

DOI: 10.5748/9788599693124-13CONTECSI/PS-4026

UM ESTUDO USANDO DATA MINING COM O ALGORITMO DENSITY BASED SCAN ALGORITHM WITH NOISE (DBSCAN) NA CLUSTERIZAÇÃO DE PERFIS DE EMPREGADOS ABSENTEÍSTAS E PRESENTEÍSTAS

Ricardo Pinto Ferreira (Universidade Nove de Julho, São Paulo, Brasil) - sedexja@ig.com.br

Andréa Martiniano (Universidade Nove de Julho, São Paulo, Brasil) - sedexja@ig.com.br

Renato José Sassi (Universidade Nove de Julho, São Paulo, Brasil) - sedexja@ig.com.br

The Advancement of computer science allows processing of large amounts of data, which motivates the study. The term data mining has arisen due to the similarities between the search for important information in a database and the act of mine the mountain to find a vein of gold. Data mining is the element responsible for the efficient extraction of implicit and useful knowledge contained in a database. The purpose of this article is to apply data mining techniques to identify by clustering the profile of absenteeist and presenteeist employees using DBSCAN algorithm. The Absenteeism is not reporting to work, as scheduled. In the Presenteeism for the employees is presence at work, although ill, however, its activities are unproductive. The DBSCAN algorithm applied in data mining to cluster. The results presented by DBSCAN algorithm showed that the application in data mining is satisfactory, confirming the use of this technique as an option to apply this kind of problem. The methodology used in the structuring of this article was defined as literature, exploratory and experimental.

Keywords: Absenteeism, Presenteeism, Data mining, DBSCAN.

UM ESTUDO USANDO DATA MINING COM O ALGORITMO DENSITY BASED SCAN ALGORITHM WITH NOISE (DBSCAN) NA CLUSTERIZAÇÃO DE PERFIS DE EMPREGADOS ABSENTEÍSTAS E PRESENTEÍSTAS

O Avanço da ciência computacional permite o processamento de grande quantidade de dados, o que motiva o estudo em questão. O termo data mining surgiu devido às semelhanças entre a procura de informação importante numa base de dados e o ato de minerar a montanha para encontrar um veio de ouro. Data mining é o elemento responsável pela extração eficiente do conhecimento implícito e útil contido em um banco de dados. O objetivo deste artigo é aplicar técnica de data mining para identificar através da clusterização o perfil de empregados absenteeistas e presenteeistas utilizando o algoritmo DBSCAN. O Absenteísmo é o não comparecimento ao trabalho, conforme o programado. No Presenteísmo há a presença do empregado no trabalho, ainda que doente, contudo, suas atividades são improdutivas. O algoritmo DBSCAN é aplicado em data mining para clusterizar. Os resultados apresentados pelo algoritmo DBSCAN mostraram que a aplicação no data mining é satisfatória, o que confirma a utilização desta técnica como uma opção a ser aplicada neste tipo de problema. A metodologia adotada na estruturação deste artigo foi definida como bibliográfica, exploratória e experimental.

Palavras-chave: Absenteísmo, Presenteísmo, Data mining, DBSCAN.

1. Introdução

As nações e as empresas vivenciam constantes avanços tecnológico, sociocultural característicos do processo de globalização, as consequências desse desenvolvimento são benéficas ao mundo moderno. Em contrapartida, esses avanços aumentam a competitividade entre nações e empresas, o que pode provocar mudanças nos comportamentos biológicos, psicológicos e sociais das pessoas provocando sobrecarga física e mental. Essas imposições e tensões desencadeiam reações adversas, interferindo diretamente na qualidade de vida e do trabalho das pessoas (JODAS & HADADD, 2009).

Nestas circunstâncias, o trabalho tornou-se mais complexo, superando o sentido tradicional, que o identifica como meio de satisfação de necessidades (LUMARE JÚNIOR, 2007). A pressão para atingir metas cada vez mais audaciosas impõe aos empregados excessos de atribuições, o que pode causar algum tipo de distúrbio em seu estado de saúde levando a ausência e ou ao baixo rendimento (BERNSTROM, 2013).

Para que as empresas consigam identificar o real motivo das ausências dos seus empregados e do baixo rendimento mesmo presente na empresa, faz-se necessário um estudo de suas causas e consequências, que podem estar relacionadas com o absenteísmo e com o presenteísmo (BRATSBERG, FEVANG & ROED, 2013).

O absenteísmo em geral é definido como o não comparecimento do empregado no trabalho, conforme o programado (JOHNS, 2010). Isso geralmente significa que o trabalho é feito de forma menos eficiente por outro empregado ou não é feito. Já o presenteísmo é definido como o fato dos empregados estarem no local de trabalho, mas, devido a problemas de ordem física ou psicológica, não cumprirem a totalidade as suas funções (HEMP, 2004).

O absenteísmo e o presenteísmo levam a perdas de produtividade (HALBESLEBEN, WHITMAN & CRAWFORD, 2014). O absenteísmo leva a perda da capacidade produtiva e o presenteísmo oculta a baixa produtividade, dificultando a gestão do trabalho e criando grupos desmotivados.

O absenteísmo e o presenteísmo apresentam uma grande variedade de comportamentos, com diversas causas. A importância do controle dos índices de absenteísmo e presenteísmo refletem em prejuízos econômicos significativos para a empresa comprometendo a sua estabilidade no mercado de trabalho competitivo (MCKEVITT TAUFFER, & COLTRE, 2007).

Em geral a detecção e a prevenção passam pela análise das informações geradas pelos empregados que devem ser registradas de forma completa e abrangente como, por exemplo: registros do Código Internacional de Doenças, exames médicos, fisioterapia, faltas injustificadas, dentre outras (BUSTILLOS, VARGAS III & GOMERO-CUADRA, 2015).

As empresas se utilizam de diversas técnicas para tentar detectar e prevenir o absenteísmo e o presenteísmo entre os seus empregados (THORWARTH, ARISHA, & HARPER, 2009).

Técnicas de *data mining* podem ser aplicadas nesta análise em busca da identificação do perfil absenteísta ou presenteísta.

Os dados gerados pelos empregados nas empresas devem ser registrados de forma completa e abrangente (registros do Código Internacional de Doenças, exames médicos,

fisioterapia, faltas injustificadas, etc.). O tratamento desses dados deve ter o devido cuidado para que as peculiaridades de cada caso sejam identificadas. É difícil a detecção do presenteísmo, mas através dos dados de absenteísmo juntamente com os dados comportamentais dos empregados podem levar a identificação deste comportamento (FONSECA, 2009). A análise das informações dos empregados armazenadas no banco de dados da empresa é fundamental para detectar o perfil do empregado absenteísta ou presenteísta.

Assim, o uso de sistemas que possam extrair o conhecimento em bases de dados denominados de *Knowledge Discovery in Databases* (KDD) ou Descoberta de Conhecimento em Bases de Dados é interessante, podem ser definidos como o processo de extração de conhecimentos válidos, novos, potencialmente úteis e compreensíveis para apoiar a tomada de decisão (FAYYAD, PIATETSKY-SHAPIRO & SMITH, 1996). O KDD é um processo iterativo e iterativo formado por diversas fases, sendo a principal o *data mining* ou a mineração de dados. O *data mining* é o elemento responsável pela extração do conhecimento implícito e útil contido em um banco de dados.

Neste sentido, adotou-se aplicar técnica de *data mining*, para identificar através da tarefa de clusterização o perfil de empregados absenteístas e/ou presenteístas.

O artigo está organizado após essa seção introdutória da seguinte forma: na seção 2 é apresentado o Referencial Teórico: Os conceitos básicos do absenteísmo, presenteísmo, *Knowledge Discovery in Databases* (KDD), *data mining* e o *algoritmo Density Based Spatial Clustering of Applications With Noise* (DBSCAN). Na seção 3 é apresentado o método de trabalho. Descrevem-se na seção 4 os resultados dos experimentos computacionais. Na seção 5 o artigo é encerrado com as Considerações finais.

2. REFERENCIAL TEÓRICO

2.1 Absenteísmo

Podem ser diversas as causas desencadeantes do absenteísmo: problemas de saúde ou doenças, acidentes do trabalho, problemas de saúde em pessoas da família, gestação e parto, licença casamento (licença gala), licença maternidade, férias, dentre outros. Podem ocorrer ainda faltas que não são justificadas legalmente, conhecidas como faltas injustificadas (PENATTI, ZAGO & QUELHAS, 2006; GEHRING JUNIOR, CORRÊA FILHO, VIEIRA NETO, FERREIRA, & VIEIRA, 2007).

O absenteísmo pode ser classificado em cinco categorias, das quais se podem investigar e obter informações necessárias de modo que as empresas consigam estruturar e administrar melhor sua produção (INOUE, MATSUDA, SILVA, UCHIMURA & MATHIAS, 2008):

- a) Absenteísmo-doença: ausência justificada por licença-saúde;
- b) Absenteísmo por patologia profissional: causado por acidente de trabalho e/ou doença profissional;
- c) Absenteísmo legal: respaldado por lei;
- d) Absenteísmo-compulsório: por suspensão imposta pelo patrão, por prisão ou por outro impedimento de comparecer ao trabalho e
- e) Absenteísmo voluntário: por razões particulares não-justificadas.

O gerenciamento do absenteísmo tem exigido muito das empresas e gerado muitas preocupações aos seus administradores, as quais estão ligadas a diversos fatores, que vão desde questões sociais, de saúde, de gestão de pessoas, além de outros problemas como

excesso de trabalho, tornando, assim, este tema complexo e difícil de ser gerenciado (PENATTI, ZAGO & QUELHAS, 2006).

Outro problema encontrado é que dificulta ainda mais o gerenciamento do absenteísmo é a pesquisa no setor público, saúde do trabalhador, perfil de saúde, absenteísmo e licença médica apontou deficientes estudos sobre este tema. Os estudos restringiram-se a alguns grupos de profissionais, tais como servidores de hospitais e de universidades. (CUNHA, BLANK & BOING, 2009).

No Brasil, os estudos sobre o perfil de doenças dos empregados não são suficientes para se conhecer adequadamente as características dessa população em relação ao absenteísmo (PENATTI, ZAGO & QUELHAS, 2006; CUNHA, BLANK & BOING, 2009).

2.2 Presenteísmo

Em paralelo ao absenteísmo, foi identificado, nas últimas décadas, um novo conceito: o presenteísmo.

De acordo com Umann (2011), nesta situação há presença do empregado, ainda que doente, no seu local de trabalho, de maneira que a realização de suas atividades e suas funções pode ocorrer de um modo não produtivo.

O presenteísmo começou a ser estudado na década de 50, na França, e tornou-se mais conhecido durante os últimos anos do século XX, quando os pesquisadores Canfield & Soash (1955) & Uris (1955) começaram a buscar as causas do absenteísmo e doenças da natureza do trabalho.

No Brasil o fenômeno era pouco conhecido e os departamentos de Recursos Humanos não o identificava como uma ameaça a produtividade e a lucratividade da empresa (GOETZEL, LONG, OZMINKOWSKI, HAWKINS, WANG & LYNCH, 2004). Porém, nas últimas décadas, pesquisas apontam para a alta incidência do presenteísmo por doenças crônicas ou episódicas (ARAÚJO, 2012).

Enquanto o absenteísmo tem sido muito pesquisado, o mesmo não ocorre com o Presenteísmo.

Também concordavam Martinez, Ferreira, Souza & Cunha (2007) que o presenteísmo ainda não era um tema dominante na investigação em gestão, apesar de abordarem os principais motivos das flutuações de produtividade individual no trabalho como:

- a) alterações no correto funcionamento fisiológico: dor de cabeça, dores crônicas e problemas respiratórios;
- b) psicossomático: ansiedade, depressão, déficit de atenção dos empregados;
- c) acesso à *internet* (redes sociais).

Esse padrão de comportamento se mostrou mais acentuado a partir dos anos 90, quando as altas taxas de desemprego levavam a um sentimento de insegurança. Os sintomas mais comuns do presenteísmo são: asma, dores de cabeça, dores nas costas, irritação, alergias, hipertensão arterial, desordens gastrintestinais, artrites e depressão (PASCHOALINO, 2008).

O presenteísmo possui duas vertentes principais, uma o empregado produz pouco e consegue simular seu rendimento rateando sua tarefa com outros empregados e, a outra vertente o empregado procura produzir acima da sua capacidade, por medo de ser

demitido, mesmo que para isso seu descanso não seja respeitado, com o passar do tempo esse empregado tende a ter afastamentos maiores devido às doenças crônicas causando prejuízos maiores para a empresa, tanto na produtividade quanto nas despesas médicas e assistenciais, ou seja, segundo Umann (2011), o presenteísmo tem uma relação entre a enfermidade e a perda de produtividade em decorrência do excesso de trabalho e a insegurança no trabalho. Isso é resultado das novas relações de trabalho, caracterizadas pelas altas taxas de desemprego, reestruturação nos setores públicos e privados, diminuição no tamanho da empresa, redução do número de empregados, aumento do número de pessoas com contratos temporários, redução dos benefícios e o avanço da tecnologia e da robótica.

Para Martins, Silva & Rocha (2009), a perda na produtividade esteve associada às faltas dos empregados ao trabalho. No entanto, a queda na produtividade pode ocorrer inclusive quando se comparece à empresa com algum problema de saúde e, doenças crônicas, tanto produzem impacto significativo, como aumentam os custos da assistência médica.

Problemas de saúde, financeiros, familiares e estresse são causas frequentes do presenteísmo. Boa parte das empresas desconhece ou ignora o problema, porque, em geral, o empregado não se ausenta.

O presenteísmo afeta a produtividade por diminuir a atenção no trabalho, levando a erros e riscos de acidentes. Em alguns casos, o presenteísmo pode ser confundido com excessiva dedicação e segundo Hemp (2009), o presenteísmo não é resultado de simulação. Muitos dos problemas de saúde que levam ao presenteísmo não são limitantes, mas se esses problemas de saúde não forem tratados a tempo podem ser cronificados e se tornarem incapacitantes. São duas as formas de presenteísmo: uma envolve a saúde ocupacional, relações e condições organizacionais do trabalho, o empregado continua trabalhando mesmo doente. Na outra, o empregado permanece no trabalho além de seus limites com medo de ser demitido ou excluído por não estar disponível quando necessário (PASCHOALINO, 2008; LARANJEIRA, 2009 & UMANN, 2011).

2.3 Absenteísmo e Presenteísmo

O absenteísmo e o presenteísmo apresentam uma grande variedade de comportamentos, com diversas causas. Segundo Johns (2003), o problema se reflete na multidisciplinaridade de conhecimentos disponíveis sobre o assunto, uma vez que esses conhecimentos tradicionalmente não se complementam isoladamente, sendo desenvolvidos à parte em cada uma das disciplinas, que vão da Psicologia à Engenharia, passando por diversas áreas.

Neste aspecto, segundo Buckerman & Laukkanen (2009), o absenteísmo e o presenteísmo são semelhantes no que diz respeito à produtividade. No caso do absenteísmo, os empregados ausentam-se do trabalho por motivo de doença e, no caso do presenteísmo, estão presentes no trabalho apesar de sua doença.

O absenteísmo e o presenteísmo levam a perdas de produtividade para a empresa. No caso, o absenteísmo leva a perda da capacidade produtiva e o presenteísmo oculta a baixa produtividade, dificultando a gestão do trabalho e criando grupos desmotivados (MARTINIANO, 2012).

As empresas que direcionam medidas excessivas à redução do absenteísmo, segundo Bergstrom, Bodin, Hagberg, Lindh, Aronsson & Josephson (2009) ocasionam um

aumento de presenteísmo, que a longo prazo podem aumentar o afastamento por doenças (absenteísmo prolongado).

Segundo Rocha (2010), o empregado doente (presenteísta) tem maior chance de se ausentar do trabalho (absenteísta) em longo prazo, além de apresentar mudanças na qualidade e expectativa de vida.

Por isso, preocupações com a Qualidade de Vida no Trabalho (QVT) é algo que está ligada à produtividade da empresa. Empregados satisfeitos e motivados são mais comprometidos com o trabalho. Afinal, com condições ótimas de saúde aliadas a um ambiente saudável, chega-se à qualidade da produção pretendida (MIRANDA, GOMES & GOMES, 2006). Daí a importância de se reduzir e prevenir o absenteísmo e o presenteísmo.

A importância do controle dos índices de absenteísmo, em que índices baixos podem representar o presenteísmo elevado, assim como a falta ao trabalho, dificulta a vida dos demais empregados, reflete em prejuízos econômicos significativos para empresa comprometendo a estabilidade da empresa no mercado de trabalho competitivo (MCKEVITT TAUFFER, & COLTRE, 2007).

Através desse controle fica claro de que o custo do absenteísmo é percebido através da produtividade do empregado que é perdida a cada dia, pois o empregado está ausente, ao passo que o presenteísmo esconde de maneira sutil a improdutividade e as doenças que estão ocultas, pelo fato do empregado estar no ambiente de trabalho, mas não consegue executar suas funções ao máximo fazendo com que o custo do presenteísmo não seja percebido (LEVIN-EPSTEIN, 2005).

Percebe-se também o lado mais negativo do presenteísmo, em que o empregado, por medo de ausentar-se do trabalho por qualquer que seja o motivo até mesmo quando a saúde apresenta fortes sinais de fragilidade, adia cirurgias, tratamentos médicos ou odontológicos, passando do presenteísmo que perdura muitas vezes por anos e anos para o absenteísmo. Com esse agravante a enfermidade muitas vezes encontra-se em estágios mais avançados e de difícil recuperação, quando há recuperação. Em outros casos a recuperação é muito mais longa, exigindo maior tempo de afastamento (MARTINIANO, 2012).

2.4 Knowledge Discovery in Databases (KDD) ou Descoberta de Conhecimento em base de Dados

O KDD foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de bases de dados. De acordo com Fayyad, Piatetsky-Shapiro & Smyth, (1996), o KDD é um processo de várias etapas não trivial, interativo (o usuário pode intervir e controlar o curso das atividades) e iterativo (sequência finita de operações onde o resultado de cada etapa é dependente dos resultados das etapas que as precedem) para identificação de padrões compreensíveis, válidos, e potencialmente úteis a partir de grandes conjuntos de dados (GOLDSCHMIDT, PASSOS & BEZERRA, 2015).

Para Freitas (2003), o processo de KDD tem por objetivo descobrir novos padrões em bases de dados de maneira automática. Esse processo de descoberta é definido como a extração de conhecimento de alto nível partindo de bases de dados reais e é composta de diversas fases, dentre as quais são utilizados métodos baseados na estatística e no aprendizado de máquina.

Goldschmidt & Passos (2005) explicaram que o termo iterativo indica a possibilidade de repetições integrais ou parciais do processo de KDD e a expressão não trivial alerta para a complexidade normalmente presente na interpretação do processo de KDD. Com relação à expressão padrão válida, ela indica que o conhecimento deve ser verdadeiro e adequado ao contexto da aplicação de KDD e o termo padrão novo deve acrescentar novos conhecimentos aos existentes para que todo esse processo gere conhecimento útil e propicie benefícios ao contexto de aplicação do KDD.

A esse respeito, Barros & Campos (2006) ressaltam que a utilização do processo de KDD vem se tornando mais comum e tendo desenvolvimentos relevantes em relação aos processos de tomada de decisão em níveis operacional, gerencial e estratégico com mais segurança. Esta segurança é proporcionada por dados estatísticos baseados em experiências anteriores. O KDD é o responsável por descobrir estes dados estatísticos dentro de bancos de dados institucionais.

Uma das etapas do KDD, o *data mining*, é considerada a etapa mais importante do processo pela existência do algoritmo minerador, que diante da tarefa especificada será capaz de extrair de modo eficiente o conhecimento implícito e útil de um banco de dados. O *data mining*, por sua vez, pode ser considerada o núcleo do processo de KDD, consistindo na aplicação de algoritmos de extração de padrões de comportamento e tendências a partir de dados (BARROS & CAMPOS, 2006).

2.4.1 Fases do KDD

As fases do KDD contêm uma série de passos que auxiliam nas mais diversas decisões a serem tomadas. Cada fase possui uma intersecção com as demais, melhorando assim a cada resultado (FAYYAD, PIATETSKY-SHAPIRO & SMYTH, 1996). O KDD é composto por cinco fases, a saber:

- a) Seleção dos dados;
- b) Pré-processamento dos dados;
- c) Transformação dos dados;
- d) *Data mining* e
- e) Interpretação/avaliação do conhecimento.

A Figura 1 ilustra as fases do processo KDD, representando suas etapas.

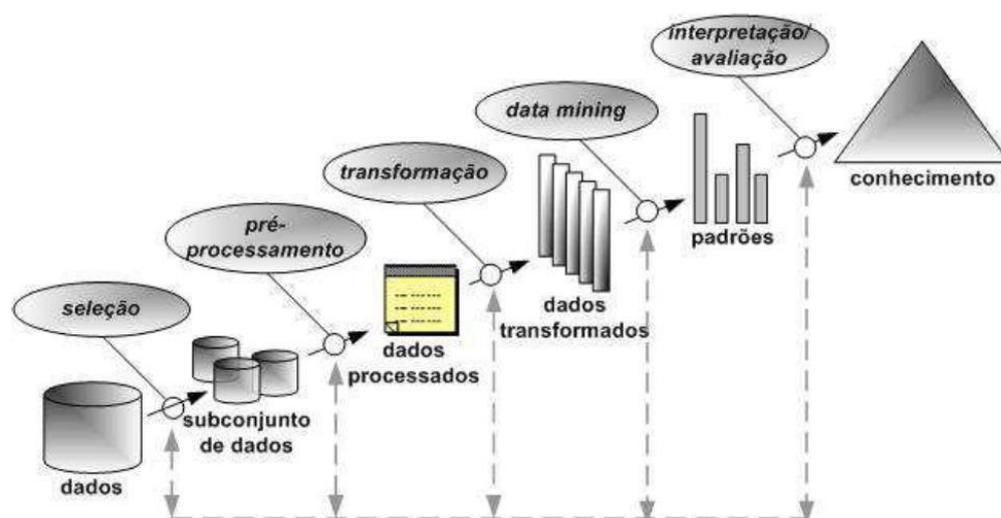


Figura 1 – Fases do processo de KDD. Fonte: Adaptado de Fayyad, Piatetsky-Shapiro & Smyth, (1996) & Sassi (2006).

O processo de KDD compreende todo o ciclo que o dado percorre até gerar conhecimento. A iteração entre as fases pode ser observada pelas setas tracejadas da (Figura 1).

Antes mesmo de se aplicar as técnicas escolhidas, é preciso definir qual será seu foco e possuir algum conhecimento dentro da área onde os dados foram captados para que seja possível fazer a interpretação.

De acordo com Barros & Campos (2006), as fases do KDD, que antecedem o *data mining*, podem levar até 80% do tempo necessário para todo o processo de análise devido às dificuldades de integração de bases de dados com estruturas variadas.

2.4.2 Tarefas de KDD

Existem várias formas de interpretação dos dados pelo KDD denominadas tarefas. Diversas tarefas podem ser realizadas, as tarefas mais comuns segundo Fayyad, Piatetsky-Shapiro & Smyth, (1996) são associação, classificação, clusterização (agrupamento) e visualização.

a) Associação

De acordo com Goldschmidt & Passos (2005), as regras de associação ou (regras associativas) fundamentam-se na ação de encontrar um grupo de itens afins, ou seja, que ocorrem simultaneamente e de forma repetida em uma base de dados. Estas regras descobertas devem considerar importantes não apenas associações triviais, mas também aquelas não óbvias onde jamais se imaginaria haver uma relação, das quais se transformam em importante fonte de informação na tomada de decisão. Uma regra de associação caracteriza o quanto a presença de um conjunto de itens nos registros de uma base de dados implica na presença de algum outro conjunto distinto de itens nos mesmos registros (COSTA & ANDRADE NETTO, 2007).

O objetivo para a tarefa de descoberta de regras de associação refere-se à possibilidade de encontrar tendências a partir de regras de associação extraídas de grandes bases de dados, podendo ser incrementada estrategicamente as vendas de um determinado segmento comercial, como promoções entre produtos, rearranjo da disposição dos produtos em prateleiras e gôndolas, entre outras (GOLDSCHMIDT & PASSOS, 2005).

b) Classificação

O processo da classificação inicia-se com a divisão do conjunto de dados em dois conjuntos totalmente distintos: o conjunto de treinamento e o conjunto de testes. De posse do conjunto de treinamento, o classificador aprende regras acessando todos os registros, sua classe e seus atributos, pertencentes a esse conjunto. O resultado desse processo é um conjunto de regras de classificação. Em seguida essas regras são avaliadas e é nesse ponto que o conjunto de testes será utilizado. O objetivo será avaliar a qualidade das regras descobertas medindo o seu poder preditivo no conjunto de testes (MITCHELL, 1997).

A tarefa de classificação, segundo Azevedo & Santos (2005), corresponde em encontrar uma função que associe um caso a uma classe dentro de diversas classes discretas de classificação. Esta tarefa pode ser definida como uma função de aprendizado, que mapeia itens de dados em uma das diversas classes pré-definidas.

Dessa forma, a classificação consiste em obter um modelo baseado em um conjunto de exemplos que descrevem uma função não-conhecida. Neste tipo de tarefa, o objetivo é descobrir um relacionamento entre um atributo meta e um conjunto de atributos previsores, de modo que o processo de classificação possa usar esse relacionamento para prever a classe de um exemplo novo e desconhecido (AZEVEDO & SANTOS, 2005).

A classificação pode ser utilizada com êxito, por exemplo, num conjunto de dados de um banco, em que se deseja descobrir qual o perfil dos clientes que aplicam seu dinheiro em fundos de investimento de longo prazo. Com classificadores pode-se prever que clientes do sexo masculino, com renda superior a R\$ 5000,00 e com idade acima de 30 anos aplicam em fundos de investimento de longo prazo. Neste caso, o atributo *aplica em fundos de investimento de longo prazo* é denominado classe, pois é o atributo alvo da classificação cujos possíveis valores, neste caso, são sim ou não (COSTA & ANDRADE NETTO, 2007).

c) Clusterização (Agrupamento)

A palavra clusterização é um neologismo do termo *clustering*, que difere da classificação, pois a primeira, visa criar os agrupamentos através da organização dos elementos, enquanto a segunda, procura alocar elementos em classes já pré-definidas. A clusterização transforma registros com grande número de atributos em conjuntos relativamente menores (segmentos) (GUELPELI, 2009).

Segundo Fayyad, Piatetsky-Shapiro & Smyth, (1996), o agrupamento ou clusterização é uma tarefa onde se procura determinar um conjunto finito de conjuntos ou agrupamentos para descrever os dados. Isso consiste na identificação de grupos semelhantes de objetos onde cada grupo é uma classe. Dentro da mesma classe os objetos são semelhantes e entre as classes são divergentes. Essa tarefa é usada para particionar os registros de uma base de dados, de forma que elementos em um *cluster* compartilhem um conjunto de propriedades comuns que os distinguem dos elementos de outros *clusters* com o objetivo de maximizar similaridade intra-*cluster* e minimizar similaridade inter-*cluster*. (FAYYAD, PIATETSKY-SHAPIRO & SMYTH, 1996; GOLDSCHMIDT & PASSOS, 2005; GOLDSCHMIDT, PASSOS & BEZERRA, 2015).

d) Visualização

A tarefa de visualização é um processo para transformar informação em uma forma visual, permitindo aos usuários observar a informação, que pode ser usada para dar suporte ao processo de decisão quando as técnicas de *data mining* requerem grande interação com o usuário de forma complexa (SILVA NETO, 2008).

A visualização, processo para transformar informação em uma forma visual, tem sido usada para analisar e mostrar grandes volumes de dados multidimensionais. A Visualização permite representar diferentemente dos métodos estatísticos os resultados sem necessariamente saber que tipo de fenômeno deve ser analisado (SILVA NETO, 2008).

No caso de técnicas de clusterização, podem ser usadas ferramentas de visualização para determinar qual, ou quais *clusters*, criados são úteis ou interessantes para os métodos de *data mining*.

As ferramentas de visualização podem ser usadas para auxiliar, ou serem auxiliadas, pelas técnicas de *data mining*. Por meio dessas ferramentas podem ser encontradas características, ou fenômenos, pouco comuns ou interessantes sem que se esteja diretamente procurando por eles (SILVA NETO, 2008).

2.5 Data Mining

Diversas técnicas de *data mining* vêm sendo utilizadas com sucesso no mundo inteiro e estão presentes em empresas nacionais e internacionais, como: SERPRO, Caixa Econômica, Dataprev, Embratel, Telefônica, Sendas, Wal-Mart, FedEx, UPS, entre muitas outras. Detecção de fraudes em arrecadações, tendências de consumo de clientes, previsão de produção na indústria, previsão de riscos no mercado financeiro e previsão de demandas são alguns exemplos dentre as inúmeras aplicações de *data mining* na atualidade (GOLDSCHMIDT, PASSOS & BEZERRA, 2015).

Todas as etapas do processo de KDD possuem grau elevado de importância para o sucesso do mesmo. Entretanto, é a etapa do *data mining* que recebe o maior destaque na literatura (BERRY & LINOFF, 1997).

Segundo Carvalho (2005), as ferramentas da nova ciência dita *data mining* são antigas conhecidas dos pesquisadores da ciência da computação e, já, há muito utilizadas no meio acadêmico e empresarial gerador de tecnologia. A popularização destas ferramentas e a sua nova designação pelo nome *data mining* deve-se a vários fatores entre os quais se pode citar a crescente utilização dos computadores, o aumento da competitividade empresarial e a disponibilidade de grandes massas de dados armazenadas em forma digital.

Conforme Berry & Linoff (1997), *data mining* é a exploração e análise, de forma automática ou semiautomática, de grandes bases de dados com objetivo de descobrir padrões, regras e fornecer informações que a permitam montar estratégias corporativas eficientes, melhorando os negócios e reduzindo custos.

A Figura 2 ilustra o processo de *data mining* iniciando com os dados brutos passando pela preparação de dados, algoritmo de mineração, terminando com a análise dos dados e a descoberta de padrões úteis, valiosos e utilizáveis.

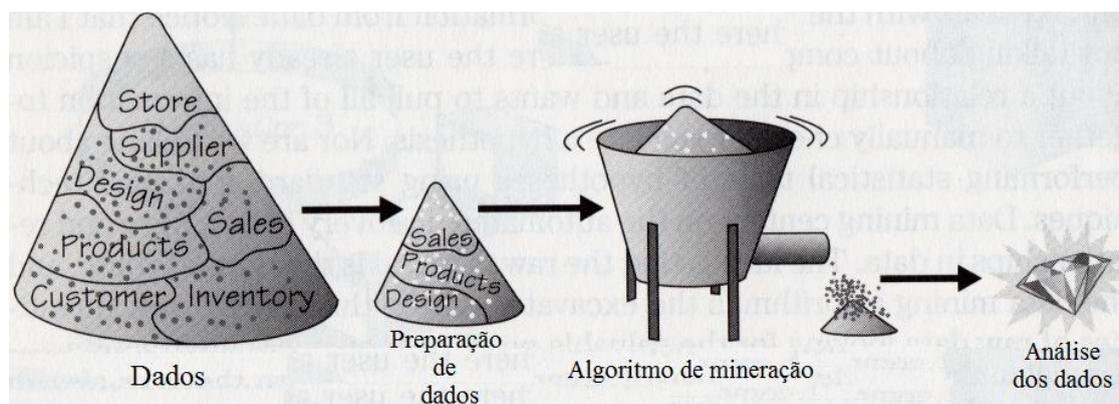


FIGURA 2 – Processo de *data mining*. Fonte: Adaptado de Bigus (1996).

O *data mining* utiliza ferramentas de análise pré-construídas para gerar automaticamente uma hipótese a respeito de padrões, tendências e anomalias encontradas nos dados e a partir desta hipótese, prever comportamentos futuros. Faz uso de técnicas estatísticas avançadas e inteligência computacional para descobrir fatos em uma grande base de dados (GOLDSCHMIDT & PASSOS, 2005; GOLDSCHMIDT, PASSOS & BEZERRA, 2015).

Segundo Fayyad, Piatetsky-Shapiro & Smyth (1996), as técnicas de *data mining* podem ser aplicadas a diversas tarefas para solucionar um problema de descoberta de conhecimento como: associação, classificação, previsão e clusterização.

A clusterização de dados agrupa uma coleção de objetos em subconjuntos, chamados de *clusters*. Um *cluster* é um subconjunto de todos os possíveis subconjuntos distintos da população (DINIZ & LOUZADA NETO, 2000).

O objeto dentro de um mesmo *cluster* significa que os elementos estão mais próximos entre si do que com qualquer outro elemento alocado em outro *cluster*.

De uma forma mais simples, *data mining* é produzir conhecimento novo escondido em grandes bases de dados.

A Figura 3 mostra um esquema simplificado de *data mining*.

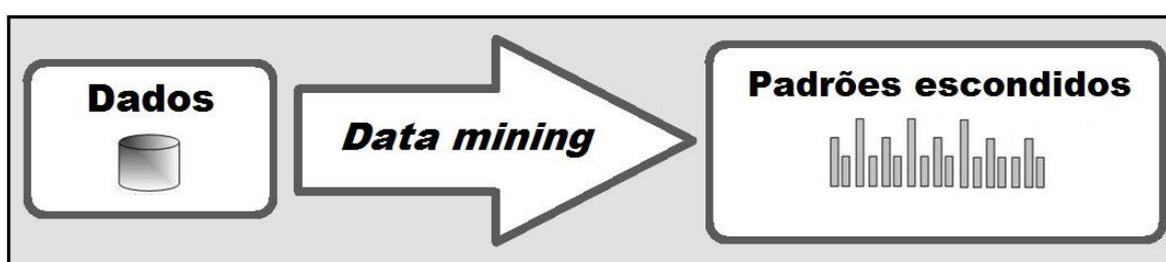


FIGURA 3 – Esquema simplificado de *data mining*. Fonte: Adaptado de Sassi (2006).

O *data mining* serve para prever tendências e comportamentos futuros, permitindo a tomada de decisão baseada em fatos e não em suposições podendo responder a questões de negócio que tradicionalmente demandariam muito tempo para resolver.

2.5.1 Análise de *cluster* com o *data mining*

Análise de *cluster* é um conjunto de métodos muito importante para classificação de itens em grupos comuns chamados *clusters*. Esses métodos são populares em biologia, medicina, genética, ciência social, antropologia, arqueologia, astronomia, reconhecimento de caracteres e até mesmo no desenvolvimento de sistemas de informações gerenciais (TURBAN, SHARDA, ARONSON & KING, 2009).

Com o aumento da popularidade do *data mining*, os métodos têm sido aplicados aos negócios, em especial ao *marketing*. A análise de *cluster* tem sido muito usado para detecção de fraude, tanto de cartão de crédito quanto de comércio eletrônico, e segmentação de mercado. Outras aplicações continuam a ser desenvolvidas à medida que a análise de *cluster* é compreendida e usada e outras técnicas de clusterização são criadas e aplicadas (TURBAN, SHARDA, ARONSON & KING, 2009).

A análise de *cluster* também é uma ferramenta de análise exploratória de dados para a solução de problemas de classificação. O objetivo é ordenar casos em grupos, ou *clusters*, de modo que o grau de associação seja forte entre os membros do mesmo *cluster* e fraco entre membros de *clusters* diferentes. Cada *cluster* descreve a classe que seus membros pertencem (FAYADD, PIATETSKY-SHAPIRO & SMYTH, 1996; TURBAN, SHARDA, ARONSON & KING, 2009; GOLDSCHMIDT, PASSOS & BEZERRA, 2015).

Na medida em que o *data mining* avança, a importância da análise de *cluster* é revelar associações e estruturas em dados que não estavam aparentes anteriormente, mas que são sensíveis e úteis uma vez encontrados.

2.6 O algoritmo *Density Based Spatial Clustering of Applications With Noise* (DBSCAN)

Em 1996 foi apresentado o algoritmo DBSCAN por Martin Ester, Hans-Peter Kriegel, Jörg Sander & Xiaowei Xu da Universidade de Munique, o DBSCAN gera *clusters* com inúmeros formatos baseando-se na vizinhança dos pontos, onde a densidade associada a um ponto é obtida por meio da contagem do número de pontos vizinhos em uma determinada região próxima desse ponto.

O DBSCAN gera *clusters* considerando as propriedades dos dados, pois não requer que seja informado antecipadamente o número de *clusters*, possibilitando a formação de *clusters* com inúmeros formatos.

O algoritmo DBSCAN identifica *outliers*, e permite o uso de diversas funções de distância.

Os dois parâmetros de entrada do algoritmo DBSCAN são:

- a) Raio de ε -vizinhança de um ponto (ε): determina o raio de vizinhança ε para cada ponto da base de dados. Dado o parâmetro ε , o algoritmo DBSCAN verifica a quantidade de pontos contidos no raio ε para cada ponto da base de dados, e se essa quantidade exceder certo número, um *cluster* é gerado;
- b) Número mínimo de pontos (η): parâmetro que especifica o número mínimo de pontos, no dado raio de ε -vizinhança, que um ponto precisa possuir para ser considerado um ponto central e consequentemente, de acordo com as definições de *cluster* baseado em densidade, gerar um *cluster*.

O algoritmo DBSCAN consiste inicialmente em construir uma estrutura denominada matriz de dissimilaridade. Em uma matriz de dissimilaridade é possível representar a distância entre pares de pontos (RIAZIFAR & SAGHAPOUR, 2015).

Essa matriz de dissimilaridade é construída empregando-se uma medida que é responsável pelo cálculo de distância para todos os pares de pontos da base de dados. A medida de distância mais utilizada é a distância euclidiana, porém outras medidas podem ser utilizadas.

A partir da matriz de dissimilaridade, o algoritmo verifica a ε -vizinhança de cada ponto da base de dados com a finalidade de identificar possíveis pontos centrais para iniciar a formação dos *clusters*.

A matriz de dissimilaridade contendo as distâncias entre os pares de pontos da base de dados é verificada e se a distância entre um ponto x_i e um ponto x_j for menor ou igual a o dado parâmetro ε , então o ponto x_j está na ε -vizinhança do ponto x_i .

Verificados quais os pontos que estão na vizinhança de x_i , o algoritmo verifica a cardinalidade desse ponto com relação aos pontos vizinhos, com a finalidade de definir a condição do ponto.

Caso a cardinalidade de x_i com relação ao raio de ε -vizinhança seja igual ou exceda o parâmetro η , então esse ponto é considerado um ponto central e os pontos contidos em sua ε -vizinhança são então diretamente alcançáveis por densidade. Se as duas condições forem satisfeitas, o algoritmo irá criar um *cluster* C_i contendo todos os pontos diretamente alcançáveis por densidade a partir de um determinado ponto x_i (DAI & LIN, 2012).

Quando o DBSCAN gera um *cluster* C_i , todos os pontos nesse *cluster* tem inúmeras vezes a sua ε -vizinhança recuperada permitindo assim que novos pontos possam ser adicionados ao *cluster*. O processo de consulta de vizinhos próximos é repetido até que todos os pontos em C_i sejam verificados (DAI & LIN, 2012).

Finalmente o algoritmo encerra o crescimento de C_i e visita o próximo ponto não visitado da base de dados. O processo é repetido até que todos os registros da base de dados tenham sido visitados e classificados como: pontos centrais, pontos de borda ou *outliers*.

a) pontos centrais: são pontos que estão no interior de uma região densa, onde existem pelo menos η pontos no raio ε desse ponto. A cardinalidade desses pontos em relação ao parâmetro de ε -vizinhança deve ser de no mínimo η pontos;

b) pontos de borda: estão na fronteira de uma região densa, são pontos que estão na ε -vizinhança de algum ponto central, contudo, não são pontos centrais, pois a cardinalidade desses pontos em relação ao raio ε não excede η ;

c) *outliers*: esses pontos não são centrais e nem de borda e assim não são conectados por densidade a nenhum outro ponto, não pertencendo a nenhum *cluster*.

A Figura 4 ilustra os pontos conforme a localização.

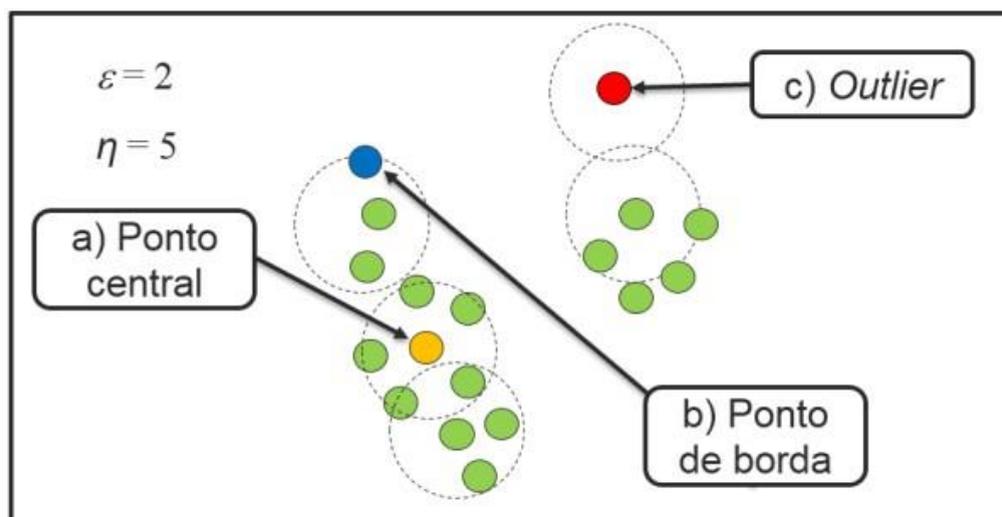


FIGURA 4 – Ponto central, ponto de borda e *outlier*. Fonte: Adaptado de Riazifar & Saghapour (2015).

Caso um ponto for classificado como *outlier* pelo algoritmo DBSCAN, depois ele pode estar na ε -vizinhança de outro ponto ainda não visitado pelo algoritmo. Sendo assim, essa classificação pode ser modificada caso o ponto seja diretamente alcançável por densidade a partir de um ponto central ainda não visitado (DAI & LIN, 2012).

3. MÉTODO DE TRABALHO

Realizou-se uma pesquisa de referencial teórico em: artigos, livros, teses, dissertações, *websites* com conteúdo sobre absentismo e presenteísmo, KDD, *data mining* e o algoritmo DBSCAN.

A base de dados utilizada no experimento é composta por 2.144 registros com 22 atributos de 41 empregados coletados durante o período de julho de 2008 a julho de 2012 na área operacional de uma organização brasileira de grande porte.

Os 22 atributos que foram utilizados para compor a base de dados são apresentados a seguir: código do nome, motivo do afastamento (Classificação Internacional de Doenças -CID), mês do afastamento, dia da semana, estação do ano, tempo de afastamento em dias, tempo de afastamento em horas, tempo de afastamento em dias e horas, custo do transporte até o trabalho em reais multiplicado pela distância do local de trabalho em quilômetros dividido por mil, tempo de serviço em anos, idade em anos, escolaridade, número de filhos, etilista, tabagista, número de animais de estimação, peso, altura, índice de massa corporal (IMC), e manequim (camisa, calça e calçado).

O *software* utilizado nos experimentos foi o WEKA versão 3.6.13 disponível no *website* (<http://www.cs.waikato.ac.nz/ml/weka>). O WEKA possui uma coleção de algoritmos de aprendizado de máquina com código aberto para tarefas de *data mining* (TURBAN, SHARDA, ARONSON & KING, 2009).

A plataforma de *hardware* utilizada nos experimentos foi um microcomputador com processador Intel® de 3,5 MHZ, com 4 GB de memória RAM e 1 TB de disco rígido.

4. RESULTADO DOS EXPERIMENTOS COMPUTACIONAIS

Uma vez apresentado nas seções anteriores o método de trabalho do artigo e o referencial teórico, nesta seção serão apresentados os resultados dos experimentos computacionais usando *data mining* com o algoritmo DBSCAN na clusterização de perfis de empregados absenteístas e presenteístas.

4.1 Experimentos Computacionais (Visualização inicial)

A Figura 5 apresenta os histogramas gerados pelo *software* WEKA com os 22 atributos.

Os histogramas evidenciam intervalos de maior representatividade que podem ajudar na interpretação inicial dos dados, por exemplo:

- a) Motivo do afastamento, apresenta grande concentração em apenas quatro motivos;
- b) Meses que apresentam maior número de afastamentos (março, agosto, outubro);
- c) Distribuição dos afastamentos nos dias da semana;
- d) Tempo de serviço de empregados com maior número de afastamentos (entre 7,5 e 10 anos);
- e) Idade dos empregados com maior número de afastamentos (entre 31 e 37 anos).

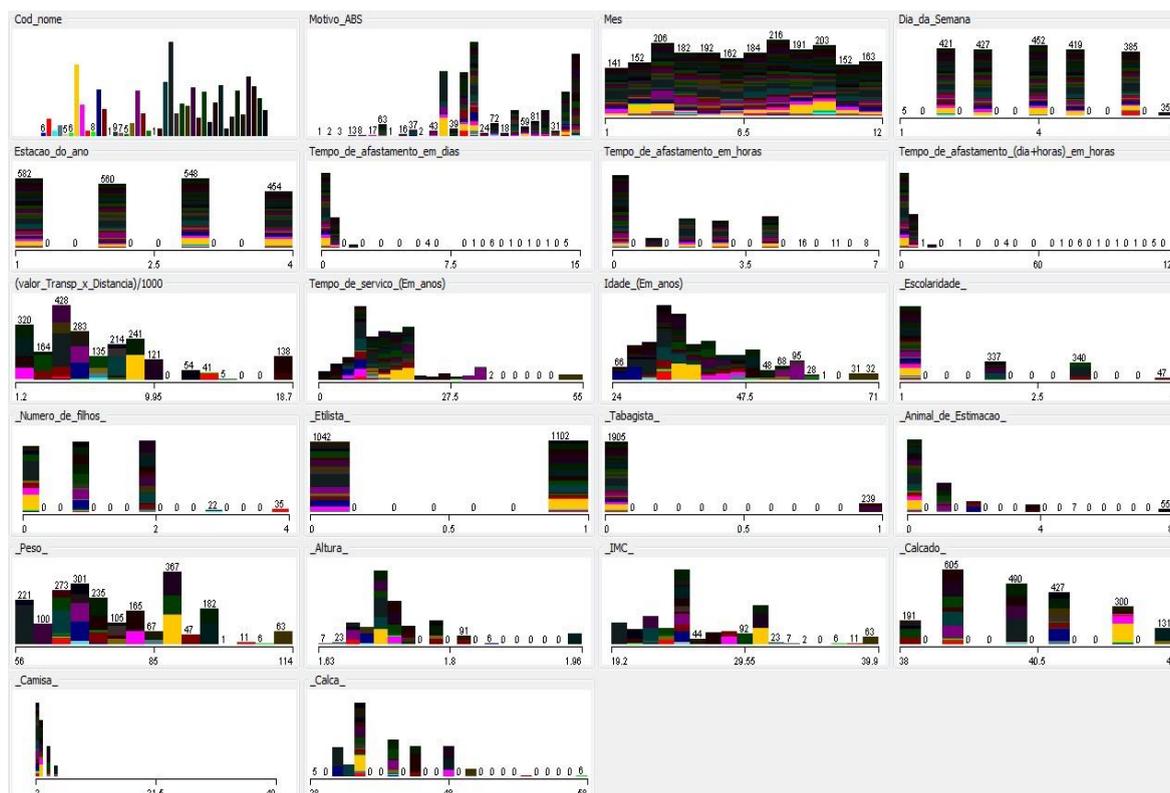


FIGURA 5 – Histogramas gerados com os 22 atributos.

4.2 Experimentos Computacionais (DBSCAN)

Os parâmetros utilizados no experimento com o DBSCAN foram: ϵ -vizinhança igual: 1,5 e mínimo pontos igual $\eta = 2$.

A função de distância utilizada foi a euclidiana. O tempo de processamento para criar o modelo (dados de treinamento completo): 1,92 segundos. A Tabela 1 mostra os registros distribuídos por *clusters*.

TABELA 1 - Registros distribuídos por *clusters*

<i>Clusters</i>	Registros	(%)
<i>Cluster 0</i>	1836	85,63%
<i>Cluster 1</i>	226	10,54%
<i>Cluster 2</i>	63	2,94%
<i>Cluster 3</i>	6	0,28%
<i>Cluster 4</i>	7	0,33%
<i>Cluster 5</i>	6	0,28%

A Figura 6 mostra os *clusters* gerados pelo algoritmo DBSCAN.

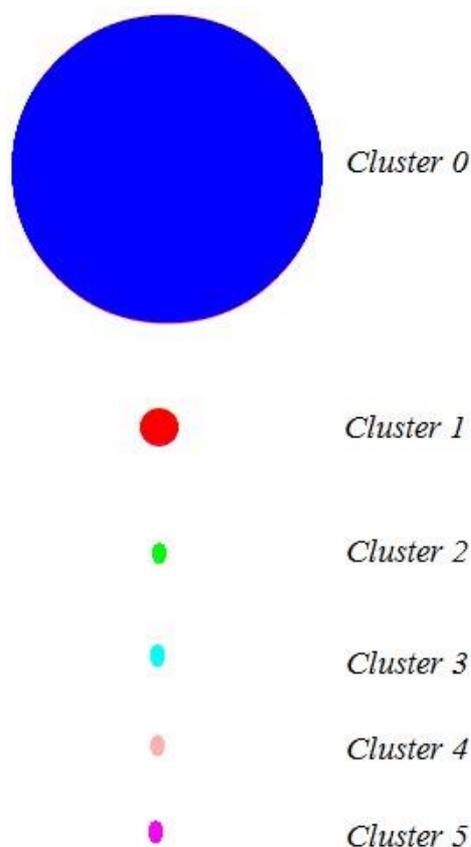
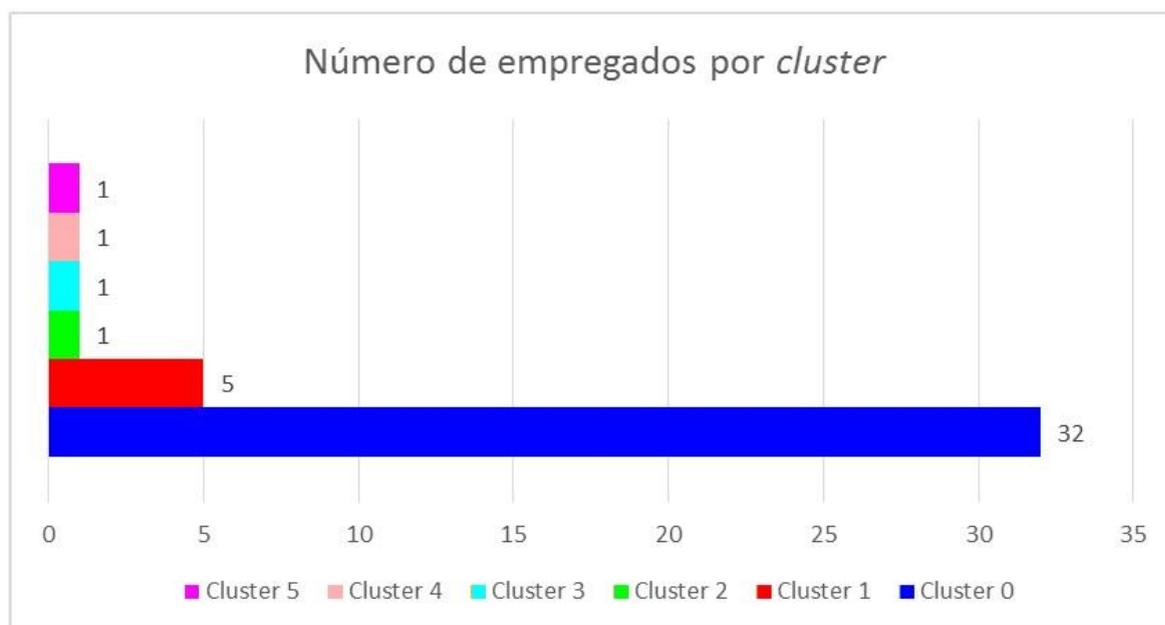


FIGURA 6 – Clusters gerados pelo algoritmo DBSCAN.

A Figura 7 mostra o número de empregados por *cluster*. Observa-se que os *clusters*: 2, 3, 4, 5, apresentam empregados presenteístas e com tendência ao presenteísmo (9,76%). O *cluster* 1 apresenta empregados absenteístas e com tendência ao absenteísmo (12,2%). O *cluster* 0 representa a maior parte dos empregados (78,05%), esses empregados apresentam em sua maioria afastamentos dentro do padrão aceitável para o empregador.

FIGURA 7 – Número de empregados por *cluster*.

A Figura 8 mostra os empregados (eixo x) e o tempo de afastamento em horas (eixo y). Observa-se que os empregados do *cluster 1* (absenteístas) apresentam afastamentos em sua maioria de curta duração e com grande frequência.

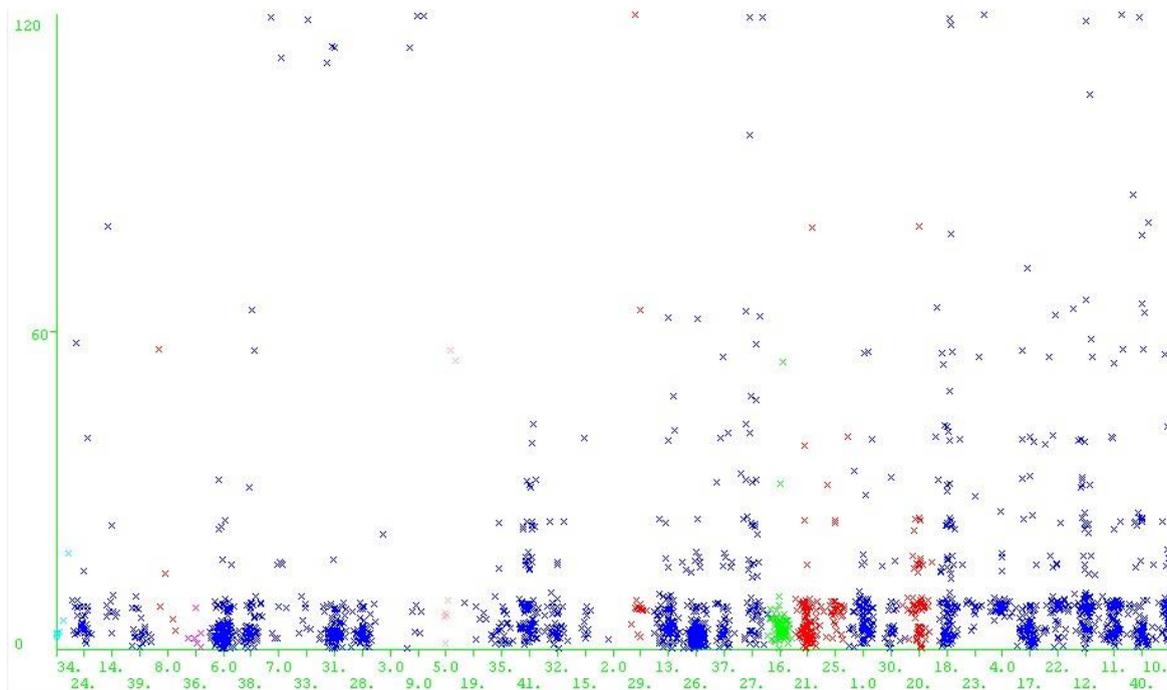


FIGURA 8 – Empregados (eixo x) e o tempo de afastamento (eixo y).

A Figura 9 mostra o tempo de afastamento (eixo x) e a idade do empregado (eixo y). Observa-se que curiosamente o empregado mais idoso (*cluster 2*) apresenta pequenos afastamentos, em sua maioria menores que quatro horas.

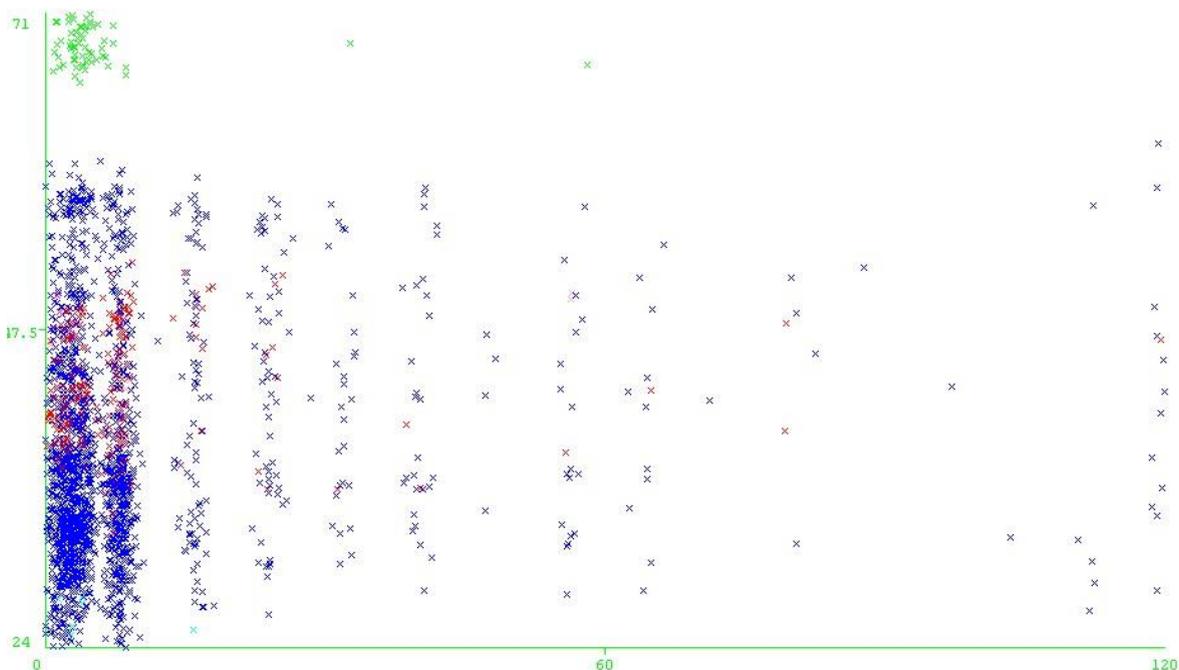


FIGURA 9 – Tempo de afastamento em horas (eixo x) e a idade do empregado (eixo y).

Percebe-se ainda na Figura 9 que os afastamentos de curta duração são

predominantes entre os empregados mais jovens, é possível perceber que o empregado mais idoso (*cluster 2*) não possui nenhum afastamento maior que 60 horas.

A Figura 10 mostra o tempo de afastamento em dias (eixo x) e o índice de massa corporal - IMC (eixo y). Observa-se que os empregados do *cluster 0* (dentro do padrão aceitável para o empregador) apresentam em sua grande maioria o índice de massa corporal classificado como saudável. Verifica-se ainda que os empregados do *cluster 2* e 5 são classificados como obesos, apesar de poucos afastamentos registrados, o que indica empregados presenteístas e com tendência ao presenteísmo.

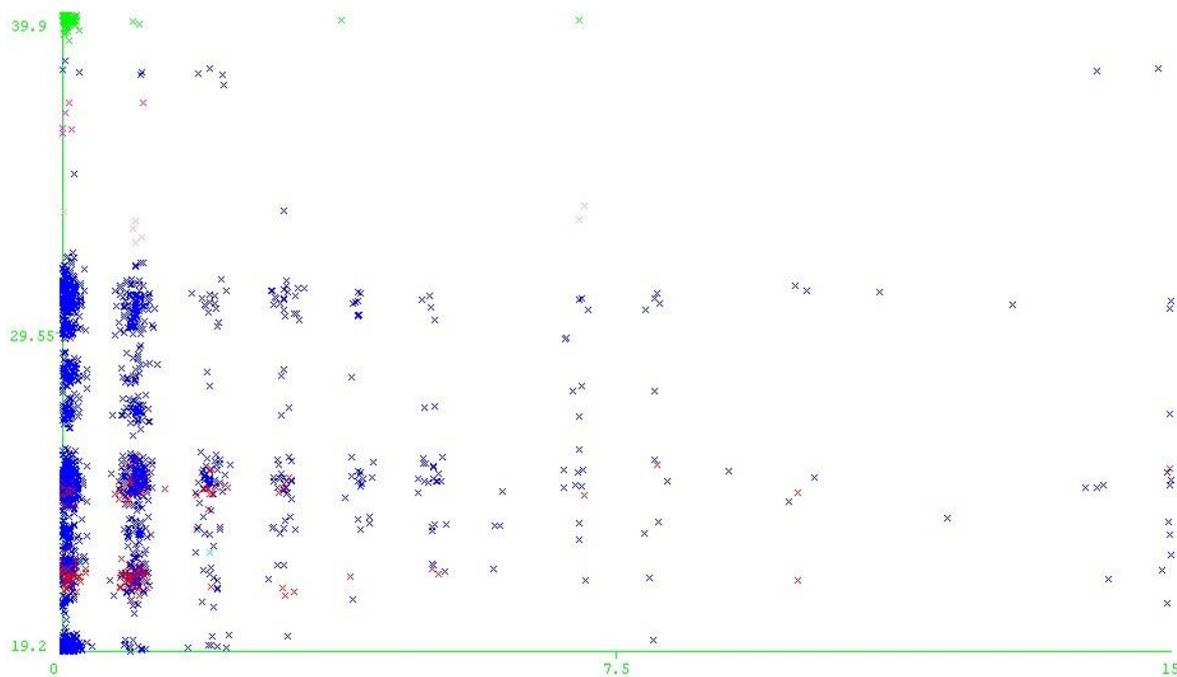


FIGURA 10 – Tempo de afastamento em dias (eixo x) e o índice de massa corporal - IMC (eixo y).

Observa-se inicialmente que os *clusters* conseguiram representar as principais características desejadas, ou seja, foram identificados os empregados com perfis absenteístas e presenteístas e também as tendências.

5. CONSIDERAÇÕES FINAIS

Os resultados apresentados mostram que é possível aplicar técnica de *data mining* para identificar através da clusterização o perfil de empregados absenteístas e presenteístas utilizando o algoritmo DBSCAN o que permite encontrar padrões de comportamento absenteísta e presenteísta em uma base de dados de uma organização.

O conhecimento útil extraído pelo *data mining* pode contribuir na tomada de decisão da organização nos seguintes aspectos:

- a) redução dos programas corporativos sem foco para dar lugar a programas localizados mais eficazes e menos custosos para a organização;
- b) auxílio no tratamento médico e social dos empregados que apresentam um dos dois perfis;
- c) implantação de programas com vistas a melhorar a qualidade de vida no trabalho, a maior produtividade e a melhor qualidade dos serviços; e
- d) redução das contas médicas e assistenciais da organização.

Os estudos aqui realizados não têm a pretensão de esgotar o assunto, pelo contrário, buscou-se realizar uma contribuição com esta aplicação. Sabe-se que existe uma clara demanda por estudos sistematizados que possam estabelecer outras aplicações para identificar o perfil dos empregados absenteístas e presenteístas em organizações de diversos setores.

REFERÊNCIAS BIBLIOGRÁFICAS

Ankerst, M., Breunig, M. M., Kriegel, Hans-Peter & Sander, J. (1999). *Optics: Ordering Points to Identify the Clustering Structure*. In Proc. ACM SIGMOD'99 Int. Conf. on Management of Data (SIGMOD'99), pages 49-60, Philadelphia.

Araújo, J. P. (2012). *Afastamento do trabalho: absenteísmo e presenteísmo em uma Instituição Federal de Ensino Superior*. Dissertação de Mestrado, programa de pós-graduação em Ciências da Saúde, Universidade de Brasília, Brasília.

Azevedo, C. S. & Santos, M. F. (2005). *Data Mining - Descoberta de Conhecimento em Bases de Dados*. Lisboa, Portugal: FCA – ed. Informática – Coleção: Sistema de Informação, 214 p.

Barros, A. S. & Campos, F. C. (2006). *Uma discussão sobre a aplicação de processo de KDD e técnicas de mineração de dados na indústria automobilística*. XIII SIMPEP - Bauru, SP, Brasil, 6 a 8 de nov.

Berry, M. J. A. & Linoff, G. (1997). *Data Mining Techniques: For Marketing, Sales, and Customer Support*. New York: Wiley Computer Publishing.

Bergstrom, G., Bodin, L., Hagberg, J., Lindh, T., Aronsson, G. & Josephson, M. (2009). *Does sickness presenteeism have an impact on future general health?* *Int Arch Occup Environ Health*.

Bernstrom, V. H. (2013). *The relationship between three stages of job change and long-term sickness absence*. *Social Science & Medicine*, v. 98, December 2013, <http://dx.doi.org/10.1016/j.socscimed.2013.10.001>. Pages 239–246.

Bigus, J. P. (1996). *Data Mining with Neural Network: Solving Business Problems from Applications Development to Decision Support*. Mcgraw-Hill, 1996.

Bratsberg, B., Fevang, E. & Roed, K. (2013). *Job loss and disability insurance*. *Labour Economics*, 24 (2013) 137–150. <http://dx.doi.org/10.1016/j.labeco.2013.08.004>.

Buckerman, P. & Laukkanen, E. (2009). *Presenteeism in Finland: Determinants by Gender and the sector of conomy*. *Ege Akademik Bakış / Ege Academic Review* 9 (3): p.1007- 1016.

Buttillos, A. S., Vargas III, K. G. & Gomero-Cuadra, R. (2015). *Work productivity among adults with varied Body Mass Index: Results from a Canadian population-based survey*. *Journal of Epidemiology and Global Health* (2015) 5, <http://dx.doi.org/10.1016/j.jegh.2014.08.001>, p.191– 199.

Canfield, G. W. & Soash, D. G. (1955). *Presenteeism - A constructive view*. *Industrial Medicine and Surgery*, vol. 24, p. 417-418.

Carvalho, L. A. V. (2005). *Data Mining: a Mineração de dados no Marketing, Medicina, Economia, Engenharia e Administração*. Editora Ciência Moderna Ltda. Rio de Janeiro.

Costa, J. A. F. & Andrade Netto, M. L. (2007). *Segmentação de Mapas Auto-Organizáveis*

- com Espaço de Saída 3-D*. Revista Controle & Automação/Vol.18 n.2/abr, mai. e Jun.
- CUNHA, J. B., BLANK, V. L. G. & BOING, A. F. (2009). *Tendência Temporal de Afastamento do Trabalho em Servidores Públicos (1995-2005)*. Revista Brasileira de Epidemiologia. São Paulo, vol.12 n. 2.
- Dai, B. & Lin, I. (2012). *Map/Reduce-based DBSCAN Algorithm with Optimized Data Partition*. IEEE Fifth International Conference on Cloud Computing, p. 59-66.
- Diniz, C. A. & Louzada Neto, F. (2000). *Data Mining: uma introdução*. São Paulo: ABE.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). *From data mining to knowledge discovery: an overview*. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press / The MIT Press, MIT, Cambridge, Massachusetts, and London, England, p.1-34.
- Fonseca, V. R. (2009). *Atividade Física, Absenteísmo e Demanda por Atendimento à Saúde de Funcionários de Indústria Automobilística de São Caetano do Sul*. Tese de Doutorado em Ciências – Universidade de São Paulo. 154 p., São Paulo.
- Freitas, A. A. (2003). *A Survey of Evolutionary Algorithms for Data Mining and knowledge Discovery*. Curitiba.
- Goetzel, R. Z, Long, S. R, Ozminkowski, R. J, Hawkins, K., Wang, S. H & Lynch, W. H. (2004). *Absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting US employers*. *Journal of Occupational And Environmental Medicine*. v.46(4): p. 398- 412.
- Goldschmidt, R. & Passos, E. (2005). *Data mining: um guia prático*, Elsevier.
- Goldschmidt, R., Passos, E. & Bezerra, E. (2015). *Data mining: Conceitos, técnicas, algoritmos, orientações e aplicações*, Rio de Janeiro, 2 ed. Elsevier.
- Guelpeli, M. V. C. (2009). *Cassiopeia: Um Modelo Baseado em Sumarização e Aprendizado Autônomo Usado em Agrupamentos para Descoberta de Conhecimento em Bases textuais*. 100 p. Tese de Doutorado em Computação – Universidade Federal Fluminens. Niterói, RJ.
- Halbesleben, J. R. B., Whitman, M. V. & Crawford, W. S. (2014). *A dialectical theory of the decision to go to work: Bringing together absenteeism and presenteeism*. *Human Resource Management Review* 24 (2014) 177–192. <http://dx.doi.org/10.1016/j.hmr.2013.09.001>.
- Hemp, P. (2004). *Presenteeism: at work – but out of it*. *Havard Business Review*. 82(10): 49-58.
- Inoue, K. C., Matsuda, L. M. , Silva, D. M. P. P., Uchimura, T. T. & Mathias, T. A. F. (2008). *Absenteísmo-doença da equipe de enfermagem em unidade de terapia intensiva*. *Rev. bras. enfermagem*. v.61 n°2. Mar./Abr., Brasília.
- Jodas, D. A. & Haddad, M. C. L. (2009). *Síndrome de Burnout em Trabalhadores de Enfermagem de um Pronto Socorro de Hospital Universitário*. *Acta Paulista de Enfermagem*. v. 22, n. 2. São Paulo.
- Johns, G. (2003). *How methodological diversity has improved our understanding of absenteeism from work*. *Human Resource Management Review*, 13 (2): p. 157-184.
- Johns, G. (2010). *Presenteeism in the workplace: A review and research agenda*. *Journal of Organizational Behavior*. vol. 31, p. 519 – 542.

Laranjeira, C. A. (2009). *O Contexto Organizacional e a Experiência de Estress: uma Perspectiva Integrativa*. Revista *salud pública* [online], vol.11, n.1, p. 123-133.

Levin-Epstein, J. (2005). *Presenteeism and Paid Sick Days*. Center for LAW and Social Policy – CLASP, ago.

Lumare Júnior, G. (2007). *Valor econômico do cliente: uma teoria das encomendas*. São Paulo: Financial Times – Prentice Hall.

Martin Ester, Hans-Peter Kriegel, Jiirg Sander & Xiaowei Xu. (1996). *A Density Based Spatial Clustering of Applications With Noise*. In Simoudis, E., Han, J. & Fayyad, U., editors, Second International Conference on Knowledge Discovery and Data Mining. Portland, Oregon. American Association for Artificial Intelligence, pages 226-231.

Martiniano, A. (2012). *Aplicação de técnica da mineração de dados na identificação do perfil de empregados absenteístas e presenteístas em uma empresa de courier da cidade de São Paulo*. 130 p. Dissertação de Mestrado em Engenharia de Produção – Universidade Nove de Julho, São Paulo.

Martinez, L. F., Ferreira, A. I., Souza, L. M. & Cunha, J. V. (2007). *A Esperança é a última a morrer? Capital Psicológico Positivo e Presentismo*. *Comportamento Organizacional e Gestão*, vol. 13, n. 1, p. 37-54.

Martins, P. L., Silva, E. M. S. & Rocha, R. C. (2009). *Fatores Prejudiciais ao bom desempenho dos Funcionários nas Diversas Organizações: com Ênfase para o Alcoolismo*. VI SEGeT – Simpósio de Excelência em Gestão e Tecnologia.

Mckevitt Tauffer, R. & Coltre, S. M. (2007). *A contribuição dos fatores intrínsecos e extrínsecos da organização para o absenteísmo nos setores de produção de uma indústria de carrocerias de ônibus*. In: Encontro Nacional de Engenharia de Produção – ENEGEP. Anais. n. 27, Foz do Iguaçu.

Miranda, A. R., Gomes, C. F. & Gomes, M. A. N. (2006). *Uma avaliação da Qualidade de Vida no Trabalho dos gerentes da construção civil*. XIII SIMPEP, nov. 2006. Bauru – SP, Brasil.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Paschoalino, J. B. Q. (2008). *O Professor Adoecido Entre o Absenteísmo e o Presenteísmo*. VII Seminário Redestrado – Nuevas Regulaciones en América Latina.

Penatti, I, Zago, J. S. & Quelhas, O. (2006). *Absenteísmo: as consequências na gestão de pessoas*. III SEGeT – Simpósio de Excelência em Gestão e Tecnologia. Resende.

Riazifar, N. & Saghapour, E. (2015). *Retinal Vessel Segmentation Using System Fuzzy and DBSCAN Algorithm*. *IEEE 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA 2015)*, p. 1-4.

Sassi, R. J. (2006). *Uma Arquitetura Híbrida para Descoberta de Conhecimento em Bases de Dados: Teoria dos Rough Sets e Redes Neurais Artificiais Mapas Auto-Organizáveis*. 169 p. Tese de Doutorado, Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia Elétrica, São Paulo.

Silva Neto, M. A. (2008). *Mineração Visual de Dados: extração do conhecimento a partir das técnicas de Visualização da Informação e mineração de dados*. 172 p. Dissertação de

Mestrado em Ciências na área de concentração Programação Matemática – Universidade Federal do Paraná, PR.

Turban, E., Sharda, R., Aronson, J. E. & King, D. (2009). *Business Intelligence: Um enfoque gerencial para a inteligência do negócio*. Porto Alegre: Bookman.

Thorwarth, M. Arisha, A. & Harper, P. (2009). *Simulation Model to Investigate Flexible Workload Management for Healthcare and Servicescape Environment*. IEEE Proceedings of the 2009 Winter Simulation Conference M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, eds. p. 1946-1956.

Umann, J. (2011). *Estresse, Coping e Presenteísmo em Enfermeiros Hospitalares*. 132 p. Dissertação de Mestrado em Enfermagem, Universidade Federal de Santa Maria, Rio Grande do Sul. RS.

URIS, A. (1955). *How to build presenteeism*. *Petroleum Refiner*, v. 34, p. 348-359.

WEKA. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka>>. Acesso em: 15 Jan. 2016.