



# Mineração de Dados em uma Central de Atendimento

## Aplicação do Algoritmo Apriori

**Leandro Dias Noceli**

JUIZ DE FORA

AGOSTO, 2013

# Mineração de Dados em uma Central de Atendimento

## Aplicação do Algoritmo Apriori

Leandro Dias Noceli

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Eduardo Barrére

JUIZ DE FORA

AGOSTO, 2013

# MINERAÇÃO DE DADOS EM UMA CENTRAL DE ATENDIMENTO
## Aplicação do Algoritmo Apriori

Leandro Dias Noceli

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

_____

Eduardo Barrére
Doutor

_____

Liamara Scortegagna
Doutora

_____

Crystiam Kelle Pereira e Silva
Especialista

JUIZ DE FORA

30 DE AGOSTO, 2013

*A minha família, pela força.*

*A minha esposa, pela esperança.*

# Resumo

Atualmente, o armazenamento de dados não é considerado um problema para as corporações, tendo em vista a diminuição dos custos dos equipamentos necessários, o que veio a permitir o aumento progressivo do volume de dados produzidos e armazenados cotidianamente. Com esta imensa geração de informações, obter dados conclusivos acerca dos mesmos é uma necessidade de qualquer corporação, no entanto, extrair informações de um grande volume de dados é uma tarefa complicada, podendo dizer, até mesmo, impossível para o ser humano sem a utilização de procedimentos automatizados. Com base nisto, o procedimento conhecido como *Knowledge Discovery in Databases* pode ser aplicado, o qual, através de tratamentos realizados nos dados e do uso de algoritmos de aprendizagem, informações legíveis são geradas a partir dos dados brutos. No presente trabalho, é apresentada uma possibilidade de aplicação do processo de KDD em uma central de chamadas, com base em análise contendo dados reais, onde foram aplicados tratamentos aos dados brutos e na sequência submetidos ao algoritmo de mineração de dados Apriori.

**Keywords:** Apriori, Central de Atendimento, *Knowledge Discovery in Databases* (KDD), Mineração de Dados.

# Abstract

Currently, the data storage is not considered a problem for corporations, in order to decrease the cost of the necessary equipment, what has come to allow the gradual increase in the volume of data produced and stored daily. With this huge generation of information, to obtain conclusive data about them is a necessity of any corporation, however, extract information from a large volume of data is a complicated task and can say even impossible for humans without the use automated procedures. Based on this, the procedure known as Knowledge Discovery in Databases can be applied, which through treatments performed on the data and using learning algorithms, readable information are generated from the raw data. In this paper, we present a possible application of the KDD process in a call center, based on an analysis containing real data, where treatments were applied to the raw data and submitted the following data mining algorithm Apriori.



**Keywords:** Apriori, Call Center, Data Mining, Knowledge Discovery in Databases (KDD).

# Agradecimentos

A Deus, pela saúde concedida e pelos ensinamentos repassados de forma indireta.

A minha família, pelo carinho, dedicação e investimento dispensado a mim. Pai, obrigado por sua presença, força e coragem. Mãe, suas palavras fortes e o carinho foram indispensáveis para a progressão de minha vida.

A minha esposa, Josilene, por seu apoio, carinho e abstenção. Sem sua ajuda este trabalho não teria sido desenvolvido.

Ao meu orientador, Professor Eduardo Barrére, por seus ensinamentos, confiança e paciência despendidos ao longo deste caminho.

E a todas as pessoas que cruzaram comigo ao longo da vida, contribuindo, de alguma forma, para fazer minha vida valer a pena.

*“Desistir... eu já pensei seriamente nisso, mas nunca me levei realmente a sério; é que tem mais chão nos meus olhos do que o cansaço nas minhas pernas, mais esperança nos meus passos, do que tristeza nos meus ombros, mais estrada no meu coração do que medo na minha cabeça”.*

*Cora Coralina*

# Sumário

# Lista de Figuras

# Lista de Abreviações

| | |
|---|---|
| DM | Data Mining |
| KDD | Knowledge Discovery Database |
| TMA | Tempo Médio de Atendimento |
| WEKA | Waikato Environment for Knowledge Analysis |

# 1 Introdução

Devido ao custo de armazenamento de informações estar cada vez menor e o poder de processamento dos computadores ter aumentado significativamente nos últimos anos, entidades públicas e privadas têm ampliado a quantidade de informações que são armazenadas diariamente em suas bases.

Tais dados, isoladamente, podem não fornecer informações suficientes para o processo de tomada de decisões no âmbito organizacional e a sobrevivência das organizações está atrelada à detenção de pequenos detalhes (ou conhecimentos) que outras entidades podem não possuir.

Neste sentido, há que se adotar técnicas que permitam obter conhecimento útil analisando-se o conjunto de dados como um todo e / ou parcialmente. Realizar esta tarefa sem auxílio de técnicas e tecnologias adequadas, para um ser humano, pode ser impossível tendo em vista a imensa gama e complexidade dos dados.

Em resposta a esta necessidade, surgiu a mineração de dados (*Data Mining*), de acordo com Fayyad et al (1996), que baseia-se na utilização de cálculos estatísticos e de algoritmos de inteligência artificial e de aprendizagem de máquina para extração de padrões de informações das bases de dados altamente complexas.

A mineração de dados está ligada ao processo de descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases* - KDD), como descrito por Han et al (2006), onde mineração de dados é apenas parte do processo de KDD, no qual estão incluídas etapas de estudo e preparação dos dados (dados desnecessários são descartados, integrados a outros e poderão ser pré-computados ou não, sendo alocados posteriormente em um cubo de dados) para que o processo de mineração possa ser executado de forma eficiente e logre êxito na extração de padrões.

Em uma central de atendimento de chamadas, empresa que se destina a centralizar ligações telefônicas e, em seguida, distribuí-las aos operadores de atendimento para que os mesmos possam comunicar-se com usuários finais ou clientes, a produção de dados é enorme. Determinar o perfil e quantidade de operadores necessários para que o

atendimento possa atender à demanda de maneira eficiente e correta é o seu principal desafio.

Desta forma, analisar os dados referentes ao atendimento é uma tarefa de suma importância para realizar a previsão das demandas futuras. É com base nisto que, aplicando as técnicas de KDD, uma central de atendimento pode presumir qual será a quantidade de ligações diárias futuras, bem como qual é o perfil de operador que melhor atende às necessidades.

## 1.1 Objetivos

### 1.1.1 Objetivo Geral

O presente trabalho tem por objetivo aplicar a técnica de mineração de dados sobre um banco de dados de uma central de atendimento, demonstrando sua aplicabilidade em reconhecer padrões nas mais variadas informações armazenadas, auxiliando na tomada de decisão no que se refere à contratação de funcionários e operação de atendimento.

Com o conhecimento obtido, a idéia é melhorar o procedimento de contratação de funcionários, fornecendo informações precisas sobre qual o perfil mais adequado para o trabalho a que se destina, bem como subsidiar a empresa com informações que ajudem-na a ministrar treinamentos a seus funcionários, atacando pontos falhos de maneira mais contundente.

### 1.1.2 Objetivos Específicos

- Estudar técnicas de pré-processamento e mineração de dados, com o intuito de criar um ambiente de mineração que possibilite sua utilização em diversas fontes de dados, sem que haja necessidade de alteração ou redesenvolvimento da ferramenta para utilização em diferentes cenários.
- Desenvolver uma ferramenta de mineração de dados que utilize o algoritmo Apriori.
- Aplicar a ferramenta desenvolvida em uma base de dados de uma central de atendimento para exemplificar sua utilização.

- Aplicar a ferramenta WEKA sobre a mesma base de dados.
- Analisar os resultados obtidos nos dois casos e verificar a confiabilidade da ferramenta criada.

# 2 Fundamentação Teórica

## 2.1 Descoberta de Conhecimento em Bases de Dados

Como citado anteriormente, empresas têm armazenado uma imensa quantidade de dados. Analisá-los com a finalidade de extrair informações inteligíveis e confiáveis sobre o negócio torna-se uma tarefa essencial para a sobrevivência das empresas, haja vista a alta competitividade do mercado atual.

Em contrapartida, o aumento substancial da capacidade de armazenamento vem dificultando cada vez mais a realização da análise dos dados por um ser humano. Para solucionar este problema, em 1989 ouviu-se, pela primeira vez, o termo *Knowledge Discovery in Databases* (KDD), que é um conjunto de técnicas computacionais modelado e, desde então aprimorado, que almeja a realização da descoberta de conhecimentos inteligíveis em bases de dados.

Segundo Fayyad et al (1996), "KDD é o processo não trivial de identificação de novos, potencialmente úteis e, finalmente, compreensíveis padrões de dados válidos".

Todo esse processo de descoberta de conhecimento abrange a aplicação de tarefas de pré-processamento (limpeza, transformação e redução) e mineração de dados (Data Mining, com a utilização de algoritmos que possam extrair de forma simples e rápida conhecimentos inteligíveis.

Todas as etapas destinadas à realização da descoberta de conhecimento em banco de dados serão descritas nas próximas seções.

## 2.2 Pré-Processamento de Dados

O pré-processamento de dados é uma das principais etapas destinadas à obtenção de dados confiáveis durante o processo de descoberta de conhecimento, tendo em vista suas principais funções de eliminação parcial e/ou integral de dados incompletos, ruidosos e/ou inconsistentes, propiciada pela execução de procedimentos de limpeza, integração,

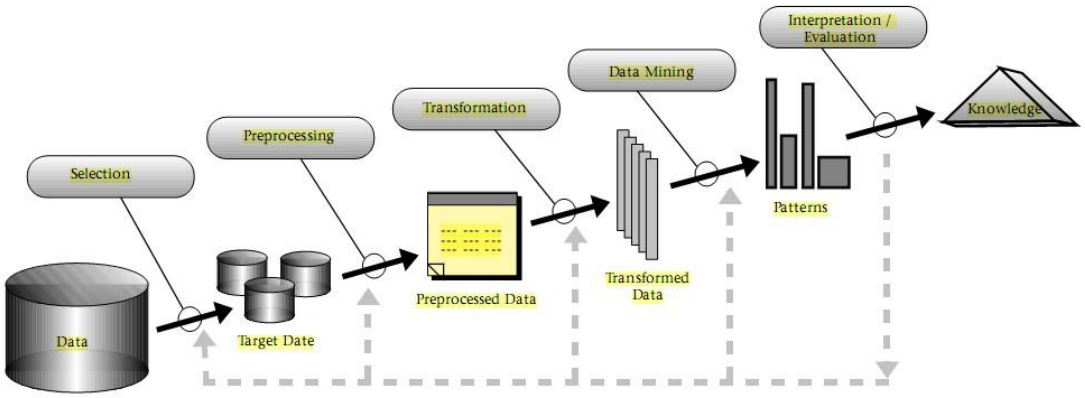


Figura 2.1: Knowledge Discovery Database
Fonte: Fayyad et al (1996)

transformação e redução de dados.

No entanto, para que a etapa de pré-processamento obtenha sucesso ao final de sua execução, é de suma importância conhecer as propriedades típicas dos dados a serem analisados. Visando o atingimento desta meta, realiza-se a "Sumarização Descritiva dos Dados", medidas que almejam identificar dados ruidosos, incompletos, enfim, dados que apresentam problemas e que necessitam de correção.

Para Han et al (2006), "dados problemáticos podem existir devido a uma série de fatores, tais como: mau funcionamento de equipamentos e *softwares*, não inclusão de certos dados em virtude de sua relevância do momento, deleções, ocorrência de erros humanos, etc.".

Em seguida, serão descritas as etapas citadas acima.

### 2.2.1 Sumarização Descritiva dos Dados

A sumarização descritiva de dados consiste na obtenção de conhecimento acerca dos dados oriundos, muitas vezes, de diversas fontes (arquivos, bancos de dados transacionais...). Esta etapa é de extrema importância para a definição dos processos que serão realizados em seguida.

Desta forma, diversas técnicas de análise estatística, descritas por Han et al

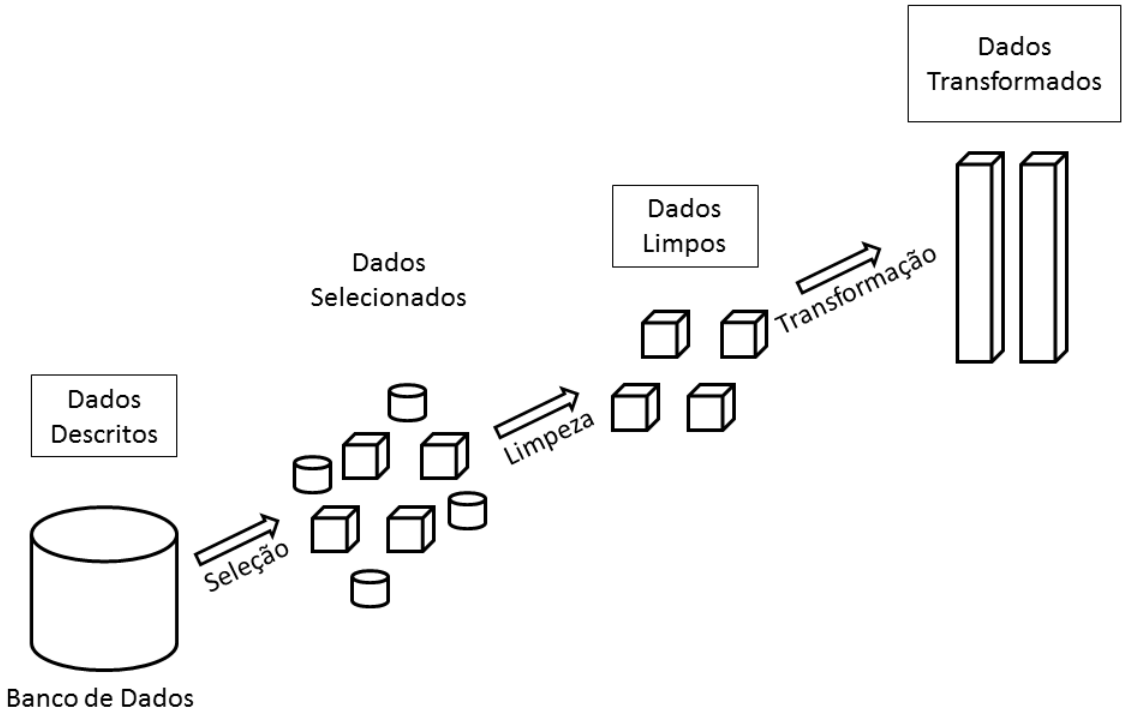


Figura 2.2: Pré-Processamento de Dados

(2006), podem ser empregadas nesta etapa, tais como: moda, média, média em faixa de valores, mediana, análise de quartis, interquartis e variância; podendo ser combinadas com análise gráfica, com o uso de histogramas, gráficos de frequência, barra, boxplot e dispersão.

A seguir serão descritas as técnicas que foram utilizadas no presente trabalho:

- **Média**: pode ser adotada em dados numéricos, permitindo o cálculo do valor médio dentre os valores presentes numa determinada coluna. Seu cálculo ocorre através da soma dos valores de determinada coluna e, em seguida, divide-se o resultado obtido pela quantidade de tuplas presentes.

- **Desvio Padrão**: desvio padrão representa o quanto há de variação nos dados em relação ao valor médio. Matematicamente falando, o desvio padrão é obtido através da raiz quadrada da variância.

- **Valor Máximo**: determinação do valor máximo de cada coluna que contém dados

numéricos.

- **Valor Mínimo**: determinação do valor mínimo de cada coluna que contém dados numéricos.

### 2.2.2 Limpeza de Dados

Atualmente, fontes de dados tendem a possuir milhares de registros e, muitas vezes, estes apresentam inconsistências, ruídos e/ou atributos incompletos. Rotinas de limpeza de dados destinam-se à realização do tratamento destes problemas, eliminando ou reduzindo-os. Neste âmbito, segundo Han et al (2006)., diversos procedimentos podem ser úteis para a extinção ou minimização dos problemas. São eles:

1. **Para registros contendo atributos indefinidos:**

   - **Não utilização da tupla:** este método não é o melhor caso a tupla a ser descartada contenha muitas colunas preenchidas. Principalmente, quando a quantidade de atributos indefinidos varia consideravelmente.
   - **Preencher o valor indefinido manualmente:** é uma tarefa demorada, principalmente em se tratando de bancos de dados de grande porte ou que contenham muitos atributos indefinidos.
   - **Substituição do atributo indefinido por um valor padrão, por exemplo, "nulo":** tal método é simples, mas não é infalível, tendo em vista que a mineração de dados poderá fornecer uma incrível informação acerca dos dados que tiveram o valor "nulo" atribuído a si, no entanto, estes não possuem qualquer ligação em comum.
   - **Substituição do atributo indefinido pela média de todos os atributos da coluna:** pode apresentar um valor distorcido dependendo do valor mínimo e máximo contido no campo e da quantidade de tuplas que os possuem.
   - **Substituição do atributo indefinido pela média dos valores da coluna, separados em grupos, de acordo com classificação através da utilização de atributos de outra coluna:** pode minimizar a variação em

relação ao item anterior, considerando-se que haverá utilização de faixas de valores baseando-se em outras colunas, garantindo um valor melhor aproximado.

- **Utilização do valor mais provável:** este pode ser obtido através de regressão, ferramentas de inferência usando o Formalismo Bayesiano ou indução por meio de árvores de decisão. Esta é a estratégia mais utilizada, pois obtém o valor indicado para substituir o valor indefinido através de diversas informações, tornando o valor mais preciso.

2. **Para registros contendo dados ruidosos (possui erro ou qualquer variação em sua medida):**

   - **Binning:** gera o valor utilizando-se dos valores próximos ao atributo ruidoso.
   - **Regressão:** divide-se em regressão linear e múltiplas regressões lineares. Na primeira o valor apropriado é encontrado mediante a pesquisa pela "melhor" linha contendo dois atributos, onde um pode predizer o outro. E a segunda, é extensão da primeira, onde mais de dois atributos são envolvidos.
   - **Clustering:** valores discrepantes são detectados utilizando-se agrupamentos de dados similares. E os valores que ficarem fora dos grupos podem ser considerados discrepantes.

### 2.2.3 Integração de Dados

É comum, na maioria das vezes, e necessária, a obtenção de dados de diversas fontes de armazenamento, tais como banco de dados relacionais, data warehouses, arquivos simples, planilhas, vídeos, imagens, entre outras.

Neste diapasão, promover a combinação dos dados oriundos destas fontes pode ser uma tarefa complicada. O importante é encontrar quais dados correspondem-se nas diversas fontes de armazenamento, eliminando dados redundantes, ou seja, que podem ser obtidos através da derivação de um ou mais atributos.

É importante atinar-se, no momento da integração dos dados, também, para aquilo que se deseja obter ao final do processo de mineração, ou seja, é imprescindível o conhecimento acerca dos dados a serem minerados. Segundo Han et al (2006), "há várias

questões a serem consideradas no processo de integração dos dados, dentre eles, o esquema de integração e o objeto correspondente".

Cabe ao usuário determinar quais os dados que deverão ser submetidos ao processo de KDD e realizar a integração dos mesmos, sendo de extrema importância, a preservação da correspondência entre os dados reais e o objeto integrado.

### 2.2.4 Transformação de Dados

Também é de suma importância adequar os dados em formatos propícios ao procedimento de descoberta do conhecimento. Como muitas vezes os dados têm origem em diversas fontes e apresentam os mais diferentes formatos e/ou valores, transformá-los em dados categóricos ou em dados numéricos (de acordo com o que se deseja) é uma necessidade.

Assim, para que isto ocorra, podem ser adotadas as seguintes técnicas, de acordo com Han et al (2006): suavização (binning, regressão e clustering), agregação (operações de síntese e agregação são aplicadas aos dados; passo normalmente utilizado para a construção de cubos de dados), generalização (onde os dados são substituídos por conceitos de alto nível através do uso de hierarquias de conceitos), normalização (atributos são dimensionados de modo a estar dentro de um pequeno intervalo especificado) e construção de atributos (novos atributos são construídos a partir de outros para ajudar no processo de mineração).

### 2.2.5 Redução de Dados

Quando se obtém dados a partir de uma única e grande fonte ou de diversas fontes, o volume desses pode ser enorme, desta forma, é importante que se promova a remoção de dados desnecessários, pois o processo de mineração poderá ser impraticável, devido ao alto custo de processamento e, inclusive, o de armazenamento. A seguir serão apresentadas técnicas que podem ser utilizadas para realizar a redução de dados, sem que se perca a integridade das informações originais:

- Agregação de cubos de dados: operações de agregação são aplicadas aos dados na construção do cubo a ser analisado.

- Seleção do conjunto de atributos: dados pouco relevantes ou totalmente irrelevantes, atributos redundantes ou dimensões podem ser detectadas e removidas.

- Redução da dimensionalidade: mecanismos são adotados para reduzir o tamanho do conjunto de dados.

- Redução de numerosidade: dados podem ser substituídos ou estimados por outros.

- Discretização e conceito de geração de hierarquia: onde os valores brutos dos atributos são substituídos por valores de alto nível. Discretização de dados é uma forma de redução de numerosidade, o que a torna muito útil para a geração automática de hierarquias de conceitos, permitindo a ocorrência da mineração em múltiplos níveis de abstração.

## 2.3 Cubo de Dados

Para a realização de mineração, dados devem estar dispostos de forma a proporcionar uma melhor performance na obtenção dos resultados. Tal fato, leva-nos ao tratamento dos dados e, também, à disposição dos mesmos em um formato que facilite a realização dos objetivos.

Os cubos de dados são estruturas de armazenamento multidimensional e são amplamente utilizados pelas ferramentas de mineração de dados na busca por informações significativas. Eles oferecem a capacidade de acessar dados de coleções de qualquer subconjunto de dimensão (chamado cubóide), sendo notável a facilidade e flexibilidade em obter informações a partir de diversos níveis de granularidade e ângulos, sem que se perca a integridade do resultado em relação ao todo.

De acordo com Gray et al (1996), “o operador cubo de dados é uma tabela contendo valores agregados e o total agregado é representado como a tupla: ALL, ALL, ALL, ..., ALL, f (*) em n-dimensões.”

É importante frisar que em um cubo, as dimensões representam um atributo do banco de dados e as células representam uma medida de interesse.

### 2.3.1 Construção

Para construir o cubo de dados a ser analisado é necessário definir o que será mais importante do ponto de vista armazenamento x performance. Assim, existem três diferentes metodologias a serem adotadas, de acordo com Han et al (2006):

- Pré-computação total das células: nesta metodologia, consultas executadas no cubo ocorrerão mais rapidamente, no entanto, haverá uma maior necessidade de espaço de armazenamento.

- Pré-computação nenhuma das células: não computando quaisquer das células, haverá um menor consumo de espaço de armazenamento e, por conseguinte, diminuição da performance de obtenção da informação de apoio à decisão, pois não computando os dados previamente, o cubo deverá ser reconstruído a cada consulta.

- Pré-computação de parte das células: visando a obtenção de um equilíbrio entre performance e armazenamento, é pertinente computar apenas parte das células, as que, verdadeiramente, serão utilizadas no processo de obtenção da informação de apoio à decisão.

### 2.3.2 Principais Operações sobre Cubo de Dados

Definindo o tipo de construção a ser adotado, é necessário definir a quais tipos de operações os dados serão submetidos para serem alocados no cubo criado. Logo abaixo, podem ser visualizadas as principais:

- Sumarização ou “Rollup”: pode ser realizada através de translação, de acordo com a hierarquia de conceitos.

  Hierarquia de conceitos refere-se ao mapeamento de um conjunto do mais baixo nível até o mais alto. Tal fato, possibilita a sumarização de informações no cubo de dados.

  À medida que os valores são combinados, cardinalidades também diminuem, consequentemente, o cubo torna-se menor.

- Drill-Down: é semelhante à sumarização, no entanto, ocorre de forma inversa, ou seja, enquanto a primeira sintetiza as informações, de baixo até o alto nível, drill-down faz o contrário. Utilizando Drill-Down, é possível tanto adicionar uma nova dimensão ao cubo de dados quanto quebrar a hierarquia de conceitos.

- Slice: é realizada a extração de um subcubo do cubo original através da seleção de dados em duas ou mais dimensões.

- Dice: Define subcubos selecionando duas dimensões, pegando uma palavra e a próxima palavra para formar parte da consulta.

## 2.4 Data Warehouse

Data Warehouse é o termo utilizado para designar o conjunto de informações que passaram pelas etapas de pré-processamento de dados para, posteriormente, serem submetidos ao processo de mineração de dados. Dados podem ser apresentados aos usuários finais na forma de relatórios e/ou conhecimento extraído após submissão dos mesmos ao processo de mineração de dados.

De acordo com Fayyad et al (1996), "Data Warehouse é um repositório de informações coletadas a partir de múltiplas fontes, alocadas sob um esquema unificado".

## 2.5 Data Mining

Data Mining é a aplicação de algoritmos específicos com o intuito de extrair padrões dos dados.

Conforme Fayyad et al (1996), "mineração de dados é um passo no processo de KDD que consiste na análise de dados e aplicação de algoritmos de descoberta de conhecimento que, sob as limitações de eficiência computacional aceitáveis, produzem um enumeração particular de padrões (ou modelos) sobre os dados."

A diferença entre mineração de dados e descoberta do conhecimento em bancos de dados está nos passos a mais que o segundo possui, de acordo com Fayyad et al (1996): "os passos adicionais no processo de KDD, tais como a preparação, a seleção, a limpeza

dos dados, a incorporação do conhecimento prévio adequado, e uma interpretação correta dos resultados da extração, é essencial para assegurar que o conhecimento útil é derivado a partir dos dados."

Ou seja, Data Mining é parte do processo de descoberta do conhecimento, ao qual se agregam as etapas de preparação, seleção e limpeza dos dados, bem como, interpretação dos resultados após extração do conhecimento através dos algoritmos de mineração.

### 2.5.1 Tipos de Algoritmos de Mineração de Dados

Para realizarmos a mineração de dados de forma automatizada, é necessário que apliquemos um ou mais algoritmos já desenvolvidos para este propósito. Desde a primeira vez que ouviu-se dizer sobre descoberta de conhecimento em bases de dados, os principais tipos de algoritmos de mineração desenvolvidos e suas respectivas especificações segundo a Microsoft (2012), são:

- Algoritmos de Classificação: responsáveis por predizer um ou mais itens discretos com base nos outros atributos do conjunto de dados. É uma das técnicas mais utilizadas, talvez pelo fato do ser humano utilizá-la para a compreensão do espaço onde vivemos.

- Algoritmos de Regressão: responsáveis por obter itens contínuos através dos outros atributos contidos no banco de dados. Um exemplo pode ser o cálculo da idade a partir da data de nascimento.

- Algoritmos de Clusterização: responsáveis por segmentar o conjunto de dados em grupos que apresentem propriedades semelhantes.

- Algoritmos de Associação: o objetivo de tais algoritmos é encontrar padrões frequentes em um determinado conjunto de dados. Em outras palavras, é verificar em varreduras sobre a base de dados, quais os itens ou conjunto de itens que possuem mais ocorrências. Um exemplo desse tipo de algoritmo é o Apriori.

- Algoritmos de Análise de Sequência: resumem sequências frequentes ou episódios em dados, como um fluxo de caminho da Web. Um exemplo de um algoritmo de

sequência é o algoritmo MSC.

# 3 Metodologia

Analisando o conjunto de dados foi possível transformá-lo (descrevê-lo) em poucos grupos de dados e, como a nossa necessidade é obter uma previsão do perfil de operador de atendimento mais adequado às funções, concluiu-se que o algoritmo Apriori, que utiliza regras de associação, atende perfeitamente, inicialmente, aos anseios propostos, tendo em vista o baixo consumo de processamento e memória no momento da realização da associação dos grupos, devido à baixa quantidade dos mesmos.

## 3.1 Algoritmo Apriori

Proposto por Agrawal e Skirant (Agrawal et al, 1994), o Apriori é um algoritmo clássico para obtenção de regras de associação em bancos de dados transacionais. Para eles o problema da obtenção de regras de associação pode ser decomposto em dois sub-problemas:

- encontrar todos os conjuntos de itens que têm suporte acima do suporte mínimo.
- e, através dos que atingiram o suporte mínimo, gerar as regras de associação verificando quais deles atendem à confiança mínima pré-estabelecida.

1) $L_1 = \{\text{large 1-itemsets}\}$;
2) **for** ( $k = 2$; $L_{k-1} \neq \emptyset$; $k$++ ) **do begin**
3) $C_k = \text{apriori-gen}(L_{k-1})$; // New candidates
4) **forall** transactions $t \in \mathcal{D}$ **do begin**
5) $C_t = \text{subset}(C_k, t)$; // Candidates contained in $t$
6) **forall** candidates $c \in C_t$ **do**
7) $c$.count++;
8) **end**
9) $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$
10) **end**
11) Answer = $\bigcup_k L_k$;

Figura 3.1: Algoritmo Apriori
Fonte : Agrawal et al, (1994)

### 3.1.1 Conceitos Envolvidos

- Suporte Mínimo: pode ser definido como a frequência com que um determinado conjunto candidato pode ser encontrado no conjunto de dados analisado.

- Confiança Mínima: pode ser definida como a relação entre a quantidade total de um determinado conjunto candidato e a quantidade total de tuplas analisadas.

### 3.1.2 Funcionamento

O funcionamento do algoritmo Apriori é extremamente simples, o que facilita o seu desenvolvimento e aplicação. Logo abaixo, serão detalhadas suas duas fases de execução:

1. Geração dos Conjuntos Candidatos com Verificação do Suporte Mínimo: nesta fase ocorre a formação de todas as possíveis combinações de itens, gerando os conjuntos candidatos. Para cada conjunto candidato formado, é verificado se o suporte mínimo é satisfeito.

   O processo ocorre da seguinte forma: inicialmente é gerado um conjunto contendo dois elementos, sendo que tais elementos já foram analisados individualmente e atenderam ao requisito de suporte mínimo. Por conseguinte, é verificado se o suporte mínimo foi satisfeito. Em seguida, gera-se um conjunto com três elementos e assim por diante, até que não existam novas combinações a serem realizadas.

   O descarte dos conjuntos candidatos que não atingiram o suporte mínimo ocorre a cada iteração do algoritmo a fim de evitar sobrecarga dos sistemas, bem como diminuir o tempo de geração do conhecimento, devido evitar o processamento do algoritmo utilizando conjuntos que não atingiram o suporte mínimo. Por tal motivo, talvez esta seja a parte mais importante do algoritmo, pois garante a eficiência de execução do mesmo.

2. Geração das Regras de Associação: após a formação de todos os possíveis conjuntos candidatos e verificação de atingimento do suporte mínimo é observado se determinado conjunto atinge a confiança mínima. Caso positivo, as regras de associação

formadas serão o resultado obtido com a execução do algoritmo Apriori.

De acordo com a definição de Agrawal et al (1994): o primeiro passo do algoritmo é simplesmente contar a quantidade de ocorrências de cada item, determinando os 1-itens mais frequentes. O passo subsequente, chamado passo k, consiste em duas fases. Primeira, os grupos Lk-1 mais frequentes encontrados no passo k-1 são usados para gerar o grupo de itens ck, usando a função apriori-gen [...]. O próximo, a base de dados é percorrida e o suporte mínimo dos candidatos no grupo Ck é verificado. Para uma rápida contagem, é necessário determinar com eficiência os candidato Ck, contidos em cada transação t.

# 4 Análise Prática

Neste capítulo é apresentado o tratamento dado a um caso específico, utilizando todo embasamento descrito nos capítulos anteriores.

Como o presente trabalho é direcionado à elaboração de uma ferramenta capaz de centralizar a mineração de dados de todas as bases de dados de uma central de atendimento, a ferramenta foi batizada como *Mining Center*. Desenvolvida em Ajax, utilizando o *Model View Controller* (MVC), a ferramenta é capaz de proporcionar, após as etapas de pré-processamento e de mineração de dados, a visualização do conhecimento oculto nos cubos gerados. Uma imagem da ferramenta pode ser vista logo abaixo.

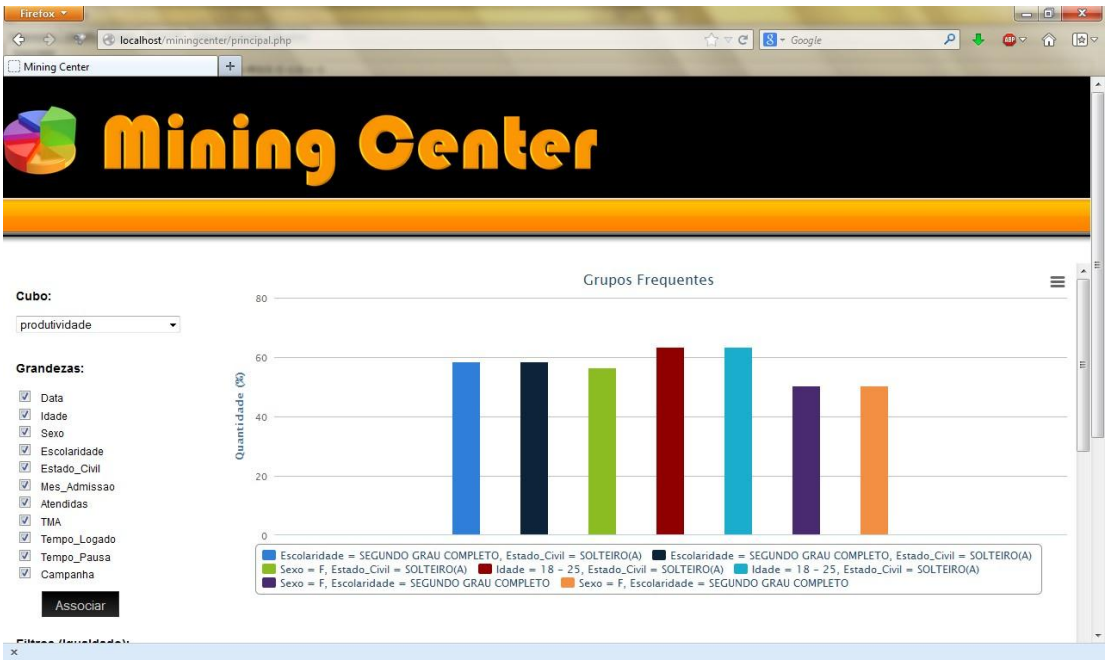


Figura 4.1: *Screenshot* do sistema de mineração de dados Mining Center

Serão analisados os dados de tempo médio de atendimento (tempo em que um operador utiliza para atender um cliente), número de chamadas atendidas, tempo total de trabalho, tempo total de pausa (horário de almoço ou lanche), setor ao qual um operador pertence (também denominado campanha), mês dos referidos dados, e, ainda, grau escolar, idade, sexo, estado civil e mês de admissão dos operadores. Aplicando a mineração de dados utilizando a ferramenta desenvolvida, será possível determinar qual o melhor perfil se enquadra às atividades de uma central de atendimento.

## 4.1 Origem dos Dados

Toda a operação de atendimento de clientes é monitorada por um sistema de atendimento, responsável por gravar dados (tempo de atendimento, qual operador realizou o atendimento, tempo das pausas do operador, etc.) acerca de cada chamada. Tais dados podem ser extraídos da ferramenta em um arquivo texto comum (txt), separados por ponto-e-vírgula (;).

Após extraídos, os mesmos podem ser utilizados para a geração de diversos relatórios, nos quais são exibidos dados simples, contendo como dados finais de análise soma, média, desvio padrão, dentre outros. No entanto, tais análises não são suficientes para gerar conhecimento acerca do conjunto.

Inicialmente, os dados abastecem uma tabela de uma base de dados MySQL, através do cruzamento com outras tabelas já contidas no banco, de onde são incluídos alguns dados do operador de atendimento referentes aos seus atendimentos, bem como de sua hierarquia (para determinação de seu superior hierárquico) e campanha. Neste momento foi gerado o que chamaremos aqui de Cubo 1.

No Cubo 1 ainda não possuímos todos os dados que desejamos minerar. Desta forma e, como a ferramenta de Mineração de Dados utiliza-se de seu próprio banco, devemos gerar outro cubo (Cubo 2) contendo parte dos dados da tabela anterior transformados em valores médios mensais, bem como os dados pessoais (idade, mês de admissão, etc.) dos operadores de atendimento, contidos em outra tabela.

O Cubo 2 conterá dados de todos os operadores de atendimento, referentes ao intervalo de tempo de dezembro de 2012 a maio de 2013.

Neste ponto, há que ser realizado o pré-tratamento de todos os dados para garantirmos a confiabilidade do processo de descoberta do conhecimento. Nas próximas seções, serão descritos todas as etapas.

## 4.2 Preparação dos Dados

Conforme a teoria elencada, é de extrema importância submeter os dados a serem minerados a procedimentos de limpeza, integração, transformação e redução dos dados.

Então, na sequência, serão descritas as fases de tratamento as quais os dados foram submetidos, com a finalidade de prepará-los para o procedimento de extração de conhecimento.

### 4.2.1 Sumarização Descritiva dos Dados

O conhecimento acerca dos dados a serem minerados pode melhorar substancialmente a obtenção das asserções, bem como, evitar que estas contenham dados falso-positivos. Também é possível diminuir o tempo de execução dos algoritmos de mineração, aumentando a performance.

Na análise em questão, todas as colunas e tuplas da tabela receptora possuem dados, no entanto tuplas que continham dados de operadores de atendimento que estavam passando por algum tipo de treinamento foram sumariamente descartadas, pois o seu atendimento não representa, fidedignamente, informações que demonstram a realidade do ambiente empresarial no que concerne à evolução do operador. Além dos últimos, tuplas contendo dados de operadores de atendimento que já não fazem parte do corpo de funcionários da empresa também tiveram o mesmo tratamento.

Os dados numéricos das colunas foram submetidos à análise de média aritmética, desvio padrão e à verificação de seus valores mínimo e máximo, com o intuito de guiar o processo de transformação de dados.

Neste sentido, os seguintes valores foram encontrados:

- Chamadas Atendidas: valor médio, 16,32; desvio padrão, 9,89; valor máximo, 39 e mínimo, 0.
- Tempo Logado: valor médio, 06:42:35h; desvio padrão, 04:38:21h; valor máximo, 08:23:44h e mínimo, 0.
- Tempo Médio de Atendimento: valor médio, 00:15:52h; desvio padrão, 0:05:30h; valor máximo, 01:56:35h e mínimo, 00:00:00h.
- Tempo de Pausa: valor médio, 01:17:46h; desvio padrão, 01:26:11h; valor máximo, 00:06:05h e mínimo, 00:00:00h.

### 4.2.2 Limpeza dos Dados

Dados de operadores de atendimento em treinamento, bem como de pessoas que não trabalham mais na empresa são dados que não participaram do procedimento de mineração por não mais servirem para a análise em questão, não representando com afinco a realidade empresarial. Portanto, tais dados foram sumariamente descartados.

Não houve necessidade de emprego de qualquer técnica de limpeza de dados, tendo vista a recepção de dados consolidados e sem qualquer tipo de problema.

### 4.2.3 Integração dos Dados

Já descrito que os dados que servirão de objeto para a primeira análise da ferramenta *Mining Center* são oriundos de dois tipos de fontes: arquivos texto e tabelas de uma base de dados MySQL, sendo a chave das relações o identificador do operador de atendimento.

Os dados relacionados foram dispostos na tabela Produtividade (Cubo 2), que recebeu este nome pela idéia de obter conhecimento sobre o processo de atendimento realizado pelos operadores de atendimento, podendo ser identificado com processo de mineração, padrões de perfis mais adequados à atividade empresarial em lide.

### 4.2.4 Transformação dos Dados

Próximo passo e talvez o mais importante, foi a realização da disposição dos dados em um formato propício à mineração. Como o algoritmo Apriori trabalha com regras de associação, houve a necessidade de adotar as técnicas de generalização, normalização e regressão.

Após a submissão dos dados numéricos às análises de média aritmética, desvio padrão e verificação dos valores máximo e mínimo de cada coluna, a seguinte representação categórica foi adotada para cada coluna:

1. Data: os dados foram dispostos contendo valores do mês e ano somente. Exemplo: “JAN - 2013”.

2. Campanha: os operadores de atendimento foram classificados em cinco grupos de-

nominados Campanha 1, Campanha 2, Campanha 3, Campanha 4 e Campanha 5, de acordo com o setor que cada um pertence.

3. Chamadas Atendidas: valor médio mensal de cada representante, disposto sob a forma de intervalos definidos da seguinte forma:

    - 0 a 5: indica número de chamadas dentro do intervalo de 0 a 5, inclusive;
    - 6 a 10: quantidade de chamadas entre 6 e 10;
    - 11 a 15: quantidade de chamadas entre 11 e 15;
    - 16 a 20: quantidade de chamadas entre 16 e 20;
    - 21 a 25: quantidade de chamadas entre 21 e 25; e,
    - Mais de 25: mais de 25 chamadas recebidas:

4. Escolaridade: na empresa, há operadores de atendimento que possuem os graus de escolaridade segundo grau incompleto, segundo grau completo, superior incompleto, superior completo e pós-graduação.

5. Estado Civil: há operadores de atendimento que possuem estado civil solteiro(a), casado(a), desq./divorc., marital e viúvo(a).

6. Idade: a empresa admite pessoas que possuam idade igual ou maior de 16 anos. Deste modo o campo idade foi subdividido em três grupos:

    - 16 a 17: de 16 a 17 anos de idade;
    - 18 a 25: de 18 a 25 anos de idade; e,
    - 25 - ?: operadores de atendimento que possuem idade maior que 25 anos.

7. Mês de Admissão: tal como o campo data, o mês de admissão possui os valores de mês e ano. Exemplo: “JAN-2013”

8. Sexo: os grupos representantes de sexo são F e M, indicando sexo feminino e masculino, respectivamente.

9. Tempo Total de Pausa: valores entre 1200 e 3600 segundos são considerados normais. Desta forma, adotando-se uma variação de 1200 segundos, os seguintes intervalos médios foram adotados:

    - 0 a 1200: tempo de pausa entre 0 e 1200 segundos;
    - 1201 a 2400: tempo de pausa entre 1201 e 2400 segundos;
    - 2401 a 3600: tempo de pausa entre 2401 e 3600 segundos;
    - 3601 a 4800: tempo de pausa entre 3601 e 4800 segundos; e,
    - Mais de 4800: tempo de pausa maior que 4800 segundos.

10. Tempo Logado (tempo de trabalho): os seguintes intervalos foram adotados na análise:

    - 0 a 22800: tempo de trabalho entre 0 e 22800;
    - 22801 a 29520: tempo de trabalho entre 22801 e 29520 segundos; e,
    - Mais de 29521: tempo de trabalho superando 29521 segundos.

11. Tempo Médio de Atendimento: os valores dos grupos foram definidos com variação de 600 (seiscentos) segundos, totalizando sete intervalos, como se segue:

    - 0 a 600: indica tempo médio de atendimento entre 0 e 600 segundos, inclusive;
    - 601 a 1200: define tempo médio de atendimento maior que 601 e menor ou igual a 1200 segundos;
    - 1201 a 1800: define tempo médio de atendimento maior que 1201 e menor ou igual a 1800 segundos;
    - 1801 a 2400: define tempo médio de atendimento maior que 1801 e menor ou igual a 2400 segundos;
    - 2401 a 3000: define tempo médio de atendimento maior que 2401 e menor ou igual a 3000 segundos;
    - 3001 a 3600: define tempo médio de atendimento maior que 3001 e menor ou igual a 3600 segundos; e,

- Mais de 3600: define tempo médio maior que 3600 segundos.

Concluindo, o cubo de dados contém 11 (onze) colunas que são: data, idade, sexo, escolaridade, estado civil, mês de admissão, chamadas atendidas, tma, tempo logado, tempo de pausa e campanha.

# 5 Mining Center

O sistema *Mining Center* foi desenvolvido em Ajax, possuindo como algoritmo de mineração o Apiori. Construído utilizando-se o paradigma Orientação a Objetos, o mesmo pode sofrer readaptação, incluindo facilidade na inclusão de novos algoritmos de mineração. O sistema possui interface amigável tendo em vista a não necessidade do usuário possuir conhecimento acerca de mineração de dados para operar a ferramenta. É possível escolher qual o cubo, bem como quais dados deste serão submetidos ao processo de mineração.

## 5.1 Funcionamento

Como informado anteriormente, a interface da aplicação é bastante simples, o que permite ao usuário, realizando o mínimo de operações obter resultados satisfatórios. Logo abaixo serão demonstrados os passos de sua utilização:

1. após realizar as verificações de segurança, a figura 5.1 será exibida.

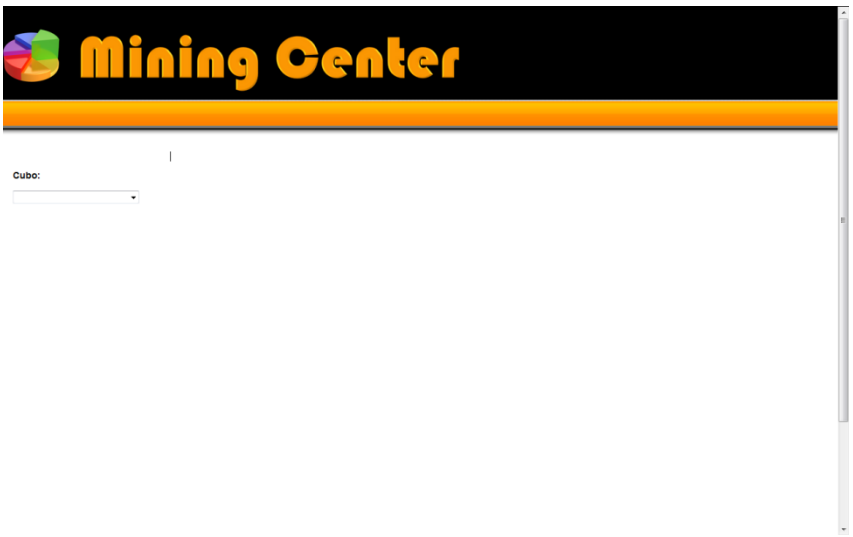


Figura 5.1: Página Principal

2. próximo passo é a escolha do cubo a ser submetido ao processo de mineração, através do campo de seleção. Após escolhido o cubo, as grandezas que o compõem serão

exibidas, conforme a figura 5.2:

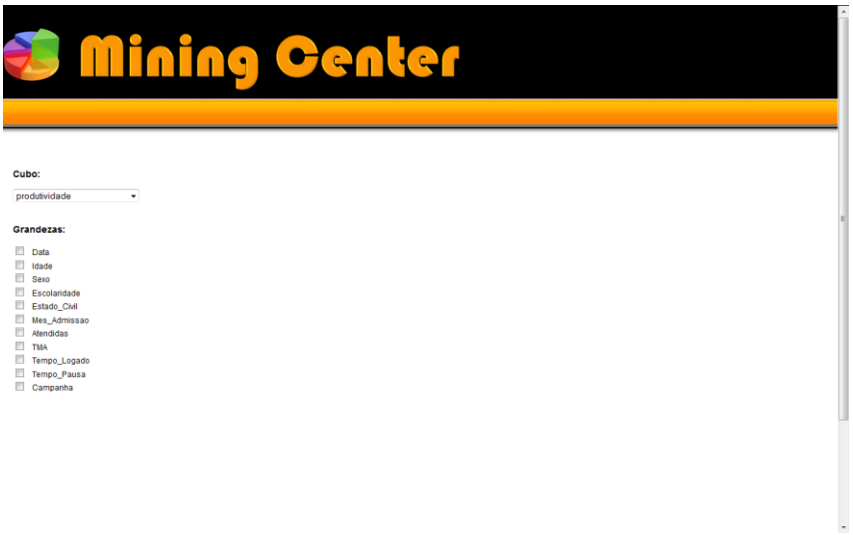


Figura 5.2: Escolha das Grandezas

3. em seguida, deve-se selecionar as grandezas e, somente a partir da escolha de pelo menos duas, que o botão Associar será exibido. Escolhidas as grandezas, caso seja necessário, deve-se realizar a escolha dos filtros, informando o tipo de seleção ("e" ou "ou") se mais de dois filtros forem utilizados. A figura 5.3 exibe os detalhes:


Figura 5.3: Escolha das Grandezas e Filtros

4. após realizar o clique sobre o botão Associar, será realizada a mineração dos dados e caso seja encontrado algum resultado, serão mostrados os grupos mais frequentes e um botão para acesso as estatísticas (regras) encontradas, conforme figura 5.4.

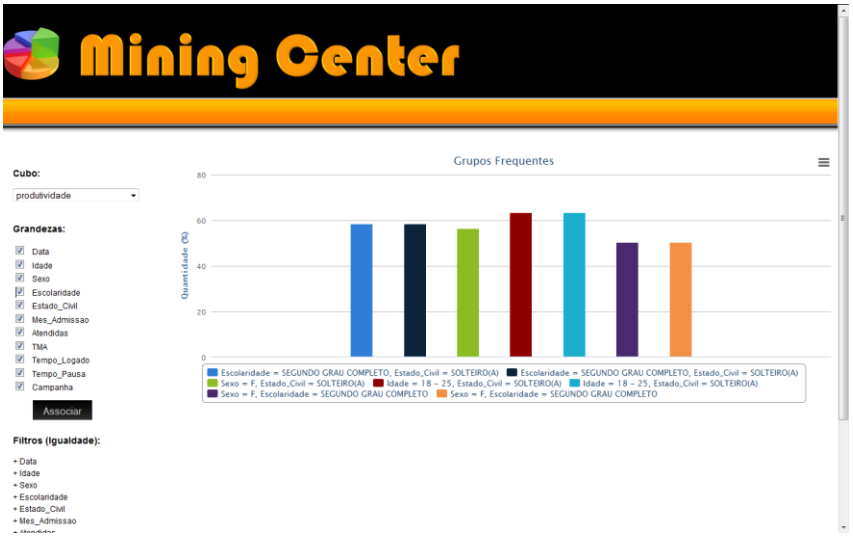


Figura 5.4: Resultado

# 6 Resultados da Análise e Comparação

## 6.1 Sistema *Mining Center* - Resultados

O sistema desenvolvido oferece a possibilidade de realizarmos diversos tipos de filtragem na busca pela extração do conhecimento. A título de demonstração de funcionamento, filtros não foram utilizados para execução desta análise. Executando-se o algoritmo Apriori sobre o conjunto de dados do Cubo 2, com suporte mínimo de 50%, os seguintes conjuntos candidatos satisfizeram o índice:

- **Conjunto 1**: 50,3% das tuplas indicam escolaridade 2º grau completo e sexo feminino.
- **Conjunto 2**: 56,4% das tuplas indicam estado civil solteiro(a) e sexo feminino.
- **Conjunto 3**: 58,4% das tuplas indicam escolaridade 2º grau completo e estado civil solteiro(a).
- **Conjunto 4**: 63,4% das tuplas indicam idade entre 18 e 25 anos e o estado civil solteiro(a).

Ao ser realizada a verificação de confiança mínima, adotando-se o padrão de 70%, definido no sistema, e, utilizando-se os conjuntos candidatos que atingiram o suporte mínimo, as seguintes regras de associação foram encontradas:

1. **Conjunto 1**:
   - Em 72,1% dos casos, se escolaridade é 2º grau completo então sexo é feminino; e,
   - Em 71,6% dos casos, se sexo é feminino então a escolaridade é 2º grau completo.
2. **Conjunto 2**:
   - Em 80,3% dos casos, se sexo é feminino então o estado civil é solteiro(a).

3. **Conjunto 3**:

   - Em 83,6% dos casos, se a escolaridade é 2º grau completo então o estado civil é solteiro(a); e,
   - Em 70,2% dos casos, se o estado civil é solteiro(a) então a escolaridade é 2º grau completo.

4. **Conjunto 4**:

   - Em 94,6% dos casos, se idade está entre 18 e 25 anos então o estado civil é solteiro(a); e,
   - Em 76,3% dos casos, se estado civil é solteiro(a) então a idade está entre 18 e 25 anos.

É importante salientar que através de realização das filtragens que a ferramenta permite, outras diversas informações podem ser obtidas, cabendo ao usuário determinar quais são os conjuntos úteis, bem como separar aquilo que é informação útil.

## 6.2 Sistema Weka e Resultados

Com o intuito de comprovar a eficiência do sistema desenvolvido, comparações de funcionamento foram realizadas utilizando o sistema Weka. A seguir, o sistema e os resultados serão brevemente apresentados.

### 6.2.1 Weka - Apresentação

Weka (*Waikato Environment for Knowledge Analysis*) é uma ferramenta detentora de uma coleção de algoritmos de pré-processamento e de aprendizagem, tal como o Apriori. Desenvolvida na linguagem JAVA pela Universidade de Waikato, da Nova Zelândia, e disponibilizada sob a licença GNU/GPL, a aplicação tornou-se mundialmente conceituada e bastante utilizada em nível acadêmico.

Segundo Witten et al (2008), "Weka pode ser executada em várias plataformas e já foi testada nos sistemas operacionais Linux, Windows e Machintosh - [...]".

Weka encontra-se disponível para download no endereço http://www.cs.wakaito.ac.nz/ml/we nas diversas plataformas citadas acima. É importante salientar que torna-se necessário possuir o Java Runtime Environment instalado.

## 6.3 Weka - Resultados

Com a aplicação da ferramenta sobre o Cubo 2, as mesmas assertivas encontradas pela ferramenta *Mining Center* foram determinadas pela Weka, utilizando os mesmos 50% e 70% de suporte e confiança mínimos, respectivamente.

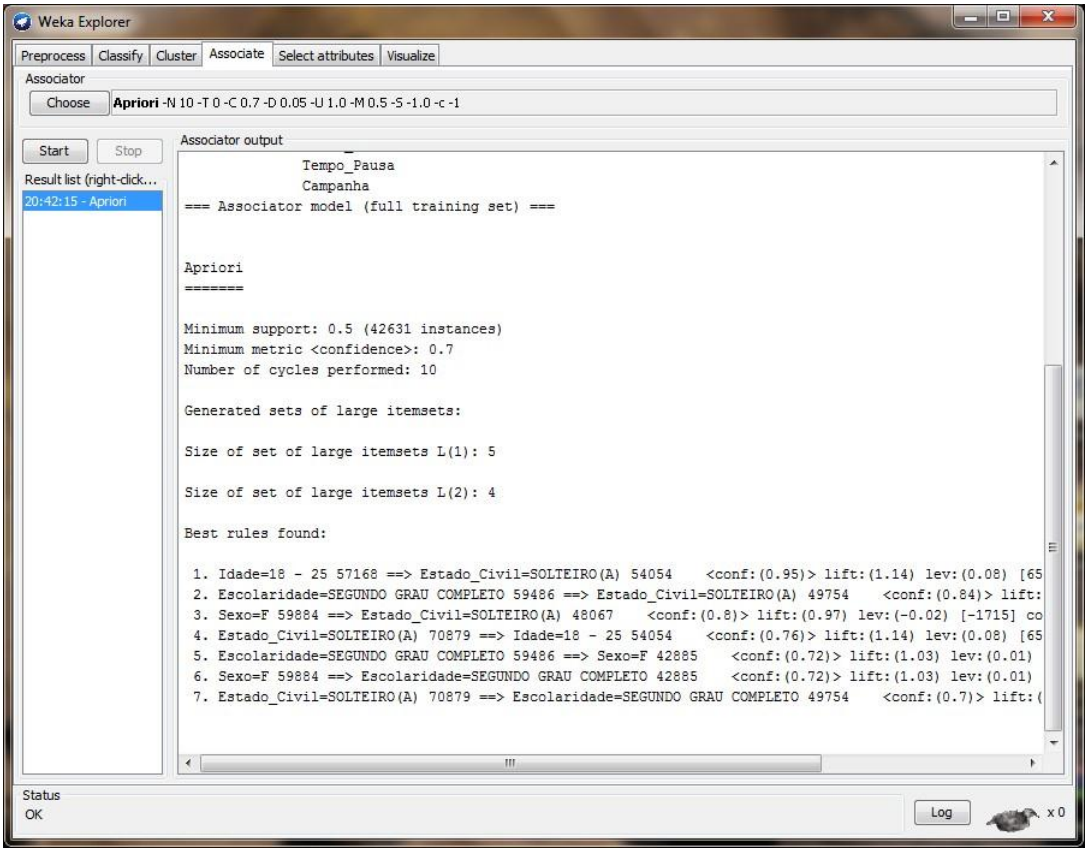


Figura 6.1: Resultado obtido durante a análise

# 7 Análises Complementares

Baseando-se nos resultados da análise anterior e, analisando-se do ponto de vista empresarial, não foram obtidas regras que atendessem ao negócio de forma satisfatória, pois para que isto ocorresse, era necessário haver correlacionamento entre dados pessoais (exemplo: idade, sexo...) e laborativos (TMA, tempo logado...), desta forma, houve a necessidade de realização de mais duas análises com adoção de filtros, bem como diminuição do suporte mínimo para 35%, com o intuito de obter regras que atendam aos objetivos propostos. As próximas análises foram realizadas utilizando-se o filtro de Estado Civil. A primeira análise, utilizou dados de pessoas solteiras, enquanto a segunda, de casadas. Nas próximas seções, os resultados de ambas as análises serão apresentados.

## 7.1 Análise I - Estado Civil Solteiro(a)

Com a execução da análise utilizando o filtro ESTADO_CIVIL = SOLTEIRO(A), as seguintes regras foram obtidas nos sistemas Mining Center e Weka:

- Em 72.93% dos casos, é possível afirmar que: se Idade = 18 - 25, Sexo = F então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 76.72% dos casos, é possível afirmar que: se Sexo = F, Escolaridade = SEGUNDO GRAU COMPLETO então Idade = 18 - 25
- Em 70.67% dos casos, é possível afirmar que: se Tempo_Logado = 0 a 22800 então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 76.45% dos casos, é possível afirmar que: se Tempo_Logado = 0 a 22800 então Idade = 18 - 25
- Em 75.69% dos casos, é possível afirmar que: se TMA = 1201 a 1800 então Idade = 18 - 25

- Em 71.62% dos casos, é possível afirmar que: se Sexo = F então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 71.60% dos casos, é possível afirmar que: se Idade = 18 - 25 então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 77.79% dos casos, é possível afirmar que: se Escolaridade = SEGUNDO GRAU COMPLETO então Idade = 18 - 25
- Em 75.35% dos casos, é possível afirmar que: se Sexo = F então Idade = 18 - 25

Os resultados da análise nos sistemas *Mining Center* e Weka, podem ser visualizados nas figuras 7.1 e 7.2, respectivamente.

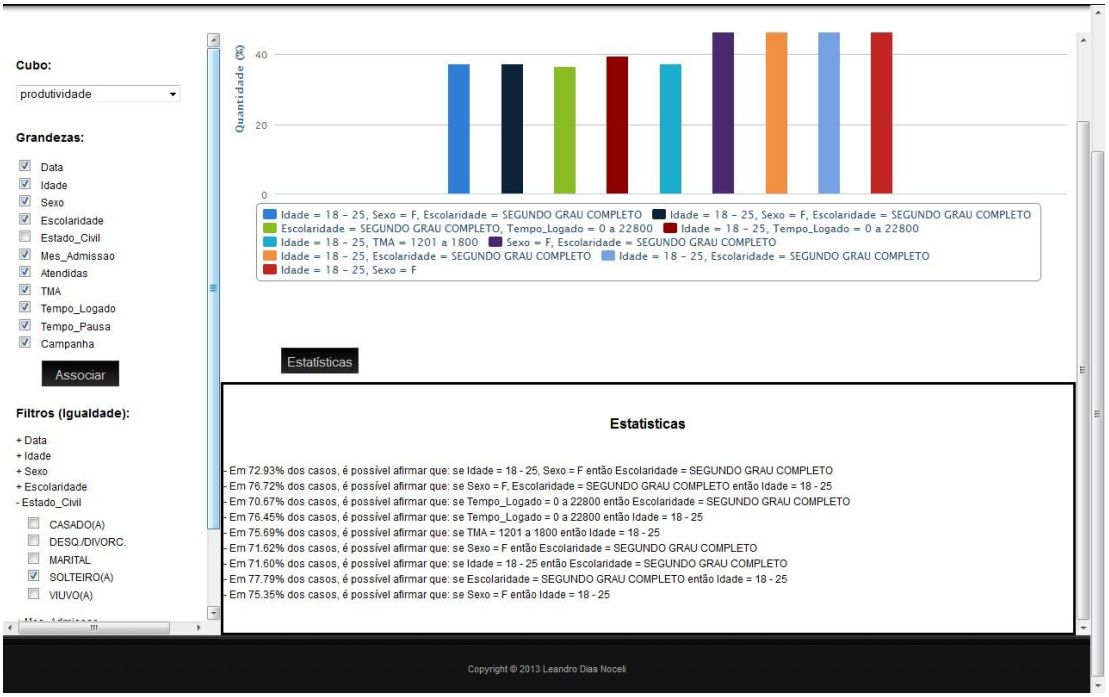


Figura 7.1: Resultado obtido no Mining Center (Solteiros)

## 7.2 Análise II - Estado Civil Casado(a)

Com a execução da análise utilizando o filtro ESTADO_CIVIL = CASADO(A), as seguintes regras foram obtidas nos sistemas Mining Center e Weka:

- Em 75.08% dos casos, é possível afirmar que: se Idade = 25 - ?, Sexo = F então Escolaridade = SEGUNDO GRAU COMPLETO

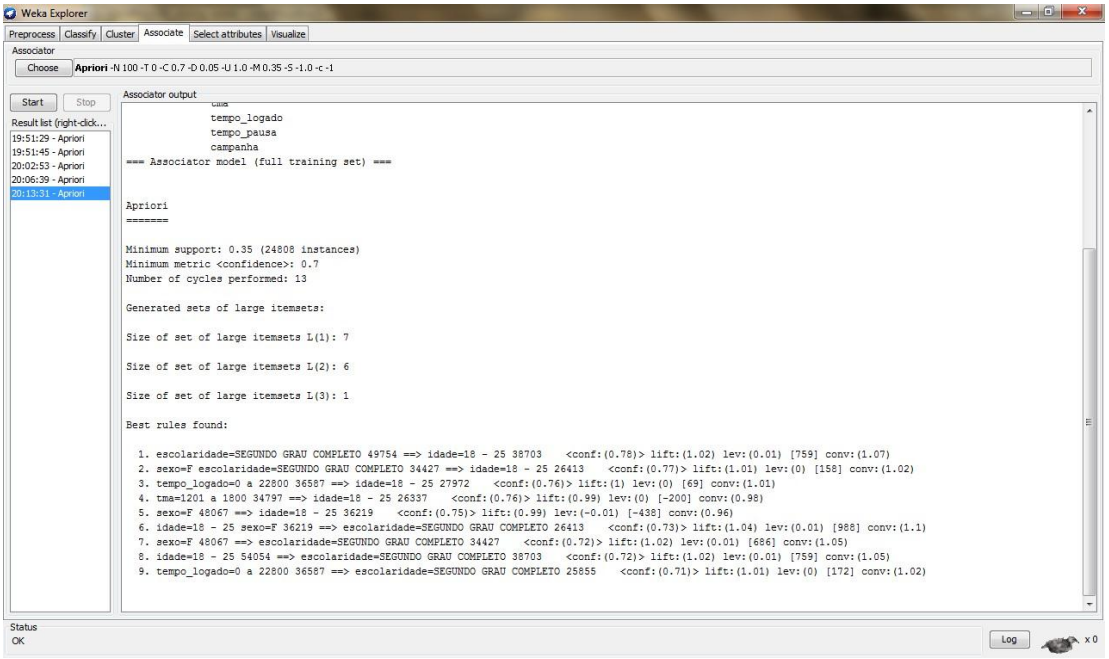


Figura 7.2: Resultado obtido na Weka (Solteiros)

- Em 82.53% dos casos, é possível afirmar que: se Idade = 25 - ?, Escolaridade = SEGUNDO GRAU COMPLETO então Sexo = F
- Em 72.42% dos casos, é possível afirmar que: se Sexo = F, Escolaridade = SEGUNDO GRAU COMPLETO então Idade = 25 - ?
- Em 70.90% dos casos, é possível afirmar que: se Tempo_Logado = 0 a 22800 então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 80.30% dos casos, é possível afirmar que: se Tempo_Logado = 0 a 22800 então Sexo = F
- Em 73.61% dos casos, é possível afirmar que: se Tempo_Logado = 0 a 22800 então Idade = 25 - ?
- Em 71.90% dos casos, é possível afirmar que: se TMA = 601 a 1200 então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 80.77% dos casos, é possível afirmar que: se TMA = 601 a 1200 então Sexo = F
- Em 76.47% dos casos, é possível afirmar que: se TMA = 601 a 1200 então Idade = 25 - ?

- Em 75.14% dos casos, é possível afirmar que: se Sexo = F então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 85.77% dos casos, é possível afirmar que: se Escolaridade = SEGUNDO GRAU COMPLETO então Sexo = F
- Em 70.70% dos casos, é possível afirmar que: se Idade = 25 - ? então Escolaridade = SEGUNDO GRAU COMPLETO
- Em 75.26% dos casos, é possível afirmar que: se Escolaridade = SEGUNDO GRAU COMPLETO então Idade = 25 - ?
- Em 77.72% dos casos, é possível afirmar que: se Idade = 25 - ? então Sexo = F
- Em 72.48% dos casos, é possível afirmar que: se Sexo = F então Idade = 25 - ?

Os resultados da análise nos sistemas *Mining Center* e Weka, podem ser visualizados nas figuras 7.3 e **??**, respectivamente.

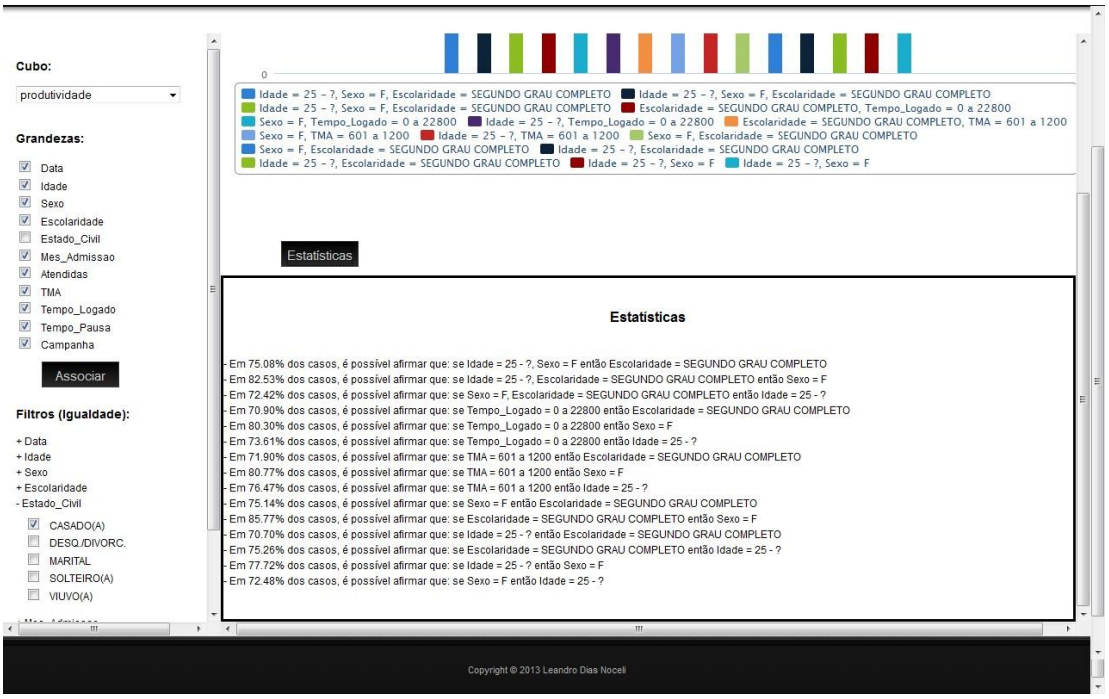


Figura 7.3: Resultado obtido no Mining Center (Casados)

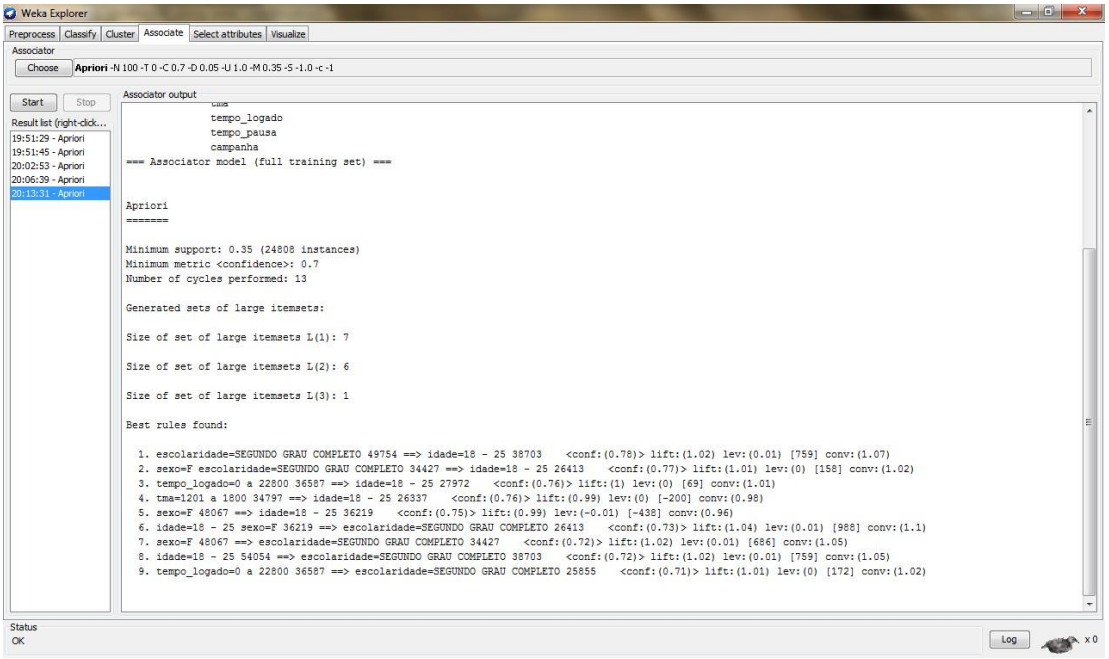


Figura 7.4: Resultado obtido na Weka (Casados)

## 7.3 Análise I x Análise II

Analisando do ponto de vista empresarial, quanto menor tempo médio de atendimento, menor será a quantidade de pessoas necessárias ao atendimento de clientes. Nesta sintonia, verificaremos, a seguir, as seguintes regras obtidas nas análises I e II:

- **Solteiros(as)**: Em 75.69% dos casos, é possível afirmar que: se TMA = 1201 a 1800 então Idade = 18 - 25
- **Casados(as)**: Em 71.90% dos casos, é possível afirmar que: se TMA = 601 a 1200 então Escolaridade = SEGUNDO GRAU COMPLETO
- **Casados(as)**: Em 80.77% dos casos, é possível afirmar que: se TMA = 601 a 1200 então Sexo = F
- **Casados(as)**: Em 76.47% dos casos, é possível afirmar que: se TMA = 601 a 1200 então Idade = 25 - ?

Baseando-se nas regras acima, é possível notar que pessoas solteiras e com idade variando entre 18 e 25 anos, possuem um TMA entre 1201 e 1800 segundos, enquanto as casadas são em sua maioria do sexo feminino, possuindo idade igual ou superior a 25 anos e o seu tempo médio de atendimento varia entre 601 e 1200 segundos.

Isto nos mostra que pessoas casadas possuem um maior comprometimento com o trabalho, pois conseguem atender aos clientes de forma mais rápida. O que não é possível ter ciência a partir desta análise é o grau de satisfação do cliente em relação ao atendimento prestado.

# 8 Conclusão

A descoberta de conhecimento em bases de dados abrange uma vasta área do conhecimento, permitindo a escolha de diversos caminhos no desenvolvimento de trabalhos e pesquisas.

O caminho adotado para o desenvolvimento da ferramenta pode não ter sido o mais adequado, no entanto norteou o enredo deste trabalho e, será, de fato, utilizado como molde para o desenvolvimento de aplicações que vislumbram a possibilidade de obter conhecimentos em bases de dados, ao menos na empresa de onde os dados foram extraídos.

O conhecimento adquirido ao longo desta investidura foi suficiente para o desenvolvimento da aplicação e estudo dos resultados, tendo sido contemplados diversos conceitos da teoria, tanto no que concerne à descoberta de conhecimento em bases de dados, como em se tratando de programação de aplicações.

Acerca do conhecimento exposto aqui, é possível concluir os seguintes itens:

- Quanto à Sumarização Descritiva dos Dados, sua necessidade é evidente, pois de outra forma, muitas vezes, não seria possível obter conhecimento em bases de dados. Como, na maioria das vezes, os dados são os mais diversos possíveis, associá-los seria um empecilho. Partiu da sumarização, os intervalos adotados para tempo médio de atendimento, chamadas atendida, datas, etc.

- Já a formação de cubos de dados é a essência da mineração, onde os mais diversos tipos de dados são correlacionados possibilitando a descoberta do conhecimento.

Já em se tratando das análises em questão, conclui-se:

- a maioria dos operadores de atendimento que possuem entre 18 e 25 anos também são solteiros;

- grande parte dos operadores de atendimento que possuem segundo grau são mulheres;

- aqueles que possuem segundo grau completo, em sua grande maioria, também são solteiros; e

- operadores de atendimento casados, apresentam tempo médio de atendimento menor que solteiros e possuem mais de 25 anos, enquanto os solteiros apresentam idade entre 18 e 25 anos.

Os dados obtidos podem ser úteis na determinação da melhor forma de tratamento interpessoal no interior da empresa, norteando a comunicação e o relacionamento de líderes com subordinados. Outras informações podem ser extraídas com a utilização do filtro, sendo a ferramenta extremamente útil na determinação do perfil, auxiliando no processo de seleção de pessoas para a ocupação de cargos.

Em se tratando da relação existente entre tempo médio de atendimento, pessoas solteiras e casadas, é possível notar que pessoas casadas despendem menor tempo para atendimento das chamadas, o que permite concluir, desconsiderando-se satisfação do cliente quanto ao atendimento prestado por pessoas casadas e solteiras (não é possível mensurar com os dados da análise), que pessoas casadas atendem mais rapidamente às solicitações dos clientes, o que demonstra empenho do trabalhador. Ainda é permitido concluir que, se todas as pessoas da central de atendimento fossem casadas, talvez a quantidade de operadores necessários seria menor.

Futuramente, serão incluídas novas formas de mineração de dados ao sistema *Mining Center*, possibilitando ao usuário, novos parâmetros e novas informações a serem interpretadas. Um exemplo, será a implementação do algoritmo K-Means. Além disso, novos cubos de dados serão preparados e incluídos na ferramenta *Mining Center*. E ainda, a ferramenta *Mining Center* poderá auxiliar na criação de uma nova ferramenta de mineração de dados, muito mais robusta e eficiente.

Este trabalho poderá auxiliar universitários e empresas que desejem obter conhecimento sobre suas bases de dados.

# Referências Bibliográficas

Agrawal, R.; Srikant, R. **Fast algorithims for mining association rules**. Santiago, Chile, 1994.

Fayyad, U.; Piatetsky-Shapiro, G. ; Smyth, P. **From data mining to knowledge discovery in databases**. Menlo Park: MIT Press, 1996. AI Magazine.

Gray, J.; Bosworth, A.; Layman, A. ; Pirahesh, H. **Data cube**. In: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals, 1996.

Han, J.; Kamber, M. **Data mining**. In: Concepts and Techniques, San Francisco, EUA, 2006. Morgan Kaufmann.

Microsoft. **Data mining algorithms (Analysis Services - Data Mining)**. EUA, 2012. [Acessado em 20/03/2013]. Disponível na Internet: http://msdn.microsoft.com/en-us/library/ms175595.aspx

Witten, I. H.; Frank, E. **Data mining**. In: Pratical Machine Learning Tools and Techniques, San Francisco, EUA, 2005. Elsevier Inc.