

O Mapa Auto-Organizável de Kohonen no Processo de Descoberta de Empregados Absenteístas e Presenteístas

Andréa Martiniano (Universidade Nove de Julho) martin@uninove.edu.br
Renato José Sassi (Universidade Nove de Julho) sassi@uninove.br
Ricardo Pinto Ferreira (Universidade Nove de Julho) kasparov@uninove.edu.br

Resumo:

Junto à evolução da sociedade, as diferentes culturas, tornaram o trabalho mais complexo, exigindo cada vez mais dos empregados que acabam adquirindo algum distúrbio na saúde. O Absenteísmo por doença e o Presenteísmo são semelhantes. No caso do Absenteísmo os empregados ausentam-se do trabalho por motivo de doença. Os empregados presenteístas, estão presentes no trabalho apesar de sua doença. Com muita informação armazenada, as organizações perceberam que não bastava apenas disponibilizá-la, era necessário interpretá-la, analisá-la e relacioná-la a fim de extrair conhecimento para apoiar a decisão. Por essa razão faz-se necessário o uso do KDD (*Knowledge Discovery in Databases*) ou Descoberta de Conhecimento em Bases de Dados. O processo KDD, implica em extrair conhecimentos, potencialmente úteis, novos, válidos e compreensíveis, visando melhorar o entendimento de um problema. O objetivo deste artigo é determinar em uma base de dados os empregados que apresentam comportamento Absenteísta e Presenteísta utilizando *Data Mining* ou Mineração de Dados e agrupá-los utilizando uma rede neural artificial do tipo Mapa Auto-Organizável de Kohonen ou *Self-Organizing Maps* (SOM). O uso da Rede SOM permite descobrir dentro de um grupo de empregados aqueles que já apresentam ou que possam apresentar estes comportamentos e antecipar um possível tratamento e recuperação.

Palavras chave: Absenteísmo, Mapa Auto-Organizável, Presenteísmo.

Self-Organizing Map of Kohonen in the Process of Discovery of Employees Absenteists and Presenteists

Abstract

Along with the evolution of society, different cultures, the work became more complex, demanding more of employees who end up getting some health disorder. The Absenteeism due to illness and Presenteeism are similar. In the case of absenteeism employees absent from work due to illness. Presenteeists employees are present at work despite his illness. With lots of stored information, organizations have realized that it was not enough just to make it available, it was necessary to interpret it, analyze it and relate it to extract knowledge to support the decision. For this reason it is necessary the use of KDD, Knowledge Discovery in Databases. The KDD process, involves extracting knowledge, potentially useful, new, valid and understandable, to improve the understanding of a problem. The aim of this paper is to determine in a database that employees exhibit behavior absenteeist and Presenteeist using Data Mining and group using an artificial neural network-type Self-Organizing Map of Kohonen (SOM). The use of SOM Network lets you discover within a group of employees who already have or who may have these behaviors and anticipate a possible treatment and recovery.

Key-words: Absenteeism, Self-Organizing Map, Presenteeism.

1. Introdução

A evolução da sociedade ressaltou a importância do relacionamento entre homens, entre as diferentes culturas e entre mercados. Nestas circunstâncias, o trabalho tornou-se mais complexo, suplantando o sentido tradicional que o identifica como meio de satisfação de necessidades (LUMARE JÚNIOR, 2007).

O Absenteísmo representa para a empresa perda da capacidade produtiva e lucrativa. Um serviço deixa de ser fornecido devido à falta do funcionário, reduzindo a receita da empresa (MARTINIANO, 2010).

Assim, o conhecimento do absenteísmo-doença em suas múltiplas determinações torna-se fundamental para a organização e planejamento das ações dos Serviços de Atenção à Saúde do Trabalhador e a determinação de seu perfil permite a identificação de excessos de ocorrências de casos de determinadas morbidades (SILVA *et al.*, 2011).

O custo do Absenteísmo por motivo de doença é, em muitas organizações, maior do que o salário pago ao trabalhador doente que deixa de comparecer ao local de trabalho Chiavenato (2004) e não havendo substitutos disponíveis, as implicações para a produtividade são bem maiores (SILVA *et al.*, 2000).

Nesse contexto o *Data Mining* (DM) surge como uma importante ferramenta que poderá ajudar a extrair informações valiosas de uma base de dados para determinação dos empregados que possam apresentar tendência ao Absenteísmo ou ao Presenteísmo auxiliando ou antecipando este auxílio no tratamento e recuperação destes empregados (MARTINIANO, 2010).

O *Data Mining* é considerado a etapa mais importante do processo de KDD. Caracteriza-se pela existência do algoritmo minerador, que diante da tarefa especificada será capaz de extrair de modo eficiente conhecimento implícito e útil de um banco de dados (SASSI, 2009a).

Segundo Cabena *et al.* (1997), o *Data Mining* é a técnica de extrair informação, previamente desconhecida e de máxima abrangência a partir de bases de dados, para usá-la na tomada de decisão, todas as técnicas que permitem extrair conhecimento de uma massa de dados que, de outra maneira, permaneceria escondido nas grandes bases.

Uma das grandes vantagens de se utilizar as técnicas de *Data Mining*, é que permitem avaliar como as perguntas se relacionam com as respostas (padrões e relações) encontradas. Achadas essas relações e padrões, é fornecida uma base de regras que servem de apoio aos processos de tomada de decisão. Para tal, utilizam-se técnicas de procura baseadas em Inteligência Artificial, como as Redes Neurais Artificiais, as Árvores de Decisões, a Teoria dos Conjuntos *Fuzzy*, os Algoritmos Genéticos ou, ainda, combinações entre técnicas gerando os chamados Sistemas Híbridos ou Arquiteturas Híbridas (Sassi, Silva e Hernandez, 2008). Nesse artigo foi utilizado o Mapa Auto-Organizável (*Self-Organizing Map*) também conhecido como rede de Kohonen ou simplesmente rede SOM, é uma arquitetura de rede neural artificial com aprendizado do tipo não-supervisionado (Haykin, 2001; Kohonen, 2001). As principais aplicações da rede SOM estão no campo de agrupamento, classificação e visualização de dados.

O artigo está organizado em 9 seções: além da Introdução (seção 1) o artigo compõe-se das seguintes partes: a seção 2 apresenta a Metodologia da pesquisa, a seção 3 apresenta a fundamentação teórica do Absenteísmo, a seção 4 apresenta o Presenteísmo, a seção 5 apresenta a *Data Mining*, a seção 6 descreve os processos de sistemas de descoberta de conhecimento, os chamados KDD, a seção 7 apresenta o conceito básico da Rede SOM, na

seção 8 são apresentados os experimentos e descrição dos resultados da pesquisa, o artigo é encerrado na seção 9 com a conclusão e discussão dos resultados obtidos e propõem-se direções para futuras investigações.

2. Metodologia

A metodologia adotada foi definida como exploratória e experimental. A realização da pesquisa bibliográfica esta embasada em consultas a fontes bibliográficas e de referencial teórico: artigos, livros, teses, dissertações, *websites* com conteúdos sobre absenteísmo, presenteísmo, KDD, DM, RNAs e rede SOM.

A base de dados do absenteísmo na empresa de Courier é composta por 740 registros de 36 empregados coletadas durante o período de Julho de 2007 a Julho de 2010.

Os motivos do absenteísmo com CID foram estratificados em 21 categorias, os registros do Absenteísmo sem CID foram estratificados em 7 categorias, totalizando 28 categorias de motivos do absenteísmo, compondo um atributo.

Os outros 20 atributos que foram utilizados para compor a base de dados são apresentados a seguir: código do nome, mês, dia da semana, estação do ano, tempo de afastamento em horas, custo do transporte até o trabalho em reais, distância do local de trabalho em quilômetros, tempo de serviço em anos, idade, carga de trabalho, meta alcançada, falta disciplinar, escolaridade, número de filhos, etilismo, tabagismo, número de animais de estimação, peso, altura e índice de massa corporal. Esta foi a base de dados com 21 atributos utilizada nos experimentos.

O CID é uma referência de classificação da Organização Mundial de Saúde (OMS) para a descrição dos estados de saúde que incluiu todas as doenças e motivos de consultas, permite o uso em um conjunto de causas capazes de produzir uma doença e atende as necessidades de informação diagnóstica para finalidades gerais (NUBILA e BUCHALLA, 2008).

3. Mineração de dados

De acordo com Berry e Linoff (1997), a mineração de dados é a exploração e análise, de forma automática ou semi-automática, de grandes bases de dados com objetivo de descobrir padrões, regras e fornecer informações que permitam montar estratégias corporativas eficientes.

Para Goldschmidt e Passos (2005), a mineração de dados 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. Este procedimento faz uso de técnicas estatísticas avançadas e Inteligência Artificial para descobrir fatos em uma grande base de dados.

É inviável a extração de conhecimentos das grandes bases de dados por recursos tradicionais da Tecnologia da Informação sem o auxílio de ferramentas computacionais apropriadas que o amparem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possam desenvolver estratégias de ação para cada processo decisório (GOLDSCHMIDT e PASSOS, 2005).

Diante deste contexto, pode-se aplicar o KDD. 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, é composta de diversas fases dentre as quais são utilizados métodos baseados na estatística e no aprendizado de máquina.

O KDD é um processo de várias fases e contém 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 *et al.*, 1996). O KDD é composto por cinco fases, a saber: a) Seleção dos dados; b) Pré-processamento dos dados; c) Transformação dos dados; d) Mineração de Dados (DM) e e) Interpretação/avaliação do conhecimento.

Existem várias formas de interpretação dos dados pela KDD denominadas tarefas. Diversas tarefas podem ser realizadas, as tarefas mais comuns são (FAYYAD *et al.*, 1996).

- a) associação – possibilidade de encontrar tendência, podendo ser incrementada estrategicamente as vendas de um determinado segmento comercial;
- b) classificação – consiste em obter um modelo baseado em um conjunto de exemplos;
- c) clusterização (agrupamento) – transforma registros com grande número de atributos em conjuntos relativamente menores (segmentos), consiste na identificação de grupos semelhantes e;
- d) visualização – processo para transformar informação em uma forma visual.

A MD é a fase mais importante da Descoberta de Conhecimento em Bases de Dados (FAYYAD *et al.*, 1996).

A MD 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. Este procedimento faz uso de técnicas estatísticas avançadas e Inteligência Artificial para descobrir fatos em uma grande base de dados (GOLDSCHMIDT e PASSOS, 2005).

A MD explora as bases de dados a procura de padrões escondidos, encontrando dados que permitem prever tendências e comportamentos futuros, que os especialistas podem não descobrir devido ao fato de essa informação sair do limite de suas expectativas, o que possibilita a tomada de decisão (FAYYAD *et al.*, 1996; CABENA *et al.*, 1998).

Uma das grandes vantagens de se utilizar as técnicas de MD, é que esta permite avaliar como as perguntas se relacionam com as respostas (padrões e relações) encontradas. Para tal, utilizam-se técnicas de procura baseadas AI, como as Redes Neurais Artificiais (RNAs), as Árvores de Decisões, a Teoria dos Conjuntos *Fuzzy*, os Algoritmos Genéticos ou, ainda, combinações entre estas técnicas, gerando os chamados Sistemas Híbridos ou Arquiteturas Híbridas (SASSI, SILVA e HERNANDEZ, 2008).

4. Mapa auto-organizável

O Mapa Auto-Organizável de Kohonen (*Self-Organizing Map*), ou rede SOM, foi desenvolvida pelo pesquisador finlandês Teuvo Kohonen no início da década de 80 (KOHONEN, 1982).

A rede SOM é uma rede neural de aprendizagem competitiva organizada em duas camadas, a primeira camada representa o vetor dos dados de entrada, a segunda corresponde a uma grade de neurônios, geralmente bidimensional, totalmente conectada aos componentes do vetor de entrada. Cada neurônio possui um vetor de código associado (KOHONEN, 2001).

As principais aplicações da rede SOM estão no campo de agrupamento e visualização de dados. Estas características fazem com que a rede SOM seja muito utilizada na geração de agrupamentos também chamados de clusters (VESANTO e ALHONIEMI, 2000).

O algoritmo responsável pela formação do Mapa Auto-Organizável começa inicializando os pesos sinápticos da grade. Isto pode ser feito atribuindo-lhes valores pequenos tomados de um

gerador de números aleatórios, fazendo dessa forma, nenhuma organização prévia imposta ao mapa de características (HAYKIN, 2001).

Embora a rede SOM seja utilizada para diversas aplicações em diferentes áreas do conhecimento; Classificação de perfis profissiográficos usando a rede SOM (ARAÚJO, 2002), Descoberta de Conhecimento em Bases de Dados (SASSI, 2006), Categorização de Imagens Médicas (SILVA, 2009); os trabalhos mais bem difundidos tratam sobre problemas que envolvem classificação de padrões e agrupamento de dados ou clusterização (SILVA *et al.*, 2010).

Esse modelo de rede neural foi inspirado no córtex sensitivo, em que regiões vizinhas respondem por partes sensitivas localizadas próximas do corpo humano (KOHONEN, 1982).

Os nós da rede SOM estão localmente interconectados por uma relação de vizinhança, determinando a topologia do mapa. Num mapa bidimensional, a vizinhança pode ser hexagonal ou retangular.

O algoritmo responsável pela formação do Mapa Auto-Organizável começa inicializando os pesos sinápticos da grade. Isto pode ser feito atribuindo-lhes valores pequenos tomados de um gerador de números aleatórios, fazendo dessa forma, nenhuma organização prévia imposta ao mapa de características (HAYKIN, 2001).

Segundo Haykin (2001) em um mapa auto-organizável, os neurônios estão colocados em nós de uma grade que é normalmente uni ou bidimensional. Mapas de dimensionalidade mais alta são também possíveis, mas não são tão comuns. Os neurônios se tornam seletivamente sintonizados a vários padrões de entrada (estímulos) ou classes de padrões de entrada no decorrer de um processo de aprendizagem. A Figura 1 exibe um arranjo hexagonal.

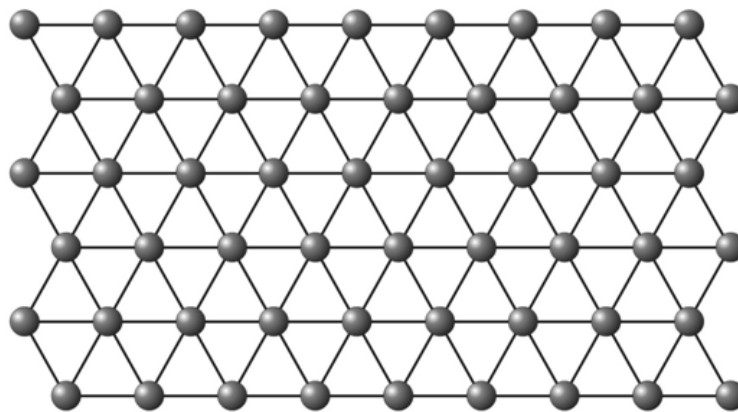


Figura 1 – Mapa com arranjo hexagonal para uma rede SOM com grade bidimensional

Na Figura 1 pode-se observar a rede com vizinhança hexagonal. O formato do arranjo influencia diretamente na adaptação da rede SOM (HAYKIN, 2001; KOHONEN, 2001).

O algoritmo de aprendizado da rede SOM é composto por três fases: competição, cooperação e adaptação:

a) competição – concorrência entre os neurônios, com o objetivo de determinar o neurônio vencedor da competição clusterização (SILVA *et al.*, 2010). Para cada padrão de entrada, os neurônios da grade calculam seus respectivos valores de uma função discriminante. Esta função discriminante fornece a base para a competição entre os neurônios. O neurônio com o

maior valor da função discriminante é declarado vencedor da competição (HAYKIN, 2001). A equação a seguir resume a essência do processo competitivo entre os neurônios:

$$i(x) = \arg \min_j \|x - w_j\|, j = 1, 2, \dots, l \quad (1)$$

b) cooperação - O neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, fornecendo assim a base para a cooperação entre os neurônios vizinhos, a equação a seguir resume a função da vizinhança topológica:

$$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right) \quad (2)$$

c) Adaptação Sináptica - permite que os neurônios excitados aumentem seus valores individuais em relação ao padrão de entrada através de ajustes adequados aplicados a seus pesos sinápticos (HAYKIN, 2001). A equação a seguir permite ajustar os pesos sinápticos:

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x - w_j(n)) \quad (3)$$

5. Resultado dos experimentos

Na primeira fase dos experimentos apresentou-se à rede SOM a toda a base de dados com os 740 registros e os 21 atributos, durante o período de Julho de 2007 a Julho de 2010, gerando todos os mapas do experimento. Na segunda fase do experimento os dados foram filtrados utilizando o *software* MS Excel, gerando as tabelas com o intuito de auxiliar a análise dos resultados.

A Figura 2 mostra os clusters gerados pela rede SOM e com seus respectivos rótulos conforme abaixo:

- P (Empregados Presenteístas)
- A (Empregados Absenteístas)
- N (Empregados dentro dos padrões Normais)

Observa-se inicialmente na Figura 2, que os absenteístas têm maior incidência, contudo, deve-se atentar para o fato de que a base de dados é composta por atestados e não por número de empregados.

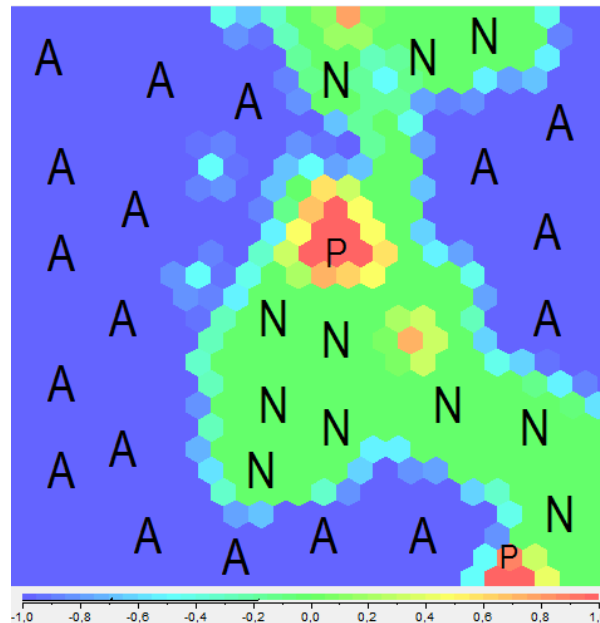


Figura 2 – Rede SOM rotula

A Tabela 1 mostra a divisão por classes (A, N, P), relacionada à frequência simples e relativa do Absenteísmo em horas e o número de empregados em cada classe, além da média de Absenteísmo em horas por empregado durante o período estudado.

Rótulo	Absenteísmo (hs)	(%)	Empregados por classe	(%)	Média de Absenteísmo (hs) por empregado
A	3.638	71,0%	12	33,33%	303
N	1.193	23,3%	21	58,33%	57
P	293	58,3%	3	8,33%	98
Total	5.124	100,0%	36		

Tabela 1 – Absenteísmo em horas

Observa-se na Figura 3 e na Tabela 2 com o atributo Tempo de Afastamento em Horas, que a Grande maioria dos registros do Absenteísmo é de períodos curtos, até 20 horas com 58,3% do Absenteísmo em horas. Constata-se que parte dos empregados com comportamento Presenteísta (P) são justamente os empregados que aparecem com longos períodos de afastamento até 120 horas. Esse fato pode estar ligado ao Presenteísmo, tendo em vista que o afastamento por longos períodos é uma característica desse grupo. Por outro lado, o Absenteísmo com curtos períodos, pode indicar inicialmente que os empregados estão preocupados com sua saúde e que a empresa dá condições para esses cuidados.

A Figura 3 mostra os clusters gerados pela rede SOM (Atributo Tempo de Afastamento em horas).

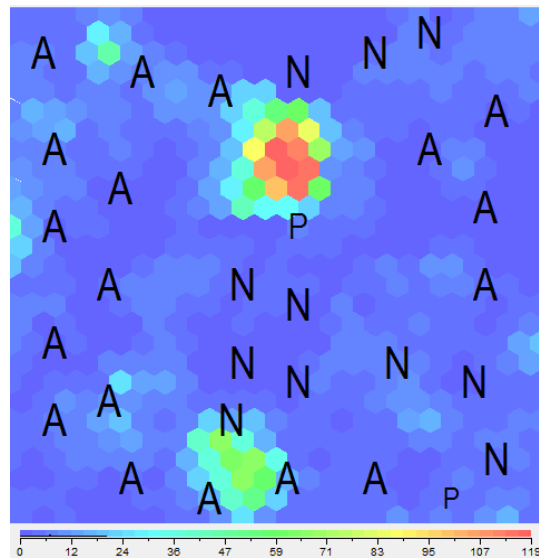


Figura 3 – Atributo tempo de afastamento em horas

Classe	Registros	(%)	Absenteísmo (hs)	(%)
Até 20 anos	660	89,2%	2988	58,3%
De 21 a 40 horas	40	5,4%	856	16,7%
De 41 a 60 horas	15	2,0%	160	3,1%
De 61 a 80 horas	11	1,5%	432	8,4%
De 81 horas em diante	14	1,9%	688	13,4%
Total	740	100,0%	5.124	100,0%

Tabela 2 – Tempo de afastamento em horas

Constata-se na Figura 4 e na Tabela 3 com o atributo Idade, que os empregados com maior idade são presenteístas (P) e a classe entre 36 a 40 anos de idade apresenta o maior absenteísmo em horas. Observa-se ainda, que a partir dos 41 anos de idade existe uma tendência ao presenteísmo.

A Figura 4 mostra os clusters gerados pela rede SOM (Atributo Idade).

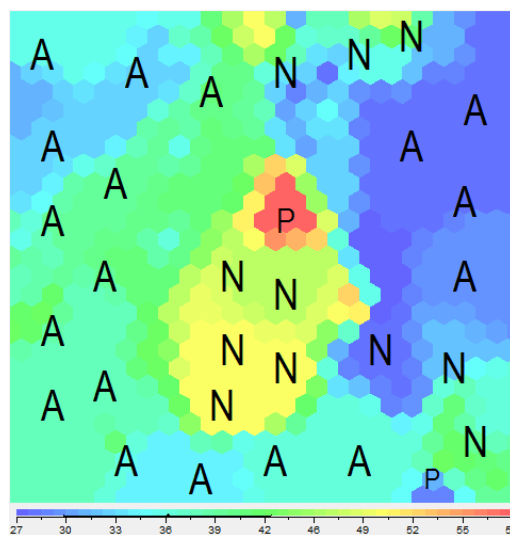


Figura 4 – Atributo idade

A Tabela 3 mostra a frequência simples e relativa relacionada a Idade, empregados, registros e o absenteísmo em horas.

Classe	Registros	(%)	Registros	(%)	Absenteísmo (hs)	(%)
Até 30 anos	6	16,7%	177	23,9%	982	18,8%
De 31 a 35 anos	8	22,2%	115	15,5%	1279	25,0%
De 36 a 40 anos	10	27,8%	307	41,5%	1702	33,2%
De 41 a 45 anos	4	11,1%	58	7,8%	462	9,0%
De 46 anos em diante	8	22,2%	83	11,2%	719	14,0%
Total	36	100,0%	740	100,0%	5124	100,0%

Tabela 3 – Idade versus Absenteísmo

Verifica-se na Figura 5 e na Tabela 4 com o atributo Distância da Residência para o Trabalho em Quilômetros, que a maior parte dos absenteístas, residem entre 11 a 30 quilômetros de distância do trabalho, o que pode indicar uma relação entre o absenteísmo e a distância do trabalho.

Constata-se ainda que os empregados presenteístas rotulados no mapa (P), não residem muito além da distância média 27,35 Km. Isso pode indicar que o tempo economizado em transporte até o local de trabalho pode ser utilizado em atendimento médico ou realização de exames, por exemplo. A Figura 5 mostra os clusters gerados pela rede SOM (Atributo Distância da Residência para o Trabalho em quilômetros).

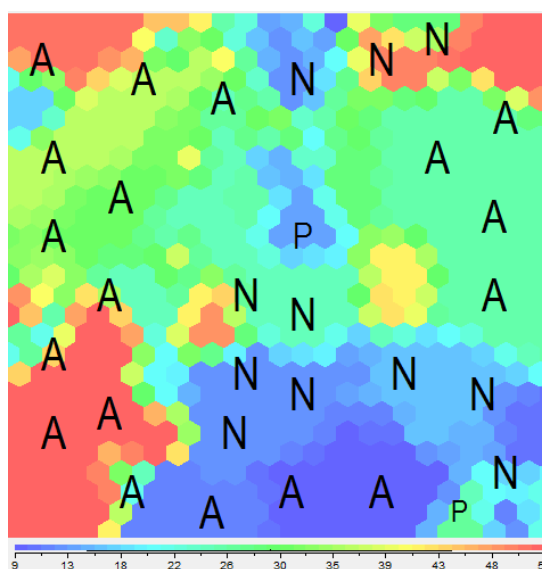


Figura 5 – Atributo Distância da Residência para o Trabalho em Quilômetros

A Tabela 4 mostra a frequência simples e relativa relacionada à distância da residência para o trabalho em quilômetros, empregados, registros e o absenteísmo em horas.

Classe	Registros	(%)	Registros	(%)	Absenteísmo (hs)	(%)
Até 10 Km	2	5,6%	61	8,2%	374	7,3%
De 11 a 20 Km	13	36,1%	167	22,6%	1582	30,9%
De 21 a 30 Km	10	27,8%	260	35,1%	1368	26,7%
De 31 a 40 Km	3	8,3%	42	5,7%	703	13,7%
De 41 Km em diante	8	22,2%	210	28,4%	1097	21,4%
Total	36	100,0%	740	100,0%	5124	100,0%

Tabela 4 – Distância da Residência para o Trabalho em Quilômetros

6. Análise dos Resultados

Analisa-se e discute-se nesta seção os resultados da mineração de dados realizada pela rede SOM. Desta forma, conseguiu-se identificar o perfil do empregado ausenteísta, presenteísta e dentro do padrão normal contido na base de dados da empresa de *courier* estudada. A Tabela 5 apresenta de forma sumarizada os resultados.

	Absenteísta (A)	Normal (N)	Presenteísta (P)
Absenteísmo em horas (%)	71%	23%	5,70%
Empregados	33%	58%	8,33%
Absenteísmo em horas (maior ocorrência de empregados)	Até 20 h	Até 20 h	Até 120 h
Absenteísmo em horas (menor ocorrência de empregados)	Até 60 h	Até 40 h	Até 20 h
Idade (maior ocorrência de empregados)	36 a 40 anos	36 a 40 anos	46 a 58 anos
Idade (menor ocorrência de empregados)	41 a 45 anos	41 a 45 anos	Até 30 anos
Distância da residência (Km / maior ocorrência de empregados)	Acima de 41 Km	Até 20 Km	Até 20 Km
Distância da residência (Km / menor ocorrência de empregados)	Até 10 Km	Até 10 Km	Até 30 Km

Tabela 5 – Distância da Residência para o Trabalho em Quilômetros

a) Síntese analítica do comportamento Absenteísta: corresponde a 71% das horas perdidas e 33,33% dos empregados. Em relação ao tempo de afastamento, esse grupo apresenta maior incidência de Absenteísmo em até 20 horas e menor incidência em até 60 horas. Com relação à idade dos empregados, esse grupo apresenta maior incidência de Absenteísmo na faixa de 36 a 40 anos de idade e menor incidência na faixa de 41 a 45 anos de idade. Em relação à distância da residência para o trabalho em quilômetros esse grupo apresenta maior incidência acima de 41 quilômetros de distância do trabalho e com menor incidência até 10 quilômetros de distância do trabalho.

b) Síntese analítica do comportamento normal: corresponde a 23,3% das horas perdidas e 58,33% dos empregados. Em relação ao tempo de afastamento, esse grupo apresenta maior incidência de Absenteísmo em até 20 horas e menor incidência em até 40 horas. Em relação à idade dos empregados, esse grupo apresenta maior incidência de Absenteísmo na faixa de 36 a 40 anos de idade e menor incidência na faixa de 41 a 45 anos de idade. Em relação à distância da residência para o trabalho em quilômetros, esse grupo apresenta maior incidência em até 20 quilômetros de distância do trabalho e com menor incidência em até 10 quilômetros de distância do trabalho.

c) Síntese analítica do comportamento Presenteísta: corresponde a 5,7% das horas perdidas e 8,33% dos empregados. Em relação ao tempo de afastamento, esse grupo apresenta maior incidência de Absenteísmo em até 120 horas e menor incidência em até 20 horas. Em relação à idade dos empregados, esse grupo apresenta maior incidência de Absenteísmo na faixa de

46 a 58 anos de idade e menor incidência até 30 anos de idade. Em relação à distância da residência para o trabalho em quilômetros, esse grupo apresenta maior incidência em até 20 quilômetros de distância do trabalho e com menor incidência em até 30 quilômetros de distância do trabalho.

7. Conclusão

O uso da rede SOM como técnica de Mineração de Dados apresentou boa capacidade de treinamento, pois permitiu determinar o perfil de empregados absenteístas e presenteístas e descobriu dentro de um grupo de empregados em uma empresa de *Courier* na cidade de São Paulo, o comportamento absenteísta ou presenteísta, permitindo a prevenção e recuperação desses empregados.

A rede SOM agrupou os registros com base na similaridade dos atributos, possibilitando a análise dos clusters dos grupos de empregados.

Com os resultados obtidos, conclui-se que a rede SOM pode auxiliar a tomada de decisão, uma vez que o comportamento dos grupos pode ser identificado e os potenciais problemas de cada grupo em relação ao Absenteísmo e Presenteísmo podem ser tratados, reduzindo os programas corporativos sem foco para dar lugar a programas localizados mais eficazes e menos custosos para a empresa. Com isso visa-se a melhor qualidade de vida no trabalho, maior produtividade e qualidade dos serviços.

Desta forma a contribuição da MD é fundamental para reduzir não apenas a baixa produtividade da empresa, mas para reduzir as contas médicas e assistenciais, além de promover melhor qualidade de vida aos empregados, primando por uma empresa saudável e produtiva.

Consideram-se como principais contribuições do trabalho:

- auxílio na tomada de decisão, uma vez que ao identificar o perfil absenteísta e presenteísta podem-se reduzir os programas corporativos sem foco para dar lugar a programas localizados mais eficazes e menos custosos para a empresa;
- auxílio no tratamento médico e social dos empregados que apresentam um dos dois perfis já citados;
- implantação de programas com vistas a melhorar a qualidade de vida no trabalho, a maior produtividade e a melhor qualidade dos serviços; e
- redução das contas médicas e assistenciais da empresa.

As empresas procuram ferramentas que melhorem não apenas seu desempenho, mas que auxiliem nas suas decisões. Dessa maneira, atuam nos interesses comuns, desenvolvendo novas formas de apoiar seus negócios e conservar o controle sobre a produtividade.

Pretende-se como continuidade dos estudos aplicar uma rede neural artificial do tipo *Multilayer Perceptron* (MLP) também como técnica de Mineração de Dados para classificar os registros que a rede SOM agrupou.

Referências

JODAS, D. A.; HADDAD, M. C. L. *Síndrome de Burnout em Trabalhadores de Enfermagem de um Pronto Socorro de Hospital Universitário*. Acta paul. Enfer. vol. 22, n. 2. São Paulo, 2009.

JOHNS, G. *Presenteeism in the workplace: A review and research agenda*. *Journal of Organizational Behavior*. vol. 31, p. 519 – 542, 2010.

HEMP, P. *Presenteeism: At Work – but out of it*. *Business Review. Journal article* vol.82, n.10, p. 49-58. Harvard Business Review. Journal article, 2004.

CUNHA, J. B.; BLANK, V. L. G.; BOING, A. F. *Tendência Temporal de Afastamento do Trabalho em Servidores Públicos (1995-2005)*. Revista Brasileira de Epidemiologia. vol.12 n. 2 São Paulo, jun., 2009.

FAYYAD, U; PIATETSKY-SHAPIRO, G.; SMITH, P. *The KDD process for extracting useful knowledge from volumes of data*. Communications of the ACM, vol. 39, p. 27-34, 1996.

MARTINIANO, A.; FERREIRA, R. P.; SASSI, R. J. *Um estudo sobre o Absenteísmo através do Código Internacional de Doenças utilizando Data Mining*. VII – EMEPRO, Encontro Mineiro de Engenharia de Produção. São João Del Rei, 2011.

MARTINIANO, A. *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*. 2012. Dissertação (Mestrado) – Universidade Nove de Julho, Engenharia de Produção, São Paulo. 120 p. São Paulo, 2012.

NUBILA, H. B. V.; BUCHALLA, C. M. *O papel das Classificações da OMS - CID e CIF nas definições de deficiência e incapacidade*. Rev. Brasileira de Epidemiologia. vol. 11, n 2: p. 324-335, 2008.

GOLDSCHMIDT, R.; PASSOS, E. *Data Mining um guia prático. Conceitos, Técnicas, Ferramentas, Orientações e Aplicações*. Ed. Campus, Rio de Janeiro, 2005.

CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; ZANASI, A. *Discovering Data Mining: from concept to implementation*. New Jersey : Prentice Hall, 1998.

SASSI, R. J.; SILVA, L. A. ; HERNANDEZ, E. M. *A Methodology using Neural Networks to Cluster Validity Discovered from a Marketing Database*. In 10th Brazilian Symposium on Artificial Neural Networks (SBRN 2008). IEEE Proceedings of SBRN 2008. IEEE Computer Society, v.1. p. 03-08, Salvador, 27-31 de out. 2008.

KOHONEN, T. *Self-Organizing Maps*. New York. Springer. 3ª Edition, 2001.

ARAÚJO, S. A. *Classificação de perfis profissiográficos usando a rede de Kohonen*. 2002. Dissertação (Mestrado) – Universidade Presbiteriana Mackenzie, Departamento de Engenharia Elétrica, São Paulo. 85 p. São Paulo, 2002.

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

SILVA, L. A. *Categorização de Imagens Médicas para Sistemas de Recuperação de Imagens por Conteúdo Baseada em Transformada Wavelet e Mapas Auto-Organizáveis*. 2009. Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia Elétrica, São Paulo. 99 p. São Paulo, 2009.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Atificiais para Engenharia e Ciências Aplicadas*. SP: Artliber, 2010.

KOHONEN, T. *Self-organized formation of topologically correct feature maps*. Biological Cy-bernetics, vol. 43, p. 59-69, 1982.

HAYKIN, S. *Redes Neurais – Princípios e Práticas*. Bookman. 2ª ed. Porto Alegre, 2001.