

# SISTEMA DE RECOMENDAÇÃO PARA CONTROLE DA DIABETES

Rodrigo Rath<sup>1</sup>, Alexandre de O. Zamberlan<sup>1</sup>, Sylvio A. G. Vieira<sup>1</sup>

<sup>1</sup>Curso de Sistemas de Informação – Centro Universitário Franciscano

rodrigo.rath@gmail.com, {alexz,sylvio}@unifra.br

**Abstract.** *This paper presents a proposal for an information system with data mining techniques for the daily control of diabetes. The software tries to find good combinations of patterns from a database of foods consumed, sleep quality, type and amount of physical activity and recommends changes in diet and physical activity. This recommendation system was built with PHP, JavaScript and jQuery; using the MySQL database and the data mining algorithm C4.5.*

**Resumo.** *Este trabalho apresenta o desenvolvimento de um sistema de informação com técnicas de mineração de dados para o controle diário da diabetes. O software em questão, a partir de uma base de dados de alimentos consumidos, qualidade de sono, tipo e quantidade de atividade física, encontra padrões de boas combinações (alimentos e exercícios, por exemplo) e recomenda variações de alimentação e atividades físicas. Esse sistema de recomendação foi construído com PHP, JavaScript e jQuery; utilizando o banco de dados MySQL e o algoritmo de mineração de dados C4.5.*

## 1. Introdução

Segundo a Sociedade Brasileira de Diabetes [SBD 2013], a diabetes *mellitus*, ou simplesmente diabetes, está associada a doenças metabólicas em que pessoas possuem níveis elevados de glicose no sangue, pelo fato do pâncreas não produzir suficientemente insulina, ou as células do corpo não responderem à insulina que é produzida. Também, conforme a Sociedade Brasileira de Diabetes [SBD 2013], a glicose é a principal fonte de energia do organismo, contudo quando em excesso, pode trazer distúrbios à saúde, como excesso de sono, cansaço e problemas físico-táticos. A doença, quando não tratada corretamente, pode gerar complicações cardíacas, cerebrais, renais, problemas na visão, entre outras. De acordo com o Portal da Saúde do Governo Federal Brasileiro [PS 2013], “diabetes está se tornando a epidemia do século e já afeta cerca de 246 milhões de pessoas em todo o mundo. Até 2025, a previsão é de que esse número chegue a 380 milhões”. No Brasil, a ocorrência média de diabetes na população adulta é de 5,2%, o que representa quase 6,4 milhões de pessoas. A prevalência aumenta com a idade, isto é, a diabetes atinge 18,6% da população com idade superior a 65 anos [PS 2013].

Novamente, de acordo com a SBD (2013) e a ADA (2013), o tratamento é baseado em conceitos básicos, como: conscientização do paciente; alimentação adequada para cada tipo de diabetes e para o perfil do paciente; vida ativa, mais do que simplesmente exercícios; uso de medicamentos; monitoração constante e regular dos níveis de glicose. Portanto, surge um desafio, qual a melhor combinação de tipo e quantidade de exercícios físicos mínimos, ingestão alimentar ideal máxima, quantidade

mínima ideal de medicamentos para se ter um controle da diabetes. Dessa forma, propõe-se o uso de descoberta de conhecimento por meio de técnicas de mineração de dados em um sistema de informação que se comporte como um sistema de aconselhamento ou recomendação, a fim de indicar ou sugerir tipos e quantidades de atividades físicas, tipos e quantidades de alimentos, por exemplo, ao paciente para o melhor controle da doença.

Dessa forma, projetou-se e implementou-se um sistema informatizado de recomendação, tendo como fluxo: registro de dados de pacientes (glicose, insulina, quilocalorias, carboidratos, qualidade sono, atividades físicas); aplicação de técnicas mineração de dados (base de dados previamente preparada); extração de padrões de boas combinações (alimentação, exercícios, medicação, etc); novamente registro de dados de pacientes; recomendação ou aconselhamento a partir dos dados registrados e dos padrões encontrados.

## **2. Diabetes**

Segundo a Sociedade Brasileira de Diabetes [SBD 2013], diabetes é uma doença caracterizada pela elevação dos níveis de glicose no sangue, ocasionada pela baixa ou nenhuma produção de insulina pelo pâncreas ou ainda causada pela resistência a ação da insulina devido à obesidade [SBD 2013].

Segundo a Sociedade Brasileira de Diabetes [SBD 2013], é fundamental a educação do paciente, o acompanhamento constante de sua dieta, exercícios físicos, monitoração de seus níveis de glicose com o intuito de conservar os níveis de glicose adequados, seja a curto ou em longo prazo. Um controle cuidadoso é necessário para reduzir os riscos das complicações em longo prazo, e isso pode ser alcançado com uma combinação de dietas, exercícios e perda de peso (para diabetes tipo 2), uso de fármacos orais (para diabetes tipo 2) e o uso de insulina (tipo 1 e tipo 2 que não esteja respondendo à medicação oral).

Sabendo-se, portanto da gravidade da doença e da inexistência de uma cura definitiva, foi desenvolvido um esquema de medições e registros de informações do paciente, de hábitos alimentares e físicos. Todas essas informações são armazenadas em uma base de dados onde são aplicadas técnicas de mineração de dados para descobrir padrões da alimentação e atividades físicas. Tais padrões auxiliarão em recomendações de alimentação e exercícios que favoreçam a redução dos níveis de glicemia do paciente.

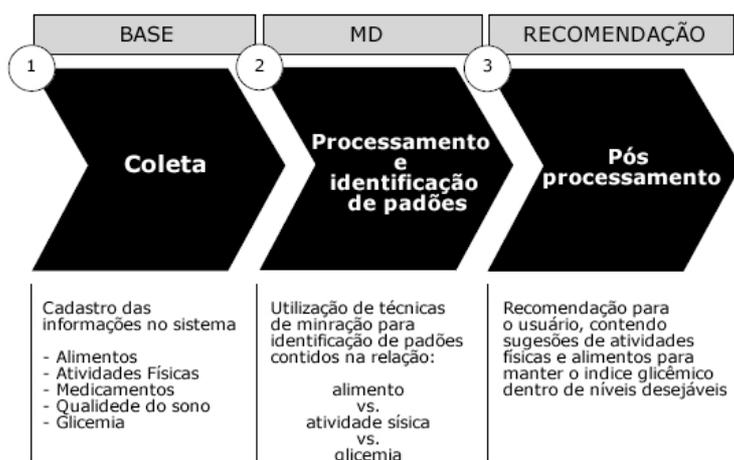
## **3. Descoberta de Conhecimento**

A Medicina, assim como outras áreas do conhecimento, produz uma grande quantidade de informação que é coletada, tratada e armazenada constituindo assim uma base de dados. Mas, apenas disponibilizar esses dados não é suficiente para um melhor aproveitamento das informações. É necessário ter ferramentas que facilitem a análise desses dados e auxiliem no desenvolvimento de estratégias de ação, isto é, a tomada de decisão sobre quadros clínicos de pacientes e/ou grupo de pacientes favorecendo o tratamento e/ou a prevenção de doenças [Prieto et al. 2004].

Segundo Prieto et al. (2004), a interpretação de dados tem sido feita por meio estatístico, e é a estatística um método matemático bem estabelecido com suporte

teórico que permite tais análises. Entretanto, há uma nova linha de pesquisa que também permite análise, a qual é conhecida como Mineração de Dados (do inglês *Data Mining* - DM). Essa metodologia objetiva por meio de métodos estatísticos ou não, a busca de modelos e/ou padrões dentro de uma base de dados que não seja possível de ser extraída em uma simples inspeção dos dados.

Alguns autores consideram os termos Mineração de Dados e Descoberta de Conhecimento em Banco de Dados (KDD - *Knowledge-Discovery in Databases*), como processos distintos [Rezende 2003]. Mas, para fins de desenvolvimento deste trabalho, os termos DM e KDD são abordados indistintamente, referindo-se ao processo de extrair conhecimento a partir de dados. Segundo Rezende (2003), a definição aceita por muitos autores foi elaborada por Fayyad, Piatetsky-Shapiro e Smyth (1996) como sendo: “extração de conhecimento de base de dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados” [Rezende 2003, p. 309]. Segundo o mesmo autor, Rezende (2003), existem variações no número de etapas que compõe o processo de mineração. Neste trabalho, a mineração foi dividida em 3 etapas: coleta de dados, processamento e identificação de padrões e pós processamento. A Figura 1 ilustra as etapas descritas.



**Figura 1 - Fases da Mineração de Dados. Adaptado de [Aranha 2007].**

Nos trabalhos de Carvalho (2005) e Tan et. al. (2009), é apresentado que os objetivos da mineração de dados são a predição (capacidade de realizar previsões a partir de uma fonte de dados) e a descrição (capacidade de encontrar padrões nos dados armazenados). Neste trabalho, assume-se a categorização adotada por Han e Kamber (2006) para descrever os principais métodos, que são associação, classificação, predição numérica e agrupamento.

Após apresentar definições e técnicas no contexto de descoberta de conhecimento e mineração de dados, registra-se que, para este trabalho, a ferramenta WEKA foi utilizada, bem como as técnicas de classificação e regressão da categoria de predição. A ferramenta WEKA foi utilizada para auxiliar na preparação dos dados a serem minerados, estabelecendo assim o formato dos dados, e por consequência a melhor forma de cadastro, armazenamento e agrupamento. A escolha dos algoritmos de testes e de produção, baseou-se na forma de funcionamento do sistema, haja visto que a mineração de dados realizada pelo sistema utiliza os dados do histórico alimentar, das

atividades físicas e dos índices de glicemia para sugerir determinados tipos de alimentos ou atividades, encaixando-se perfeitamente no contexto do sistema.

#### **4. Algoritmo C4.5**

O algoritmo C4.5 é um aprimoramento do algoritmo ID3, que foi um dos primeiros algoritmos de árvore de decisão. Seu aprimoramento consiste na capacidade e trabalhar com registros que apresentam valores indisponíveis (que não foram cadastrados, ou que foram considerados não relevantes para um determinado caso).

O algoritmo C4.5 foi desenvolvido na linguagem C, do qual derivou a versão em Java denominada J48 [Waikato 2014] [Tsunoda 2004] (que foi usado na fase de testes com o WEKA).

#### **5. Trabalhos Correlatos**

Nesta seção, são apresentados, sumariamente, alguns trabalhos científicos que envolvem Sistemas de Descoberta de Conhecimento relacionados à saúde.

Freitas (2012) buscou através da mineração de dados identificar padrões de respiração bucal e nasal de crianças, para auxiliar o diagnóstico precoce de alterações posturais que estão relacionadas à forma de respiração. No trabalho foi utilizado o algoritmo Perceptron Multicamadas de Rede Neural Artificial em ambiente WEKA, obtendo um resultado de 100% de acertos sobre a base de dados usada [Freitas 2012].

Vieira (2011) fez uso de mineração de dados através da Ferramenta R e classificador SVM, utilizando para tal as informações de 32 pacientes contidos na base de dados biológicos GEO, mantida pelo NCBI. Buscou identificar os genes com maior atividade que poderiam estar envolvidos com o adenoma de colo de intestino. Como resultado, percebeu-se que as características dos genes apesar de bem distintas e grande variação de nível de atividade, seguia um padrão permitindo então que o algoritmo identificasse estes padrões e sugerir a participação do(s) gene(s) como relacionado(s) ao adenoma.

Em Prieto et al. (2004), apresenta-se a elaboração de um software para coleta de dados de 186 pacientes com diabetes e/ou hipertensão arterial. E utilizando métodos de mineração aplicados aos dados colhidos durante as consultas ambulatoriais, observou a prevalência de sobrepeso e obesidade nos pacientes consultados, estando esses valores muito próximos de valores percentuais encontrados em pesquisas de âmbito nacional. Em seu trabalho foi implementado um algoritmo de árvore de decisão, Oracle Forms para o desenvolvimento das telas e Reports Builder para relatórios, além da linguagem PL/SQL para banco de dados.

Nesta seção, foram apresentados alguns trabalhos relacionados, destacando-se o trabalho de [Freitas 2012], pois utilizou o ambiente WEKA. Ressalta-se, também, que diferente dos trabalhos correlatos, neste trabalho não haverá a etapa de pré-processamento (fase de limpeza e preparação dos dados), uma vez que a base de dados foi construída para receber os dados da melhor forma para o processamento da mineração.

O diferencial deste trabalho frente aos trabalhos correlatos está no fato de que ele não será usado para o diagnóstico de alguma doença ou elemento que leve a uma doença, mas sim em sua proposta, de auxiliar o usuário (diabético e portando já

conhecedor de sua doença) a ter uma vida mais saudável, baseado no histórico alimentar e de atividades físicas, sugerindo a este combinações de alimentos e atividades que propiciem a ele um nível de glicemia mais adequado, e com isso a possível redução do uso de medicação.

## **6. Solução desenvolvida**

Este *software* foi concebido para um amplo acesso, tanto por computadores quanto por dispositivos móveis. Para alcançar esse propósito, foi utilizada a plataforma Web. Desta forma, Apache como servidor Web, PHP como linguagem de programação, MySQL como gerenciador de banco de dados, HTML e CSS para interface gráfica e o Framework jQuery para recursos visuais e funcionais da interface. Além dessas ferramentas tecnológicas, utilizou-se uma classe para renderização dos arquivos de interface, de forma a separar o sistema em três camadas: modelo, interface e controle (*Model, View, Controller* - MVC).

O funcionamento do *software* é considerado simples, exigindo do usuário poucos passos para sua utilização. Após o cadastro, todos os acessos feitos pelo usuário serão autenticados mediante um nome de *login* e uma senha, dando assim privacidade e segurança para seus dados. Após a autenticação, o usuário terá as opções para inserir os dados de sua alimentação, exercícios físicos, medicamentos, qualidade de sono e os valores obtidos de sua glicemia.

Para cada tipo de informação a ser inserida, foi construído um formulário, em que o usuário irá fornecer os dados solicitados, que ao serem preenchidos são enviados para o servidor, onde são organizados e armazenados no banco de dados. Esse cadastro, possibilitará consultas futuras a esses dados, bem como à execução da mineração de dados necessária para realizar as recomendações, objetivo principal deste trabalho.

### **6.1. Interface de Glicemia**

Através deste formulário, o usuário irá informar os dados dos valores da sua medida de glicemia, bem como: se a medida é em jejum ou não; e a data e hora em que foi obtida a medida. Nesta tela, também é apresentado um breve histórico das medições dos últimos dias (Figura 2).

Com o uso constante ou diário do *software*, será criado um histórico dos dados. Sobre este histórico, é empregada a mineração de dados com a utilização do algoritmo C4.5. Permitindo assim, extrair das informações os padrões que resultam em um bom índice de glicemia, contribuindo assim para o controle da diabetes.

### **6.2. Interface de Refeições**

Através desta interface, o usuário vai inserir os dados de sua alimentação. Uma vez selecionado o alimento, serão preenchidos automaticamente os campos referentes aos valores de quilo-calorias e carboidratos presentes no alimento. Cabe ressaltar que esta lista de sugestões de alimentos foi previamente cadastrada no *software*, sendo obtida junto a Sociedade Brasileira de Diabetes. Essa lista contém 900 tipos de alimentos naturais ou industrializados, que já tiveram seus valores nutricionais pesquisados.

The image shows a web interface for recording blood glucose measurements. At the top, there is a blue header with the SIACD logo and the text 'SIACD » SISTEMA DE INFORMAÇÃO PARA AUXILIAR O CONTROLE DO DIABETES'. Below this is a breadcrumb trail 'Glicemia'. The main content area is a form titled 'Medição de Glicemia'. It contains several fields: a 'Jejum' (Fasting) toggle switch currently set to 'OFF'; a 'Data' (Date) field containing '14-06-2014' with a calendar icon; a 'Hora' (Time) field containing '03:18:36' with a clock icon; and a 'Medida de glicemia' (Blood glucose measurement) input field. At the bottom of the form is a blue 'Salvar' (Save) button.

**Figura 2 - interface de cadastro de glicemia**

Juntamente com os dados dos alimentos, o usuário deverá informar dados complementares como: o tipo de refeição (café da manhã, almoço, lanche, jantar ou outro) e a data e hora em que foi realizada. Os campos que contêm a informação de data e hora da ingestão são campos automaticamente preenchidos. No entanto, essa informação pode ser alterada, caso o usuário não possa inserir os dados no momento da refeição. Nesta interface, também é exibido um breve histórico das últimas refeições em modo resumido, contendo apenas os totais ingeridos (quilo-calorias e carboidratos).

### **6.3. Interface de Atividades Físicas**

Nesta interface, é apresentado um formulário para o cadastramento das atividades físicas que o usuário realizar durante o dia. Para tal, foram dispostos campos de texto para a inserção do tipo de atividade (futebol, vôlei, padel, corrida, caminhada, etc...), tempo de atividade, data e hora em que foram realizadas. Nesta tela, igualmente as demais, é exibido um breve histórico das últimas atividades realizadas.

### **6.5. Interface de Recomendações**

Nesta tela, é exibido ao usuário o resultado final do *software*. Através da ação do usuário ao clicar no botão para solicitar a Recomendação, o *software* irá reunir os dados cadastrados de forma organizada, construindo três arquivos semelhantes ao arquivo ARFF (Attribute-Relation File Format) utilizado pelo WEKA. O primeiro “.names” contém o cabeçalho com o nome dos campos que serão utilizados. O segundo arquivo “.test” apresenta cerca de 60% do conteúdo presente no terceiro arquivo, “.data”. Sobre estes dois o algoritmo irá treinar e encontrar os padrões contidos nos dados, encontrando assim o padrão existente que contribui para um bom ou mal resultado da medida de glicemia. Na Figura 3, observa-se a estrutura destes arquivos.



Este procedimento foi necessário, pois os valores de glicemia podem oscilar, tornando praticamente impossível para o algoritmo encontrar um padrão em uma base de dados com poucos registros. Também, como se assumiu que os valores ideais de glicemia estão entre 80 e 90 (faixa), o algoritmo também os trata como valores distintos, gerando possíveis regras de associações diferentes para cada um destes valores.

A interface de recomendação é o elemento final do *software*, em que o usuário poderá solicitar a recomendação. Conforme mostrado na Figura 6, o padrão encontrado e na sequencia a recomendação, é composta de uma lista de alimentos e atividades físicas, que objetivam alcançar um melhor resultado de glicemia.

```
if($musculacao_R <= 30 ){
  if($seleptico <= 0 ){
    if($corrida > 0 ){ return "ideal (5.0/2.0)";}
    if($corrida <= 0 ){
      if($Carb <= 256 ){
        if($natacao > 0 ){ return "baixo (3.0/2.0)";}
        if($natacao <= 0 ){
          if($bike <= 0 ){ return "ideal (123.0/34.0)";}
          if($bike > 0 ){
            if($padel <= 0 ){ return "ideal (2.0)";}
            if($padel > 0 ){ return "alto (3.0/1.0)";}
          }
        }
      }
    }
    if($Carb > 256 ){
      if($tenis > 0 ){ return "ideal (25.0/1.0)";}
      if($tenis <= 0 ){
        if($volei_de_areia > 0 ){ return "ideal (2.0/1.0)";}
        if($volei_de_areia <= 0 ){
          if($atividade <= 80 ){ return "ideal (63.0/10.0)";}
        }
      }
    }
  }
}
```

Figura 5 – algoritmo gerado

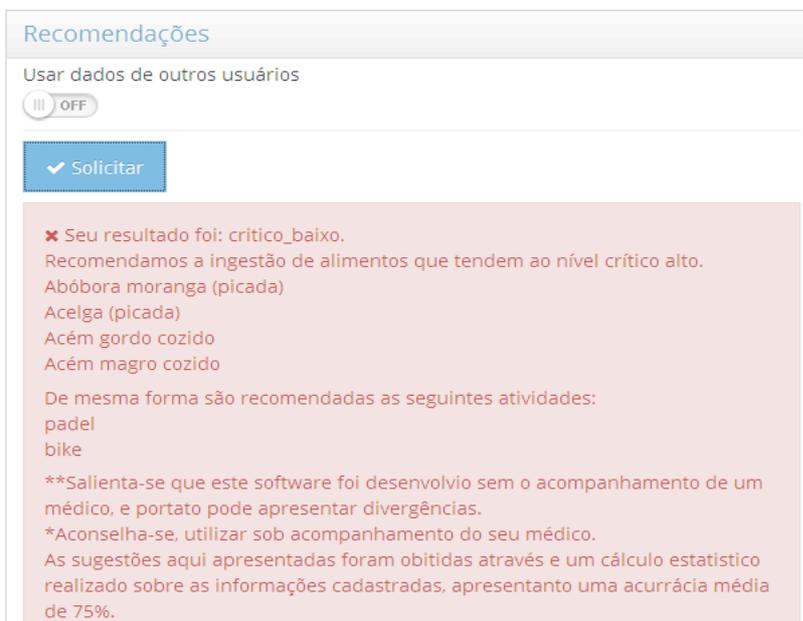


Figura 6 - Interface de Recomendações

### 6.6. Obtenção das Recomendações

Para as recomendações, foi necessário realizar a comunicação entre processos, haja visto que o *software* desenvolvido utilizou uma linguagem de programação diferente da linguagem em que foi desenvolvido o algoritmo de mineração de dados (C4.5). Após a

execução do algoritmo, a resposta é retornada em forma de texto para o sistema. Este então, processa a resposta obtida para localizar os padrões encontrados pelo algoritmo de mineração, como descrito anteriormente. Em posse dos padrões, o sistema volta ao banco de dados e busca os registros que correspondem a esse padrão. Gerando assim uma lista dos alimentos, atividades físicas que resultaram no índice de glicemia ideal.

## 7. Resultados

Foi possível projetar e implementar um sistema de recomendação, com processos de cadastros de novos usuários, valores de glicemia, alimentos ingeridos, atividades físicas praticadas e medicamentos. Nele, é possível executar a gestão dessas informações e do próprio processo de recomendação.

Um dos objetivos planejados inicialmente, era obter dentro da relação de alimentos, quais destes seria melhor indicado para o consumo. No entanto, percebeu-se que sugerir alimentos apenas levando em consideração seus valores nutricionais poderia causar a sugestão de itens que o usuário não gosta ou não pode ingerir por alguma restrição alimentar. Para isso, propõe-se a inclusão de um item no sistema em que o usuário possa determinar quais alimentos não pode ou não quer consumir. Se o usuário consumir sempre os itens indicados, o sistema começará a possuir somente os mesmos dados, e portanto o algoritmo não irá gerar novos resultados, ocasionando a repetição de uma única sugestão. E como é necessário variar o cardápio para obter uma gama de nutrientes necessários para manter a saúde e não somente os valores de glicemia. Para essa situação, foi incorporada uma nota juntamente à sugestão, alertando que ele deve possuir uma dieta variada e balanceada.

A busca de padrões realizada pelo algoritmo de mineração de dados apenas encontra o padrão estatístico que faz gerar o resultado presente na coluna alvo, não distinguindo na coluna alvo (coluna de valores para os quais o algoritmo deve obter o padrão) se o valor de glicemia é baixo, ideal ou crítico alto. Para um usuário recém cadastrado no *software*, não será possível realizar a recomendação, visto que ele ainda não apresenta um número mínimo de registros necessários para o algoritmo de mineração conseguir extrair um padrão. Para suprir esta falta de dados iniciais, planejou-se a opção de uso de dados de outros usuários. Entretanto, como muitos fatores colaboram para os índices de glicemia, e que cada pessoa reage de forma diferenciada, a recomendação obtida com estes dados pode não ser correta. Para tal, é adicionada à recomendação uma mensagem de alerta informando esse fato.

## 8. Conclusões

Este trabalho teve como objetivo o projeto e desenvolvimento de um *software* que fosse capaz de realizar recomendações de alimentação e atividades físicas a portadores de diabetes. O *software* possui formulários de coleta de dados, para que o algoritmo de mineração de dados execute suas operações e encontre os padrões que resultam nos valores de glicemia. Com a obtenção desse padrão, o *software* realiza uma busca nos registros da base de dados, gerando uma relação dos alimentos e atividades físicas que, atendendo aos padrões, resultam em um valor ideal de glicemia.

No decorrer da construção do *software*, foram constatadas algumas dificuldades, tais como: integração do algoritmo de mineração de dados, quantidade de registros necessários para obter os primeiros resultados de recomendação, recomendar alimentos

dentro das preferências do paladar do usuário. Para trabalhos futuros, sugere-se identificar critérios e realizar avaliações de: desempenho da recomendação; precisão da recomendação; tempo mínimo de cadastro de atividades e/ou refeições para que o sistema gere recomendações.

## Referências

- ADA (2013), “American Diabetes Association”, <http://www.diabetes.org>, Agosto.
- Aranha, Christian Nunes. (2007), “Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional”, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ.
- Carvalho, Luís Alfredo Vidal de. (2005), “Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração”, Rio de Janeiro, RJ, Ciência Moderna.
- Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P. (1996), “From data mining to knowledge Discovery: an overview”, Em *Advances in Knowledge Discovery and Data mining*.
- Freitas, Vania Cristina Brodin. (2012), “Auxílio da Mineração de Dados para Diagnosticar Respiradores Bucais e Nasais”, Centro Universitário Franciscano, Santa Maria, RS.
- Han, J; Kamber, M. (2006), “Data Mining: Concepts and Techniques”, San Francisco, CA: Elsevier Inc.
- PS (2013), “Portal da Saúde”, [http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=29793](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=29793), Agosto.
- Rezende, Solange Oliveira. (2003) *Sistemas inteligentes: fundamentos e aplicações*, Barueri-SP, Manole.
- Prieto, Rodrigo Gomes; Linhares, Kathya Collazos; Pinto, Luiz Gustavo; Ortiz, Jorge Roel Ortiz. (2004), “Programa de Mineração de Dados para Análise de Diabetes e Hipertensão”, Universidade do Vale do Itajaí (UNIVALI), Brasil.
- SBD (2013), “Sociedade Brasileira de Diabetes”, <http://www.diabetes.org.br>, Setembro.
- Tan, Pang-Ning; Steinbach, Michael; Kumar, Vipin. (2009), “Introdução ao datamining: mineração de dados”, Rio de Janeiro, RJ, Ciência Moderna.
- Tsunoda, Denise Fukumi. (2004) “Abordagens Evolucionárias para a Descoberta de Padrões e Classificação de Proteínas”, Centro Federal de Educação Tecnológica do Paraná, Curitiba, PR.
- Vieira, Sylvio André. (2011), “Aplicação de Maquinas de Vetores de Suporte na Investigação da Atividade Gênica do Câncer do Colo de Intestino”, Centro Universitário Franciscano, Santa Maria, RS.
- Waikato, U. O. (2013), “WEKA” <http://www.cs.waikato.ac.nz/ml/weka/>, Outubro.