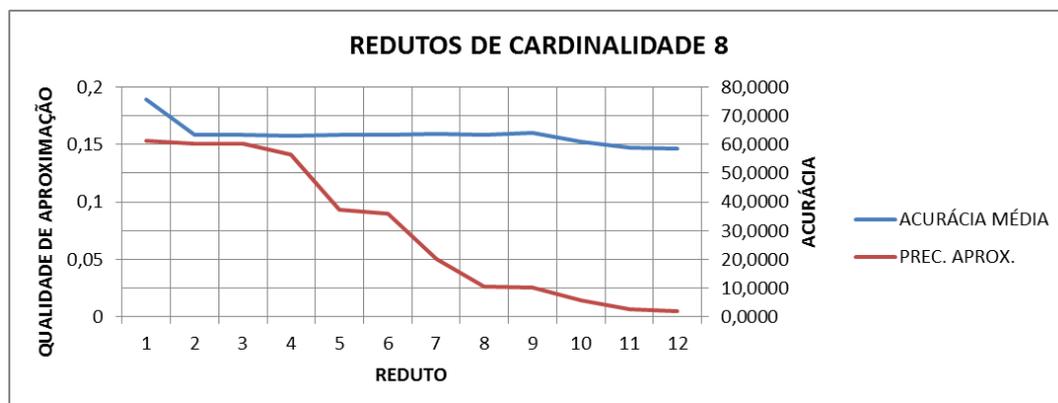


Figura 23 – Gráfico comparativo para redutos de tamanho 8

5.1.8 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 9)

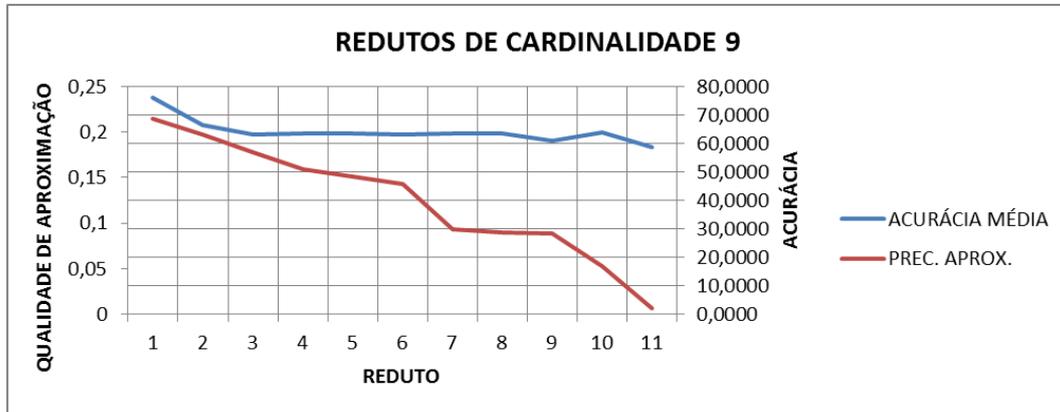


Figura 24 – Gráfico comparativo para redutos de tamanho 9

5.1.9 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 10)

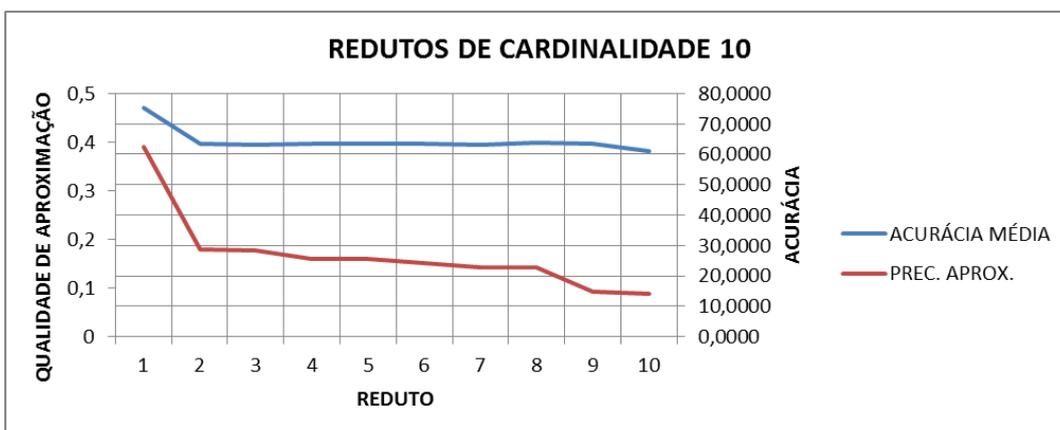


Figura 25 – Gráfico comparativo para redutos de tamanho 10

5.1.10 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 11)

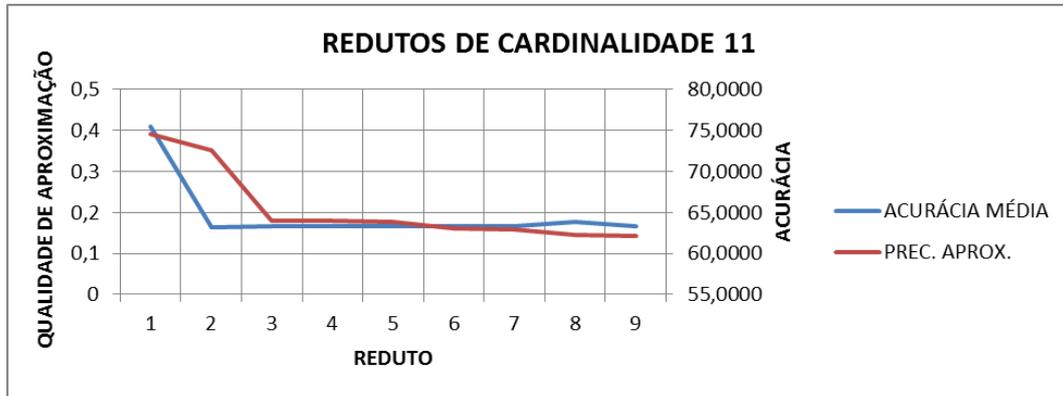


Figura 26 – Gráfico comparativo para redutos de tamanho 11

5.1.11 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 12)

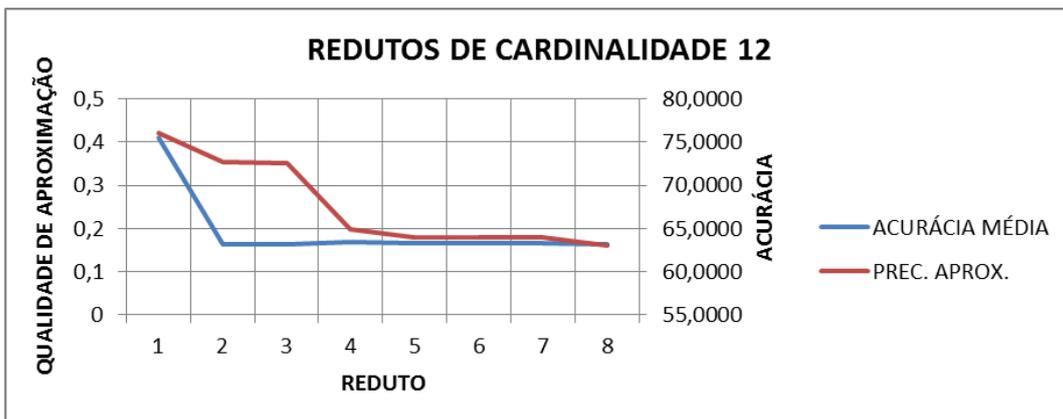


Figura 27 – Gráfico comparativo para redutos de tamanho 12

5.1.12 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 13)

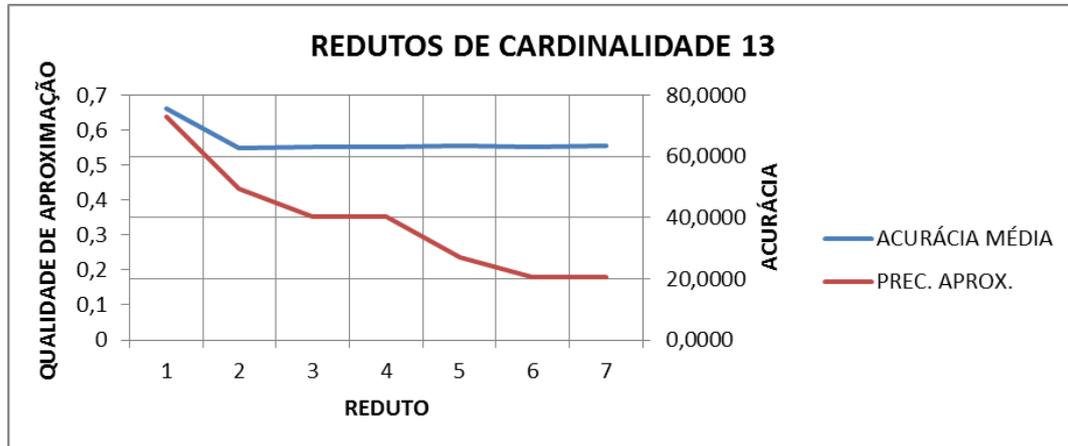


Figura 28 – Gráfico comparativo para redutos de tamanho 13

5.1.13 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 14)

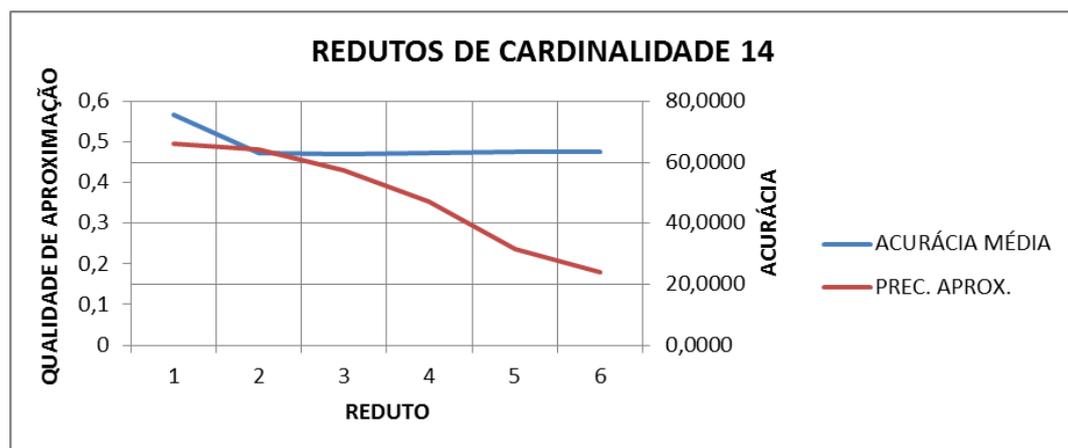


Figura 29 – Gráfico comparativo para redutos de tamanho 14

5.1.14 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 15)

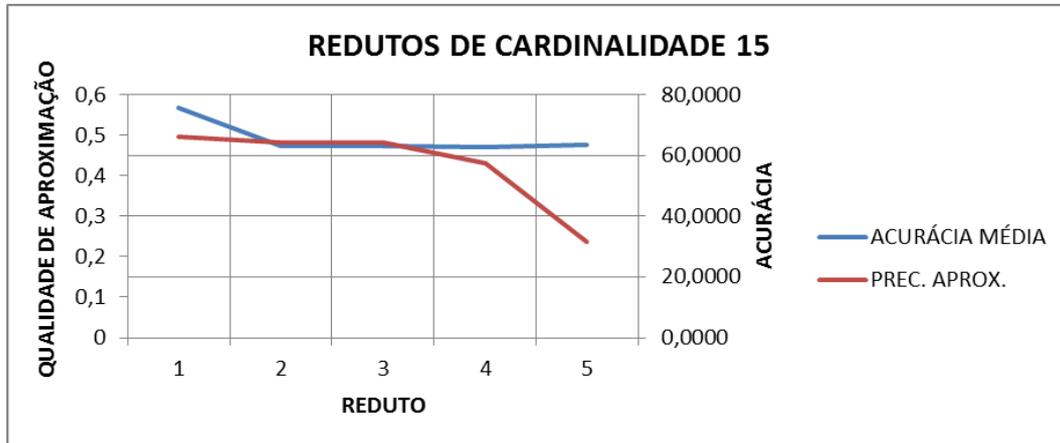


Figura 30 – Gráfico comparativo para redutos de tamanho 15

5.1.15 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 16)

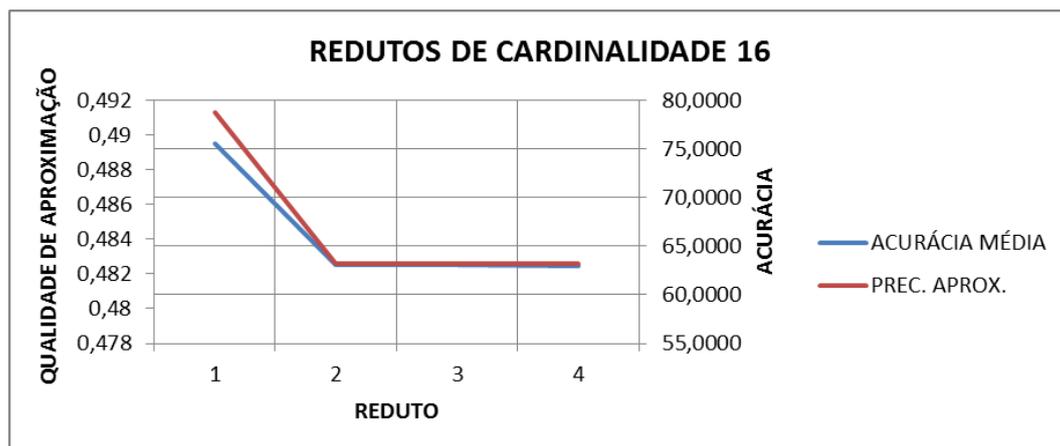


Figura 31 – Gráfico comparativo para redutos de tamanho 16

5.1.16 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 17)

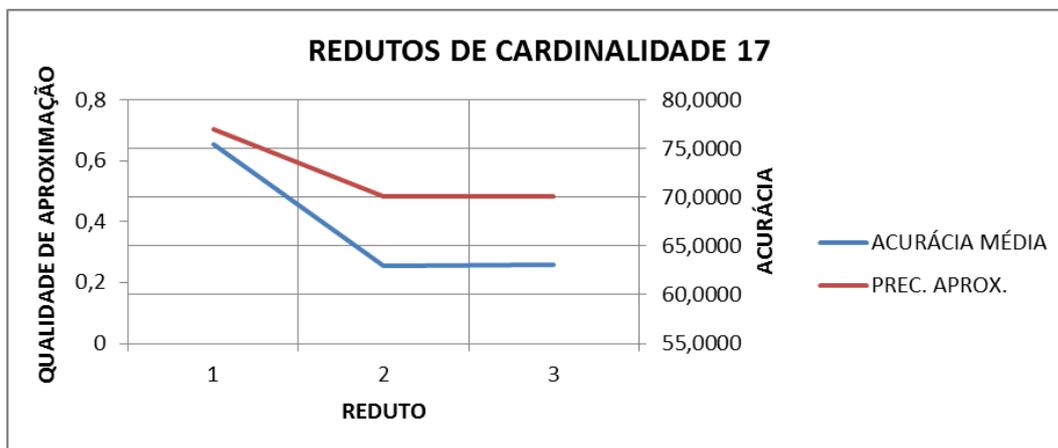


Figura 32 – Gráfico comparativo para redutos de tamanho 17

5.1.17 Comparativo – Qualidade de Aproximação X Acurácia Média (Redutos de Cardinalidade 18)

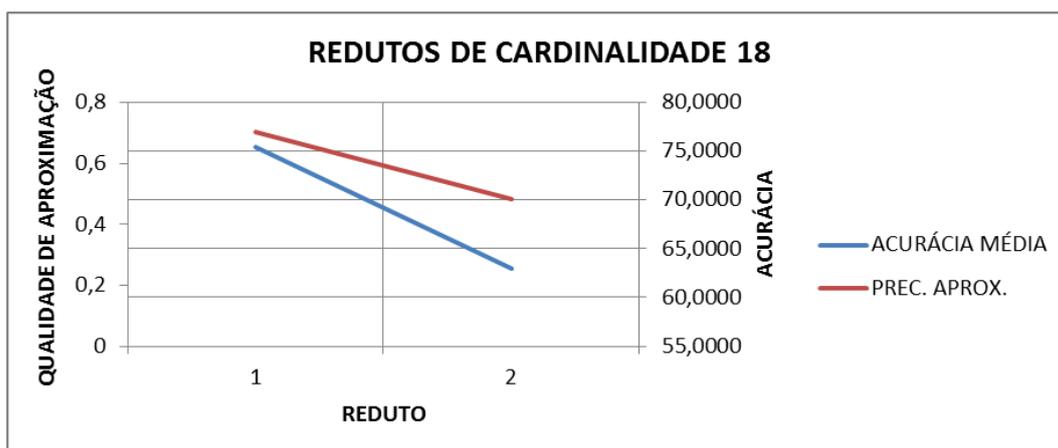


Figura 33 – Gráfico comparativo para redutos de tamanho 18

De modo geral os gráficos mostram o diagnóstico provê uma referência sobre a acurácia. Embora as duas medidas estejam configuradas em unidades diferentes, verifica-se que a curva de acurácia decresce em todos os casos.

Observando em detalhe o Anexo B é possível ver que há muitos casos em que um reduto de menor cardinalidade alcança maior acurácia que outro de maior cardinalidade. Embora esses casos sejam contraintuitivos, são reais. De modo geral entende-se que quanto maior o subconjunto de dados, maior será a sua acurácia. Porém, nem sempre isso acontece e é justamente nesse cenário que RSAPP ajuda o minerador, na escolha de um subconjunto (reduto) ótimo, ou seja, aqueles redutos com maior acurácia e menor custo computacional. Vejamos um exemplo do Anexo B. Um dos redutos de cardinalidade 9 pontuou 0,21 de qualidade de aproximação e um dos redutos de cardinalidade 14 pontuou 0,17. O reduto de tamanho 9 obteve acurácia média de 76% e o de tamanho 14 de 63%. Nesse caso, o reduto menor possui custo computacional inferior e acurácia superior ao reduto de cardinalidade 14. Com o diagnóstico em mãos o técnico tem subsídios de balancear a cardinalidade versus a qualidade de aproximação.

5.2 Diagnóstico RSAPP – Melhores e Piores Redutos e a Variação da Acurácia

Nessa seção fez-se uma heurística diferente da seção anterior, a fim de verificar o quanto o diagnóstico RSAPP pode contribuir na indicação de um subconjunto de dados mais promissor para tarefas de mineração. Para tanto, foram identificados os “melhores” e “piores” redutos segundo a qualidade de aproximação do diagnóstico RSAPP, para cada cardinalidade. Feito isso,

aplicou-se tarefas de mineração a fim de obter a acurácia desses redutos. Ao final, calculou-se a variação de acurácia entre o melhor e o pior reduto.

Vejamos o um dos casos da Tabela 9, onde as variações estão dispostas. Dos redutos de cardinalidade 10, o reduto que tem a melhor qualidade de aproximação pontuou 0,39 e o reduto com pior qualidade de aproximação pontuou 0,08 (como pode ser visto no Anexo A). Nesse caso obteve-se uma variação 23,5663%. Ou seja, para o melhor reduto de cardinalidade 10 se obteve acurácia média⁴ de 75,4104% e para o pior reduto de mesma cardinalidade se obteve 61,0283% de acurácia média. Portanto, um ganho de 23,5663% de variação entre os dois redutos. Neste caso, e nos demais, o diagnóstico mostrou-se eficaz, pois houve uma variação considerável entre o melhor e o pior reduto. Especialmente sobre os redutos de cardinalidade menor como é mostrado na Tabela 9.

Tabela 9 – Variação de acurácia entre os melhores e piores redutos

Cardinalidade do Reduto	Média de Acurácia		
	Melhor Reduto	Pior Reduto	Variação Acurácia (%)
1	75,8500	57,7338	31,3789
2	75,8591	57,7288	31,4061
3	75,8440	57,7338	31,3685
4	75,8641	57,7288	31,4148
5	75,8812	57,6825	31,5498
6	75,9175	58,9671	28,7454
7	75,8621	58,6160	29,4222
8	75,8551	58,6603	29,3125

⁴ Média do valor de acurácia dos algoritmo de mineração J48, *Randon Forest*, *Simple Cart* e *Naive Bayes*.

9	76,0029	58,7810	29,2985
10	75,4104	61,0283	23,5663
11	75,4305	63,2706	19,2190
12	75,4557	63,1800	19,4297
13	75,4507	63,3148	19,1675
14	75,4376	63,3792	19,0258
15	75,4608	63,5311	18,7777
16	75,5764	62,9899	19,9818
17	75,4014	63,0573	19,5760

No Anexo C essa heurística pode ser vista por completa, para cada um dos quatro algoritmos de mineração utilizados.

5.3 Considerações do Capítulo

Neste capítulo foram percorridos os resultados dos testes de mineração, bem como a relação da acurácia obtida nesses testes com a qualidade de aproximação apurada por RSAPP. Duas heurísticas foram usadas para avaliar a relação entre a qualidade de aproximação e acurácia de mineração.

6- CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Os resultados apontados nas seções 5.1 e 5.2 sugerem uma relação entre a qualidade de aproximação, gerada por RSAPP e a acurácia dos modelos de mineração. Embora essa relação não tenha se mantido homogênea para todos os redutos, ainda assim observa-se que quando a qualidade de aproximação foi ordenada, a curva de acurácia acompanhou o decréscimo. Portanto, nesse contexto de pesquisa, a TCA através de RSAPP efetivamente diferenciou os “melhores” e “piores” redutos para tarefas de mineração.

A intensão deste trabalho foi desenvolver um algoritmo capaz de produzir um diagnóstico de auxílio ao técnico minerador. Tal diagnóstico informa algumas métricas para cada um dos redutos do conjunto de dados (neste trabalho, um DW). Visto a proposta deste trabalho e os testes executados, entende-se que RSAPP pode ser usado para produzir uma base de conhecimento sobre os dados, mesmo que o conhecimento de domínio seja superficial. RSAPP vem a contribuir como um artifício sob o contexto de KDD, permitindo que, não somente o conhecimento de domínio possa indicar quais dados tem maior potencial de produzir resultados mais precisos.

O diagnóstico de RSAPP usou apenas métricas da TCA, porém outras métricas, usando outras abordagens poderiam complementar o diagnóstico. Ou seja, o diagnóstico pode ser enriquecido através de outras técnicas disponíveis. Por exemplo, na literatura diversos trabalhos indicam a técnica de Análise de Componentes Principais como uma proposta semelhante à TCA. Nesse

contexto, em trabalhos futuros essa possibilidade pode ser explorada. Ainda, como uma possível evolução deste trabalho, RSAPP pode ser melhorado, quanto a sua complexidade e sua interface. Nesse cenário, outras abordagens de algoritmos/linguagem podem ser desenvolvidas, bem como uma interface GUI para seleção dos atributos dos redutos.

REFERÊNCIAS

- [ALV03] Alves R. C. O. "Mineração de Dados em Sistemas Multidimensionais". Relatório Técnico, Departamento de Informática, Escola de Engenharia, Universidade do Minho, 2003, 10p.
- [AST12] Astah Community. Disponível em: <<http://astah.net/editions/community>>. Acessado em novembro de 2012.
- [BAR10] Barbetta P. A., Reis M. M., Bornia A. C. "Estatística: para cursos de engenharia e informática". São Paulo: Editora Atlas S.A., 2010, 3 ed., 410p.
- [BAR01] Barbieri C., "BI – Business Intelligence: Modelagem & Tecnologia". Rio de Janeiro: Editora Axcel Books, 2001, 1 ed., 244p.
- [CRI03] Wirth R., Hipp J.. "CRISP-DM: Towards a Standard Process Model for Data Mining". Wilhelm-Schickard-Institute, University of Tübingen and DaimlerChrysler Research & Technology FT3/KL, ECML/PKDD, 2003, 11p.
- [COL11] Colares P. "Processo de Indução e Ranqueamento de Árvores de Decisão Sobre Modelos OLAP". Dissertação de Mestrado, Programa de Pós-Graduação em Ciência da Computação,

Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, 2011, 132p

- [DAT99] Date, C. “Introduction To Database Systems”. Boston: Addison Wesley, 1999, 7 ed., 938p.
- [FAY96] Fayyad U., Piatetsky G., Smyth P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, vol. 39-11, p27-34, 1996.
- [GON08] González-Aranda P., Menasalvas E., Millán S., Ruiz C. “Towards a Methodology for Data Mining Project Development: The Importance of Abstraction”. *Data Mining: Foundations and Practice, Studies in Computational Intelligence (SCI)* vol. 118, p. 165–178, 2008.
- [GRE01] Greco S., Matarazzo B., Slowinski R. “Rough sets theory for multicriteria decision analysis”. *European Journal of Operational Research*, vol. 129, p1-47, 2001.
- [HAN97] Han J., Chiang J. Y., Chee S., Chen J., Chen Q., Cheng S., Gong W., Kamber M., Koperski K., Liu G., Lu Y., Stefanovic N., Winstone L., Xia B. B., Zaiane O. R., Zhang S., and Zhu H. “DBMiner: A System for Data Mining in Relational Databases and Data Warehouses”. In: 7th Conference of the Centre For Advanced Studies on Collaborative Research, 1997, 12p.
- [HAN11] Han J., Kamber M., Pei J. “Data Mining: concepts and techniques”. Burlington: Morgan Kaufmann Publishers, 2011, 3 ed., 703p.

- [INM05] Inmon W. H. "Building the Data Warehouse". Indianapolis: Wiley Publishing, 2005, 4 ed., 543p.
- [KIM02] Kimball R., Ross M. "The Data Warehouse Toolkit: the complete guide to dimensional modeling". New York: John Wiley & Sons, 2002, 2 ed., 436p.
- [KLÖ02] Klösgen W. (editor), Zytkow J. M. (editor). "Handbook of Data Mining and Knowledge Discovery". New York: Oxford University Press, 2002, 1 ed., 1026p.
- [KOP02] Kopanas I., Avouris N. M., Daskalaki S. "The Role of Domain Knowledge in a Large Scale Data Mining Project". I.P Vlahavas, C.D. Spyropoulos, *Methods and Applications of Artificial Intelligence, Lecture Notes in AI*, LNAI no. 2308, p. 288–299, Springer-Verlag, 2002.
- [LEU06] Leung Y., Wu W.Z., Zhang W. X., Knowledge Acquisition in Incomplete Information Systems: A Rough Set Approach, *European Journal of Operational Research*, Elsevier, vol. 168, p. 164–180, 2006.
- [NAV05] Navathe S., Elmasri R. "Sistemas de Banco de Dados". São Paulo: Pearson, 2005, 4 ed., 744p.
- [ORA12] Oracle Corporation. Disponível em: <<http://www.oracle.com/index.html>>. Acessado em julho de 2012.
- [PAW82] Pawlak Z., "Rough Sets". *International Journal of Computer and Information Sciences*, vol. 11-5, p. 341–356, 1982.

- [PAW02] Pawlak Z. “Rough Set Theory and its Applications”. *Journal of Telecommunications and Information Technology*, vol. 3, p. 7–10, 2002.
- [SIL09] Silveira J. G., Scherer A. P. Z. “Soluções Business Intelligence, Tecnologia e Utilização: Um estudo de Caso de Sistema de Apoio à Decisão para Gestão de Cartões de Crédito”. Trabalho de Conclusão de Curso, Faculdade Dom Bosco de Porto Alegre, 2009, 93p.
- [SPS00] SPSS, CRISP-DM 1.0: Step-by-step Data Mining Guide. Disponível em: <<http://www.crisp-dm.org>>. Acessado em outubro de 2012.
- [SUR04] Zbigniew S., “An Introduction to Rough Set Theory and Its Applications: A tutorial”. 1st International Computer Engineering Conference. Faculty of Engineering, Cairo University. Cairo, 2004, 39p.
- [TAN06] Tan P. N., Steinbach M., Kumar V. “Introduction to Data Mining”. Boston: Person Education, 2006, 1 ed., 769p.
- [THA09] Thangavel K., Pethalakshmi A. “Dimensionality Reduction Based on Rough Set Theory: A Review”. *Applied Soft Computing*, Elsevier, vol. 9, p. 1–12, 2009.
- [TUR08] Turban E., Sharda R., Aronson J. E., King D. “Business Intelligence: um enfoque gerencial para a inteligência do negócio”. Porto Alegre: Bookman, 2009, 1 ed., 256p.

- [URE11] Regina U. “Overview of the KDD Process”. Disponível em: <http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html> . Acessado em outubro de 2012.
- [WEK12] WEKA: Waikato environment for knowledge analysis. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acessado em julho de 2012.
- [WIT05] Witten I. H., Frank E. “Data Mining: Pratical Machine Learning Tools and Tchniques”. San Francisco: Morgan Kaufmann Publishers, 2005. 2 ed., 525p.

ANEXO A

Diagnóstico de RSAPP da pesquisa:

Reduto	Cardinalidade Reduto	Aproximação Inferior	Qualidade da Aproximação
Atributo_18	1	0	0,00000000000000000000
Atributo_17	1	0	0,00000000000000000000
Atributo_1	1	58	0,0023338161918557900
Atributo_3	1	1	0,0000402382102044101
Atributo_16	1	0	0,00000000000000000000
Atributo_6	1	0	0,00000000000000000000
Atributo_7	1	0	0,00000000000000000000
Atributo_4	1	2	0,0000804764204088202
Atributo_5	1	4	0,0001609528408176400
Atributo_11	1	0	0,00000000000000000000
Atributo_2	1	1	0,0000402382102044101
Atributo_10	1	24	0,0009657170449058430
Atributo_15	1	0	0,00000000000000000000
Atributo_12	1	0	0,00000000000000000000
Atributo_9	1	36	0,0014485755673587600
Atributo_19	1	0	0,00000000000000000000
Atributo_14	1	0	0,00000000000000000000
Atributo_8	1	0	0,00000000000000000000
Atributo_13	1	0	0,00000000000000000000
Atributo_7 + Atributo_8	2	8	0,0003219056816352810
Atributo_3 + Atributo_4	2	10	0,0004023821020441010
Atributo_16 + Atributo_17	2	57	0,0022935779816513800
Atributo_8 + Atributo_9	2	526	0,0211652985675197000
Atributo_13 + Atributo_14	2	0	0,00000000000000000000
Atributo_6 + Atributo_7	2	4	0,0001609528408176400

Atributo_9 + Atributo_10	2	268	0,0107838403347819000
Atributo_2 + Atributo_3	2	7	0,0002816674714308710
Atributo_15 + Atributo_16	2	0	0,0000000000000000000
Atributo_18 + Atributo_19	2	0	0,0000000000000000000
Atributo_17 + Atributo_18	2	0	0,0000000000000000000
Atributo_12 + Atributo_13	2	0	0,0000000000000000000
Atributo_4 + Atributo_5	2	6	0,0002414292612264610
Atributo_14 + Atributo_15	2	0	0,0000000000000000000
Atributo_10 + Atributo_11	2	43	0,0017302430387896300
Atributo_5 + Atributo_6	2	11	0,0004426203122485110
Atributo_11 + Atributo_12	2	0	0,0000000000000000000
Atributo_1 + Atributo_2	2	1518	0,0610816030902945000
Atributo_16 + Atributo_17 + Atributo_18	3	57	0,0022935779816513800
Atributo_7 + Atributo_8 + Atributo_9	3	93	0,0037421535490101400
Atributo_8 + Atributo_9 + Atributo_10	3	447	0,0179864799613713000
Atributo_10 + Atributo_11 + Atributo_12	3	73	0,0029373893449219400
Atributo_1 + Atributo_2 + Atributo_3	3	978	0,0393529695799131000
Atributo_2 + Atributo_3 + Atributo_4	3	16	0,0006438113632705620
Atributo_14 + Atributo_15 + Atributo_16	3	0	0,0000000000000000000
Atributo_6 + Atributo_7 + Atributo_8	3	20	0,0008047642040882020
Atributo_3 + Atributo_4 + Atributo_5	3	22	0,0008852406244970220
Atributo_5 + Atributo_6 + Atributo_7	3	43	0,0017302430387896300
Atributo_9 + Atributo_10 + Atributo_11	3	373	0,0150088524062450000

Atributo_13 + Atributo_14 + Atributo_15	3	0	0,000000000000000000
Atributo_4 + Atributo_5 + Atributo_6	3	15	0,0006035731530661520
Atributo_17 + Atributo_18 + Atributo_19	3	0	0,000000000000000000
Atributo_12 + Atributo_13 + Atributo_14	3	0	0,000000000000000000
Atributo_15 + Atributo_16 + Atributo_17	3	57	0,0022935779816513800
Atributo_11 + Atributo_12 + Atributo_13	3	0	0,000000000000000000
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	4	0	0,000000000000000000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	4	569	0,0228955416063094000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5	4	31	0,0012473845163367100
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	4	89	0,0035812007081925000
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	4	0	0,000000000000000000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	4	83	0,0033397714469660400
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4	4	2567	0,1032914855947210000
Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	4	57	0,0022935779816513800
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	4	202	0,0081281184612908400

Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	4	99	0,0039835828102366000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6	4	41	0,0016497666183808100
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	4	2	0,0000804764204088202
Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	4	57	0,0022935779816513800
Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	4	57	0,0022935779816513800
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	4	447	0,0179864799613713000
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	4	1650	0,0663930468372767000
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	5	0	0,0000000000000000000
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	5	180	0,0072428778367938200
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	5	1767	0,0711009174311927000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	5	640	0,0257524545308225000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5	5	2143	0,0862304844680509000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 +	5	47	0,0018911958796072800

Atributo_6			
Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	5	57	0,0022935779816513800
Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	5	57	0,0022935779816513800
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	5	139	0,0055931112184130000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	5	590	0,0237405440206020000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	5	400	0,0160952840817640000
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	5	6	0,0002414292612264610
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	5	1650	0,0663930468372767000
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	5	66	0,0026557218734910700
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	5	158	0,0063576372122968000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	6	590	0,0237405440206020000

Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	6	1767	0,0711009174311927000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6	6	3670	0,1476742314501850000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	6	149	0,0059954933204571100
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	6	66	0,0026557218734910700
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	6	113	0,0045469177530983400
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	6	2198	0,0884435860292934000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	6	233	0,0093755029776275600
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	6	2858	0,1150008047642040000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	6	594	0,0239014968614196000
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	6	341	0,0137212296797038000

Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	6	7	0,0002816674714308710
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	6	660	0,0265572187349107000
Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	6	57	0,0022935779816513800
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	7	66	0,0026557218734910700
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	7	113	0,0045469177530983400
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	7	626	0,0251891195879607000
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	7	166	0,0066795428939320800
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 +	7	246	0,0098985997102848900

Atributo_8			
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	7	2285	0,0919443103170771000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	7	660	0,0265572187349107000
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	7	2222	0,0894093030741993000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	7	590	0,0237405440206020000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	7	3977	0,1600273619829390000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	7	2858	0,1150008047642040000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	7	3743	0,1506116207951070000
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	7	342	0,0137614678899083000

+ Atributo_15 + Atributo_16			
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	8	349	0,0140431353613391000
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	8	113	0,0045469177530983400
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	8	3811	0,1533478190890070000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	8	3743	0,1506116207951070000
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	8	166	0,0066795428939320800
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	8	640	0,0257524545308225000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	8	660	0,0265572187349107000

+ Atributo_15			
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	8	2309	0,0929100273619829000
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	8	2222	0,0894093030741993000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	8	1249	0,0502575245453082000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	8	3752	0,1509737646869470000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	8	3518	0,1415580234991150000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	9	3542	0,1425237405440210000

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	9	4898	0,1970867535812010000
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	9	2195	0,0883228713986802000
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	9	2309	0,0929100273619829000
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	9	2222	0,0894093030741993000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	9	1303	0,0524303878963464000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	9	3952	0,1590214067278290000

Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	9	3754	0,1510542411073560000
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	9	166	0,0066795428939320800
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	9	4413	0,1775712216320620000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	9	5340	0,2148720424915500000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	10	3968	0,1596652180910990000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	10	3754	0,1510542411073560000

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	10	4418	0,1777724126830840000
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	10	2195	0,0883228713986802000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	10	3952	0,1590214067278290000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	10	3542	0,1425237405440210000
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	10	3519	0,1415982617093190000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	10	4437	0,1785369386769680000

+ Atributo_12 + Atributo_13			
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	10	9721	0,3911556413970710000
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	10	2309	0,0929100273619829000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	11	4418	0,1777724126830840000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	11	3542	0,1425237405440210000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	11	3968	0,1596652180910990000

Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	11	8741	0,3517221953967490000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	11	4442	0,1787381297279900000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	11	9721	0,3911556413970710000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	11	4437	0,1785369386769680000
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	11	3952	0,1590214067278290000

Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	11	3582	0,1441332689521970000
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	12	3968	0,1596652180910990000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	12	4442	0,1787381297279900000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	12	4442	0,1787381297279900000
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	12	8741	0,3517221953967490000

+ Atributo_16 + Atributo_17 + Atributo_18			
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	12	4925	0,1981731852567200000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	12	10460	0,4208916787381300000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	12	4437	0,1785369386769680000
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	12	8765	0,3526879124416550000

Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	13	8741	0,3517221953967490000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	13	5884	0,2367616288427490000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	13	4442	0,1787381297279900000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	13	4442	0,1787381297279900000

Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	13	8765	0,3526879124416550000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	13	10710	0,4309512312892320000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	13	15870	0,6385803959439880000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	14	11993	0,4825768549814900000

Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	14	12343	0,4966602285530340000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	14	5886	0,2368421052631580000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	14	10710	0,4309512312892320000
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	14	8765	0,3526879124416550000

Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	14	4442	0,1787381297279900000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	15	11993	0,4825768549814900000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	15	12321	0,4957749879285370000
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	15	10710	0,4309512312892320000

Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	15	5886	0,2368421052631580000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	15	11993	0,4825768549814900000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	16	12209	0,4912683083856430000
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	16	11993	0,4825768549814900000

+ Atributo_16 + Atributo_17 + Atributo_18			
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	16	11993	0,4825768549814900000
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	16	11993	0,4825768549814900000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 +	17	17486	0,7036053436343150000

Atributo_17			
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	17	11993	0,4825768549814900000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	17	11993	0,4825768549814900000
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 +	18	11993	0,4825768549814900000

Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19			
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	18	17486	0,7036053436343150000
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	19	17486	0,7036053436343150000

ANEXO B

Tarefas de mineração aplicadas sobre os redutos.

Reduto	Cardinalidade Reduto	Aproximação Inferior	Qualidade de Aproximação	J48	Randon Forest	Simple Cart	Naive Bayes	Média
Atributo_1	1	58	0,0023338161918557900	75,8490	75,8490	75,8531	75,8490	75,8500
Atributo_9	1	36	0,0014485755673587600	64,9525	64,9364	64,9525	57,7338	63,1438
Atributo_10	1	24	0,0009657170449058430	61,8582	61,7455	61,8421	57,7056	60,7879
Atributo_5	1	4	0,0001609528408176400	57,7338	57,6976	57,7338	57,7016	57,7167
Atributo_4	1	2	0,0000804764204088202	57,7338	57,7418	57,7338	57,7418	57,7378
Atributo_2	1	1	0,0000402382102044101	57,7257	57,7418	57,7499	57,7338	57,7378
Atributo_3	1	1	0,0000402382102044101	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_11	1	0	0,0000000000000000000	57,7137	57,7137	57,6654	57,7137	57,7016
Atributo_18	1	0	0,0000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_17	1	0	0,0000000000000000000	57,7338	57,5165	57,7338	57,5567	57,6352
Atributo_13	1	0	0,0000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_14	1	0	0,0000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_8	1	0	0,0000000000000000000	57,6493	57,7096	57,6654	57,7096	57,6835

Atributo_16	1	0	0,00000000000000000000	58,8323	58,8323	58,8323	58,8323	58,8323
Atributo_19	1	0	0,00000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_15	1	0	0,00000000000000000000	57,7338	57,7338	57,7338	58,6995	57,9752
Atributo_6	1	0	0,00000000000000000000	57,8746	57,8746	57,8746	57,8746	57,8746
Atributo_7	1	0	0,00000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_12	1	0	0,00000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_1 + Atributo_2	2	1518	0,0610816030902945000	75,8410	75,8772	75,8490	75,8691	75,8591
Atributo_8 + Atributo_9	2	526	0,0211652985675197000	65,3911	65,3670	65,3589	57,4722	63,3973
Atributo_9 + Atributo_10	2	268	0,0107838403347819000	65,0370	64,7634	65,0089	57,7096	63,1297
Atributo_16 + Atributo_17	2	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,6995	58,6925
Atributo_10 + Atributo_11	2	43	0,0017302430387896300	61,9709	61,8180	61,9145	57,7137	60,8543
Atributo_5 + Atributo_6	2	11	0,0004426203122485110	57,8625	57,8384	57,8505	57,8465	57,8495
Atributo_3 + Atributo_4	2	10	0,0004023821020441010	57,7338	57,7338	57,7338	57,7418	57,7358
Atributo_7 + Atributo_8	2	8	0,0003219056816352810	57,6614	57,7459	57,7660	57,6412	57,7036
Atributo_2 + Atributo_3	2	7	0,0002816674714308710	57,7257	57,7338	57,7459	57,7338	57,7348
Atributo_4 + Atributo_5	2	6	0,0002414292612264610	57,7338	57,6895	57,6935	57,7096	57,7066
Atributo_6 + Atributo_7	2	4	0,0001609528408176400	57,8465	57,8907	57,8465	57,8746	57,8646
Atributo_15 + Atributo_16	2	0	0,00000000000000000000	58,8323	58,8323	58,8323	58,8323	58,8323
Atributo_13 + Atributo_14	2	0	0,00000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_17 + Atributo_18	2	0	0,00000000000000000000	57,7338	57,5165	57,7338	57,5567	57,6352

Atributo_11 + Atributo_12	2	0	0,00000000000000000000	57,6815	57,7016	57,6573	57,7177	57,6895
Atributo_18 + Atributo_19	2	0	0,00000000000000000000	57,6815	57,7338	57,7338	57,7338	57,7207
Atributo_14 + Atributo_15	2	0	0,00000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_12 + Atributo_13	2	0	0,00000000000000000000	57,7338	57,7217	57,7257	57,7338	57,7288
Atributo_1 + Atributo_2 + Atributo_3	3	978	0,0393529695799131000	75,8370	75,8209	75,8531	75,8651	75,8440
Atributo_8 + Atributo_9 + Atributo_10	3	447	0,0179864799613713000	65,4635	65,1457	65,5400	57,4682	63,4044
Atributo_9 + Atributo_10 + Atributo_11	3	373	0,0150088524062450000	65,4233	65,1376	65,5883	57,4642	63,4034
Atributo_7 + Atributo_8 + Atributo_9	3	93	0,0037421535490101400	65,3790	65,1859	65,3790	59,5365	63,8701
Atributo_10 + Atributo_11 + Atributo_12	3	73	0,0029373893449219400	61,9709	61,6369	61,9186	57,7177	60,8110
Atributo_16 + Atributo_17 + Atributo_18	3	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,6995	58,6925
Atributo_15 + Atributo_16 + Atributo_17	3	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,6995	58,6925
Atributo_5 + Atributo_6 + Atributo_7	3	43	0,0017302430387896300	58,7236	57,8907	57,8344	57,8304	58,0698

Atributo_3 + Atributo_4 + Atributo_5	3	22	0,0008852406244970220	57,7338	57,6734	57,6935	57,7177	57,7046
Atributo_6 + Atributo_7 + Atributo_8	3	20	0,0008047642040882020	57,7740	57,8867	57,9108	57,7700	57,8354
Atributo_2 + Atributo_3 + Atributo_4	3	16	0,0006438113632705620	57,7257	57,7338	57,7459	57,7418	57,7368
Atributo_4 + Atributo_5 + Atributo_6	3	15	0,0006035731530661520	57,8625	57,8183	57,8465	57,8545	57,8455
Atributo_12 + Atributo_13 + Atributo_14	3	0	0,0000000000000000000	57,7338	57,7217	57,7257	57,7338	57,7288
Atributo_11 + Atributo_12 + Atributo_13	3	0	0,0000000000000000000	57,6815	57,6895	57,6412	57,7177	57,6825
Atributo_17 + Atributo_18 + Atributo_19	3	0	0,0000000000000000000	57,7338	57,5165	57,7338	57,3877	57,5930
Atributo_14 + Atributo_15 + Atributo_16	3	0	0,0000000000000000000	58,8323	58,8323	58,8323	58,8323	58,8323
Atributo_13 + Atributo_14 + Atributo_15	3	0	0,0000000000000000000	57,7338	57,7338	57,7338	57,7338	57,7338
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4	4	2567	0,1032914855947210000	75,8571	75,8812	75,8450	75,8732	75,8641

Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	4	1650	0,0663930468372767000	65,4515	63,3873	65,2624	59,5767	63,4195
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	4	569	0,0228955416063094000	65,4434	64,7030	65,5923	57,4602	63,2997
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	4	447	0,0179864799613713000	65,4193	65,1497	65,5400	57,7217	63,4577
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	4	202	0,0081281184612908400	65,5078	65,2986	65,4676	59,0415	63,8289
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	4	99	0,0039835828102366000	58,0879	58,0597	57,9631	57,8344	57,9863
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	4	89	0,0035812007081925000	61,9709	61,6208	61,9186	57,7177	60,8070
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	4	83	0,0033397714469660400	58,1120	58,1845	58,1643	58,1160	58,1442
Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	4	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,3213	58,5979
Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	4	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,6995	58,6925
Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	4	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,6995	58,6925

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6	4	41	0,0016497666183808100	57,8625	57,8102	57,8424	57,8585	57,8434
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5	4	31	0,0012473845163367100	57,7217	57,6775	57,7298	57,7298	57,7147
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	4	2	0,0000804764204088202	58,8162	58,8283	58,8202	58,8283	58,8233
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	4	0	0,0000000000000000000	57,6815	57,6895	57,6412	57,7177	57,6825
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	4	0	0,0000000000000000000	57,7338	57,7217	57,7257	57,7338	57,7288
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5	5	2143	0,0862304844680509000	75,8490	75,7967	75,8128	75,8812	75,8349
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	5	1767	0,0711009174311927000	65,5521	63,4033	65,3549	59,6371	63,4869
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	5	1650	0,0663930468372767000	65,4354	63,4033	65,2342	59,6853	63,4396
Atributo_8 + Atributo_9 +	5	640	0,0257524545308225000	65,4314	64,7594	65,5601	57,7217	63,3682

Atributo_10 + Atributo_11 + Atributo_12								
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	5	590	0,0237405440206020000	65,4434	64,6910	65,5883	57,4602	63,2957
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	5	400	0,0160952840817640000	65,4676	65,2382	65,4676	59,7095	63,9707
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	5	180	0,0072428778367938200	61,9709	61,6208	61,9186	57,7177	60,8070
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	5	158	0,0063576372122968000	58,2005	58,2126	58,1764	58,0396	58,1573
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	5	139	0,0055931112184130000	58,1201	58,1080	58,1965	58,1281	58,1382
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	5	66	0,0026557218734910700	58,7236	58,5868	58,8323	58,6955	58,7096

Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	5	57	0,0022935779816513800	58,7236	58,5989	58,7478	58,6955	58,6915
Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	5	57	0,0022935779816513800	58,7236	58,5828	57,7338	58,3213	58,3404
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6	5	47	0,0018911958796072800	57,8625	57,8263	57,8585	57,8706	57,8545
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	5	6	0,0002414292612264610	58,8001	57,8263	58,8041	58,8363	58,5667
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	5	0	0,0000000000000000000	57,6815	57,6895	57,6412	57,7177	57,6825
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6	6	3670	0,1476742314501850000	75,9456	75,8531	75,9738	75,8973	75,9175
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	6	2858	0,1150008047642040000	65,4595	62,9487	65,3589	59,2427	63,2525

Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	6	2198	0,0884435860292934000	65,3589	63,0573	65,2342	59,6612	63,3279
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	6	1767	0,0711009174311927000	65,5561	63,4798	65,3549	59,6894	63,5201
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	6	660	0,0265572187349107000	65,4273	64,7513	65,5601	57,7217	63,3651
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	6	594	0,0239014968614196000	65,4837	65,2020	65,4676	60,0596	64,0532
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	6	590	0,0237405440206020000	65,4434	64,6829	65,5883	57,4602	63,2937
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	6	341	0,0137212296797038000	61,9709	61,6127	61,9186	57,7177	60,8050
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	6	233	0,0093755029776275600	58,2086	58,1160	58,2166	58,0557	58,1492

Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	6	149	0,0059954933204571100	58,1040	58,1201	58,1845	58,1563	58,1412
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	6	113	0,0045469177530983400	58,8484	58,6915	58,7719	58,6955	58,7518
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	6	66	0,0026557218734910700	58,7236	58,5868	58,8323	58,6955	58,7096
Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	6	57	0,0022935779816513800	58,7236	58,5949	58,7478	58,6955	58,6905
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	6	7	0,0002816674714308710	59,0979	59,0737	59,0697	58,6271	58,9671
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7	7	3977	0,1600273619829390000	75,9295	75,6961	75,9255	75,8973	75,8621

Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	7	3743	0,1506116207951070000	65,6607	62,5141	65,5239	59,8543	63,3883
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	7	2858	0,1150008047642040000	65,4515	63,0010	65,3589	59,3393	63,2877
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	7	2285	0,0919443103170771000	65,5440	63,1177	65,3308	59,6371	63,4074
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	7	2222	0,0894093030741993000	65,3670	63,0734	65,2422	59,6692	63,3380
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	7	660	0,0265572187349107000	65,4273	64,7513	65,5601	57,7217	63,3651

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	7	626	0,0251891195879607000	65,4796	65,1416	65,4676	60,0475	64,0341
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	7	590	0,0237405440206020000	65,4434	64,7151	65,5883	58,5305	63,5693
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	7	342	0,0137614678899083000	62,6469	61,5363	62,6791	58,5305	61,3482
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	7	246	0,0098985997102848900	58,1885	58,1402	58,2126	58,0839	58,1563
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	7	166	0,0066795428939320800	59,1260	58,8524	58,9892	58,5506	58,8796

Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	7	113	0,0045469177530983400	58,8484	58,6915	58,7719	58,6955	58,7518
Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	7	66	0,0026557218734910700	58,7236	58,5909	58,8323	58,3172	58,6160
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8	8	3811	0,1533478190890070000	75,9295	75,6599	75,9255	75,9054	75,8551
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	8	3752	0,1509737646869470000	65,6366	62,5583	65,5239	59,8503	63,3923
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	8	3743	0,1506116207951070000	65,6647	62,5342	65,5239	59,9308	63,4134

Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	8	3518	0,1415580234991150000	65,3388	62,6388	65,3106	59,3312	63,1549
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	8	2309	0,0929100273619829000	65,5641	63,0653	65,3308	59,6250	63,3963
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	8	2222	0,0894093030741993000	65,3670	63,0291	65,2422	59,6692	63,3269
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	8	1249	0,0502575245453082000	65,7331	64,7433	65,8659	58,4420	63,6961
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	8	660	0,0265572187349107000	65,4273	64,7352	65,5601	57,7217	63,3611

Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	8	640	0,0257524545308225000	65,4796	65,1296	65,4635	60,0716	64,0361
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	8	349	0,0140431353613391000	62,7394	60,5424	62,6469	58,4420	61,0927
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	8	166	0,0066795428939320800	59,1260	58,8524	58,9892	58,5506	58,8796
Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	8	113	0,0045469177530983400	58,8484	58,6915	58,7719	58,3293	58,6603
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9	9	5340	0,2148720424915500000	76,4727	75,3621	76,4204	75,7565	76,0029

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	9	4898	0,1970867535812010000	65,5963	75,4788	65,5239	59,9227	66,6304
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	9	4413	0,1775712216320620000	65,4072	62,3209	65,5239	59,9348	63,2967
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	9	3952	0,1590214067278290000	65,6607	63,0734	65,6446	59,0697	63,3621
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	9	3754	0,1510542411073560000	65,6366	62,5624	65,5239	59,8463	63,3923

Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	9	3542	0,1425237405440210000	65,3549	62,5664	65,3106	59,3353	63,1418
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	9	2309	0,0929100273619829000	65,5641	63,0130	65,3308	59,6250	63,3832
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	9	2222	0,0894093030741993000	65,3670	63,0855	65,2422	59,6692	63,3410
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	9	2195	0,0883228713986802000	62,7394	60,5223	62,6469	58,4420	61,0877

Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	9	1303	0,0524303878963464000	65,6889	64,7875	65,8056	58,7116	63,7484
Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	9	166	0,0066795428939320800	59,1260	58,8524	58,9892	58,1563	58,7810
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10	10	9721	0,3911556413970710000	76,5170	73,1209	76,3077	75,6961	75,4104
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	10	4437	0,1785369386769680000	65,4394	62,3370	65,5239	59,9388	63,3098

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	10	4418	0,1777724126830840000	65,4233	62,2807	65,5239	59,9429	63,2927
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	10	3968	0,1596652180910990000	65,6527	63,0010	65,6245	59,0455	63,3309
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	10	3952	0,1590214067278290000	65,6607	62,9929	65,6446	59,0697	63,3420
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	10	3754	0,1510542411073560000	65,5963	62,5583	65,5239	59,9509	63,4074

Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	10	3542	0,1425237405440210000	65,3549	62,5865	65,3106	59,3353	63,1468
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	10	3519	0,1415982617093190000	65,6406	62,9889	65,6205	60,6309	63,7202
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	10	2309	0,0929100273619829000	65,5641	63,0050	65,3308	59,6250	63,3812
Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	10	2195	0,0883228713986802000	62,7394	60,5344	62,6469	58,1925	61,0283

Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11	11	9721	0,3911556413970710000	76,5371	73,1933	76,3077	75,6840	75,4305
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	11	8741	0,3517221953967490000	65,6567	60,9408	65,5078	60,3935	63,1247
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	11	4442	0,1787381297279900000	65,4032	62,3209	65,5239	59,9469	63,2987
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 +	11	4437	0,1785369386769680000	65,4394	62,3008	65,5239	59,9388	63,3007

Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14								
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	11	4418	0,1777724126830840000	65,4112	62,3209	65,4998	59,9469	63,2947
Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	11	3968	0,1596652180910990000	65,6527	63,0573	65,6245	59,0455	63,3450
Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	11	3952	0,1590214067278290000	65,6245	63,0975	65,6446	58,7236	63,2726

Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	11	3582	0,1441332689521970000	65,7130	63,0895	65,6245	60,6390	63,7665
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	11	3542	0,1425237405440210000	65,3549	62,4779	65,3106	59,9388	63,2706
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12	12	10460	0,4208916787381300000	76,6095	73,0726	76,4486	75,6921	75,4557

Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	12	8765	0,3526879124416550000	65,5843	60,9528	65,4917	60,4499	63,1197
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	12	8741	0,3517221953967490000	65,6567	60,9327	65,5078	60,3935	63,1227
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	12	4925	0,1981731852567200000	65,5279	62,2686	65,6406	60,3050	63,4355

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	12	4442	0,1787381297279900000	65,4032	62,2686	65,5239	59,9469	63,2857
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	12	4442	0,1787381297279900000	65,3951	62,3209	65,4998	59,9590	63,2937
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	12	4437	0,1785369386769680000	65,4394	62,4215	65,5239	59,9388	63,3309

Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	12	3968	0,1596652180910990000	65,6245	63,0090	65,6245	58,4621	63,1800
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13	13	15870	0,6385803959439880000	76,6136	73,0484	76,4486	75,6921	75,4507
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	13	10710	0,4309512312892320000	65,4314	60,1803	65,4837	60,0998	62,7988

Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	13	8765	0,3526879124416550000	65,5843	61,0253	65,4917	60,4499	63,1378
Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	13	8741	0,3517221953967490000	65,6607	60,7758	65,5078	59,9308	62,9688
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	13	5884	0,2367616288427490000	65,7211	61,8421	65,7170	60,7838	63,5160

Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	13	4442	0,1787381297279900000	65,3951	62,2244	65,4998	59,9590	63,2696
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	13	4442	0,1787381297279900000	65,4032	62,3853	65,5239	59,9469	63,3148
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14	14	12343	0,4966602285530340000	76,6136	72,9961	76,4486	75,6921	75,4376

Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	14	11993	0,4825768549814900000	65,4515	60,5223	65,5561	60,6511	63,0453
Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	14	10710	0,4309512312892320000	65,4314	60,3734	65,4837	60,0998	62,8471
Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	14	8765	0,3526879124416550000	65,6084	60,9287	65,4917	59,9268	62,9889

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	14	5886	0,2368421052631580000	65,6084	61,9226	65,6969	60,8040	63,5080
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15	14	4442	0,1787381297279900000	65,7412	62,3169	65,4998	59,9590	63,3792
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 +	15	12321	0,4957749879285370000	76,6136	73,0887	76,4486	75,6921	75,4608

Atributo_15								
Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	15	11993	0,4825768549814900000	65,4515	60,5102	65,5561	60,6511	63,0422
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	15	11993	0,4825768549814900000	65,4756	60,4941	65,5561	60,6752	63,0503

Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	15	10710	0,4309512312892320000	65,4233	60,4660	65,4837	59,7859	62,7897
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16	15	5886	0,2368421052631580000	65,7412	61,8421	65,7452	60,7959	63,5311
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 +	16	12209	0,4912683083856430000	76,9556	72,6340	76,8067	75,9094	75,5764

Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16								
Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	16	11993	0,4825768549814900000	65,4756	60,3855	65,5561	60,6752	63,0231
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	16	11993	0,4825768549814900000	65,4676	60,5625	65,5561	60,6752	63,0654

Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	16	11993	0,4825768549814900000	65,4515	60,6108	65,5561	60,3412	62,9899
Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17	17	17486	0,7036053436343150000	76,7061	72,3443	76,6659	75,8893	75,4014

Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	17	11993	0,4825768549814900000	65,4756	60,6511	65,5561	60,3573	63,0100
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	17	11993	0,4825768549814900000	65,4676	60,5303	65,5561	60,6752	63,0573

Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18	18	17486	0,7036053436343150000	76,7061	72,3322	76,6659	75,8893	75,3984
Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	18	11993	0,4825768549814900000	65,4676	60,5666	65,5561	60,3452	62,9839

Atributo_1 + Atributo_2 + Atributo_3 + Atributo_4 + Atributo_5 + Atributo_6 + Atributo_7 + Atributo_8 + Atributo_9 + Atributo_10 + Atributo_11 + Atributo_12 + Atributo_13 + Atributo_14 + Atributo_15 + Atributo_16 + Atributo_17 + Atributo_18 + Atributo_19	19	17486	0,7036053436343150000	76,7101	71,9379	76,6659	75,7967	60,3628
---	----	-------	-----------------------	---------	---------	---------	---------	---------

ANEXO C

Comparativo dos resultados de mineração para melhor/pior reduto apontado pelo diagnóstico RSAPP, usando o algoritmo de mineração J48.

Cardinalidade do Reduto	J48			
	Melhor	Pior	Varição Acurácia	Varição Acurácia (%)
1	75,8490	57,7338	18,1152	31,3771
2	75,8410	57,7338	18,1072	31,3633
3	75,8370	57,7338	18,1032	31,3563
4	75,8571	57,7338	18,1233	31,3911
5	75,8490	57,6815	18,1675	31,4962
6	75,9456	59,0979	16,8477	28,5081
7	75,9295	58,7236	17,2059	29,2998
8	75,9295	58,8484	17,0811	29,0256
9	76,4727	59,1260	17,3467	29,3385
10	76,5170	62,7394	13,7776	21,9600
11	76,5371	65,3549	11,1822	17,1100
12	76,6095	65,6245	10,9850	16,7392
13	76,6136	65,4032	11,2104	17,1404
14	76,6136	65,7412	10,8724	16,5382
15	76,6136	65,7412	10,8724	16,5382
16	76,9556	65,4515	11,5041	17,5765
17	76,7061	65,4676	11,2385	17,1665

Comparativo dos resultados de mineração para melhor/pior reduto apontado pelo diagnóstico RSAPP, usando o algoritmo de mineração *Randon Forest*.

Cardinalidade do Reduto	Randon Forest			
	Melhor	Pior	Varição Acurácia	Varição Acurácia (%)
1	75,8490	57,7338	18,1152	31,3771
2	75,8772	57,7217	18,1555	31,4535
3	75,8209	57,7338	18,0871	31,3284
4	75,8812	57,7217	18,1595	31,4604
5	75,7967	57,6895	18,1072	31,3873
6	75,8531	59,0737	16,7794	28,4042
7	75,6961	58,5909	17,1052	29,1943
8	75,6599	58,6915	16,9684	28,9112
9	75,3621	58,8524	16,5097	28,0527
10	73,1209	60,5344	12,5865	20,7923
11	73,1933	62,4779	10,7154	17,1507
12	73,0726	63,0090	10,0636	15,9717
13	73,0484	62,3853	10,6631	17,0923
14	72,9961	62,3169	10,6792	17,1369
15	73,0887	61,8421	11,2466	18,1860
16	72,6340	60,6108	12,0232	19,8367
17	72,3443	60,5303	11,8140	19,5175

Comparativo dos resultados de mineração para melhor/pior reduto apontado pelo diagnóstico RSAPP, usando o algoritmo de mineração *Simple Cart*.

Cardinalidade do Reduto	Simple Cart			
	Melhor	Pior	Varição Acurácia	Varição Acurácia (%)
1	75,8531	57,7338	18,1193	31,3842
2	75,8490	57,7257	18,1233	31,3955
3	75,8531	57,7338	18,1193	31,3842
4	75,8450	57,7257	18,1193	31,3886
5	75,8128	57,6412	18,1716	31,5254
6	75,9738	59,0697	16,9041	28,6172
7	75,9255	58,8323	17,0932	29,0541
8	75,9255	58,7719	17,1536	29,1867
9	76,4204	58,9892	17,4312	29,5498
10	76,3077	62,6469	13,6608	21,8060
11	76,3077	65,3106	10,9971	16,8382
12	76,4486	65,6245	10,8241	16,4940
13	76,4486	65,5239	10,9247	16,6728
14	76,4486	65,4998	10,9488	16,7158
15	76,4486	65,7452	10,7034	16,2801
16	76,8067	65,5561	11,2506	17,1618
17	76,6659	65,5561	11,1098	16,9470

Comparativo dos resultados de mineração para melhor/pior reduto apontado pelo diagnóstico RSAPP, usando o algoritmo de mineração *Neive Bayes*.

Cardinalidade do Reduto	Naive Bayes			
	Melhor	Pior	Varição Acurácia	Varição Acurácia (%)
1	75,8490	57,7338	18,1152	31,3771
2	75,8691	57,7338	18,1353	31,4119
3	75,8651	57,7338	18,1313	31,4050
4	75,8732	57,7338	18,1394	31,4190
5	75,8349	57,7177	18,1172	31,3894
6	75,8973	58,6271	17,2702	29,4577
7	75,8973	58,3172	17,5801	30,1457
8	75,9054	58,3293	17,5761	30,1325
9	75,7565	58,1563	17,6002	30,2636
10	75,6961	58,1925	17,5036	30,0788
11	75,6840	59,9388	15,7452	26,2688
12	75,6921	58,4621	17,2300	29,4721
13	75,6921	59,9469	15,7452	26,2652
14	75,6921	59,9590	15,7331	26,2398
15	75,6921	60,7959	14,8962	24,5020
16	75,9094	60,3412	15,5682	25,8003
17	75,8893	60,6752	15,2141	25,0747

Comparativo dos resultados de mineração para melhor/pior reduto apontado pelo diagnóstico RSAPP, usando a média dos resultados dos testes de mineração de J48, *Randon Forest*, *Simple Cart* e *Naive Bayes*.

Cardinalidade do Reduto	Média			
	Melhor	Pior	Varição Acurácia	Varição Acurácia (%)
1	75,8500	57,7338	18,1162	31,3789
2	75,8591	57,7288	18,1303	31,4061
3	75,8440	57,7338	18,1102	31,3685
4	75,8641	57,7288	18,1354	31,4148
5	75,8812	57,6825	18,1987	31,5498
6	75,9175	58,9671	16,9504	28,7454
7	75,8621	58,6160	17,2461	29,4222
8	75,8551	58,6603	17,1948	29,3125
9	76,0029	58,7810	17,2220	29,2985
10	75,4104	61,0283	14,3821	23,5663
11	75,4305	63,2706	12,1600	19,2190
12	75,4557	63,1800	12,2757	19,4297
13	75,4507	63,3148	12,1359	19,1675
14	75,4376	63,3792	12,0584	19,0258
15	75,4608	63,5311	11,9297	18,7777
16	75,5764	62,9899	12,5865	19,9818
17	75,4014	63,0573	12,3441	19,5760