

## MINERAÇÃO DE DADOS PARA DESCOBERTA DE CONHECIMENTO EM DADOS DE PROMOÇÃO À SAÚDE

Guilherme Gregory<sup>1</sup>, Fabrício Pretto<sup>2</sup>

**Resumo:** Atualmente, as corporações se deparam com um problema que são os grandes volumes de informações em seus bancos de dados. Muitas vezes, essas corporações não possuem ferramentas adequadas para extrair essas informações e transformá-las em conhecimento. Sendo assim, esse estudo tem como objetivo auxiliar uma empresa de promoção à saúde na tomada de decisões em relação aos seus clientes. Para que isso fosse possível utilizou-se as técnicas de Data Mining, que por sua vez aborda a sua aplicabilidade, quais áreas de atuação mais comuns da mineração e qual a real importância dessa mineração para a tomada das decisões. Dentre os métodos permitidos de Data Mining, foram utilizados os métodos por regras de associação e clusterização. As informações necessárias para a mineração dos dados foram extraídas do software de gestão de clientes da empresa analisada. Os principais campos que foram analisados e relacionados são, região de atendimento do paciente, gênero, nível de glicose, pressão arterial, tabagismo, consumidores de bebidas alcoólicas, pessoas que praticam atividades físicas, se a alimentação dos mesmos são saudáveis ou não, se participam de programas de acompanhamento na qualidade de vida ou não. Com base no referencial teórico e aplicando as técnicas de Data Mining, foi possível descobrir as relações e correlações da saúde das pessoas dos Vales do Taquari e Rio Pardo.

**Palavras-chave:** Mineração de Dados. Data Mining. Mineração na saúde. Extração de dados.

### 1 INTRODUÇÃO

Atualmente, o volume de dados armazenados pelas empresas vem aumentando consideravelmente, com isso, surgem os problemas de como tratar da melhor forma esses dados armazenados. Hoje é fundamental para os gestores, administradores ou até mesmo os funcionários das empresas, saberem de resultados que estão sendo gerados pelos serviços.

---

1 Graduado em Sistemas de Informação – Univates.

2 Mestre em Ciência da Computação – Professor no Centro Universitário UNIVATES.

Na área de saúde, de forma específica, há certa precariedade nas informações que estão sendo geradas para pacientes em tratamento, ou até mesmo para consultas rotineiras. O que se sabe muitas vezes é o diagnóstico do paciente naquele período e não dados históricos desse paciente.

Pensando nisso, a motivação para a realização desse estudo, foi auxiliar uma empresa da área de promoção a saúde a extrair as informações de um de seus sistemas de informação e acompanhamento de clientes, de forma que esses dados sejam tratados e analisados pelos profissionais da área. De posse dessas informações, torna-se possível que os profissionais tomem algumas decisões e também que possam melhorar os seus programas de prevenção de doenças, de qualidade de vida, doenças crônicas, entre outros.

O grande problema enfrentado pela área hoje, é conseguir controlar o histórico dos usuários, pois o sistema utilizado pela empresa de saúde é terceirizado, onde o fornecedor que desenvolveu não implantou um módulo de relatórios robusto ou até mesmo um módulo para as tais análises dos prontuários desses usuários. Com esse problema, outra dificuldade da área é tratar os seus pacientes de forma proativa, pois como o sistema não contempla um módulo de avaliação desses pacientes, é muito complexo saber com que tipo de paciente estão trabalhando. Apenas pacientes que já possuem algum problema ou doença, trabalhando-se esses reativamente.

Os dados contemplados no sistema atualmente estão distribuídos de forma incoerente, pois muitos deles não estão preenchidos ou estão informados de forma errada. Dificultando assim a manipulação dessas informações. Pensando então em uma estratégia para solucionar esse problema, observou-se a possibilidade de trabalhar com a mineração desses dados (Data Mining) para a geração de resultados concretos e de muita valia para ações futuras, de forma proativa ao tratamento.

O Data Mining por sua vez, é uma técnica de descoberta de conhecimento que se aplica a um grande volume de dados que possam estar escondidos dentro dos bancos de dados das corporações. A mineração de dados vem crescendo cada vez mais entre as soluções aplicadas nas médias e grandes corporações.

O objetivo principal desse trabalho é buscar descobrir conhecimento com base nos dados de paciente armazenados em um sistema de informação de uma operadora de planos de saúde. A partir dessa informação minerada, profissionais da área de saúde (médicos, psicólogos, enfermeiros e afins) poderão verificar como se encontra a saúde das pessoas dos Vales do Taquari e Rio Pardo ao longo do tempo, assim como, acompanhar a tendência com base nos dados do passado, para que possam ser tomadas algumas ações em relação a essa situação, sejam elas ações proativas ou reativas.

Para que esse objetivo seja alcançado, foi utilizada a metodologia de Estudo de Caso. Este projeto se trata de caso um estudo específico com dados extraídos dos usuários da área de promoção à saúde de uma empresa de saúde

dos Vales do Taquari e Rio Pardo. É uma pesquisa aplicada, uma vez que envolve interesses locais, ou específicos de empresas/pessoas determinadas. Por fim, a partir dos resultados gerados pelo sistema WEKA, realizou-se toda a análise desses resultados, apresentando os mesmos para os envolvidos na empresa estudada.

## 2 FUNDAMENTAÇÃO TEÓRICA

Data Mining é uma técnica que se aplica a uma grande quantidade de dados e informações que muitas vezes estão escondidas nos bancos de dados das corporações. Essa técnica pode ser aplicada em qualquer segmento (medicina, vendas, marketing, entre outras) que trabalhe com um grande volume de dados armazenados, porém para que essas informações possam ser analisadas de forma coerente, um especialista no assunto não pode ser dispensado (CARVALHO, 2002).

Ainda segundo Carvalho (2002), apesar de ser uma técnica antiga, a utilização da mesma passou a ser aplicada por muitas empresas devido a alguns motivos agora citados:

- **Grande volume de dados:** O Data Mining apenas se aplica nas grandes massas de dados, isso para que os algoritmos de mineração de dados tenham a sua real contribuição e, por fim, extrair as informações necessárias de forma confiável.
- **Organização dos dados:** As organizações atuais então padronizando seus dados utilizando DW (Data Warehouse) auxiliando na tomada de decisões, mas muitas vezes ainda não são extraídas as informações da forma mais inteligente, por isso a partir dessa organização e padronização dos dados, aplicam-se as técnicas de Data Mining, buscando de forma inteligente os melhores resultados.
- **Recursos computacionais muito mais potentes:** O Data Mining exige um poder computacional muito elevado para executar os seus algoritmos, porém com a evolução da microeletrônica e o baixo custo dos equipamentos, isso possibilita a aplicação do Data Mining. Outro fator que favoreceu muito foram os bancos de dados bem distribuídos, onde as informações estão mais coerentes e robustas.
- **Competição de mercado exige técnicas para tomada de decisão:** Como já foram citadas no item acima, as empresas geram um grande volume de dados nos seus sistemas, porém não sabem como utilizar a informação. O Data Mining auxilia essas empresas nas tomadas de decisões, tais como promoção de vendas, CRM (sistema de relacionamento com o cliente), pró-atividade no tratamento de saúde, entre vários outros segmentos, tornando assim a empresa mais forte perante o mercado extremamente competitivo.

- **Aquisição dos produtos Data Mining:** Alguns produtos já podem ser adquiridos em pontos comerciais, porém alguns ainda são de estado acadêmico, com isso há alguma limitação na escolha da melhor solução.

## 2.1 Data Warehouse e Data Mining

Conforme Machado (2009), o primeiro passo para o sucesso de uma mineração de dados é trabalhar esses dados que vão ser tratados. Para isso deve-se criar um Data Warehouse, que por sua vez é o pré-processamento das informações. O DW normalmente se dá por meio da criação de um novo banco de dados, constando então as informações já organizadas e tratadas.

De acordo com Machado (2009, p. 35):

a extração, organização e integração dos dados devem ser realizadas com o propósito de garantir a consistência e integridade das informações, construindo desta forma uma base de dados de alta qualidade e confiabilidade, que retrate efetivamente a realidade de negócios da empresa.

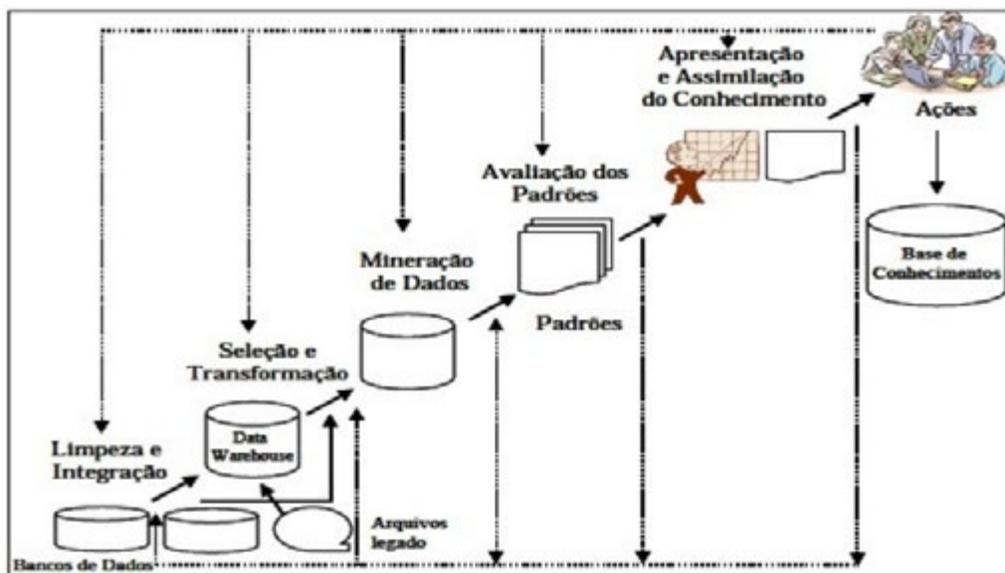
O repositório de dados proporcionado pelo Data Warehouse, faz com que o Data Mining possa atuar de forma muito mais acertada e com muito mais chances de sucesso. Afirma-se que para a execução da técnica Data Mining não é necessário uma criação de DW, porém as chances de êxito são muito menores sem essa prévia organização (CARVALHO, 2002).

Para Machado (2009) o sucesso de um bom Data Warehouse é um levantamento de requisitos muito criterioso, pois esses dados serão a base de todo o processo de mineração de dados. Caso o projeto comece errado, não terá o seu objetivo alcançado.

## 2.2 Fases de uma mineração de dados

Como afirma Cortês (2002), o processo de busca de conhecimento KDD (Knowledge Discovery in Databases), é um dos passos mais importantes na mineração de dados, pois é o processo que será seguido para atingir o objetivo final, sendo ele, o conhecimento. As etapas de um processo de Data Mining estão ilustradas na Figura 1 e explicadas na sequência.

Figura 1 – Fases de um processo de mineração



Fonte: CORTÉS, p. 25.

**1. Limpeza dos dados (*cleaning data*):** na primeira etapa é feita a limpeza de todos os dados inconsistentes, nulos ou fora dos padrões a serem analisados.

**2. Integração dos dados:** nessa etapa entra a tecnologia DW (Data Warehouse), onde serão integrados, mantendo a consistência e coerência dos mesmos.

**3. Seleção dos dados:** esta fase selecionará os dados necessários para a fase de transformação de dados, ainda assim descartando algum resquício de dados irrelevantes.

**4. Transformação dos dados:** na etapa de transformação dos dados, é o momento da preparação dos mesmos para a fase de mineração (Data Mining), essas transformações podem ser feitas utilizando várias operações de mineração, essas serão vistas nas próximas seções.

**5. Mineração dos dados:** nessa serão aplicadas as técnicas de mineração de dados, verificando qual delas se aplica a necessidade da extração.

**6. Avaliação dos padrões:** a sexta fase é uma pré-filtragem para a análise final, aqui serão avaliadas as informações verdadeiramente interessantes conforme o objetivo da mineração. Isso será feito através de algumas medidas de interesses.

**7. Apresentação e assimilação do conhecimento:** Na sétima e última fase da busca de conhecimento, é o momento de apresentar o resultado de forma visual e amigável ao usuário final.

### 2.3 Técnicas de Data Mining

Atualmente existem diversas técnicas aplicadas ao Data Mining, todas elas são genéricas e podem ser implementadas por diversas ferramentas de mineração, sendo através de redes neurais artificiais, estatística ou inteligência artificial. Contudo, existem cinco técnicas básicas que atendem a visão global da mineração de dados, lembrando que se deve analisar e verificar qual delas se enquadra melhor com a situação a ser implementada (SILVA, 2004).

As referidas técnicas são descritas como segue:

- **Mineração por Classificação:** é uma das mais utilizadas atualmente pelo fato de parecer muito com tarefas do cotidiano humano, sendo que o ser humano é muito induzido em classificar suas ações e pensamentos, definindo redes sociais, classificação de ambientes como moradia, classes sociais, entre outros (CARVALHO, 2002).
- **Regras de Associação:** semelhante à técnica por classificação, porém a característica que as diferem é que por regras de associação não são classificados os grupos previamente, mas sim criados a partir da associação e relação entre os dados extraídos, à medida que o algoritmo for encontrando novos dados iguais aos já criados, ele agrega esse dado ao grupo relacionado.
- **Previsão:** ocorre com base em dados históricos, ou seja, trabalha em cima de estatísticas, isso se aplica, por exemplo, em dados da bolsa de valores, determinando se um índice vai subir ou descer nos próximos dias. Outra área que se aplica essa técnica poderia ser na previsão de crescimento de uma cidade em 25 anos, tudo isso buscando de dados do passado (CARVALHO, 2002).
- **Agrupamentos ou *Clustering*:** tem como objetivo criar segmentos do banco de dados em subconjuntos, com base na semelhança entre os elementos estudados, sendo que um elemento pode estar em um ou mais conjuntos (SILVA, 2004).
- **Mineração por Afinidade:** segundo (CARVALHO, 2002), determina se uma massa de dados tem alguma correlação por fatos simultâneos. Um exemplo que facilita o entendimento seria as compras de uma pessoa no supermercado, essa pessoa completa seu carrinho de compras com vários itens, posteriormente as vendas de tais itens podem ser diretamente direcionadas a esses clientes, por afinidade das compras anteriores, essas regras são geradas através de inteligência artificial.

### 3 PROCEDIMENTOS METODOLÓGICOS

Para desenvolvimento do processo de mineração, reuniões foram realizadas com alguns profissionais envolvidos com o sistema de promoção à saúde. Pelo fato de não haver um problema específico, a busca se deu de forma exploratória, buscando descobrir conhecimento por meio da aplicação de alguns algoritmos de Data Mining e pelo cruzamento de dados definidos como importantes, pelos profissionais da área.

De posse de seus anseios, partiu-se para o processo de aquisição dos dados, ainda em formato bruto, ou seja, extraídos da base de dados do sistema de informação. Os dados passaram por um processo de limpeza e normalização, para que então pudessem ser utilizados no aplicativo WEKA, próprio para Mineração de Dados.

Esse trabalho caracterizou-se como Estudo de Caso por manipular dados de uma empresa específica, de perfil definido. De acordo com Yin (2001, p. 21):

O estudo de caso permite uma investigação para se preservar as características holísticas e significativas dos eventos da vida real, tais como ciclos de vida individuais, processos organizacionais e administrativos, mudanças ocorridas em regiões urbanas, relações internacionais e maturação de alguns setores.

Por estar direcionado a um problema concreto, a pesquisa classifica-se ainda como aplicada, que segundo Vergara (1998, p.45):

A pesquisa aplicada é fundamentalmente motivada pela necessidade de resolver problemas concretos, mais imediatos, ou não. Tem, portanto, finalidade de prática, ao contrário da pesquisa pura, motivada basicamente pela curiosidade intelectual do pesquisador e situada sobre tudo no nível da especulação.

### 4 TRABALHOS RELACIONADOS

A área de Mineração de Dados busca descobrir conhecimento em diferentes segmentos de negócio, principalmente em situações nas quais há um grande volume de dados envolvido. Áreas de negócio como: Marketing, Biologia, Medicina, Ensino e Saúde, são exemplos de trabalhos que vem sendo desenvolvidos. Nesse trabalho, em particular, buscou-se trabalhos relacionados com a área da saúde, foco desse estudo.

Conforme cita Maciel et al. (2015), a área de saúde é uma das áreas mais importantes para fazer esse tipo de trabalho, a mineração de dados. Com isso, pode-se prevenir ou tratar diversas doenças, infecções e até mesmo tratar através de um diagnóstico. Em seu trabalho (MACIEL, 2015), foi aplicada a técnica de mineração por classificação em triagem de risco de vida em uma

Unidade Ponto de Atendimento (UPA) do Sistema Único de Saúde (SUS). Essa é uma técnica muito interessante quando se quer trabalhar com grau de confiança de um resultado. O conjunto de dados utilizado para esse estudo estavam relacionados aos sinais vitais das pessoas. Depois de obtido os dados de referência, o autor aplicou classificações para os graus de riscos, numerando de 1 a 4, denominado Eletiva, Baixa, Média e Alta respectivamente. Através de alguns campos obtidos na consulta do indivíduo, tais como, pressão, peso, frequência respiratória, entre outros, calculou-se o grau de risco dessa pessoa e então se classificou onde esse indivíduo se enquadraria. Os resultados obtidos pelos autores foram os seguintes, foram feitos testes com 10.997 pacientes e desses então foram constatados. Eletivo = 3.802, Baixo = 6.173, Médio = 981 e Alto = 41. Nesse caso pode-se afirmar que a maioria dessa amostra está com um risco baixo ou nem isso.

Funchal et al. (2015), também fizeram um estudo em relação à saúde pública, porém as suas classificações foram um pouco além das descritas pelo autor Maciel (2015). Funchal et al. (2015) classificaram os riscos por cores, denominando quais as doenças estavam relacionadas a tais cores. Como exemplo: Alteração no estado mental = Vermelho, Glicemia > 250 mg/dl e sinais de desidratação = Amarelo e assim por diante. Aplicando então as regras de classificação, os autores foram distribuindo os pacientes na sua devida classificação. Esse é um estudo que pode prevenir muitas mortes precoces por não se ter a informação correta. Tratando essas pessoas em programas de qualidade de vida, certamente se tem bons resultados.

Conforme estudo de Mattos et al. (2004), em técnicas de Data Mining aplicadas à saúde bucal, obtiveram bons resultados utilizando a técnica de clusters dividindo crianças entre 6 a 12 anos de idade, moradores das capitais dos estados do Rio Grande do Sul, Santa Catarina e Paraná. Os autores separaram essas crianças pelas suas regiões e avaliaram o grau de cáries que as mesmas continham, foi avaliada também a questão de dentes permanentes ou não. Por fim chegaram à conclusão que as regiões se equiparavam muito na quantidade de cáries das crianças, porém verificaram que as crianças de seis anos do sexo feminino possuíam mais cáries que os demais.

Kuretzki (2009), utilizou meios eletrônicos para a descoberta de conhecimento. A pesquisa do autor esteve ligada a área de Nutrição, para isso utilizou o Algoritmo Apriori. Foram utilizados protocolos de Nutrição e suas boas práticas para que assim houvesse relacionamento com o estado de saúde uma pessoa. O autor concluiu que se todos os processos nutricionais forem seguidos rigorosamente, as pessoas terão uma longevidade muito maior.

Santos (2007), realizou uma pesquisa relacionada a resultados de WHOQOL, que por sua vez é um questionário aplicado nos quesitos Físico, Psicológico, Nível de Independência, Relações Sociais, Meio Ambiente e Espiritualidade / Crenças Pessoais. Os dados aplicados são traduzidos através

da escala de likert, pois as respostas variavam de 1 a 5. Por fim concluiu Santos (2007, p.89):

Através desses resultados comprova-se a importância do uso de técnicas de Mineração de Dados, no monitoramento de fatores que possam interferir nos indicadores de Qualidade de Vida, pois com a utilização apenas dos métodos estatísticos propostos no instrumento WHOQOL, seria possível apenas a identificação dos índices de QV da amostra ou de determinados segmentos, enquanto através do uso das técnicas de Data Mining, foi permitido identificar entre os dados disponíveis, fatores que contribuem na predição dos indicadores de qualidade de vida.

## 5 DESENVOLVIMENTO DA PESQUISA

A estrutura aplicada nesse estudo de caso, foi construída junto com os profissionais da área de promoção à saúde da empresa estudada. Foram realizadas reuniões iniciais para a estruturação dos dados desejados e seus os relacionamentos até reuniões de acompanhamento desses relacionamentos e seus resultados.

Inicialmente, foram reunidos profissionais dos três tipos de atividades realizadas pela área, sendo eles enfermeiros, nutricionistas e psicólogos. Nessa reunião, os envolvidos apresentaram um grande interesse em buscar o conhecimento entre hábitos rotineiros e sintomas psicológicos relacionados com as doenças crônicas mais frequentes entre os seus pacientes, sendo elas a hipertensão e o diabetes.

Os profissionais, escolheram a hipertensão e o diabetes como pontos principais, pois principalmente a hipertensão é a doença que mais causa mortes no país e é muito comum em todas as idades, não somente em pessoas mais idosas como se tinha o conhecimento. O diabetes está entre as primeiras causadoras de óbitos também, porém em menos quantidade que a hipertensão.

Após as reuniões iniciais concluídas, foram relacionados, conforme Tabela 1, os dados para cruzamento, todos os campos necessários para a realização da extração das informações e conseqüentemente o cruzamento dos mesmos. Esses dados foram classificados em dois grandes grupos: o primeiro grupo contendo todos campos aferidos ou que possuem valor exato; já o segundo grupo são campos autoavaliativos, ou seja, são valores onde o próprio paciente vai avaliar a sua situação conforme o que lhe for questionado. Essas perguntas são aplicadas de duas formas: pessoalmente ou por telefone. O questionário pessoal se aplica na primeira consulta do paciente, sendo que a cada 6 meses é aplicado o mesmo questionário via telefone, para fins de telemonitoramento e atualização dos dados.

Tabela 1 – Dados para cruzamento

| <b>Aferidos ou Valores Exatos</b> | <b>Autoavaliativos</b>          |
|-----------------------------------|---------------------------------|
| Peso                              | Estresse Diário                 |
| IMC                               | Ansiedade                       |
| Pressão arterial                  | Excesso de Sal                  |
| Glicemia                          | Excesso de Frituras             |
| Triglicerídeos                    | Alimentação Saudável            |
| Sexo                              | Frequenta Médico Periodicamente |
| Idade                             | Consome Frutas                  |
| Região de Atendimento             |                                 |
| Fumantes                          |                                 |
| Consumem Álcool                   |                                 |
| Atividade Física                  |                                 |
| Diabetes                          |                                 |
| Hipertensão                       |                                 |

Fonte: dos autores.

As informações foram oriundas de um sistema de acompanhamento de pacientes em promoção a saúde nas regiões dos Vales do Taquari e Rio Pardo. A massa de dados foi gerada em um arquivo do tipo de valores separados por vírgulas (CSV), extraído de forma manual diretamente do banco de dados Oracle. Essa consulta foi realizada através da linguagem SQL utilizando-se a ferramenta PL/SQL Developer (programa para manipular dados na linguagem PL/SQL). Frente a isso, foi realizada a segunda etapa onde os dados são melhores organizados eliminando valores nulos, irrelevantes ou incompletos para a análise futura, essa etapa é chamada de pré-processamento.

Durante o processo de captura e transferência dos dados, muitos registros necessitaram de ajuste manual, pois estavam digitados incorretamente ou em branco. Muitos dados foram informados em campos abertos, do tipo texto, apesar de existirem campos específicos para registro da informação.

Após a limpeza e pré-seleção dos dados, no mesmo período de amostragem, restaram 5.700 registros no arquivo CSV, reduzindo consideravelmente a quantidade desses registros. A partir de então, o próximo passo realizado foi a conversão do arquivo CSV no formato interpretado pelo programa de mineração de dados WEKA, software desenvolvido na Universidade de Waikato na Nova Zelândia. Essa conversão foi realizada através de um script em PHP que transformou uma entrada CSV em um arquivo no formato com extensão .ARFF.

### **5.1 Aplicação das técnicas - Agrupamento**

Os campos envolvidos nos primeiros cruzamentos, são bem comuns para a avaliação clínica de um paciente, por isso foi pensado em agregar os

valores psicológicos para analisar se tais campos têm alguma influência nos resultados. Para esse primeiro experimento, o algoritmo K-means foi aplicado. De um total de 5.700 registros, 4.560 apresentaram relacionamentos com algum dos 5 clusters criados. Os relacionamentos resultantes foram os seguintes:

**1º Cluster:** 16 Até 25 anos, Excesso\_sal = Não, Atividade\_fisica = Nao, Diabetes = Sim. Total de 852 relacionamentos

**2º Cluster:** 26 Até 35 anos, Estresse\_diario = Sim, Alimentacao\_saudavel = Não, Pressão arterial média = 150,90, Hipertensos = Sim. Total de 2558 relacionamentos

**3º Cluster:** 36 Até 45 anos, Excesso\_sal = Sim, Alimentacao\_saudavel = Sim, Atividade\_fisica = Sim, Hipertensos = Não. Total 380 relacionamentos

**4º Cluster:** 46 Até 55 anos, Alimentacao\_saudavel = Não, Atividade\_fisica = Sim, Diabetes = Não. Total 630 relacionamentos

**5º Cluster:** Acima 56 anos, Atividade\_fisica = Não, Diabetes = Sim, Hipertensos = Sim. Total 140 relacionamentos.

O 2º Cluster chamou a atenção dos profissionais por dois motivos distintos: um pela quantidade total de relacionamentos (2558) e outro pelo campo Estresse\_diario estar presente nessa massa de registros. Lembrando que o Estresse\_diario é um campo autoavaliativo, sendo assim ele pode não estar informado corretamente.

## 5.2 Aplicação das técnicas - Apriori

Frente a esses resultados, aplicou-se a segunda técnica avaliada nesse estudo, as regras por associação. Os profissionais solicitaram que fosse realizado cruzamentos com todos os campos previamente separados para utilização na mineração, procurando assim uma associação entre esses campos com os resultados encontrados no 2º Cluster. Para a realização desses cruzamentos, foi utilizado o algoritmo Apriori das regras por associação. Conforme referencial teórico, esse algoritmo trata as informações por frequência, por isso os dados foram ajustados para contemplar o algoritmo, ajustando os valores dos atributos para tipos mais normalizados como: Sim/Não para os campos Estresse, Ansiedade, Tabagismo e demais; bem como Masculino/Femino para gênero.

Os dados utilizados nos cruzamentos foram: Região, Gênero, Estresse, Ansiedade, Atividade\_fisica, Tabagismo, Consumo\_alcool, Alimentacao\_saudavel, Hipertenso, Diabetico, Excesso\_sal, Excesso\_fritura, Consume\_fruta e Frequenta\_medico, com 90% de confiança.

Nesses resultados buscou-se quais os campos que poderiam influenciar nas pessoas hipertensas. Frente a isso, observou-se os seguintes resultados:

1º - {26 até 35 anos, mulheres, vale taquari, estresse\_diario = SIM, atividade\_fisica = NAO, alimentacao\_saudável = SIM} -> Hipertensos -> 92,57% confiança

2º - {26 até 35 anos, mulheres, vale taquari, estresse\_diario = SIM, excesso\_sal = NAO, atividade\_fisica = NAO} -> Hipertensos -> 90,36% confiança

3º - {26 até 35 anos, homens, vale rio pardo, fumantes = SIM, consomem álcool = SIM, Excesso\_sal = NAO} -> Hipertensos -> 91,27% confiança

Dos resultados analisados, o 2º e 3º cruzamento chamaram a atenção dos profissionais por dois motivos: o 2º se manteve com o campo estresse\_diario = SIM, onde apresentou excesso\_sal = NAO (normalmente é um fator que está ligado diretamente a hipertensão), porém o campo de atividade física = NAO (fator que se não realizado pode elevar a pressão arterial). Frente a isso, os profissionais solicitaram que fosse retirado o campo estresse diário da relação.

Após a retirada do campo estresse diário, foi a realizado novamente a mineração desses dados e com isso apresentou o seguinte resultado:

{26 até 35 anos, mulheres, vale taquari, excesso\_sal = NAO, atividade\_fisica = NAO} -> Hipertensos -> 74,50% confiança

Com esse resultado, percebeu-se que o estresse diário está muito ligado a hipertensão. Os profissionais tiveram uma surpresa no resultado, pois não tratavam os sintomas psicológicos com as doenças crônicas mais comuns entre os brasileiros. Essas doenças estão relacionadas normalmente a costumes ou hábitos alimentares e falta de exercícios.

No resultado do 3º cruzamento, avaliou-se o relacionamento entre pessoas fumantes e que consomem bebidas alcoólicas. Nessa análise relatou-se que o problema de hipertensão está ligado ao álcool, dado que o excesso de sal = NAO. Para que esse cruzamento fosse avaliado como plausível, foi retirado o campo de pessoas fumantes. Nesse caso o resultado apresentado foi o seguinte:

{26 até 35 anos, homens, vale rio pardo, consomem álcool = SIM, Excesso\_sal = NAO} -> Hipertensos -> 68,70 confiança

Com o relacionamento aplicado, o índice foi reduzido consideravelmente. Frente a isso, foi avaliado a hipótese de que o tabagismo estava ligado a hipertensão, com isso foi realizado um novo cruzamento retirando o consumo de álcool e adicionado o tabagismo. Após gerada a nova mineração, o resultado foi o seguinte:

{26 até 35 anos, homens, vale rio pardo, consomem álcool = SIM, Excesso\_sal = NAO} -> Hipertensos -> 76,30 confiança

Os resultados apresentados ainda não foram satisfatórios, dado que o grau de confiança aplicado é de 90%. Perante isso, descartou-se esses novos relacionamentos, que não se mostraram confiantes. Com isso a análise foi realizada utilizando o cruzamento que apresentou o índice de 91,27% de confiança.

O cruzamento analisado apresentou uma quantidade muito elevada de homens do Vale do Rio Pardo que são fumantes; por consequência, os mesmos consomem bebidas alcoólicas, elevando assim a pressão arterial e tornando-os hipertensos.

### **5.3 Proposta de Implementação de Melhorias**

O processo de mineração de dados apresenta durante o processo de descoberta de conhecimento, situações pertinentes ao problema mas não diretamente relacionadas. É comum constatar que o processo de registro das informações está incorreto, bem como, o tipo de informação que está sendo armazenada. De maneira inevitável, os processos são realizados por pessoas, o que leva conseqüentemente a necessidade de mudança na forma que as mesmas interagem com o sistema, referindo-se a sua postura e comprometimento. Dentre as sugestões de melhoria no processo de aquisição e tratamento de dados para uma melhor mineração em um futuro projeto, foram apontadas as seguintes ações:

- Troca do sistema de gestão
- Treinamento de pessoal
- Revisão do questionário aplicado
- Automação do processo de extração de dados

## **6 CONSIDERAÇÕES FINAIS**

Conforme avaliações realizadas no estudo, o mesmo apresentou um grande problema relacionado as informações extraídas do sistema. Os dados estavam muito falhos, sendo que grande parte deles preenchidos de forma incorreta e muitos nem são informados, devido a não obrigatoriedade dos mesmos.

Apesar de muitas informações se apresentarem de forma incorreta, foi possível realizar a limpeza e seleção de alguns dados para realizar os cruzamentos e conseqüentemente a mineração desses dados. Após a realização da mineração, foram avaliadas duas descobertas que se destacaram com um grau de confiança elevado, sendo elas:

A) {26 até 35 anos, mulheres, vale taquari, estresse\_diario = SIM} -> Hipertensos -> 92,57% confiança: O resultado apresentou uma ligação entre estresse e a hipertensão. Essa relação aplicou-se a mulheres jovens,

fomentando uma preocupação por parte dos profissionais da área. Com a essa preocupação, os mesmos tomaram algumas decisões para a criar um programa de prevenção, tudo isso relacionado as informações encontradas.

B) {26 até 35 anos, homens, vale rio pardo, fumantes = SIM, consomem álcool = SIM, Excesso\_sal = NAO} -> Hipertensos -> 91,27% confiança: Essa relação apresentou ligações entre pessoas fumantes ao consumo de álcool, resultando em pacientes hipertensos. Com isso avaliou-se a hipótese de que o consumo de cigarros na região do Vale do Rio Pardo é maior devido as empresas fumageiras e fabricantes de cigarros que comporta essa região.

Foi realizado um plano de ação para tratar melhor as informações do sistema atual, ou até mesmo a troca do sistema buscando um software mais confiável, com consistência de dados, obrigatoriedade nos campos em informações cruciais para a futura análise, entro outros.

Contudo, os principais trabalhos futuros propostos nesse estudo foram dois: a primeira proposta é reavaliar as condições do sistema atual (caso não houver o interesse de trocar) e treinar os operadores do mesmo para o correto preenchimento dos formulários; a segunda proposta é a automatização do processo de mineração dos dados, independente do sistema de gestão implantado.

Por fim, o estudo foi apresentado para os profissionais envolvidos e os demais da área de promoção a saúde da empresa estudada. Esses profissionais se mostraram satisfeitos perante os resultados que a mineração de dados pode agregar aos trabalhos rotineiros, principalmente servir como subsídio para a prevenção de doenças.

## REFERÊNCIAS

CARVALHO, Luís A. V. de. **DATAMINING: A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. 2a Edição. São Paulo-SP: Érica, 2002.

CÔRTEZ, Sérgio da Costa. **Mineração de Dados – Funcionalidades, Técnicas e Abordagens**. PUC-RJ. Artigo, 2002 Disponível em: [ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02\\_10\\_cortes.pdf](ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf)

FUNCHAL, João P.; MADSEN, Carlos A.; ADAMATTI, Diana F. **Classificação automática de dados para a descoberta de conhecimento: um estudo de caso para a classificação de risco na área da saúde**. Passo Fundo-RS. Revista Brasileira de Computação Aplicada, 2015.

KURETZKI, Carlos Henrique. **Técnicas de Mineração de Dados Aplicadas em Bases de Dados da Saúde a Partir de Protocolos Eletrônicos**. Curitiba-PR. Dissertação, 2009.

MACHADO, Felipe N.R. **Tecnologia de Projetos de Data Warehouse**. 4a Edição, 63 São Paulo-SP: Érica, 2009.

MACIEL, Thales V.; SEUS, Vinicius R.; MACHADO, Karina dos Santos; BORGES, Eduardo N. **Mineração de dados em triagem de risco de saúde**. Passo Fundo-RS. Revista Brasileira de Computação Aplicada, 2015.

MATTOS, Merisandra C.; SELINGER, Tarcísio C.; SIMÕES, Priscyla W. T. de Azevedo. **Data Mining em Saúde Bucal por meio da Técnica de Clusterização e do Algoritmo K-means**. Unesc-SC. Congresso Brasileiro de Informática em Saúde. 2004.

SANTOS, Celso Bilynkiewicz dos. **Análise dos Resultados do WHOQOL-100 Utilizando Data Mining**. Ponta Grossa. Dissertação, 2007.

SILVA, Glauco C. **Mineração de Regras de Associação Aplicada a Dados da Secretaria Municipal de Saúde de Londrina – PR**. Porto Alegre-RS. Dissertação, 2004. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/8696/000586835.pdf?sequence=1>

VERGARA, Sylvia Constant. **Projetos e Relatórios de Pesquisa em Administração**. 2a Edição. São Paulo-SP: Atlas, 1998.

YIN, Robert K. **Estudo de Caso – Planejamento de Métodos**. 2 a Edição: Porto Alegre-RS: Bookmann, 2001.